WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

May 29, 2020 – 06:27 am BST

PDB ID	:	4YGA
Title	:	CDPK1, from Toxoplasma gondii, bound to inhibitory VHH-1B7
Authors	:	Knockenhauer, K.E.; Schwartz, T.U.
Deposited on	:	2015-02-26
Resolution	:	2.94 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 following versions of software and data (see references (1)) were used in the production of this report:

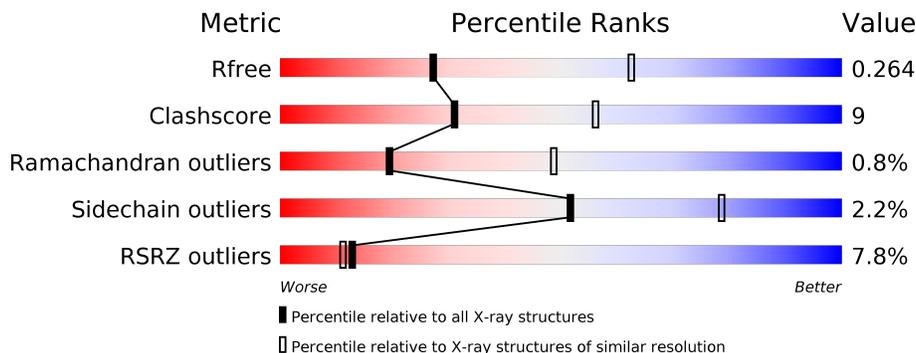|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.94 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		0.264
Clashscore		9
Ramachandran outliers		0.8%
Sidechain outliers		2.2%
RSRZ outliers		7.8%

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	2969 (2.98-2.90)
Clashscore	141614	3218 (2.98-2.90)
Ramachandran outliers	138981	3122 (2.98-2.90)
Sidechain outliers	138945	3124 (2.98-2.90)
RSRZ outliers	127900	2902 (2.98-2.90)

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	516	6% / 69% · 16% · 13%
1	C	516	8% / 70% · 16% · 13%
1	E	516	7% / 70% · 15% · 14%
1	G	516	6% / 67% · 19% · 13%
2	B	141	15% / 70% · 18% · 12%
2	D	141	7% / 64% · 23% · 13%

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	141	4% ▮ 67% ▮ 20% ▮ • 13%
2	H	141	3% ▮ 56% ▮ 30% ▮ • 13%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 17348 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 Calmodulin-domain protein kinase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	448	Total 3419	C 2177	N 559	O 667	S 16	0	0	0
1	C	449	Total 3404	C 2170	N 553	O 664	S 17	0	0	0
1	E	442	Total 3395	C 2163	N 560	O 656	S 16	0	0	0
1	G	447	Total 3409	C 2170	N 558	O 664	S 17	0	0	0

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP Q9BJF5
A	508	LEU	-	expression tag	UNP Q9BJF5
A	509	GLU	-	expression tag	UNP Q9BJF5
A	510	HIS	-	expression tag	UNP Q9BJF5
A	511	HIS	-	expression tag	UNP Q9BJF5
A	512	HIS	-	expression tag	UNP Q9BJF5
A	513	HIS	-	expression tag	UNP Q9BJF5
A	514	HIS	-	expression tag	UNP Q9BJF5
A	515	HIS	-	expression tag	UNP Q9BJF5
C	0	MET	-	initiating methionine	UNP Q9BJF5
C	508	LEU	-	expression tag	UNP Q9BJF5
C	509	GLU	-	expression tag	UNP Q9BJF5
C	510	HIS	-	expression tag	UNP Q9BJF5
C	511	HIS	-	expression tag	UNP Q9BJF5
C	512	HIS	-	expression tag	UNP Q9BJF5
C	513	HIS	-	expression tag	UNP Q9BJF5
C	514	HIS	-	expression tag	UNP Q9BJF5
C	515	HIS	-	expression tag	UNP Q9BJF5
E	0	MET	-	initiating methionine	UNP Q9BJF5
E	508	LEU	-	expression tag	UNP Q9BJF5
E	509	GLU	-	expression tag	UNP Q9BJF5

