wwPDB X-ray Structure Validation Summary Report (i)

Jan 31, 2016 – 11:52 PM GMT

PDB ID	:	1YNT
Title	:	Structure of the immunodominant epitope displayed by the surface antigen 1 (SAG1) of Toxoplasma gondii complexed to a monoclonal antibody
Authors	:	Graille, M.; Stura, E.A.; Bossus, M.; Muller, B.H.; Letourneur, O.; Battail-Poirot, N; Sibai, G.; Rolland, D.; Le Du, M.H.; Ducancel, F.
Deposited on	:	2005-01-25
Resolution	:	3.10 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
http://wwpdb.org/validation/2016/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:

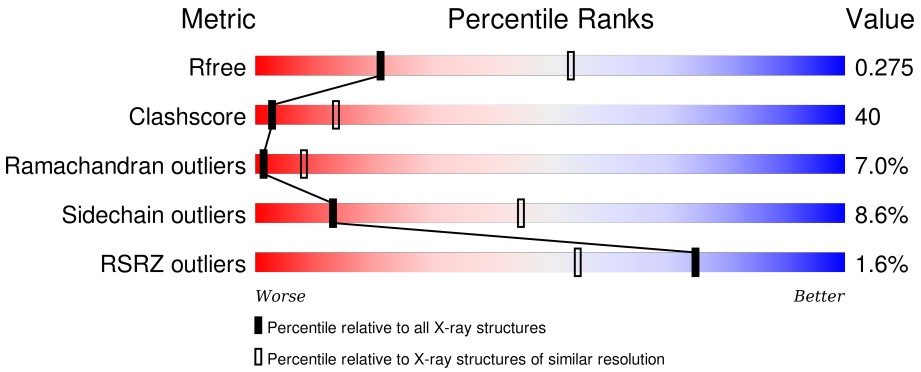
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.10 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.



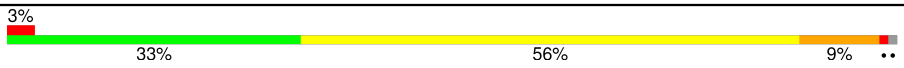
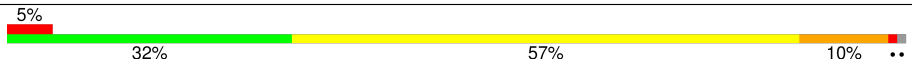
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1114 (3.14-3.06)
Clashscore	102246	1222 (3.14-3.06)
Ramachandran outliers	100387	1174 (3.14-3.06)
Sidechain outliers	100360	1174 (3.14-3.06)
RSRZ outliers	91569	1119 (3.14-3.06)

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	213	% 34% 59% 8%
1	C	213	32% 61% 8%
2	B	218	45% 44% 10%
2	D	218	47% 42% 10%
3	E	61	31% 64% 5%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	F	254	
4	G	254	

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 10802 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 4F11E12 Fab variable light chain region.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	213	Total	C	N	O	S	0	0	0
			1658	1027	280	346	5			
1	C	213	Total	C	N	O	S	0	0	0
			1658	1027	280	346	5			

- Molecule 2 is a protein called 4F11E12 Fab variable heavy chain region.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	218	Total	C	N	O	S	0	0	0
			1657	1048	269	332	8			
2	D	218	Total	C	N	O	S	0	0	0
			1657	1048	269	332	8			

- Molecule 3 is a protein called protein L.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	61	Total	C	N	O	S	0	0	0
			476	302	76	97	1			

- Molecule 4 is a protein called Major surface antigen p30.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	F	252	Total	C	N	O	S	0	0	0
			1847	1151	308	374	14			
4	G	252	Total	C	N	O	S	0	0	0
			1847	1151	308	374	14			

- Molecule 5 is CADMIUM ION (three-letter code: CD) (formula: Cd).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	B	1	Total	Cd	0	0
			1	1		

Continued on next page...

Continued from previous page...

