WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 06:45 PM GMT

PDB ID	:	4MNQ
Title	:	TCR-peptide specificity overrides affinity enhancing TCR-MHC interactions
Authors	:	Rizkallah, P.J.; Cole, D.K.; Sewell, A.K.; Jakobsen, B.K.
Deposited on	:	2013-09-11
Resolution	:	2.74 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	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
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)	:	trunk26865

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 2.74 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		0.285
Clashscore		13
Ramachandran outliers		2.0%
Sidechain outliers		13.1%
RSRZ outliers		3.4%

Worse *Better*

■ Percentile relative to all X-ray structures
▯ Percentile relative to X-ray structures of similar resolution
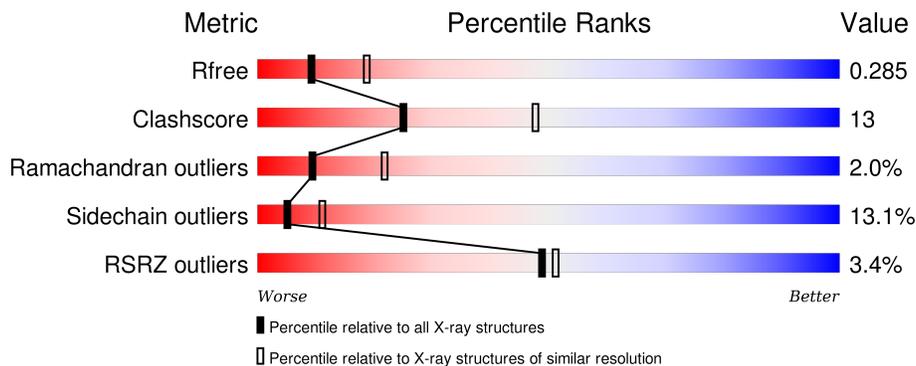
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	3050 (2.78-2.70)
Clashscore	102246	3424 (2.78-2.70)
Ramachandran outliers	100387	3367 (2.78-2.70)
Sidechain outliers	100360	3368 (2.78-2.70)
RSRZ outliers	91569	3055 (2.78-2.70)

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 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 <=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	276	2% 79% 17% •
2	B	100	% 77% 21% •
3	C	9	56% 44%
4	D	200	5% 56% 31% 12% •
5	E	240	5% 65% 31% •

2 Entry composition (i)

There are 8 unique types of molecules in this entry. The entry contains 6684 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 HLA class I histocompatibility antigen, A-2 alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	276	Total	C	N	O	S	0	0	0
			2254	1408	410	427	9			

- Molecule 2 is a protein called Beta-2-microglobulin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	100	Total	C	N	O	S	0	0	0
			837	533	141	159	4			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	EXPRESSION TAG	UNP P61769

- Molecule 3 is a protein called Telomerase reverse transcriptase.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	9	Total	C	N	O	0	0	0
			82	59	13	10			

- Molecule 4 is a protein called Uncharacterized protein, T-cell receptor, sp3.4 alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	200	Total	C	N	O	S	0	0	0
			1565	985	258	315	7			

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	27	ASP	ALA	CONFLICT	UNP H2RG00
D	33	TRP	GLN	CONFLICT	UNP H2RG00
D	89	VAL	-	LINKER	UNP H2RG00

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	90	ASP	-	LINKER	UNP H2RG00
D	91	SER	-	LINKER	UNP H2RG00
D	92	ALA	-	LINKER	UNP H2RG00
D	93	THR	-	LINKER	UNP H2RG00
D	94	ALA	-	LINKER	UNP H2RG00
D	95	LEU	-	LINKER	UNP H2RG00
D	96	PRO	-	LINKER	UNP H2RG00
D	97	TYR	-	LINKER	UNP H2RG00
D	98	GLY	-	LINKER	UNP H2RG00
D	99	TYR	-	LINKER	UNP H2RG00
D	100	ILE	-	LINKER	UNP H2RG00
D	101	PHE	-	LINKER	UNP H2RG00
D	102	GLY	-	LINKER	UNP H2RG00
D	103	THR	-	LINKER	UNP H2RG00
D	104	GLY	-	LINKER	UNP H2RG00
D	105	THR	-	LINKER	UNP H2RG00
D	106	ARG	-	LINKER	UNP H2RG00
D	107	LEU	-	LINKER	UNP H2RG00
D	108	LYS	-	LINKER	UNP H2RG00
D	109	VAL	-	LINKER	UNP H2RG00
D	110	LEU	-	LINKER	UNP H2RG00
D	111	ALA	-	LINKER	UNP H2RG00
D	112	ASN	-	LINKER	UNP H2RG00

