wwPDB X-ray Structure Validation Summary Report (i)

Sep 10, 2024 – 12:17 PM JST

PDB ID	:	9IHS
Title	:	Microbial transglutaminase mutant - D3C/G283C
Authors	:	Suzuki, M.; Date, M.; Kashiwagi, T.; Takahashi, K.; Nakamura, A.; Tanokura, M.; Suzuki, E.; Yokoyama, K.
Deposited on	:	2024-06-18
Resolution	:	2.00 Å(reported)

This is a wwPDB X-ray Structure Validation Summary Report for a publicly released PDB entry.

We welcome your comments at *validation@mail.wwpdb.org*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

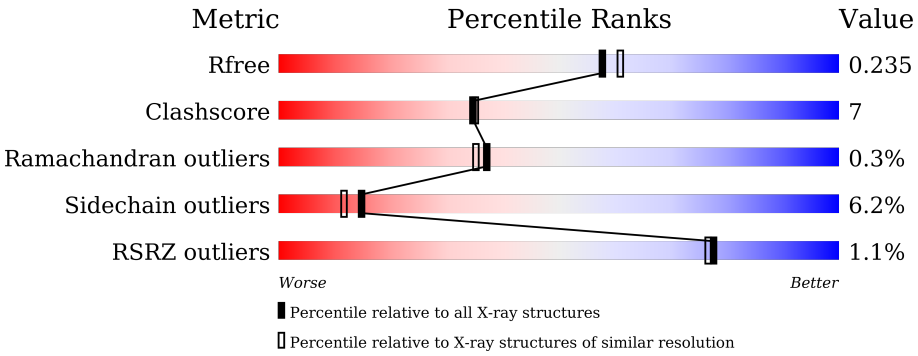
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.002 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.38.2

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.00 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}	164625	9409 (2.00-2.00)
Clashscore	180529	10737 (2.00-2.00)
Ramachandran outliers	177936	10628 (2.00-2.00)
Sidechain outliers	177891	10627 (2.00-2.00)
RSRZ outliers	164620	9409 (2.00-2.00)

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 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	331	86% 11%
1	B	331	85% 14%
1	C	331	90% 8%
1	D	331	3% 76% 21%

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 11885 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.

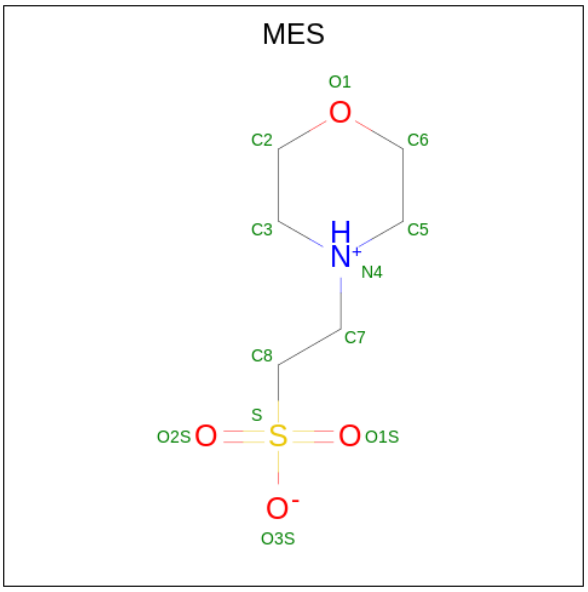
- Molecule 1 is a protein called Protein-glutamine gamma-glutamyltransferase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	331	Total	C	N	O	S	0	0	0
			2678	1659	495	515	9			
1	B	331	Total	C	N	O	S	0	0	0
			2678	1659	495	515	9			
1	C	331	Total	C	N	O	S	0	1	0
			2686	1664	498	515	9			
1	D	331	Total	C	N	O	S	0	0	0
			2678	1659	495	515	9			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	3	CYS	ASP	engineered mutation	UNP P81453
A	283	CYS	GLY	engineered mutation	UNP P81453
B	3	CYS	ASP	engineered mutation	UNP P81453
B	283	CYS	GLY	engineered mutation	UNP P81453
C	3	CYS	ASP	engineered mutation	UNP P81453
C	283	CYS	GLY	engineered mutation	UNP P81453
D	3	CYS	ASP	engineered mutation	UNP P81453
D	283	CYS	GLY	engineered mutation	UNP P81453

- Molecule 2 is 2-(N-MORPHOLINO)-ETHANESULFONIC ACID (three-letter code: MES) (formula: $C_6H_{13}NO_4S$).

