WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jun 15, 2024 – 10:02 PM EDT

PDB ID	:	2PFY
Title	:	Crystal structure of DctP7, a Bordetella pertussis extracytoplasmic solute receptor binding pyroglutamic acid
Authors	:	Rucktooa, P.
Deposited on	:	2007-04-06
Resolution	:	1.95 Å(reported)

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

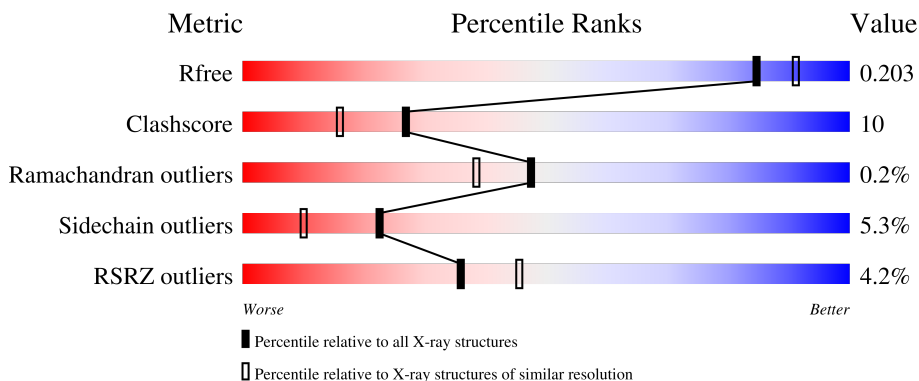|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:
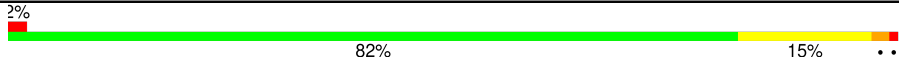*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.95 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		0.203
Clashscore		10
Ramachandran outliers		0.2%
Sidechain outliers		5.3%
RSRZ outliers		4.2%

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	2580 (1.96-1.96)
Clashscore	141614	2705 (1.96-1.96)
Ramachandran outliers	138981	2678 (1.96-1.96)
Sidechain outliers	138945	2678 (1.96-1.96)
RSRZ outliers	127900	2539 (1.96-1.96)

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	301	2% / 82% / 15% / ••
1	B	301	% / 82% / 15% / •
1	C	301	11% / 82% / 16% / •
1	D	301	3% / 81% / 16% / •

2 Entry composition ⓘ

There are 3 unique types of molecules in this entry. The entry contains 10009 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 exported protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	301	Total	C	N	O	Se	0	0	0
			2308	1462	401	439	6			
1	B	301	Total	C	N	O	Se	0	0	0
			2308	1462	401	439	6			
1	C	301	Total	C	N	O	Se	0	0	0
			2308	1462	401	439	6			
1	D	301	Total	C	N	O	Se	0	0	0
			2308	1462	401	439	6			

There are 24 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	6	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
A	152	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
A	156	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
A	180	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
A	260	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
A	280	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
B	6	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
B	152	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
B	156	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
B	180	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
B	260	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
B	280	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
C	6	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
C	152	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
C	156	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
C	180	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
C	260	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
C	280	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
D	6	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
D	152	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
D	156	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	180	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
D	260	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1
D	280	MSE	MET	MODIFIED RESIDUE	UNP Q7VXB1