- Molecule 5 is a protein called V_segment translation product, T-cell receptor beta-1 chain C region.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	E	240	1911	1203	329	370	9	0	0	0

There are 26 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	50	ILE	VAL	CONFLICT	UNP A0A585
E	51	HIS	GLY	CONFLICT	UNP A0A585
E	52	PRO	ALA	CONFLICT	UNP A0A585
E	53	GLU	GLY	CONFLICT	UNP A0A585
E	54	TYR	ILE	CONFLICT	UNP A0A585
E	96	GLN	-	LINKER	UNP A0A585
E	97	GLY	-	LINKER	UNP A0A585
E	98	THR	-	LINKER	UNP A0A585

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	99	GLU	-	LINKER	UNP A0A585
E	100	ALA	-	LINKER	UNP A0A585
E	101	PHE	-	LINKER	UNP A0A585
E	102	PHE	-	LINKER	UNP A0A585
E	103	GLY	-	LINKER	UNP A0A585
E	104	GLN	-	LINKER	UNP A0A585
E	105	GLY	-	LINKER	UNP A0A585
E	106	THR	-	LINKER	UNP A0A585
E	107	ARG	-	LINKER	UNP A0A585
E	108	LEU	-	LINKER	UNP A0A585
E	109	THR	-	LINKER	UNP A0A585
E	110	VAL	-	LINKER	UNP A0A585
E	111	VAL	-	LINKER	UNP A0A585
E	112	GLU	-	LINKER	UNP A0A585
E	148	TYR	PHE	CONFLICT	UNP P01850
E	168	CYS	SER	CONFLICT	UNP P01850
E	186	ALA	CYS	CONFLICT	UNP P01850
E	200	ASP	ASN	CONFLICT	UNP P01850

- Molecule 6 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).
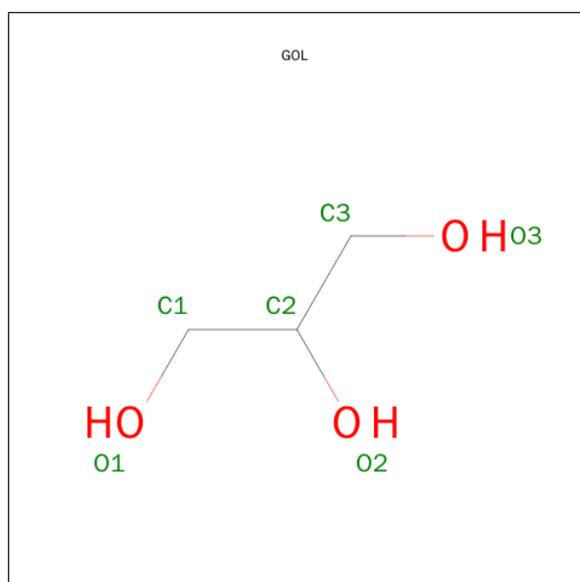
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
6	A	1	6	3	3	0	0

- Molecule 7 is 1,2-ETHANEDIOL (three-letter code: EDO) (formula: $C_2H_6O_2$).

