Full wwPDB X-ray Structure Validation Report ⓘ

May 29, 2020 – 06:02 am BST

PDB ID	:	4UDU
Title	:	Crystal structure of staphylococcal enterotoxin E in complex with a T cell receptor
Authors	:	Rodstrom, K.E.J.; Regenthal, P.; Lindkvist-Petersson, K.
Deposited on	:	2014-12-11
Resolution	:	2.50 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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 ⓘ symbol.

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

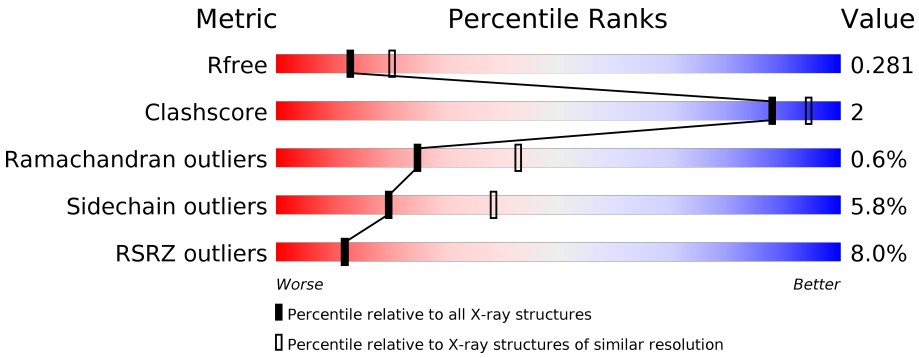
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 2.50 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	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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 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	206	14% 82% 7% 11%
2	B	243	6% 88% 9%
3	C	233	3% 85% 7% 6%

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 4915 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 T CELL RECEPTOR ALPHA CHAIN, T-CELL RECEPTOR ALPHA CHAIN C REGION.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	184	Total	C	N	O	S	0	0	0
			1315	836	218	254	7			

There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	160	CYS	THR	engineered mutation	UNP P01848

- Molecule 2 is a protein called PROTEIN TRBV7-9, T-CELL RECEPTOR BETA-2 CHAIN C REGION.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	238	Total	C	N	O	S	0	0	0
			1848	1166	331	345	6			

There are 22 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	1	MET	-	initiating methionine	UNP A0A5A3
B	98	GLY	-	linker	UNP A0A5A3
B	99	GLY	-	linker	UNP A0A5A3
B	100	TYR	-	linker	UNP A0A5A3
B	101	GLU	-	linker	UNP A0A5A3
B	102	GLN	-	linker	UNP A0A5A3
B	103	TYR	-	linker	UNP A0A5A3
B	104	PHE	-	linker	UNP A0A5A3
B	105	GLY	-	linker	UNP A0A5A3
B	106	PRO	-	linker	UNP A0A5A3
B	107	GLY	-	linker	UNP A0A5A3
B	108	THR	-	linker	UNP A0A5A3
B	109	ARG	-	linker	UNP A0A5A3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	110	LEU	-	linker	UNP A0A5A3
B	111	THR	-	linker	UNP A0A5A3
B	112	VAL	-	linker	UNP A0A5A3
B	113	THR	-	linker	UNP A0A5A3
B	114	GLU	-	linker	UNP A0A5A3
B	128	VAL	GLU	engineered mutation	UNP A0A5B9
B	170	CYS	SER	engineered mutation	UNP A0A5B9
B	188	ALA	CYS	engineered mutation	UNP A0A5B9
B	202	ASP	ASN	engineered mutation	UNP A0A5B9

- Molecule 3 is a protein called ENTEROTOXIN TYPE E.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	218	Total	C	N	O	S	0	0	0
			1699	1084	292	320	3			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	1	Total	Zn	0	0
			1	1		

- Molecule 5 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	C	2	Total	Na	0	0
			2	2		

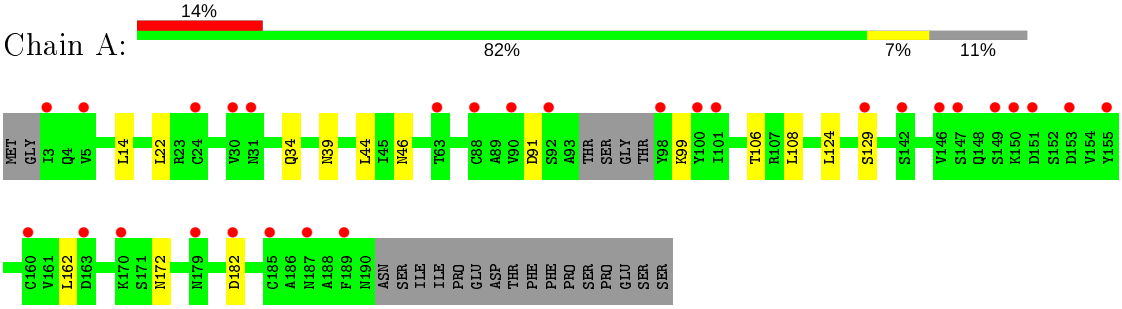
- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	A	9	Total	O	0	0
			9	9		
6	B	15	Total	O	0	0
			15	15		
6	C	26	Total	O	0	0
			26	26		