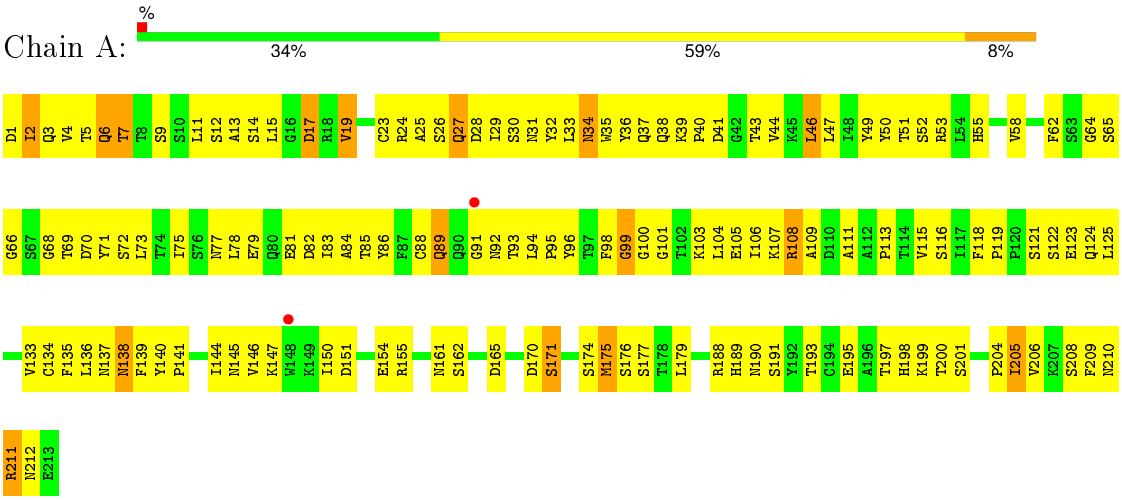
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cd		
5	D	1	1	1	0	0

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.

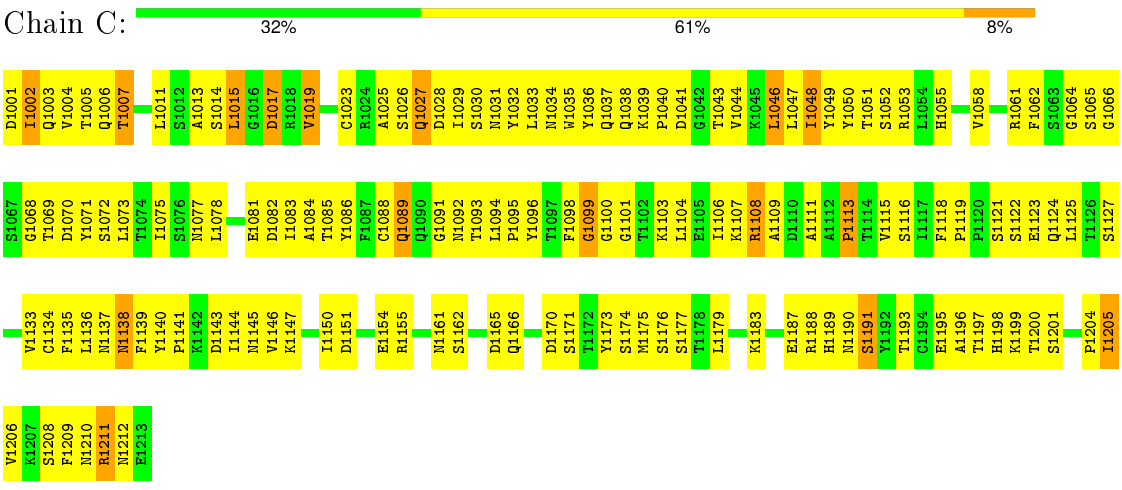
- Molecule 1: 4F11E12 Fab variable light chain region

Chain A:



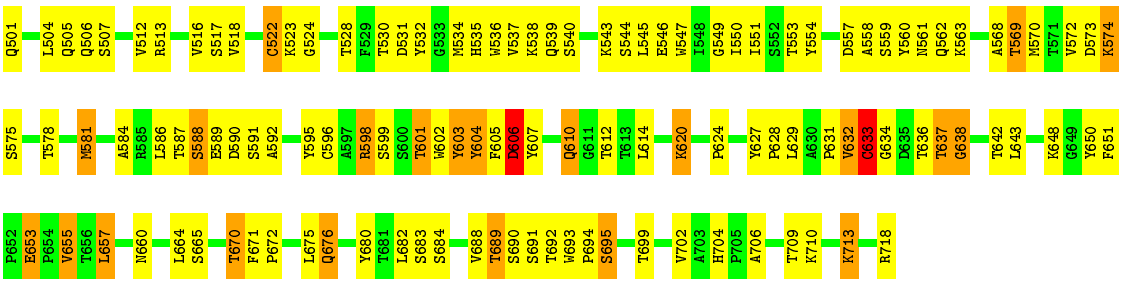
- Molecule 1: 4F11E12 Fab variable light chain region

Chain C:

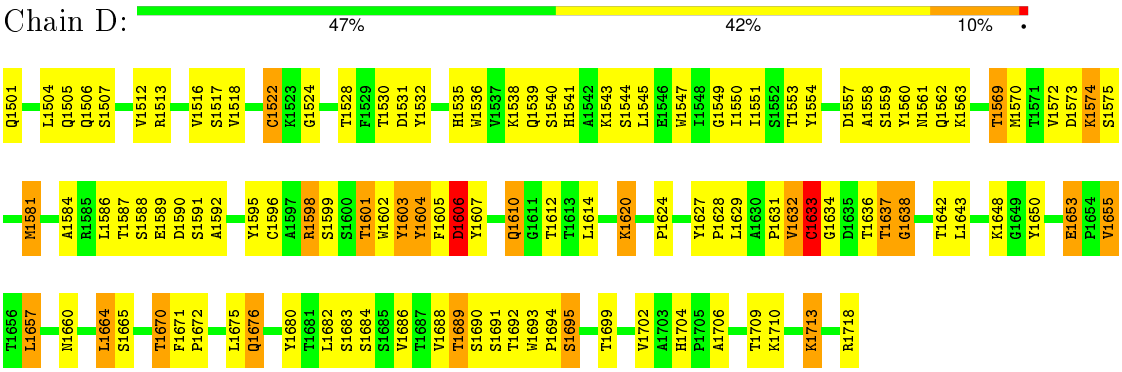


- Molecule 2: 4F11E12 Fab variable heavy chain region

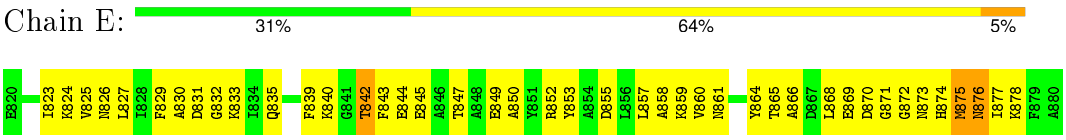
Chain B:



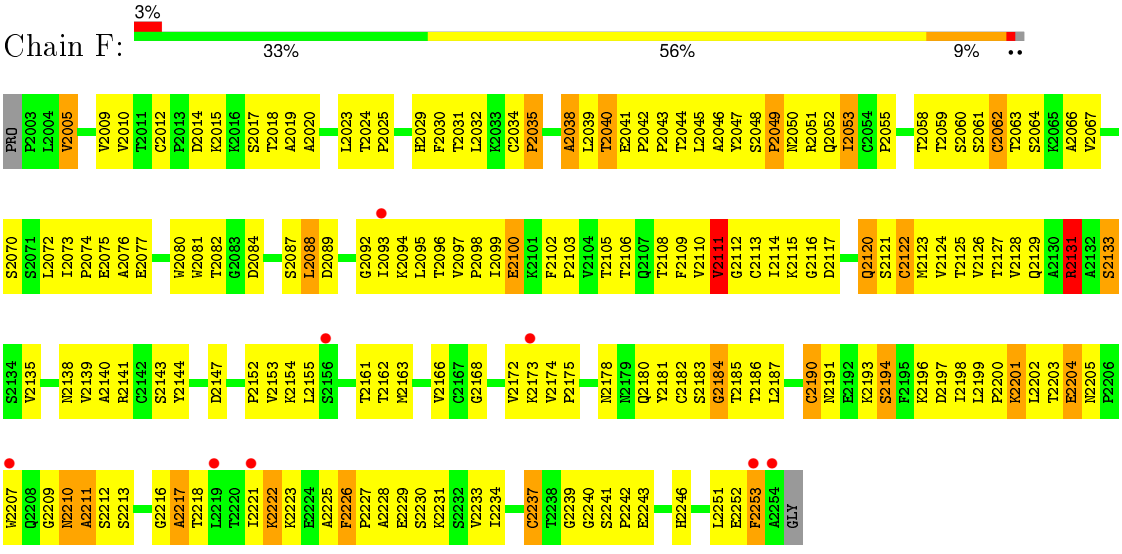
• Molecule 2: 4F11E12 Fab variable heavy chain region