- Molecule 2 is PYROGLUTAMIC ACID (three-letter code: PCA) (formula: $C_5H_7NO_3$).
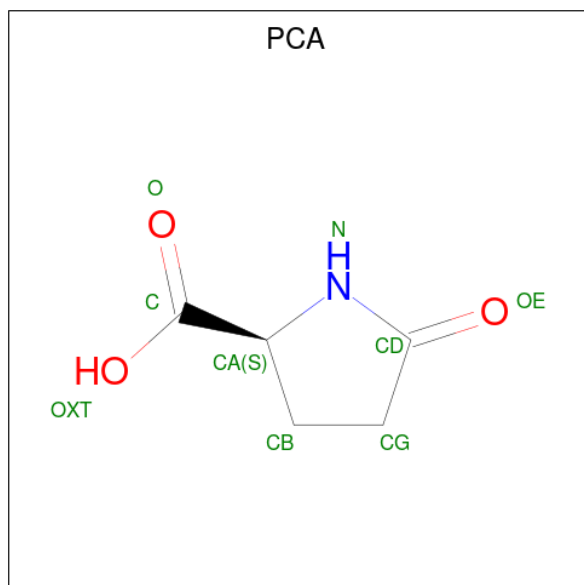
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	N	O	0	0
			9	5	1	3		
2	B	1	Total	C	N	O	0	0
			9	5	1	3		
2	C	1	Total	C	N	O	0	0
			9	5	1	3		
2	D	1	Total	C	N	O	0	0
			9	5	1	3		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	223	Total	O	0	0
			223	223		
3	B	189	Total	O	0	0
			189	189		
3	C	174	Total	O	0	0
			174	174		
3	D	155	Total	O	0	0
			155	155		

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.
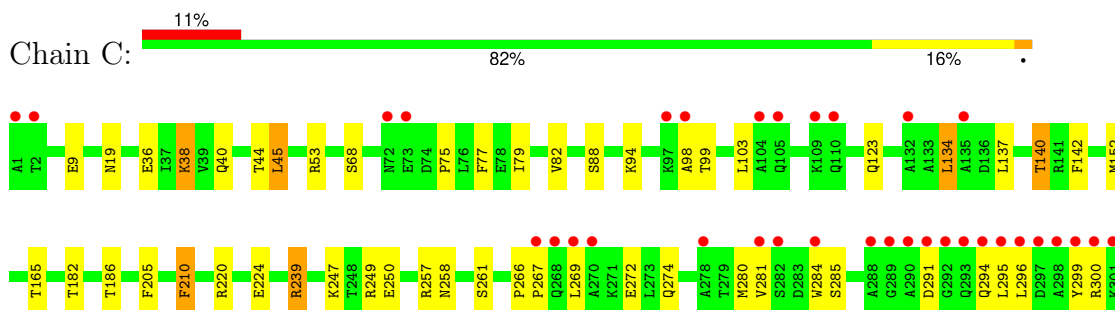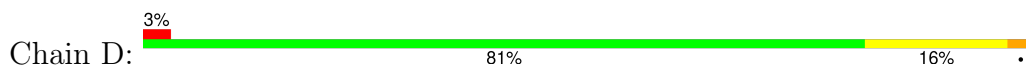
- Molecule 1: Putative exported protein
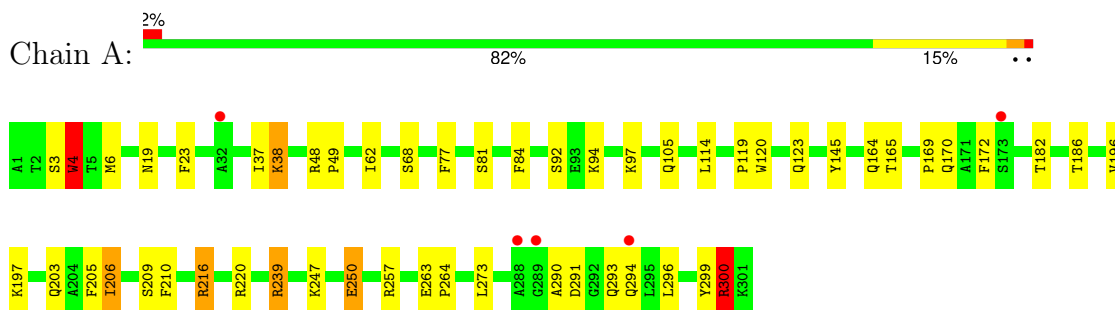
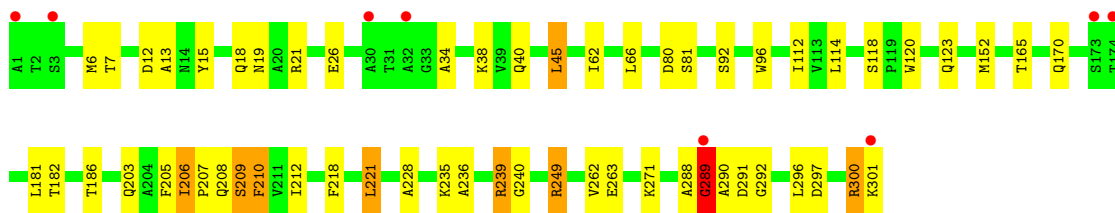