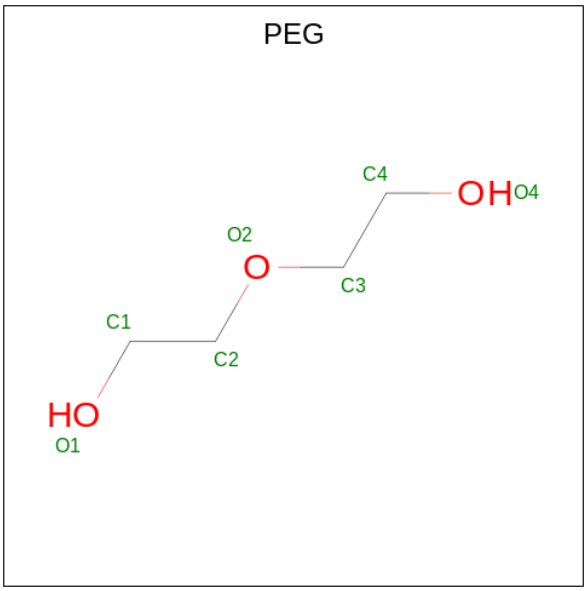


Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	1	Total C N O S 12 6 1 4 1	0	0
2	B	1	Total C N O S 12 6 1 4 1	0	0
2	C	1	Total C N O S 12 6 1 4 1	0	0
2	D	1	Total C N O S 12 6 1 4 1	0	0

- Molecule 3 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total Cl 1 1	0	0
3	B	1	Total Cl 1 1	0	0
3	C	1	Total Cl 1 1	0	0
3	D	1	Total Cl 1 1	0	0

- Molecule 4 is DI(HYDROXYETHYL)ETHER (three-letter code: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	B	1	Total C O 7 4 3	0	0

- Molecule 5 is water.

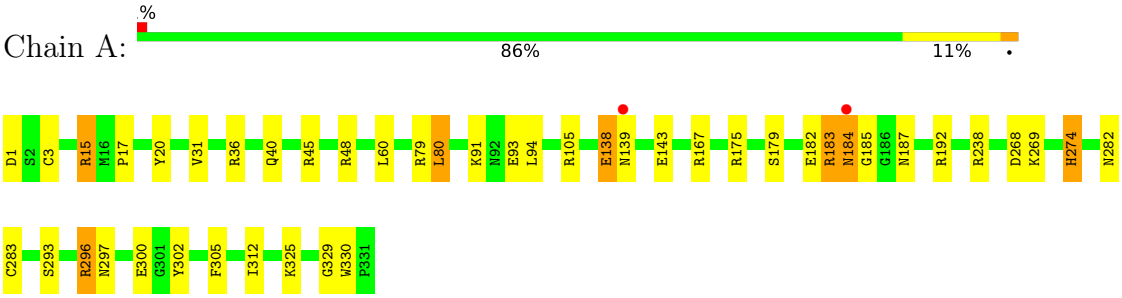
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	292	Total O 292 292	0	0
5	B	306	Total O 306 306	0	0
5	C	320	Total O 320 320	0	0
5	D	188	Total O 188 188	0	0

3 Residue-property plots (i)

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.

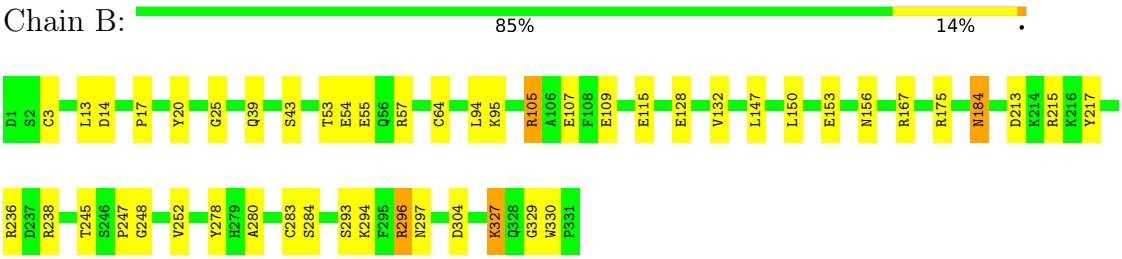
• Molecule 1: Protein-glutamine gamma-glutamyltransferase

Chain A:



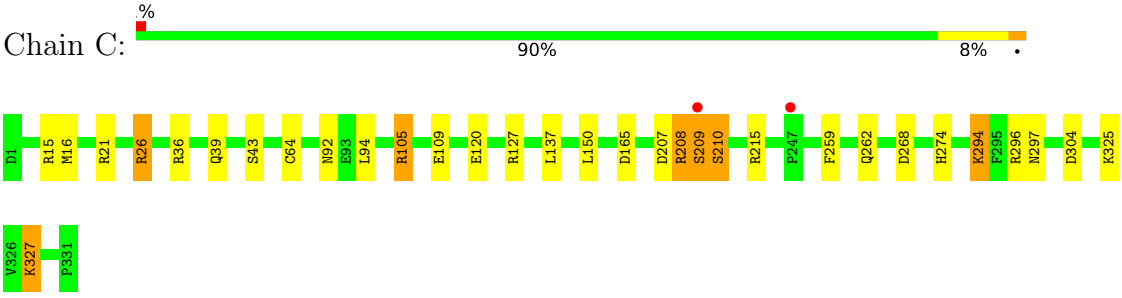
• Molecule 1: Protein-glutamine gamma-glutamyltransferase