EDO
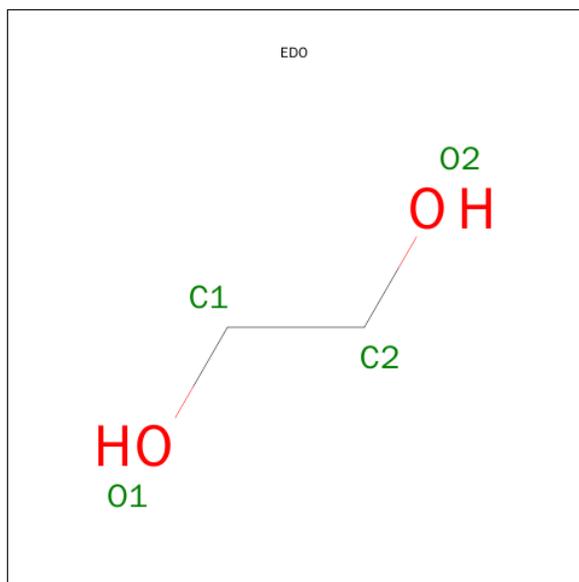
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
7	A	1	Total 4	C 2	O 2	0	0
7	E	1	Total 4	C 2	O 2	0	0

- Molecule 8 is water.

| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	11	Total 11	O 11	0	0
8	B	5	Total 5	O 5	0	0
8	D	2	Total 2	O 2	0	0
8	E	3	Total 3	O 3	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: HLA class I histocompatibility antigen, A-2 alpha chain
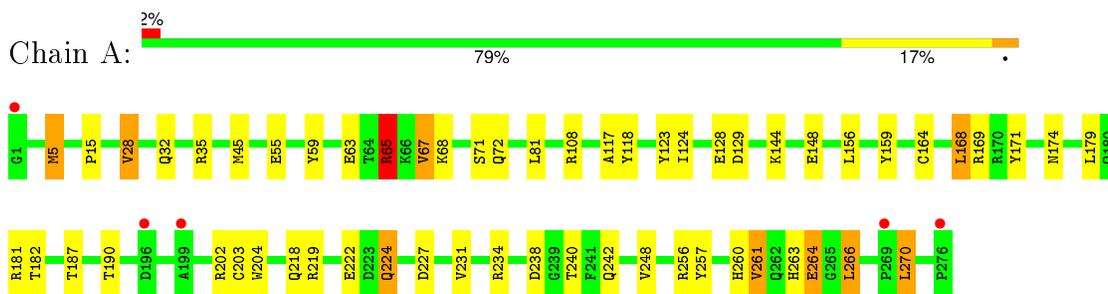

• Molecule 2: Beta-2-microglobulin
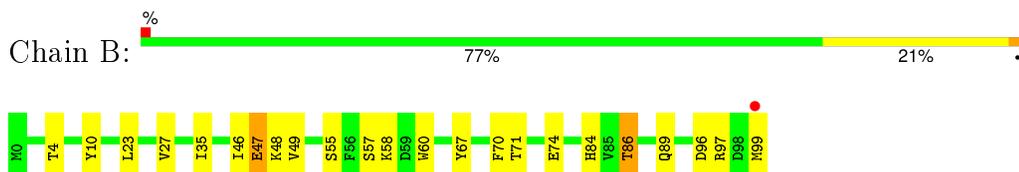
• Molecule 3: Telomerase reverse transcriptase

• Molecule 4: Uncharacterized protein, T-cell receptor, sp3.4 alpha chain
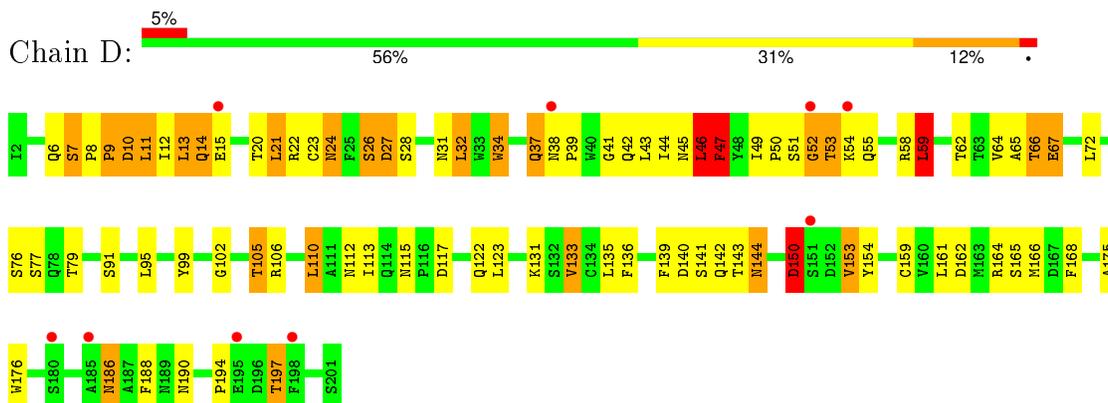
• Molecule 5: V_segment translation product, T-cell receptor beta-1 chain C region