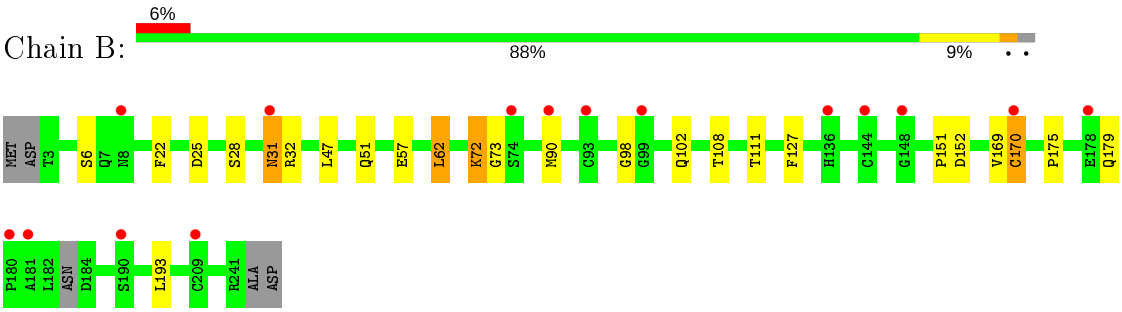
3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of 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.

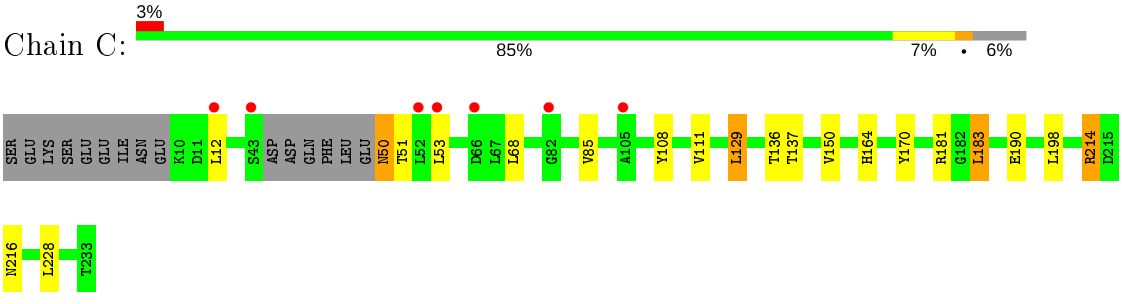
• Molecule 1: T CELL RECEPTOR ALPHA CHAIN, T-CELL RECEPTOR ALPHA CHAIN C REGION



• Molecule 2: PROTEIN TRBV7-9, T-CELL RECEPTOR BETA-2 CHAIN C REGION



• Molecule 3: ENTEROTOXIN TYPE E



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	63.13Å 78.54Å 180.81Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	32.33 – 2.50 47.48 – 2.50	Depositor EDS
% Data completeness (in resolution range)	97.4 (32.33-2.50) 97.5 (47.48-2.50)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.14 (at 2.51Å)	Xtriage
Refinement program	BUSTER 2.10.0	Depositor
R, R_{free}	0.246 , 0.255 0.261 , 0.281	Depositor DCC
R_{free} test set	1567 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å^2)	52.9	Xtriage
Anisotropy	0.563	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 48.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	4915	wwPDB-VP
Average B, all atoms (Å^2)	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.99% 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: NA, ZN

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	0.35	0/1341	0.59	0/1835
2	B	0.33	0/1899	0.57	0/2587
3	C	0.34	0/1735	0.60	0/2349
All	All	0.34	0/4975	0.58	0/6771

There are no bond length outliers.

There are no bond angle outliers.

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	1315	0	1156	4	0
2	B	1848	0	1702	12	0
3	C	1699	0	1608	6	0
4	C	1	0	0	0	0
5	C	2	0	0	0	0
6	A	9	0	0	0	0
6	B	15	0	0	0	0
6	C	26	0	0	0	0
All	All	4915	0	4466	19	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 2.

All (19) 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)
3:C:164:HIS:HD2	3:C:170:TYR:H	1.34	0.75
3:C:111:VAL:H	3:C:216:ASN:HD21	1.51	0.59
2:B:47:LEU:HB3	2:B:62:LEU:HD23	1.92	0.51
1:A:162:LEU:HD12	2:B:170:CYS:HB2	1.92	0.51
1:A:22:LEU:HD12	1:A:106:THR:HG21	1.93	0.51
2:B:152:ASP:HB2	2:B:175:PRO:HG2	1.95	0.48
2:B:22:PHE:CD1	2:B:108:THR:HG21	2.49	0.47
3:C:50:ASN:HD22	3:C:51:THR:HG23	1.79	0.47
2:B:169:VAL:HG22	2:B:193:LEU:HD23	1.98	0.46
2:B:111:THR:HG21	2:B:151:PRO:HB3	1.98	0.45
2:B:31:ASN:HD22	2:B:31:ASN:H	1.65	0.45
1:A:34:GLN:HE22	2:B:102:GLN:HB2	1.83	0.43
2:B:32:ARG:HG2	2:B:51:GLN:HG3	1.99	0.42
2:B:6:SER:HB2	2:B:25:ASP:HB2	2.01	0.42
3:C:12:LEU:HD11	3:C:183:LEU:HG	2.02	0.42
1:A:124:LEU:HB3	2:B:127:PHE:HB3	2.01	0.41
3:C:108:TYR:CD1	3:C:214:ARG:HG2	2.55	0.41
3:C:129:LEU:HD23	3:C:228:LEU:HB3	2.02	0.40
2:B:72:LYS:HA	2:B:73:GLY:HA2	1.61	0.40