Chain B:



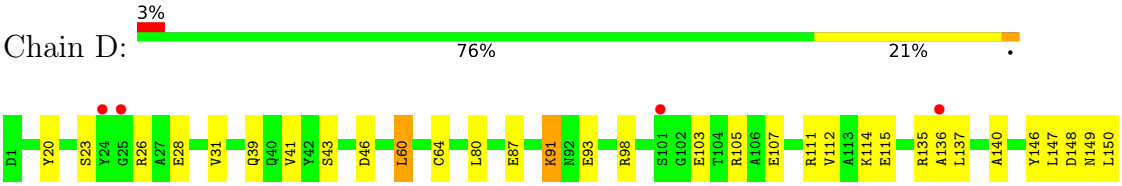
• Molecule 1: Protein-glutamine gamma-glutamyltransferase

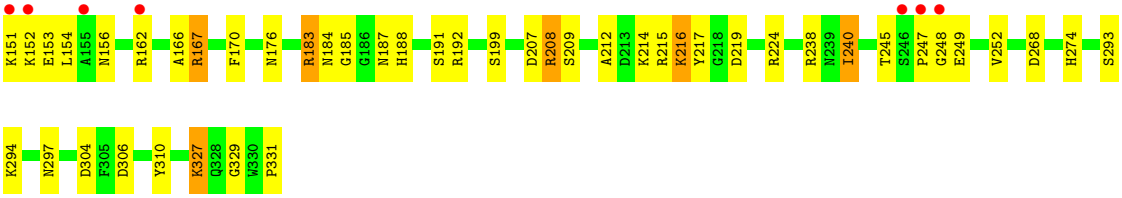
Chain C:



• Molecule 1: Protein-glutamine gamma-glutamyltransferase

Chain D:




K151
K152
E153
L154
A155
N156
R162
A166
R167
F170
N176
R183
N184
G185
G186
N187
H188
S191
R192
S199
D207
R208
S209
A212
D213
K214
R215
K216
Y217
G218
D219
R224
R238
N239
I240
T245
S246
P247
G248
E249
V252
D268
H274
S293

K294
N297
D304
F305
D306
Y310
K327
Q328
G329
W330
P331

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	78.76Å 117.36Å 85.47Å 90.00° 113.03° 90.00°	Depositor
Resolution (Å)	47.08 – 2.00 47.08 – 2.00	Depositor EDS
% Data completeness (in resolution range)	97.2 (47.08-2.00) 97.2 (47.08-2.00)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	5.95 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.8.0267	Depositor
R, R_{free}	0.181 , 0.234 0.181 , 0.235	Depositor DCC
R_{free} test set	4822 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	25.5	Xtriage
Anisotropy	0.027	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 55.7	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	11885	wwPDB-VP
Average B, all atoms (Å^2)	30.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.77% 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: CL, MES, PEG

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	0.40	0/2755	0.78	0/3726
1	B	0.44	0/2755	0.87	6/3726 (0.2%)
1	C	0.44	0/2766	0.78	0/3740
1	D	0.38	0/2755	0.78	0/3726
All	All	0.42	0/11031	0.80	6/14918 (0.0%)

There are no bond length outliers.

The worst 5 of 6 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	175	ARG	NE-CZ-NH2	-9.12	115.74	120.30
1	B	215	ARG	CG-CD-NE	-7.78	95.47	111.80
1	B	175	ARG	NE-CZ-NH1	7.50	124.05	120.30
1	B	215	ARG	NE-CZ-NH2	-6.71	116.95	120.30
1	B	105	ARG	CB-CG-CD	5.76	126.58	111.60

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	2678	0	2487	43	0
1	B	2678	0	2487	37	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	C	2686	0	2500	33	0
1	D	2678	0	2487	55	0
2	A	12	0	13	2	0
2	B	12	0	13	1	0
2	C	12	0	13	0	0
2	D	12	0	13	0	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	B	7	0	10	0	0
5	A	292	0	0	8	0
5	B	306	0	0	10	0
5	C	320	0	0	9	0
5	D	188	0	0	3	0
All	All	11885	0	10023	153	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 7.