Chain E:
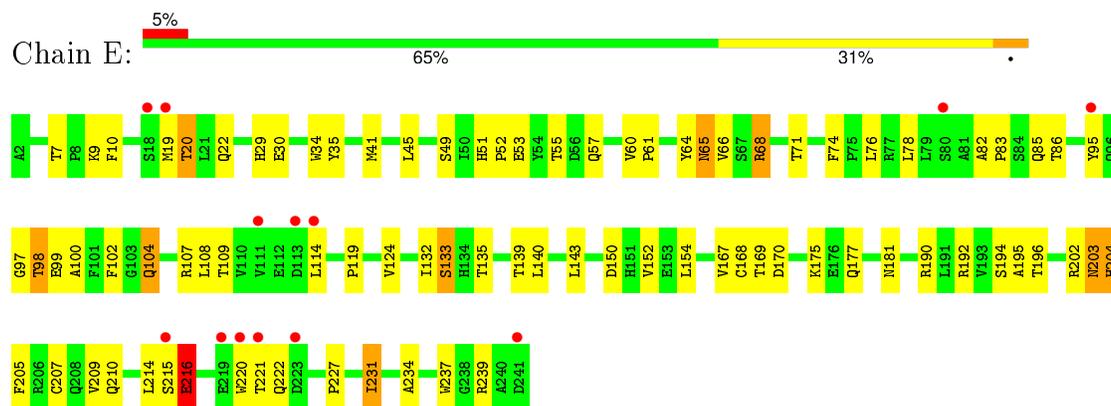
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	94.22Å 48.49Å 118.07Å 90.00° 107.86° 90.00°	Depositor
Resolution (Å)	42.15 – 2.74 42.15 – 2.74	Depositor EDS
% Data completeness (in resolution range)	97.7 (42.15-2.74) 97.7 (42.15-2.74)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.04 (at 2.73Å)	Xtriage
Refinement program	REFMAC 5.7.0032	Depositor
R, R_{free}	0.214 , 0.284 0.217 , 0.285	Depositor DCC
R_{free} test set	1344 reflections (5.33%)	DCC
Wilson B-factor (Å^2)	59.1	Xtriage
Anisotropy	0.290	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 41.5	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Outliers	0 of 26563 reflections	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	6684	wwPDB-VP
Average B, all atoms (Å^2)	65.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.32% 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, 0.375 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, EDO

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	0.64	0/2320	0.78	1/3149 (0.0%)
2	B	0.70	0/860	0.78	0/1162
3	C	0.64	0/85	0.64	0/113
4	D	0.59	0/1602	0.81	2/2183 (0.1%)
5	E	0.53	0/1965	0.70	1/2679 (0.0%)
All	All	0.61	0/6832	0.77	4/9286 (0.0%)

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
4	D	0	2

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	D	28	SER	N-CA-C	5.74	126.51	111.00
4	D	59	LEU	CA-CB-CG	5.32	127.53	115.30
5	E	97	GLY	N-CA-C	-5.26	99.96	113.10
1	A	65	ARG	NE-CZ-NH1	5.17	122.89	120.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	D	47	PHE	Peptide

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Group
4	D	65	ALA	Peptide

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	2254	0	2103	30	0
2	B	837	0	803	19	0
3	C	82	0	90	3	0
4	D	1565	0	1480	72	0
5	E	1911	0	1796	51	0
6	A	6	0	8	0	0
7	A	4	0	6	0	0
7	E	4	0	6	0	0
8	A	11	0	0	1	0
8	B	5	0	0	0	0
8	D	2	0	0	4	0
8	E	3	0	0	0	0
All	All	6684	0	6292	164	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 13.