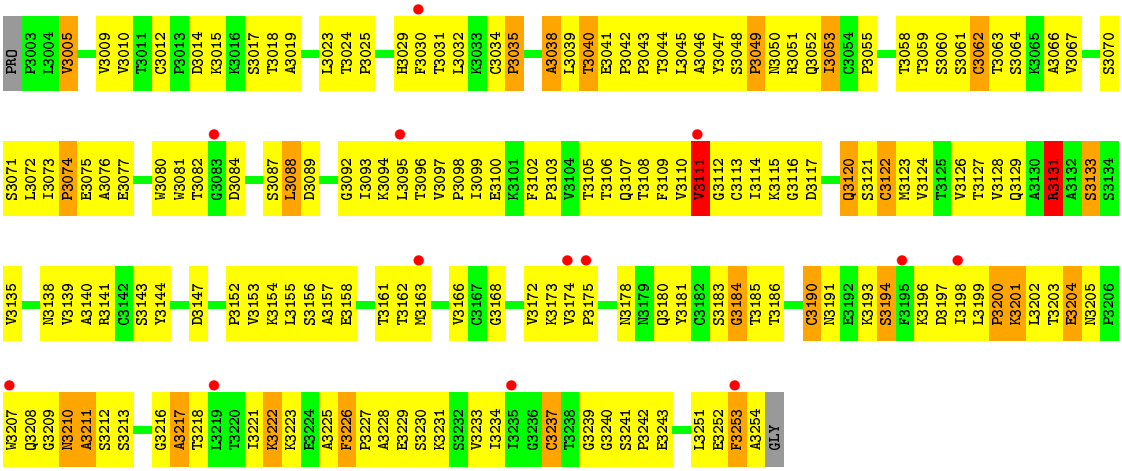
• Molecule 3: protein L



• Molecule 4: Major surface antigen p30



• Molecule 4: Major surface antigen p30


PRO P3003 L3004 V3005 V3009 V3010 T3011 C3012 P3013 D3014 K3015 K3016 S3017 T3018 A3019 L3023 T3024 P3025 H3029 F3030 T3031 L3032 K3033 C3034 P3035 A3038 L3039 T3040 E3041 P3042 P3043 T3044 L3045 A3046 Y3047 S3048 P3049 N3050 R3051 Q3052 I3053 C3054 P3055 T3058 T3059 S3060 S3061 C3062 T3063 S3064 K3065 A3066 V3067 S3070
S3071 L3072 I3073 P3074 E3075 A3076 E3077 W3080 W3081 T3082 G3083 D3084 S3087 L3088 D3089 G3092 I3093 K3094 L3095 T3096 V3097 P3098 I3099 E3100 K3101 F3102 P3103 V3104 T3105 T3106 Q3107 T3108 F3109 V3110 V3111 G3112 C3113 I3114 K3115 G3116 D3117 Q3120 S3121 C3122 M3123 V3124 T3125 V3126 T3127 V3128 Q3129 A3130 R3131 A3132 S3133 S3134
V3135 N3138 V3139 A3140 R3141 C3142 S3143 Y3144 D3147 P3152 V3153 K3154 L3155 S3156 A3157 E3158 T3161 T3162 M3163 V3166 C3167 G3168 V3172 K3173 V3174 P3175 N3178 N3179 Q3180 Y3181 C3182 S3183 G3184 T3185 T3186 C3190 N3191 E3192 K3193 S3194 F3195 K3196 D3197 I3198 L3199 P3200 K3201 L3202 T3203 E3204 N3205 P3206
W3207 Q3208 G3209 N3210 A3211 S3212 S3213 G3216 A3217 T3218 L3219 T3220 I3221 K3222 K3223 E3224 A3225 F3226 P3227 A3228 E3229 S3230 K3231 S3232 V3233 I3234 I3235 G3236 C3237 T3238 G3239 G3240 S3241 P3242 E3243 L3251 E3252 F3253 A3254 GLY

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	71.03Å 198.28Å 128.37Å 90.00° 89.97° 90.00°	Depositor
Resolution (Å)	20.00 – 3.10 39.29 – 3.08	Depositor EDS
% Data completeness (in resolution range)	96.7 (20.00-3.10) 95.1 (39.29-3.08)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.57 (at 3.06Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.241 , 0.283 0.237 , 0.275	Depositor DCC
R_{free} test set	1534 reflections (4.95%)	DCC
Wilson B-factor (Å^2)	71.2	Xtriage
Anisotropy	0.335	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 34.4	EDS
Estimated twinning fraction	0.447 for h,-k,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Outliers	2 of 32176 reflections (0.006%)	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	10802	wwPDB-VP
Average B, all atoms (Å^2)	73.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.06% 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

5.1 Standard geometry

Bond lengths and bond angles in the following residue types are not validated in this section: CD