The worst 5 of 153 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:92:ASN:HB3	5:C:745:HOH:O	1.26	1.30
1:D:327:LYS:HE2	1:D:331:PRO:OXT	1.55	1.05
1:A:184:ASN:ND2	1:A:184:ASN:H	1.64	0.94
1:B:3:CYS:CB	1:D:245:THR:HG21	1.99	0.93
1:A:184:ASN:H	1:A:184:ASN:HD22	1.18	0.88

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	329/331 (99%)	319 (97%)	10 (3%)	0	100	100
1	B	329/331 (99%)	321 (98%)	7 (2%)	1 (0%)	37	35
1	C	330/331 (100%)	318 (96%)	10 (3%)	2 (1%)	22	17
1	D	329/331 (99%)	313 (95%)	15 (5%)	1 (0%)	37	35
All	All	1317/1324 (100%)	1271 (96%)	42 (3%)	4 (0%)	37	35

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	208	ARG
1	C	210	SER
1	B	247	PRO
1	D	247	PRO

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	280/280 (100%)	264 (94%)	16 (6%)	17	14
1	B	280/280 (100%)	264 (94%)	16 (6%)	17	14
1	C	281/280 (100%)	271 (96%)	10 (4%)	30	30
1	D	280/280 (100%)	252 (90%)	28 (10%)	6	3
All	All	1121/1120 (100%)	1051 (94%)	70 (6%)	15	12

5 of 70 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	183	ARG
1	D	187	ASN
1	D	249	GLU
1	B	167	ARG
1	B	156	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13

such sidechains are listed below:

Mol	Chain	Res	Type
1	C	96	ASN
1	C	297	ASN
1	D	187	ASN
1	D	163	ASN
1	D	176	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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 9 ligands modelled in this entry, 4 are monoatomic - leaving 5 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	PEG	B	402	-	6,6,6	0.24	0	5,5,5	0.20	0
2	MES	C	401	-	12,12,12	0.81	0	14,16,16	1.55	4 (28%)
2	MES	A	401	-	12,12,12	0.69	0	14,16,16	0.59	0
2	MES	D	401	-	12,12,12	0.66	0	14,16,16	0.86	0
2	MES	B	401	-	12,12,12	0.68	0	14,16,16	0.61	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
4	PEG	B	402	-	-	1/4/4/4	-
2	MES	C	401	-	-	1/6/14/14	0/1/1/1
2	MES	A	401	-	-	1/6/14/14	0/1/1/1
2	MES	D	401	-	-	2/6/14/14	0/1/1/1
2	MES	B	401	-	-	1/6/14/14	0/1/1/1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	C	401	MES	C5-N4-C3	3.14	115.90	108.83
2	C	401	MES	C2-C3-N4	3.03	114.69	110.10
2	C	401	MES	C6-C5-N4	2.61	114.05	110.10
2	C	401	MES	O2S-S-C8	-2.09	104.40	106.92

There are no chirality outliers.

5 of 6 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	401	MES	N4-C7-C8-S
2	C	401	MES	C8-C7-N4-C5
2	D	401	MES	C8-C7-N4-C3
2	D	401	MES	C8-C7-N4-C5
4	B	402	PEG	C4-C3-O2-C2

There are no ring outliers.

2 monomers are involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	401	MES	2	0
2	B	401	MES	1	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, 95^{th} 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	331/331 (100%)	-0.33	2 (0%) 85 85	16, 26, 47, 77	0
1	B	331/331 (100%)	-0.45	0 100 100	14, 23, 42, 64	0
1	C	331/331 (100%)	-0.44	2 (0%) 85 85	13, 23, 44, 67	1 (0%)
1	D	331/331 (100%)	0.22	11 (3%) 49 47	20, 35, 62, 94	0
All	All	1324/1324 (100%)	-0.25	15 (1%) 77 76	13, 27, 52, 94	1 (0%)

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	247	PRO	4.3
1	D	155	ALA	4.0
1	D	248	GLY	3.3
1	D	151	LYS	2.8
1	C	209	SER	2.6

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, 95^{th} 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	MES	B	401	12/12	0.78	0.15	40,51,68,69	0
2	MES	A	401	12/12	0.80	0.15	41,46,72,75	0
2	MES	D	401	12/12	0.85	0.13	42,47,58,59	0
4	PEG	B	402	7/7	0.85	0.14	41,46,54,54	0
3	CL	B	403	1/1	0.95	0.10	32,32,32,32	0
2	MES	C	401	12/12	0.97	0.06	25,27,32,34	0
3	CL	A	402	1/1	0.97	0.10	39,39,39,39	0
3	CL	D	402	1/1	0.99	0.04	27,27,27,27	0
3	CL	C	402	1/1	0.99	0.04	21,21,21,21	0

6.5 Other polymers (i)

There are no such residues in this entry.