Full wwPDB X-ray Structure Validation Report ⓘ

Jun 15, 2024 – 06:55 PM EDT

|		
PDB ID	:	4P8I
Title	:	Tgl - a bacterial spore coat transglutaminase
Authors	:	Brito, J.A.; Placido, D.; Fernandes, C.; Lousa, D.; Isidro, A.; Soares, C.M.; Pohl, J.; Carrondo, M.A.; Henriques, A.O.; Archer, M.
Deposited on	:	2014-03-31
Resolution	:	1.85 Å(reported)

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 1.85 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.
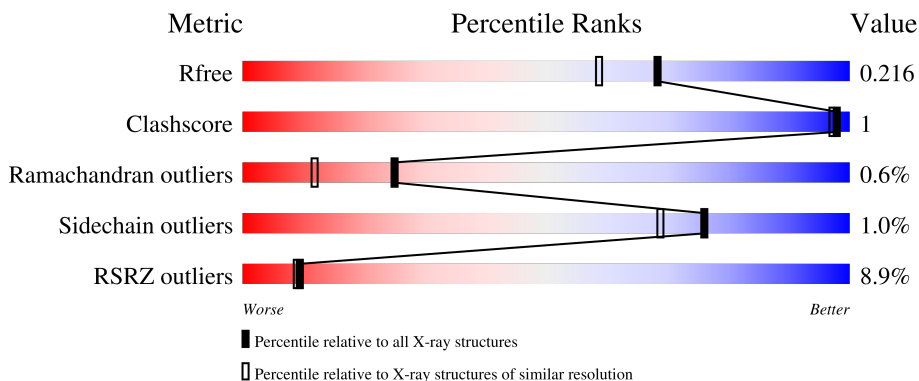
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2469 (1.86-1.86)
Clashscore	141614	2625 (1.86-1.86)
Ramachandran outliers	138981	2592 (1.86-1.86)
Sidechain outliers	138945	2592 (1.86-1.86)
RSRZ outliers	127900	2436 (1.86-1.86)

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	260	9% ▮ 92% · 7%
1	B	260	8% ▮ 88% · 7%

2 Entry composition ⓘ

There are 6 unique types of molecules in this entry. The entry contains 4264 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	242	Total	C	N	O	S	0	1	0
			1982	1294	329	351	8			
1	B	242	Total	C	N	O	S	0	1	0
			1986	1297	331	351	7			

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	246	VAL	-	expression tag	UNP P40746
A	247	ASP	-	expression tag	UNP P40746
A	248	LYS	-	expression tag	UNP P40746
A	249	LEU	-	expression tag	UNP P40746
A	250	ALA	-	expression tag	UNP P40746
A	251	ALA	-	expression tag	UNP P40746
A	252	ALA	-	expression tag	UNP P40746
A	253	LEU	-	expression tag	UNP P40746
A	254	GLU	-	expression tag	UNP P40746
A	255	HIS	-	expression tag	UNP P40746
A	256	HIS	-	expression tag	UNP P40746
A	257	HIS	-	expression tag	UNP P40746
A	258	HIS	-	expression tag	UNP P40746
A	259	HIS	-	expression tag	UNP P40746
A	260	HIS	-	expression tag	UNP P40746
B	246	VAL	-	expression tag	UNP P40746
B	247	ASP	-	expression tag	UNP P40746
B	248	LYS	-	expression tag	UNP P40746
B	249	LEU	-	expression tag	UNP P40746
B	250	ALA	-	expression tag	UNP P40746
B	251	ALA	-	expression tag	UNP P40746
B	252	ALA	-	expression tag	UNP P40746
B	253	LEU	-	expression tag	UNP P40746
B	254	GLU	-	expression tag	UNP P40746
B	255	HIS	-	expression tag	UNP P40746

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	256	HIS	-	expression tag	UNP P40746
B	257	HIS	-	expression tag	UNP P40746
B	258	HIS	-	expression tag	UNP P40746
B	259	HIS	-	expression tag	UNP P40746
B	260	HIS	-	expression tag	UNP P40746

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

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

- Molecule 3 is TRIETHYLENE GLYCOL (three-letter code: PGE) (formula: $C_6H_{14}O_4$).

PGE

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total 20	C 12	O 8	0	1
3	B	1	Total 10	C 6	O 4	0	0

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

PEG

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

- Molecule 5 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).