- Molecule 1: Putative exported protein
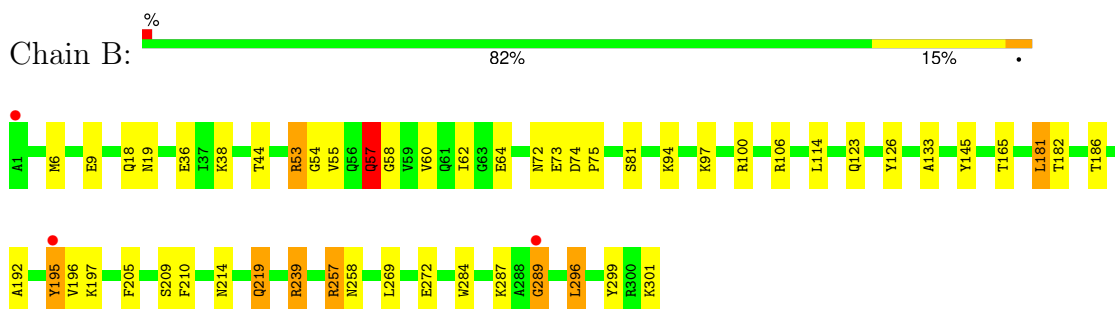
- Molecule 1: Putative exported protein

- Molecule 1: Putative exported protein

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	64.84Å 149.75Å 169.58Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	112.51 – 1.95 5.00 – 1.95	Depositor EDS
% Data completeness (in resolution range)	91.8 (112.51-1.95) 97.9 (5.00-1.95)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.26 (at 1.95Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, R_{free}	0.168 , 0.201 0.170 , 0.203	Depositor DCC
R_{free} test set	5579 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å^2)	28.4	Xtriage
Anisotropy	0.078	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	1.25 , 112.1	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	10009	wwPDB-VP
Average B, all atoms (Å^2)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows:

---

[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.

WORLDWIDE
wwPDB
PROTEIN DATA BANK

5 Model quality ⓘ

5.1 Standard geometry ⓘ

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

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	0.84	1/2350 (0.0%)	0.81	7/3187 (0.2%)
1	B	0.81	2/2350 (0.1%)	0.77	4/3187 (0.1%)
1	C	0.61	0/2350	0.69	6/3187 (0.2%)
1	D	0.63	0/2350	0.70	1/3187 (0.0%)
All	All	0.73	3/9400 (0.0%)	0.75	18/12748 (0.1%)

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	B	0	1
1	D	0	1
All	All	0	2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	219	GLN	CG-CD	5.66	1.64	1.51
1	A	4	TRP	CE3-CZ3	-5.36	1.29	1.38
1	B	57	GLN	CG-CD	5.24	1.63	1.51

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	239	ARG	NE-CZ-NH2	-12.93	113.84	120.30
1	B	239	ARG	NE-CZ-NH2	-10.51	115.04	120.30
1	A	239	ARG	NE-CZ-NH1	10.44	125.52	120.30
1	B	239	ARG	NE-CZ-NH1	8.86	124.73	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	257	ARG	NE-CZ-NH2	-7.42	116.59	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	289	GLY	Peptide
1	D	289	GLY	Peptide

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	2308	0	2322	46	0
1	B	2308	0	2322	45	0
1	C	2308	0	2322	44	0
1	D	2308	0	2322	62	0
2	A	9	0	5	1	0
2	B	9	0	5	1	0
2	C	9	0	5	1	0
2	D	9	0	5	1	0
3	A	223	0	0	17	0
3	B	189	0	0	20	0
3	C	174	0	0	23	0
3	D	155	0	0	30	0
All	All	10009	0	9308	189	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 10.