The worst 5 of 164 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)
8:A:411:HOH:O	4:D:27:ASP:HB2	1.71	0.89
4:D:67:GLU:HA	8:D:402:HOH:O	1.74	0.86
4:D:49:ILE:O	4:D:64:VAL:HG21	1.79	0.81
4:D:67:GLU:CA	8:D:402:HOH:O	2.29	0.77
4:D:51:SER:O	4:D:53:THR:N	2.18	0.77

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	274/276 (99%)	254 (93%)	18 (7%)	2 (1%)	26	54
2	B	98/100 (98%)	94 (96%)	4 (4%)	0	100	100
3	C	7/9 (78%)	4 (57%)	3 (43%)	0	100	100
4	D	198/200 (99%)	167 (84%)	18 (9%)	13 (7%)	1	2
5	E	238/240 (99%)	212 (89%)	25 (10%)	1 (0%)	39	68
All	All	815/825 (99%)	731 (90%)	68 (8%)	16 (2%)	9	22

5 of 16 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	7	SER
4	D	27	ASP
4	D	52	GLY
4	D	66	THR
4	D	46	LEU

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	232/232 (100%)	212 (91%)	20 (9%)	13	28
2	B	95/95 (100%)	87 (92%)	8 (8%)	14	29
3	C	8/8 (100%)	8 (100%)	0	100	100
4	D	179/179 (100%)	141 (79%)	38 (21%)	1	3

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
5	E	209/209 (100%)	180 (86%)	29 (14%)	4	9
All	All	723/723 (100%)	628 (87%)	95 (13%)	5	11

5 of 95 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	38	ASN
4	D	105	THR
5	E	203	ASN
4	D	43	LEU
4	D	54	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 28 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	6	GLN
4	D	31	ASN
5	E	104	GLN
4	D	14	GLN
4	D	24	ASN

5.3.3   RNA (i)

There are no RNA molecules in this entry.

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

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

5.5   Carbohydrates (i)

There are no carbohydrates in this entry.

5.6   Ligand geometry (i)

3 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and

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

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	#$|Z| > 2$	Counts	RMSZ	#$|Z| > 2$
6	GOL	A	301	-	5,5,5	0.37	0	5,5,5	0.23	0
7	EDO	A	302	-	3,3,3	0.56	0	2,2,2	0.47	0
7	EDO	E	301	-	3,3,3	0.44	0	2,2,2	0.35	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
6	GOL	A	301	-	-	0/4/4/4	0/0/0/0
7	EDO	A	302	-	-	0/1/1/1	0/0/0/0
7	EDO	E	301	-	-	0/1/1/1	0/0/0/0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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 (i)

6.1 Protein, DNA and RNA chains (i)

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	276/276 (100%)	0.17	5 (1%)	71	74	27, 56, 101, 119	0
2	B	100/100 (100%)	-0.09	1 (1%)	84	86	26, 47, 75, 95	0
3	C	9/9 (100%)	0.52	0	100	100	42, 46, 51, 54	0
4	D	200/200 (100%)	0.42	9 (4%)	37	37	42, 76, 110, 123	0
5	E	240/240 (100%)	0.26	13 (5%)	29	29	35, 68, 112, 129	0
All	All	825/825 (100%)	0.23	28 (3%)	49	51	26, 63, 107, 129	0

The worst 5 of 28 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	52	GLY	6.3
2	B	99	MET	4.6
4	D	180	SER	4.2
5	E	223	ASP	4.1
4	D	38	ASN	3.9

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

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

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($Å^2$)	Q<0.9
7	EDO	A	302	4/4	0.74	0.27	1.92	59,65,67,68	0
6	GOL	A	301	6/6	0.78	0.21	0.82	70,73,75,78	0
7	EDO	E	301	4/4	0.82	0.24	-	61,61,65,70	0

6.5 Other polymers (i)

There are no such residues in this entry.