wwPDB X-ray Structure Validation Summary Report (i)

Jun 22, 2024 – 12:09 PM EDT

PDB ID	:	4XA9
Title	:	Crystal structure of the complex between the N-terminal domain of RavJ and LegL1 from Legionella pneumophila str. Philadelphia
Authors	:	Stogios, P.J.; Cuff, M.E.; Nocek, B.; Evdokimova, E.; Di Leo, R.; Yim, V.; Savchenko, A.; Joachimiak, A.; Midwest Center for Structural Genomics (MCSG)
Deposited on	:	2014-12-13
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:

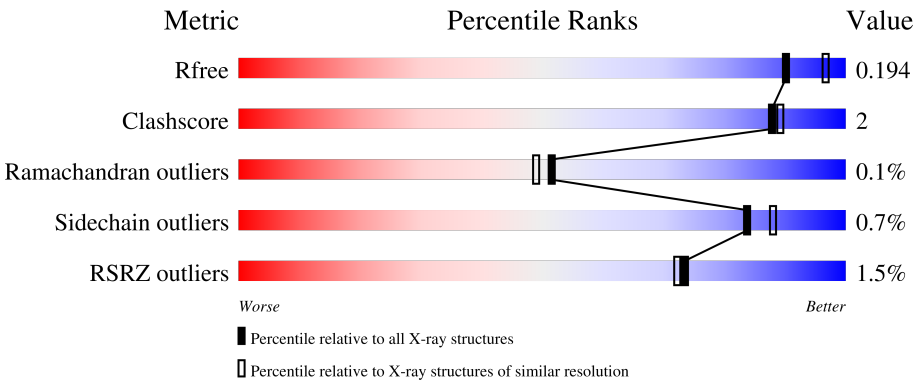
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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}	130704	8085 (2.00-2.00)
Clashscore	141614	9178 (2.00-2.00)
Ramachandran outliers	138981	9054 (2.00-2.00)
Sidechain outliers	138945	9053 (2.00-2.00)
RSRZ outliers	127900	7900 (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	297	81% • 17%
1	B	297	77% 6% 17%
1	C	297	80% • 17%
1	D	297	81% • 17%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	E	297	79% 17%
1	F	297	79% 17%
1	G	297	82% 16%
1	H	297	81% 17%
2	a	231	4% 97%
2	b	231	4% 97%
2	c	231	4% 96%
2	d	231	2% 97%
2	e	231	2% 97%
2	f	231	2% 97%
2	g	231	2% 97%
2	h	231	2% 98%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 34121 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 Gala protein type 1, 3 or 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	247	1916	1209	319	384	4	0	10	0
1	B	247	1907	1200	317	386	4	0	8	0
1	C	247	1889	1189	316	379	5	0	3	0
1	D	247	1897	1195	316	382	4	0	6	0
1	E	247	1898	1192	317	385	4	0	6	0
1	F	247	1904	1198	319	383	4	0	6	0
1	G	248	1900	1195	319	382	4	0	4	0
1	H	247	1894	1188	318	384	4	0	5	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	GLY	-	expression tag	UNP Q5ZWY8
B	0	GLY	-	expression tag	UNP Q5ZWY8
C	0	GLY	-	expression tag	UNP Q5ZWY8
D	0	GLY	-	expression tag	UNP Q5ZWY8
E	0	GLY	-	expression tag	UNP Q5ZWY8
F	0	GLY	-	expression tag	UNP Q5ZWY8
G	0	GLY	-	expression tag	UNP Q5ZWY8
H	0	GLY	-	expression tag	UNP Q5ZWY8