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	180/206 (87%)	171 (95%)	7 (4%)	2 (1%)	14 26
2	B	234/243 (96%)	222 (95%)	10 (4%)	2 (1%)	17 31
3	C	214/233 (92%)	204 (95%)	10 (5%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	628/682 (92%)	597 (95%)	27 (4%)	4 (1%)	25 43

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	129	SER
1	A	182	ASP
2	B	72	LYS
2	B	98	GLY

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	125/185 (68%)	117 (94%)	8 (6%)	17 33
2	B	189/213 (89%)	182 (96%)	7 (4%)	34 60
3	C	172/207 (83%)	159 (92%)	13 (8%)	13 25
All	All	486/605 (80%)	458 (94%)	28 (6%)	20 38

All (28) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	14	LEU
1	A	39	ASN
1	A	44	LEU
1	A	46	ASN
1	A	91	ASP
1	A	99	LYS
1	A	108	LEU
1	A	172	ASN
2	B	28	SER
2	B	31	ASN
2	B	57	GLU
2	B	62	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	90	MET
2	B	170	CYS
2	B	179	GLN
3	C	50	ASN
3	C	53	LEU
3	C	68	LEU
3	C	85	VAL
3	C	129	LEU
3	C	136	THR
3	C	137	THR
3	C	150	VAL
3	C	181	ARG
3	C	183	LEU
3	C	190	GLU
3	C	198	LEU
3	C	214	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (11) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	7	GLN
1	A	34	GLN
1	A	39	ASN
2	B	31	ASN
2	B	55	GLN
2	B	81	GLN
3	C	50	ASN
3	C	158	GLN
3	C	164	HIS
3	C	204	GLN
3	C	216	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

There are no carbohydrates in this entry.

5.6 Ligand geometry

Of 3 ligands modelled in this entry, 3 are monoatomic - leaving 0 for Mogul analysis.

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

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, 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	184/206 (89%)	1.13	29 (15%) 2 1	45, 75, 105, 125	6 (3%)
2	B	238/243 (97%)	0.67	15 (6%) 20 21	38, 61, 86, 118	10 (4%)
3	C	218/233 (93%)	0.54	7 (3%) 47 51	36, 60, 87, 111	0
All	All	640/682 (93%)	0.76	51 (7%) 12 12	36, 63, 98, 125	16 (2%)

All (51) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	129	SER	6.3
1	A	149	SER	6.0
1	A	63	THR	5.8
1	A	31	ASN	5.1
2	B	93	CYS	5.1
1	A	24	CYS	5.0
1	A	146	VAL	4.9
3	C	53	LEU	4.8
2	B	74	SER	4.4
1	A	88	CYS	4.1
2	B	209	CYS	3.8
1	A	3	ILE	3.8
3	C	43	SER	3.8
2	B	170	CYS	3.7
2	B	190	SER	3.6
1	A	98	TYR	3.6
3	C	82	GLY	3.5
2	B	144	CYS	3.3
3	C	52	LEU	3.0
1	A	179	ASN	3.0
1	A	151	ASP	2.9
1	A	90	VAL	2.9
1	A	160	CYS	2.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	182	ASP	2.8
1	A	5	VAL	2.8
1	A	187	ASN	2.7
1	A	185	CYS	2.7
1	A	100	TYR	2.7
3	C	66	ASP	2.7
1	A	142	SER	2.6
1	A	147	SER	2.6
2	B	8	ASN	2.6
1	A	170	LYS	2.5
2	B	178	GLU	2.5
1	A	101	ILE	2.5
2	B	90	MET	2.5
1	A	92	SER	2.4
2	B	136	HIS	2.4
2	B	181	ALA	2.4
2	B	99	GLY	2.4
2	B	31	ASN	2.3
1	A	155	TYR	2.3
1	A	30	VAL	2.3
1	A	153	ASP	2.2
1	A	163	ASP	2.2
3	C	12	LEU	2.2
1	A	189	PHE	2.1
2	B	180	PRO	2.1
2	B	148	GLY	2.1
3	C	105	ALA	2.0
1	A	150	LYS	2.0

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. 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
4	ZN	C	1234	1/1	0.91	0.18	69,69,69,69	0
5	NA	C	1236	1/1	0.94	0.16	59,59,59,59	0
5	NA	C	1235	1/1	0.96	0.20	47,47,47,47	0

6.5 Other polymers (i)

There are no such residues in this entry.