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	510	HIS	-	expression tag	UNP Q9BJF5
E	511	HIS	-	expression tag	UNP Q9BJF5
E	512	HIS	-	expression tag	UNP Q9BJF5
E	513	HIS	-	expression tag	UNP Q9BJF5
E	514	HIS	-	expression tag	UNP Q9BJF5
E	515	HIS	-	expression tag	UNP Q9BJF5
G	0	MET	-	initiating methionine	UNP Q9BJF5
G	508	LEU	-	expression tag	UNP Q9BJF5
G	509	GLU	-	expression tag	UNP Q9BJF5
G	510	HIS	-	expression tag	UNP Q9BJF5
G	511	HIS	-	expression tag	UNP Q9BJF5
G	512	HIS	-	expression tag	UNP Q9BJF5
G	513	HIS	-	expression tag	UNP Q9BJF5
G	514	HIS	-	expression tag	UNP Q9BJF5
G	515	HIS	-	expression tag	UNP Q9BJF5

- Molecule 2 is a protein called VHH-1B7.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	124	Total 924	C 573	N 160	O 186	S 5	0	0	0
2	D	123	Total 932	C 581	N 163	O 183	S 5	0	0	0
2	F	123	Total 925	C 573	N 162	O 185	S 5	0	0	0
2	H	123	Total 924	C 576	N 162	O 181	S 5	0	0	0

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	G	4	Total 4	Ca 4	0	0
3	A	4	Total 4	Ca 4	0	0
3	C	4	Total 4	Ca 4	0	0
3	E	4	Total 4	Ca 4	0	0

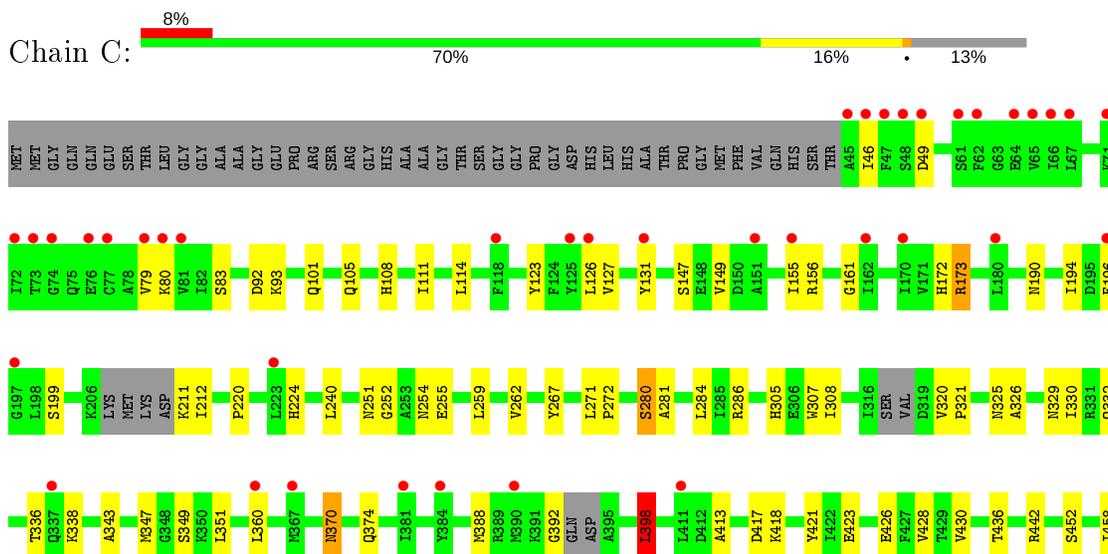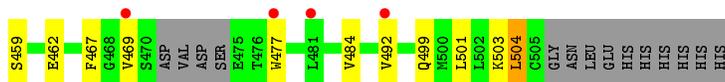# 3 Residue-property plots ⓘ

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.