- Molecule 2 is a protein called Uncharacterized protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
2	a	228	1796	1139	312	340	2	3	0	3	0
2	b	227	1779	1128	310	337	2	2	0	1	0
2	c	224	1779	1132	309	334	2	2	0	3	0
2	d	228	1808	1149	313	341	2	3	0	5	0
2	e	227	1787	1134	311	337	2	3	0	2	0
2	f	226	1810	1151	314	340	2	3	0	7	0
2	g	225	1782	1132	310	335	2	3	0	4	0
2	h	229	1794	1137	312	340	2	3	0	2	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
a	0	GLY	-	expression tag	UNP Q5ZWY9
b	0	GLY	-	expression tag	UNP Q5ZWY9
c	0	GLY	-	expression tag	UNP Q5ZWY9
d	0	GLY	-	expression tag	UNP Q5ZWY9
e	0	GLY	-	expression tag	UNP Q5ZWY9
f	0	GLY	-	expression tag	UNP Q5ZWY9
g	0	GLY	-	expression tag	UNP Q5ZWY9
h	0	GLY	-	expression tag	UNP Q5ZWY9

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	284	287	287	0	3
3	a	230	234	234	0	4
3	B	288	295	295	0	7
3	b	243	248	248	0	5
3	C	326	332	332	0	6
3	c	256	261	261	0	5

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	D	300	306	306	0	6
			Total	O		
3	d	249	254	254	0	5
			Total	O		
3	E	288	294	294	0	6
			Total	O		
3	e	306	311	311	0	5
			Total	O		
3	F	297	301	301	0	4
			Total	O		
3	f	338	342	342	0	4
			Total	O		
3	G	293	295	295	0	2
			Total	O		
3	g	273	274	274	0	1
			Total	O		
3	H	260	262	262	0	2
			Total	O		
3	h	281	285	285	0	4

3 Residue-property plots (i)

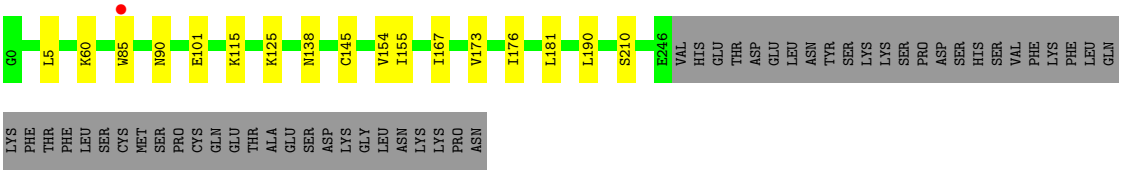
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Gala protein type 1, 3 or 4

Chain A:

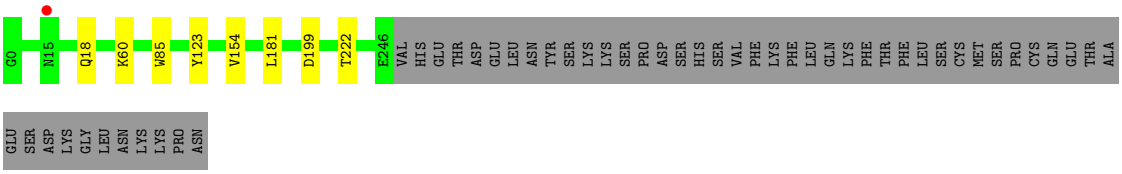
• Molecule 1: Gala protein type 1, 3 or 4

Chain B:



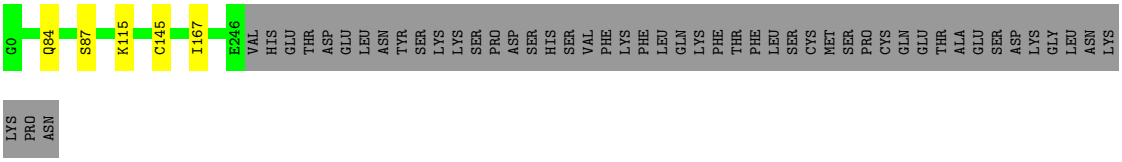
• Molecule 1: Gala protein type 1, 3 or 4

Chain C:



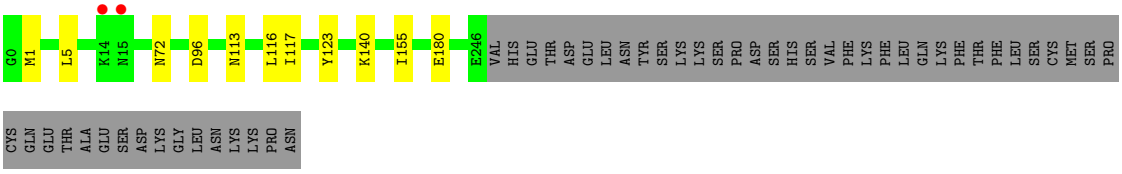
• Molecule 1: Gala protein type 1, 3 or 4

Chain D:



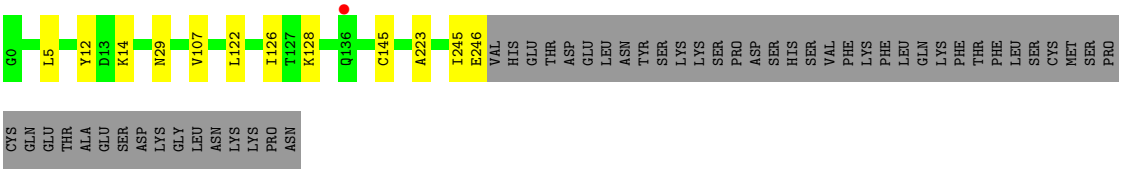
• Molecule 1: Gala protein type 1, 3 or 4

Chain E:



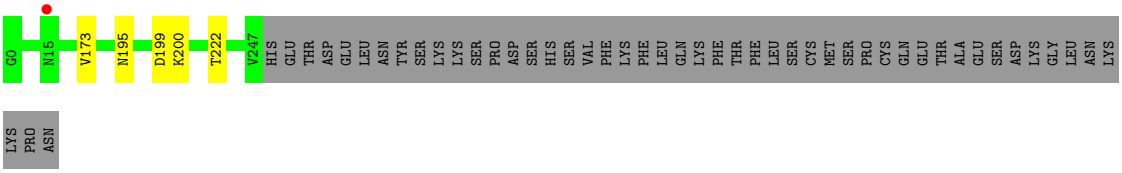
• Molecule 1: Gala protein type 1, 3 or 4

Chain F:



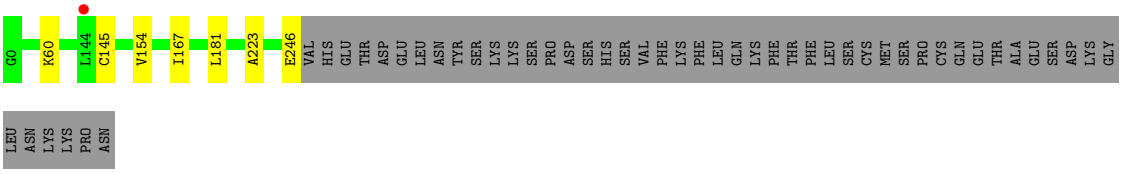
• Molecule 1: Gala protein type 1, 3 or 4

Chain G:



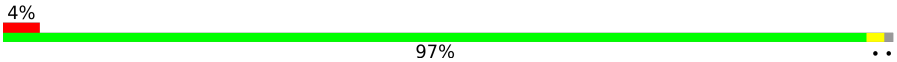
• Molecule 1: Gala protein type 1, 3 or 4

Chain H:



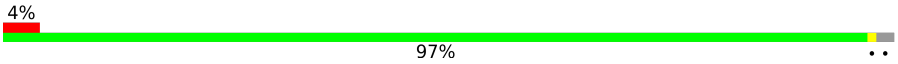
• Molecule 2: Uncharacterized protein

Chain a:



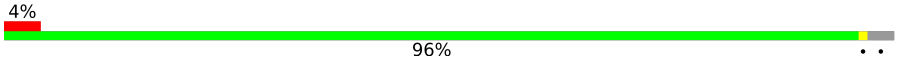
• Molecule 2: Uncharacterized protein

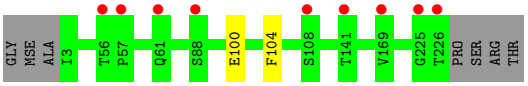
Chain b:



• Molecule 2: Uncharacterized protein

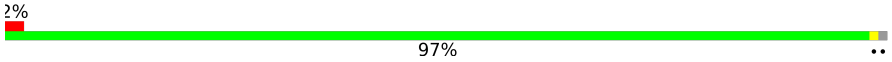
Chain c:

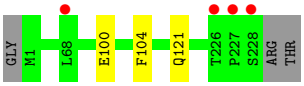




• Molecule 2: Uncharacterized protein

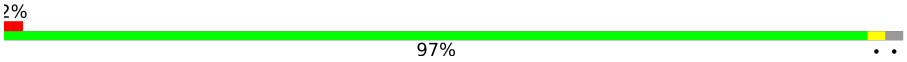
Chain d:

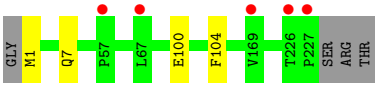




• Molecule 2: Uncharacterized protein

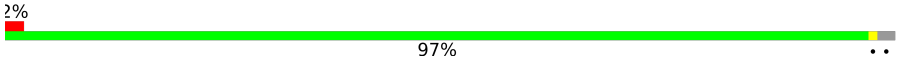
Chain e:

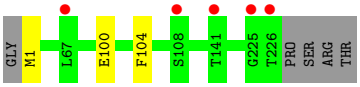




• Molecule 2: Uncharacterized protein

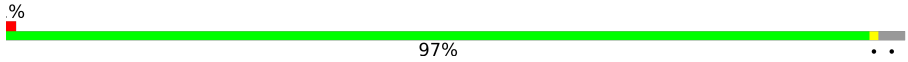
Chain f:

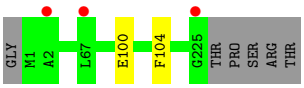




• Molecule 2: Uncharacterized protein

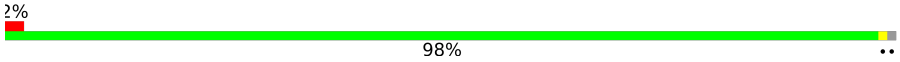
Chain g:

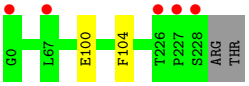




• Molecule 2: Uncharacterized protein

Chain h:





4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	219.61Å 120.89Å 149.29Å 90.00° 92.05° 90.00°	Depositor
Resolution (Å)	29.94 – 2.00 29.94 – 1.99	Depositor EDS
% Data completeness (in resolution range)	86.6 (29.94-2.00) 95.6 (29.94-1.99)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.22 (at 2.00Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, R_{free}	0.158 , 0.195 0.158 , 0.194	Depositor DCC
R_{free} test set	12932 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	23.0	Xtriage
Anisotropy	0.067	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.33 , 62.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.46$, $< L^2 > = 0.29$	Xtriage
Estimated twinning fraction	0.001 for -1/2*h-3/2*k,-1/2*h+1/2*k,-l 0.004 for -1/2*h+3/2*k,1/2*h+1/2*k,-l 0.010 for 1/2*h-3/2*k,-1/2*h-1/2*k,-l 0.003 for 1/2*h+3/2*k,1/2*h-1/2*k,-l 0.027 for -h,-k,l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	34121	wwPDB-VP
Average B, all atoms (Å^2)	29.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 43.81 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.6723e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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)