The worst 5 of 189 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:A:4:TRP:CZ2	3:A:502:HOH:O	1.86	1.26
1:D:206:ILE:HG21	3:D:318:HOH:O	1.08	1.23

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:D:289:GLY:HA3	1:D:292:GLY:H	1.03	1.16
1:A:4:TRP:CE2	3:A:502:HOH:O	2.00	1.09
1:A:250:GLU:HG2	3:A:349:HOH:O	1.53	1.07

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	299/301 (99%)	297 (99%)	2 (1%)	0	100	100
1	B	299/301 (99%)	295 (99%)	4 (1%)	0	100	100
1	C	299/301 (99%)	292 (98%)	7 (2%)	0	100	100
1	D	299/301 (99%)	295 (99%)	2 (1%)	2 (1%)	22	11
All	All	1196/1204 (99%)	1179 (99%)	15 (1%)	2 (0%)	47	38

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	289	GLY
1	D	290	ALA

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	240/234 (103%)	225 (94%)	15 (6%)	18	7
1	B	240/234 (103%)	230 (96%)	10 (4%)	30	17
1	C	240/234 (103%)	228 (95%)	12 (5%)	24	11
1	D	240/234 (103%)	226 (94%)	14 (6%)	20	8
All	All	960/936 (103%)	909 (95%)	51 (5%)	22	10

5 of 51 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	140	THR
1	C	291	ASP
1	D	296	LEU
1	C	165	THR
1	C	261	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 14 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	123	GLN
1	C	274	GLN
1	D	293	GLN
1	D	123	GLN
1	D	203	GLN

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 ⓘ

4 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	PCA	B	302	-	9,9,9	1.56	1 (11%)	12,12,12	4.24	6 (50%)
2	PCA	D	302	-	9,9,9	1.68	2 (22%)	12,12,12	4.91	6 (50%)
2	PCA	C	302	-	9,9,9	1.83	1 (11%)	12,12,12	4.31	5 (41%)
2	PCA	A	302	-	9,9,9	1.66	3 (33%)	12,12,12	4.02	3 (25%)

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	PCA	B	302	-	-	0/4/13/13	0/1/1/1
2	PCA	D	302	-	-	0/4/13/13	0/1/1/1
2	PCA	C	302	-	-	0/4/13/13	0/1/1/1
2	PCA	A	302	-	-	0/4/13/13	0/1/1/1

The worst 5 of 7 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	302	PCA	CD-N	4.92	1.46	1.34
2	B	302	PCA	CD-N	4.21	1.44	1.34
2	D	302	PCA	CD-N	3.82	1.43	1.34
2	A	302	PCA	CD-N	3.31	1.42	1.34
2	A	302	PCA	OXT-C	-2.66	1.22	1.30

The worst 5 of 20 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	302	PCA	CA-N-CD	-12.79	103.65	114.48
2	C	302	PCA	CA-N-CD	-10.41	105.67	114.48

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	302	PCA	CA-N-CD	-10.40	105.67	114.48
2	A	302	PCA	CA-N-CD	-9.98	106.03	114.48
2	D	302	PCA	CB-CA-N	9.72	110.43	102.70

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	B	302	PCA	1	0
2	D	302	PCA	1	0
2	C	302	PCA	1	0
2	A	302	PCA	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(Å²)	Q<0.9
1	A	295/301 (98%)	-0.43	5 (1%)	70	77	15, 26, 46, 59	0
1	B	295/301 (98%)	-0.37	3 (1%)	82	87	14, 27, 45, 54	0
1	C	295/301 (98%)	0.51	34 (11%)	4	8	22, 41, 64, 81	0
1	D	295/301 (98%)	-0.02	8 (2%)	54	63	24, 36, 54, 63	0
All	All	1180/1204 (98%)	-0.08	50 (4%)	36	45	14, 33, 56, 81	0

The worst 5 of 50 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	1	ALA	8.9
1	C	281	VAL	4.6
1	C	288	ALA	4.4
1	C	299	TYR	4.2
1	C	295	LEU	3.9

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
2	PCA	C	302	9/9	0.95	0.09	31,31,32,32	0
2	PCA	D	302	9/9	0.97	0.09	29,30,31,31	0
2	PCA	A	302	9/9	0.98	0.08	20,22,24,25	0
2	PCA	B	302	9/9	0.98	0.06	23,24,25,28	0

6.5 Other polymers ⓘ

There are no such residues in this entry.