GOL

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
5	A	1	Total 6	C 3	O 3	0	0

• Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	132	Total 133	O 133	0	1
6	B	107	Total 108	O 108	0	1

3 Residue-property plots (i)
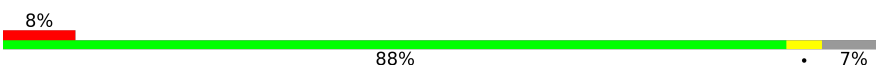
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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:
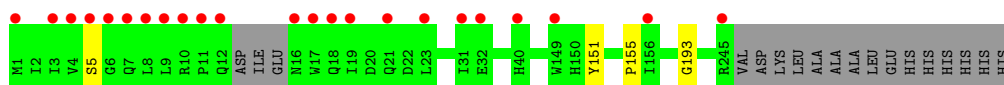
• Molecule 1: Protein-glutamine gamma-glutamyltransferase

Chain B:

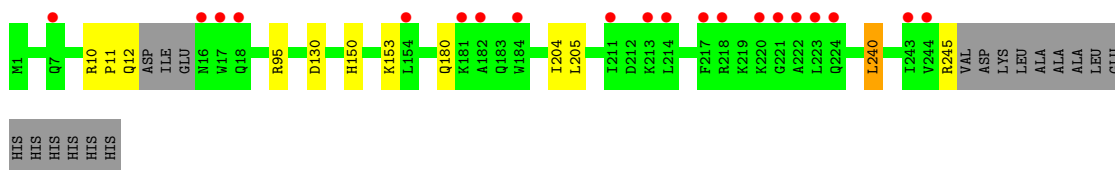# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, α, β, γ	97.94Å 97.94Å 66.18Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	47.85 – 1.85 47.85 – 1.85	Depositor EDS
% Data completeness (in resolution range)	88.6 (47.85-1.85) 88.6 (47.85-1.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.92 (at 1.86Å)	Xtriage
Refinement program	BUSTER-TNT 2.10, PHENIX 1.8.4-dev1565	Depositor
R, R_{free}	0.158 , 0.208 0.172 , 0.216	Depositor DCC
R_{free} test set	2358 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å^2)	30.9	Xtriage
Anisotropy	0.842	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 57.2	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.034 for h,-k,-l	Xtriage
F_o,F_c correlation	0.97	EDS
Total number of atoms	4264	wwPDB-VP
Average B, all atoms (Å^2)	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.46% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: GOL, PGE, EDO, 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	0.56	0/2033	0.62	0/2750
1	B	0.53	0/2038	0.62	1/2757 (0.0%)
All	All	0.55	0/4071	0.62	1/5507 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	240	LEU	CA-CB-CG	5.13	127.11	115.30

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	1982	0	1986	2	0
1	B	1986	0	1988	5	0
2	A	8	0	12	0	0
2	B	4	0	6	0	0
3	A	20	0	28	0	0
3	B	10	0	14	0	0
4	A	7	0	10	0	0
5	A	6	0	8	1	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
6	A	133	0	0	0	0
6	B	108	0	0	2	0
All	All	4264	0	4052	6	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 1.

All (6) 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:95:ARG:NH1	6:B:1101:HOH:O	2.25	0.68
1:B:10:ARG:O	1:B:12:GLN:N	2.31	0.63
1:A:151:TYR:O	1:B:150:HIS:HA	2.10	0.50
1:B:130:ASP:OD2	6:B:1205:HOH:O	2.20	0.48
1:A:193:GLY:HA3	5:A:1005:GOL:H2	2.00	0.43
1:B:204:ILE:C	1:B:205:LEU:HD12	2.41	0.40

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	239/260 (92%)	223 (93%)	15 (6%)	1 (0%)	34	19
1	B	239/260 (92%)	226 (95%)	11 (5%)	2 (1%)	19	7
All	All	478/520 (92%)	449 (94%)	26 (5%)	3 (1%)	25	12

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	11	PRO

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	B	153	LYS
1	A	155	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	210/224 (94%)	209 (100%)	1 (0%)	88	86
1	B	210/224 (94%)	207 (99%)	3 (1%)	67	55
All	All	420/448 (94%)	416 (99%)	4 (1%)	76	69

All (4) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	5	SER
1	B	180	GLN
1	B	240	LEU
1	B	245	ARG