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	0.37	0/1964	0.47	1/2663 (0.0%)
1	B	0.33	0/1952	0.45	0/2650
1	C	0.30	0/1919	0.45	0/2603
1	D	0.28	0/1936	0.44	0/2627
1	E	0.28	0/1934	0.43	0/2625
1	F	0.29	0/1941	0.44	0/2635
1	G	0.30	0/1934	0.45	0/2626
1	H	0.28	0/1930	0.44	0/2620
2	a	0.37	0/1850	0.49	0/2516
2	b	0.35	0/1827	0.46	0/2486
2	c	0.32	0/1833	0.46	0/2492
2	d	0.36	0/1868	0.47	0/2541
2	e	0.40	0/1838	0.50	0/2500
2	f	0.36	0/1876	0.47	0/2551
2	g	0.37	0/1838	0.47	0/2497
2	h	0.40	0/1844	0.48	0/2506
All	All	0.34	0/30284	0.46	1/41138 (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	A	85	TRP	CA-CB-CG	-5.49	103.27	113.70

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	1916	0	1982	4	0
1	B	1907	0	1951	10	0
1	C	1889	0	1931	3	0
1	D	1897	0	1946	3	0
1	E	1898	0	1934	6	0
1	F	1904	0	1942	9	0
1	G	1900	0	1936	3	0
1	H	1894	0	1926	4	0
2	a	1796	0	1778	0	0
2	b	1779	0	1753	0	0
2	c	1779	0	1761	0	0
2	d	1808	0	1799	0	0
2	e	1787	0	1768	0	0
2	f	1810	0	1803	0	0
2	g	1782	0	1772	0	0
2	h	1794	0	1773	0	0
3	A	287	0	0	0	0
3	B	295	0	0	2	0
3	C	332	0	0	0	0
3	D	306	0	0	1	0
3	E	294	0	0	0	0
3	F	301	0	0	2	0
3	G	295	0	0	1	0
3	H	262	0	0	1	0
3	a	234	0	0	0	0
3	b	248	0	0	0	0
3	c	261	0	0	0	0
3	d	254	0	0	0	0
3	e	311	0	0	0	0
3	f	342	0	0	0	0
3	g	274	0	0	0	0
3	h	285	0	0	0	0
All	All	34121	0	29755	40	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.

The worst 5 of 40 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:F:107:VAL:HG11	1:F:126:ILE:HD13	1.71	0.70
1:A:167:ILE:HD12	1:A:191:ILE:HD12	1.86	0.57
1:B:5:LEU:HD21	1:E:5:LEU:HD21	1.89	0.55
1:A:5:LEU:HD21	1:F:5:LEU:HD21	1.91	0.52
1:H:60:LYS:HD3	3:H:560:HOH:O	2.10	0.52

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	255/297 (86%)	244 (96%)	11 (4%)	0	100	100
1	B	253/297 (85%)	243 (96%)	10 (4%)	0	100	100
1	C	248/297 (84%)	239 (96%)	8 (3%)	1 (0%)	34	30
1	D	251/297 (84%)	240 (96%)	11 (4%)	0	100	100
1	E	251/297 (84%)	243 (97%)	7 (3%)	1 (0%)	34	30
1	F	251/297 (84%)	241 (96%)	10 (4%)	0	100	100
1	G	250/297 (84%)	239 (96%)	11 (4%)	0	100	100
1	H	250/297 (84%)	240 (96%)	10 (4%)	0	100	100
2	a	229/231 (99%)	226 (99%)	3 (1%)	0	100	100
2	b	226/231 (98%)	224 (99%)	2 (1%)	0	100	100
2	c	225/231 (97%)	224 (100%)	1 (0%)	0	100	100
2	d	231/231 (100%)	227 (98%)	4 (2%)	0	100	100
2	e	227/231 (98%)	224 (99%)	3 (1%)	0	100	100
2	f	231/231 (100%)	229 (99%)	2 (1%)	0	100	100
2	g	227/231 (98%)	222 (98%)	5 (2%)	0	100	100
2	h	229/231 (99%)	227 (99%)	2 (1%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
All	All	3834/4224 (91%)	3732 (97%)	100 (3%)	2 (0%)	51 49