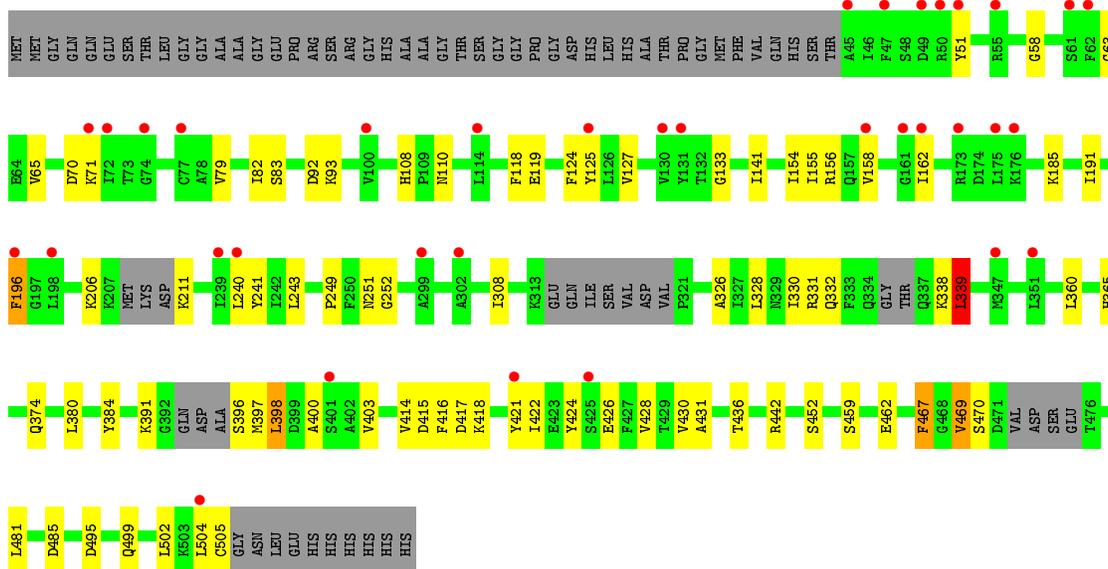- Molecule 1: Calmodulin-domain protein kinase 1
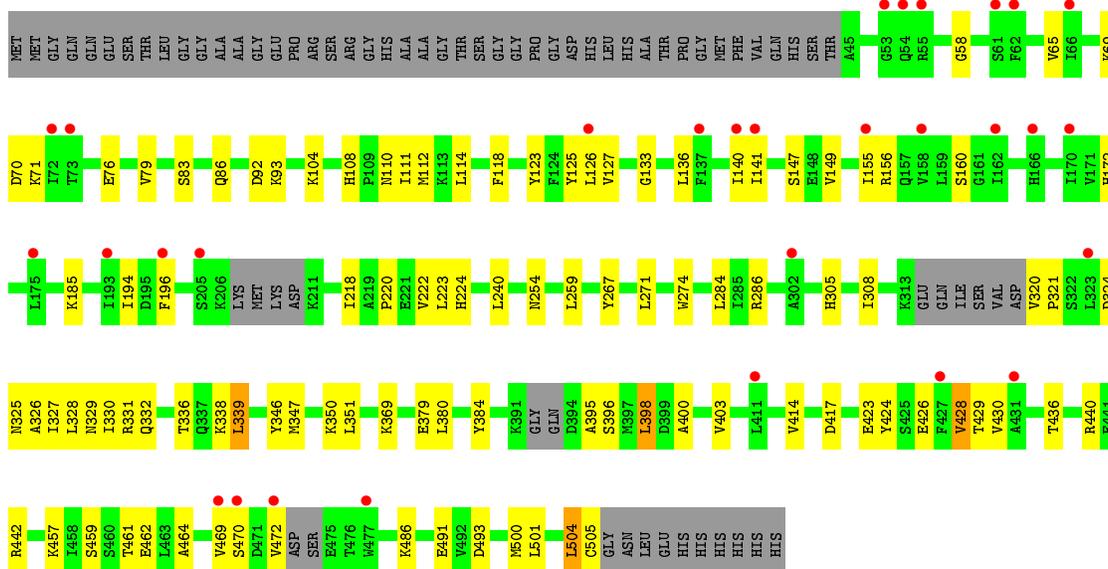

- Molecule 1: Calmodulin-domain protein kinase 1

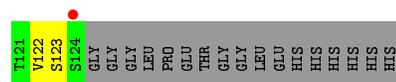• Molecule 1: Calmodulin-domain protein kinase 1
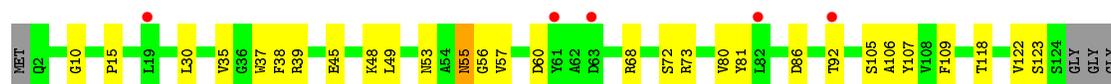
Chain E:

• Molecule 1: Calmodulin-domain protein kinase 1

Chain G:

• Molecule 2: VHH-1B7

Chain B:

M1 T8 G9 G10 G11 L12 P15 L19 R20 L21 A25 L30 D31 H32 W37 F38 R39 K48 L49 N53 A54 N55 G56 Y61 I65 S72 R73 N78 Y81 L82 Q83 M84 D91 T92 A93 S105 A106 Y107 V108 F109 L110 T118 Q119 V120

T121 V122 S123 S124 GLY GLY GLY LEU PRO GLU THR GLY GLY LEU GLU HIS HIS HIS HIS HIS HIS

● **Molecule 2: VHH-1B7**

Chain D:

7%

64%   23%   13%

MET Q2 V3 P15 G27 F28 T29 L30 A34 W37 F38 R39 L49 L50 C51 I52 N53 A54 N55 G56 L59 I65 K66 G67 R68 I71 S72 R73 Y81 D86 L87 K88 P89 T92 Y95 S96 C104 S105 A106 Y107 V108 F109 L110 V120 T121 V122

S123 S124 GLY GLY GLY LEU PRO GLU THR GLY GLY LEU GLU HIS HIS HIS HIS HIS HIS

● **Molecule 2: VHH-1B7**

Chain F:

4%

67%   20%   13%

MET Q2 G10 P15 L19 L30 V35 G36 W37 F38 R39 E45 K48 L49 N53 A54 N55 G56 V57 D60 Y61 A62 D63 R68 S72 R73 V80 Y81 L82 D86 T92 S105 A106 Y107 V108 F109 T118 V122 S123 S124 GLY GLY GLY

LEU PRO GLU THR GLY GLY LEU GLU HIS HIS HIS HIS HIS HIS

● **Molecule 2: VHH-1B7**

Chain H:

3%

56%   30%   13%

MET Q2 V3 V13 Q14 P15 L19 F28 T29 L30 S33 W37 F38 R39 L49 L50 C51 I52 N53 A54 N55 G56 L59 A62 D63 S64 I65 K66 G67 R68 I71 S72 R73 V80 Y81 M84 N85 D86 L87 K88 P89 T92 A93 T94 A98 C104

S105 A106 L110 Y111 Q119 V122 S123 S124 GLY GLY GLY LEU PRO GLU THR GLY GLY LEU GLU HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	76.92Å　91.07Å　106.64Å 88.71°　108.27°　90.26°	Depositor
Resolution (Å)	70.67　–　2.94 70.67　–　2.94	Depositor EDS
% Data completeness (in resolution range)	92.3 (70.67-2.94) 86.4 (70.67-2.94)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$<I/\sigma(I)>$ [1]	1.97 (at 2.96Å)	Xtriage
Refinement program	PHENIX 1.9_1692	Depositor
R, R_{free}	0.226　,　0.262 0.228　,　0.264	Depositor DCC
R_{free} test set	1718 reflections (3.18%)	wwPDB-VP
Wilson B-factor (Å^2)	86.0	Xtriage
Anisotropy	0.172	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 101.6	EDS
L-test for twinning[2]	$<|L|> = 0.49, <L^2> = 0.32$	Xtriage
Estimated twinning fraction	0.287 for -h,k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	17348	wwPDB-VP
Average B, all atoms (Å^2)	113.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.59% 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: CA

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	0.23	0/3476	0.42	0/4702
1	C	0.23	0/3460	0.44	1/4682 (0.0%)
1	E	0.24	0/3449	0.46	2/4657 (0.0%)
1	G	0.23	0/3465	0.43	1/4689 (0.0%)
2	B	0.24	0/941	0.43	0/1279
2	D	0.23	0/949	0.46	0/1287
2	F	0.24	0/942	0.44	0/1279
2	H	0.24	0/941	0.47	0/1278
All	All	0.24	0/17623	0.44	4/23853 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	E	398	LEU	CA-CB-CG	6.25	129.67	115.30
1	C	398	LEU	CA-CB-CG	6.00	129.09	115.30
1	E	339	LEU	CA-CB-CG	5.83	128.72	115.30
1	G	339	LEU	CA-CB-CG	5.07	126.95	115.30

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	3419	0	3233	57	0
1	C	3404	0	3200	58	0
1	E	3395	0	3238	47	0
1	G	3409	0	3214	68	0
2	B	924	0	853	15	0
2	D	932	0	886	24	0
2	F	925	0	859	17	0
2	H	924	0	871	29	0
3	A	4	0	0	0	0
3	C	4	0	0	0	0
3	E	4	0	0	0	0
3	G	4	0	0	0	0
All	All	17348	0	16354	305	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 9.