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

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 monosaccharides 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	PGE	A	1003[B]	-	9,9,9	0.29	0	8,8,8	0.32	0
2	EDO	B	1001	-	3,3,3	0.48	0	2,2,2	0.35	0
2	EDO	A	1002	-	3,3,3	0.44	0	2,2,2	0.58	0
4	PEG	A	1004	-	6,6,6	0.47	0	5,5,5	0.33	0
3	PGE	B	1002	-	9,9,9	0.36	0	8,8,8	0.55	0
5	GOL	A	1005	-	5,5,5	0.44	0	5,5,5	0.48	0
3	PGE	A	1003[A]	-	9,9,9	0.32	0	8,8,8	0.35	0
2	EDO	A	1001	-	3,3,3	0.37	0	2,2,2	0.49	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	PGE	A	1003[B]	-	-	3/7/7/7	-
2	EDO	B	1001	-	-	1/1/1/1	-
2	EDO	A	1002	-	-	1/1/1/1	-
4	PEG	A	1004	-	-	0/4/4/4	-
3	PGE	B	1002	-	-	2/7/7/7	-
5	GOL	A	1005	-	-	2/4/4/4	-
3	PGE	A	1003[A]	-	-	0/7/7/7	-
2	EDO	A	1001	-	-	1/1/1/1	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (10) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
5	A	1005	GOL	C1-C2-C3-O3
5	A	1005	GOL	O2-C2-C3-O3
3	B	1002	PGE	O1-C1-C2-O2
3	B	1002	PGE	O2-C3-C4-O3
2	A	1001	EDO	O1-C1-C2-O2
2	A	1002	EDO	O1-C1-C2-O2
3	A	1003[B]	PGE	C1-C2-O2-C3
3	A	1003[B]	PGE	O1-C1-C2-O2
3	A	1003[B]	PGE	O2-C3-C4-O3
2	B	1001	EDO	O1-C1-C2-O2

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
5	A	1005	GOL	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	242/260 (93%)	0.27	23 (9%)	8	7	26, 40, 95, 150	0
1	B	242/260 (93%)	0.21	20 (8%)	11	11	26, 46, 103, 129	0
All	All	484/520 (93%)	0.24	43 (8%)	9	9	26, 42, 102, 150	0

All (43) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	223	LEU	7.0
1	A	10	ARG	6.1
1	B	221	GLY	5.2
1	A	8	LEU	5.2
1	A	9	LEU	4.4
1	A	17	TRP	4.4
1	A	18	GLN	4.0
1	B	220	LYS	4.0
1	B	18	GLN	3.9
1	A	16	ASN	3.6
1	A	245	ARG	3.6
1	A	31	ILE	3.6
1	A	11	PRO	3.4
1	B	7	GLN	3.4
1	B	222	ALA	3.4
1	A	1	MET	3.4
1	A	5	SER	3.3
1	B	217	PHE	3.0
1	B	182	ALA	2.9
1	A	21	GLN	2.9
1	B	16	ASN	2.9
1	A	12	GLN	2.8
1	B	214	LEU	2.8
1	A	7	GLN	2.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	243	ILE	2.7
1	A	4	VAL	2.6
1	B	218	ARG	2.6
1	A	19	ILE	2.5
1	B	224	GLN	2.5
1	A	3	ILE	2.5
1	A	40	HIS	2.5
1	A	6	GLY	2.5
1	A	156	ILE	2.4
1	B	154	LEU	2.4
1	A	23	LEU	2.4
1	B	211	ILE	2.4
1	B	244	VAL	2.2
1	B	17	TRP	2.2
1	B	181	LYS	2.2
1	B	184	TRP	2.1
1	A	149	TRP	2.1
1	A	32	GLU	2.0
1	B	213	LYS	2.0

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
5	GOL	A	1005	6/6	0.70	0.23	65,74,80,84	0
2	EDO	B	1001	4/4	0.84	0.12	52,58,59,62	0
2	EDO	A	1001	4/4	0.87	0.16	48,56,58,60	0
2	EDO	A	1002	4/4	0.90	0.15	52,54,58,59	0

Continued on next page...

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
3	PGE	B	1002	10/10	0.91	0.18	43,51,67,71	0
3	PGE	A	1003[A]	10/10	0.94	0.32	38,46,58,61	10
4	PEG	A	1004	7/7	0.94	0.12	42,49,64,70	0
3	PGE	A	1003[B]	10/10	0.94	0.32	39,45,52,54	10

6.5　Other polymers ⓘ

There are no such residues in this entry.