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	123	TYR
1	C	123	TYR

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	222/260 (85%)	222 (100%)	0	100 100
1	B	220/260 (85%)	220 (100%)	0	100 100
1	C	215/260 (83%)	214 (100%)	1 (0%)	88 92
1	D	218/260 (84%)	218 (100%)	0	100 100
1	E	218/260 (84%)	218 (100%)	0	100 100
1	F	218/260 (84%)	218 (100%)	0	100 100
1	G	217/260 (84%)	217 (100%)	0	100 100
1	H	217/260 (84%)	217 (100%)	0	100 100
2	a	198/194 (102%)	194 (98%)	4 (2%)	55 58
2	b	195/194 (100%)	193 (99%)	2 (1%)	76 81
2	c	195/194 (100%)	193 (99%)	2 (1%)	76 81
2	d	200/194 (103%)	196 (98%)	4 (2%)	55 58
2	e	196/194 (101%)	191 (97%)	5 (3%)	46 48
2	f	200/194 (103%)	197 (98%)	3 (2%)	65 69
2	g	196/194 (101%)	194 (99%)	2 (1%)	76 81
2	h	197/194 (102%)	195 (99%)	2 (1%)	76 81
All	All	3322/3632 (92%)	3297 (99%)	25 (1%)	84 86

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	e	7[A]	GLN
2	e	104	PHE
2	h	104	PHE
2	e	100	GLU
2	f	1	MSE

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)

There are no ligands in this entry.

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(Å^2)	Q<0.9
1	A	247/297 (83%)	-0.43	1 (0%)	92	92	14, 23, 38, 64	0
1	B	247/297 (83%)	-0.47	1 (0%)	92	92	14, 23, 37, 62	0
1	C	247/297 (83%)	-0.53	1 (0%)	92	92	11, 21, 36, 74	0
1	D	247/297 (83%)	-0.48	0	100	100	14, 24, 36, 61	0
1	E	247/297 (83%)	-0.44	2 (0%)	86	85	11, 23, 41, 77	0
1	F	247/297 (83%)	-0.47	1 (0%)	92	92	13, 23, 42, 72	0
1	G	248/297 (83%)	-0.40	1 (0%)	92	92	13, 22, 42, 89	0
1	H	247/297 (83%)	-0.37	1 (0%)	92	92	15, 26, 40, 67	0
2	a	225/231 (97%)	0.05	9 (4%)	38	37	17, 31, 60, 91	1 (0%)
2	b	225/231 (97%)	0.08	10 (4%)	34	33	15, 31, 62, 133	1 (0%)
2	c	222/231 (96%)	-0.01	9 (4%)	37	36	12, 29, 55, 110	0
2	d	225/231 (97%)	-0.11	4 (1%)	68	66	16, 28, 55, 75	0
2	e	224/231 (96%)	-0.22	5 (2%)	62	60	13, 24, 48, 79	0
2	f	223/231 (96%)	-0.30	5 (2%)	62	60	14, 23, 42, 83	0
2	g	222/231 (96%)	-0.14	3 (1%)	75	74	15, 27, 49, 80	0
2	h	226/231 (97%)	-0.16	5 (2%)	62	60	14, 26, 47, 62	1 (0%)
All	All	3769/4224 (89%)	-0.28	58 (1%)	73	72	11, 25, 47, 133	3 (0%)

The worst 5 of 58 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	a	226	THR	7.4
2	e	67	LEU	6.6
2	b	2	ALA	6.2
2	c	226	THR	5.7
2	a	227	PRO	5.3

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 monosaccharides in this entry.

6.4 Ligands (i)

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.