The worst 5 of 305 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:467:PHE:HB2	2:B:109:PHE:HB2	1.42	0.99
1:C:332:GLN:HE22	1:C:504:LEU:H	1.30	0.79
1:C:155:ILE:HG13	1:C:240:LEU:HD13	1.68	0.76
2:H:30:LEU:O	2:H:73:ARG:NH2	2.19	0.74
1:E:467:PHE:HB2	2:F:109:PHE:HB2	1.68	0.74

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	440/516 (85%)	416 (94%)	19 (4%)	5 (1%)	14	40

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	439/516 (85%)	420 (96%)	16 (4%)	3 (1%)	22	52
1	E	430/516 (83%)	415 (96%)	14 (3%)	1 (0%)	47	76
1	G	437/516 (85%)	416 (95%)	18 (4%)	3 (1%)	22	52
2	B	122/141 (86%)	114 (93%)	7 (6%)	1 (1%)	19	49
2	D	121/141 (86%)	112 (93%)	8 (7%)	1 (1%)	19	49
2	F	121/141 (86%)	114 (94%)	5 (4%)	2 (2%)	9	29
2	H	121/141 (86%)	110 (91%)	9 (7%)	2 (2%)	9	29
All	All	2231/2628 (85%)	2117 (95%)	96 (4%)	18 (1%)	19	49

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	334	GLN
1	A	336	THR
1	A	337	GLN
1	A	473	ASP
1	C	503	LYS

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	350/448 (78%)	345 (99%)	5 (1%)	67	86
1	C	345/448 (77%)	337 (98%)	8 (2%)	50	78
1	E	349/448 (78%)	340 (97%)	9 (3%)	46	75
1	G	348/448 (78%)	342 (98%)	6 (2%)	60	83
2	B	96/116 (83%)	93 (97%)	3 (3%)	40	71
2	D	99/116 (85%)	97 (98%)	2 (2%)	55	80
2	F	97/116 (84%)	94 (97%)	3 (3%)	40	71
2	H	97/116 (84%)	93 (96%)	4 (4%)	30	61
All	All	1781/2256 (79%)	1741 (98%)	40 (2%)	52	78

5 of 40 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	83	SER
1	E	365	HIS
2	H	19	LEU
1	E	328	LEU
1	E	398	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	332	GLN
2	H	55	ASN
2	D	55	ASN
2	B	55	ASN
2	F	55	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)

Of 16 ligands modelled in this entry, 16 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 (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(Å²)	Q<0.9
1	A	448/516 (86%)	0.32	30 (6%)	17	16	60, 107, 192, 296	1 (0%)
1	C	449/516 (87%)	0.45	43 (9%)	8	7	59, 104, 198, 345	1 (0%)
1	E	442/516 (85%)	0.41	35 (7%)	12	11	59, 106, 185, 287	1 (0%)
1	G	447/516 (86%)	0.33	30 (6%)	17	16	60, 104, 182, 289	1 (0%)
2	B	124/141 (87%)	0.72	21 (16%)	1	1	65, 102, 171, 199	0
2	D	123/141 (87%)	0.50	10 (8%)	12	10	64, 109, 192, 251	0
2	F	123/141 (87%)	0.33	5 (4%)	37	36	65, 88, 184, 267	0
2	H	123/141 (87%)	0.28	4 (3%)	46	45	62, 99, 159, 219	0
All	All	2279/2628 (86%)	0.39	178 (7%)	13	11	59, 105, 189, 345	4 (0%)

The worst 5 of 178 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	72	ILE	12.7
1	C	73	THR	7.8
1	E	72	ILE	7.6
1	A	72	ILE	7.4
1	C	131	TYR	7.1

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, 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
3	CA	C	601	1/1	0.82	0.14	96,96,96,96	0
3	CA	C	603	1/1	0.89	0.18	89,89,89,89	0
3	CA	G	604	1/1	0.91	0.15	90,90,90,90	0
3	CA	E	603	1/1	0.91	0.17	80,80,80,80	0
3	CA	A	604	1/1	0.96	0.15	94,94,94,94	0
3	CA	A	603	1/1	0.96	0.11	80,80,80,80	0
3	CA	G	603	1/1	0.96	0.20	79,79,79,79	0
3	CA	G	602	1/1	0.97	0.20	75,75,75,75	0
3	CA	C	604	1/1	0.97	0.12	82,82,82,82	0
3	CA	E	604	1/1	0.98	0.23	82,82,82,82	0
3	CA	G	601	1/1	0.98	0.15	100,100,100,100	0
3	CA	C	602	1/1	0.98	0.13	81,81,81,81	0
3	CA	E	602	1/1	0.98	0.17	75,75,75,75	0
3	CA	E	601	1/1	0.98	0.18	79,79,79,79	0
3	CA	A	602	1/1	0.99	0.21	81,81,81,81	0
3	CA	A	601	1/1	0.99	0.20	80,80,80,80	0

6.5 Other polymers (i)

There are no such residues in this entry.