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	0.48	0/1692	0.74	0/2297
1	C	0.49	0/1692	0.73	0/2297
2	B	0.51	0/1700	0.76	1/2318 (0.0%)
2	D	0.51	0/1700	0.77	1/2318 (0.0%)
3	E	0.46	0/483	0.65	0/649
4	F	0.32	0/1882	0.63	0/2568
4	G	0.32	0/1882	0.63	0/2568
All	All	0.44	0/11031	0.71	2/15015 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	653	GLU	N-CA-C	5.54	125.96	111.00
2	D	1653	GLU	N-CA-C	5.41	125.60	111.00

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	1658	0	1583	137	0
1	C	1658	0	1580	143	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	1657	0	1607	118	0
2	D	1657	0	1607	115	0
3	E	476	0	456	40	0
4	F	1847	0	1844	168	0
4	G	1847	0	1844	176	0
5	B	1	0	0	0	0
5	D	1	0	0	0	0
All	All	10802	0	10521	850	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 40.

The worst 5 of 850 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)
4:G:3133:SER:HB3	4:G:3143:SER:H	1.14	1.11
4:F:2133:SER:HB3	4:F:2143:SER:H	1.14	1.08
2:D:1624:PRO:HB3	2:D:1650:TYR:HB3	1.35	1.05
2:B:624:PRO:HB3	2:B:650:TYR:HB3	1.36	1.02
1:C:1002:ILE:O	1:C:1002:ILE:HG12	1.59	1.01

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	211/213 (99%)	165 (78%)	40 (19%)	6 (3%)	6 30
1	C	211/213 (99%)	166 (79%)	37 (18%)	8 (4%)	4 22
2	B	216/218 (99%)	177 (82%)	26 (12%)	13 (6%)	2 11
2	D	216/218 (99%)	177 (82%)	26 (12%)	13 (6%)	2 11

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
3	E	59/61 (97%)	44 (75%)	9 (15%)	6 (10%)	1 4
4	F	250/254 (98%)	177 (71%)	47 (19%)	26 (10%)	1 3
4	G	250/254 (98%)	173 (69%)	50 (20%)	27 (11%)	0 3
All	All	1413/1431 (99%)	1079 (76%)	235 (17%)	99 (7%)	1 8

5 of 99 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	17	ASP
1	A	27	GLN
2	B	575	SER
2	B	633	CYS
1	C	1017	ASP

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	191/191 (100%)	172 (90%)	19 (10%)	10 34
1	C	191/191 (100%)	174 (91%)	17 (9%)	12 42
2	B	188/188 (100%)	167 (89%)	21 (11%)	7 29
2	D	188/188 (100%)	167 (89%)	21 (11%)	7 29
3	E	47/47 (100%)	43 (92%)	4 (8%)	13 45
4	F	214/215 (100%)	202 (94%)	12 (6%)	26 62
4	G	214/215 (100%)	202 (94%)	12 (6%)	26 62
All	All	1233/1235 (100%)	1127 (91%)	106 (9%)	13 45

5 of 106 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	1089	GLN
2	D	1569	THR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
4	G	3062	CYS
1	C	1108	ARG
1	C	1165	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 31 such sidechains are listed below:

Mol	Chain	Res	Type
2	D	1539	GLN
3	E	861	ASN
4	G	3138	ASN
2	D	1660	ASN
3	E	876	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 2 ligands modelled in this entry, 2 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	213/213 (100%)	-0.22	2 (0%) 85 72	24, 54, 77, 89	0
1	C	213/213 (100%)	-0.27	0 100 100	23, 54, 76, 89	0
2	B	218/218 (100%)	-0.29	0 100 100	16, 50, 78, 102	0
2	D	218/218 (100%)	-0.26	0 100 100	16, 50, 78, 102	0
3	E	61/61 (100%)	0.11	0 100 100	48, 77, 95, 103	0
4	F	252/254 (99%)	0.12	8 (3%) 51 27	67, 106, 142, 150	0
4	G	252/254 (99%)	0.21	13 (5%) 31 13	67, 106, 139, 150	0
All	All	1427/1431 (99%)	-0.09	23 (1%) 74 55	16, 68, 134, 150	0

The worst 5 of 23 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	G	3253	PHE	4.4
4	F	2156	SER	4.4
4	G	3174	VAL	3.9
4	G	3235	ILE	3.7
4	F	2221	ILE	3.6

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 carbohydrates 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. 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, 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	LLDF	B-factors(Å^2)	Q<0.9
5	CD	B	4001	1/1	0.95	0.10	-	111,111,111,111	0
5	CD	D	4002	1/1	0.98	0.10	-	115,115,115,115	0

6.5 Other polymers

There are no such residues in this entry.