Full wwPDB X-ray Structure Validation Report (i)

Dec 18, 2023 – 04:05 am GMT

PDB ID	:	2YO3
Title	:	Salmonella enterica SadA 1185-1386 fused to GCN4 adaptors (SadAK14)
Authors	:	Hartmann, M.D.; Hernandez Alvarez, B.; Lupas, A.N.
Deposited on	:	2012-10-20
Resolution	:	2.00 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 (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 (1)) were used in the production of this report:

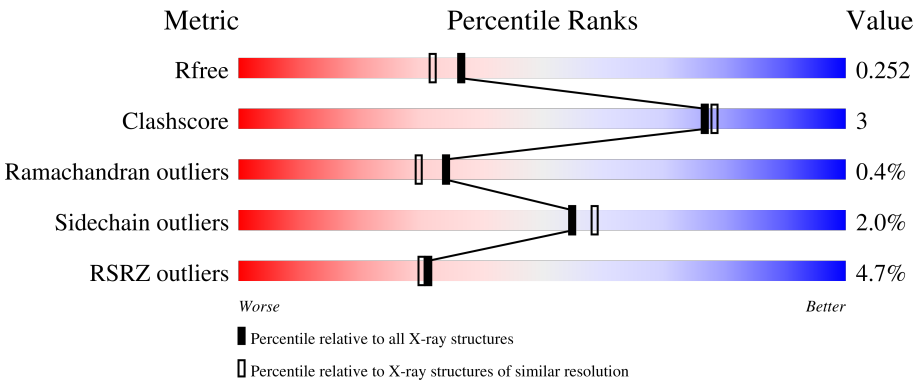
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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	268	4% 89% 7% •
1	B	268	3% 90% 6% •
1	C	268	6% 88% 8% •

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 6246 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 GENERAL CONTROL PROTEIN GCN4, PUTATIVE INNER MEMBRANE PROTEIN, GENERAL CONTROL PROTEIN GCN4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	257	1909	1156	341	409	3	0	1	0
			Total	C	N	O	S			
1	B	258	1906	1157	334	412	3	0	3	0
			Total	C	N	O	S			
1	C	257	1857	1129	330	395	3	0	1	0

There are 72 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1159	ILE	LEU	engineered mutation	UNP P03069
A	1163	ILE	VAL	engineered mutation	UNP P03069
A	1166	ILE	LEU	engineered mutation	UNP P03069
A	1170	ILE	ASN	engineered mutation	UNP P03069
A	1173	ILE	LEU	engineered mutation	UNP P03069
A	1177	ILE	VAL	engineered mutation	UNP P03069
A	1180	ILE	LEU	engineered mutation	UNP P03069
A	1184	ILE	VAL	engineered mutation	UNP P03069
A	1390	ILE	LEU	engineered mutation	UNP P03069
A	1394	ILE	VAL	engineered mutation	UNP P03069
A	1397	ILE	LEU	engineered mutation	UNP P03069
A	1401	ILE	ASN	engineered mutation	UNP P03069
A	1404	ILE	LEU	engineered mutation	UNP P03069
A	1408	ILE	VAL	engineered mutation	UNP P03069
A	1411	ILE	LEU	engineered mutation	UNP P03069
A	1415	ILE	VAL	engineered mutation	UNP P03069
A	1416	LYS	-	expression tag	UNP P03069
A	1417	LEU	-	expression tag	UNP P03069
A	1418	HIS	-	expression tag	UNP P03069
A	1419	HIS	-	expression tag	UNP P03069
A	1420	HIS	-	expression tag	UNP P03069
A	1421	HIS	-	expression tag	UNP P03069

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	1422	HIS	-	expression tag	UNP P03069
A	1423	HIS	-	expression tag	UNP P03069
B	1159	ILE	LEU	engineered mutation	UNP P03069
B	1163	ILE	VAL	engineered mutation	UNP P03069
B	1166	ILE	LEU	engineered mutation	UNP P03069
B	1170	ILE	ASN	engineered mutation	UNP P03069
B	1173	ILE	LEU	engineered mutation	UNP P03069
B	1177	ILE	VAL	engineered mutation	UNP P03069
B	1180	ILE	LEU	engineered mutation	UNP P03069
B	1184	ILE	VAL	engineered mutation	UNP P03069
B	1390	ILE	LEU	engineered mutation	UNP P03069
B	1394	ILE	VAL	engineered mutation	UNP P03069
B	1397	ILE	LEU	engineered mutation	UNP P03069
B	1401	ILE	ASN	engineered mutation	UNP P03069
B	1404	ILE	LEU	engineered mutation	UNP P03069
B	1408	ILE	VAL	engineered mutation	UNP P03069
B	1411	ILE	LEU	engineered mutation	UNP P03069
B	1415	ILE	VAL	engineered mutation	UNP P03069
B	1416	LYS	-	expression tag	UNP P03069
B	1417	LEU	-	expression tag	UNP P03069
B	1418	HIS	-	expression tag	UNP P03069
B	1419	HIS	-	expression tag	UNP P03069
B	1420	HIS	-	expression tag	UNP P03069
B	1421	HIS	-	expression tag	UNP P03069
B	1422	HIS	-	expression tag	UNP P03069
B	1423	HIS	-	expression tag	UNP P03069
C	1159	ILE	LEU	engineered mutation	UNP P03069
C	1163	ILE	VAL	engineered mutation	UNP P03069
C	1166	ILE	LEU	engineered mutation	UNP P03069
C	1170	ILE	ASN	engineered mutation	UNP P03069
C	1173	ILE	LEU	engineered mutation	UNP P03069
C	1177	ILE	VAL	engineered mutation	UNP P03069
C	1180	ILE	LEU	engineered mutation	UNP P03069
C	1184	ILE	VAL	engineered mutation	UNP P03069
C	1390	ILE	LEU	engineered mutation	UNP P03069
C	1394	ILE	VAL	engineered mutation	UNP P03069
C	1397	ILE	LEU	engineered mutation	UNP P03069
C	1401	ILE	ASN	engineered mutation	UNP P03069
C	1404	ILE	LEU	engineered mutation	UNP P03069
C	1408	ILE	VAL	engineered mutation	UNP P03069
C	1411	ILE	LEU	engineered mutation	UNP P03069
C	1415	ILE	VAL	engineered mutation	UNP P03069

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	1416	LYS	-	expression tag	UNP P03069
C	1417	LEU	-	expression tag	UNP P03069
C	1418	HIS	-	expression tag	UNP P03069
C	1419	HIS	-	expression tag	UNP P03069
C	1420	HIS	-	expression tag	UNP P03069
C	1421	HIS	-	expression tag	UNP P03069
C	1422	HIS	-	expression tag	UNP P03069
C	1423	HIS	-	expression tag	UNP P03069

- Molecule 2 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	Cl		
2	B	1	1	1	0	0

- Molecule 3 is water.

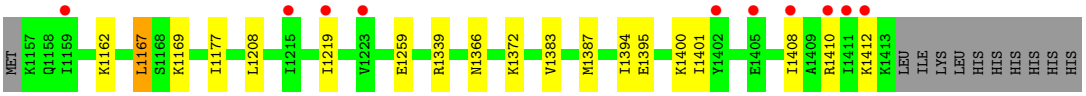
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	244	244	244	0	0
			Total	O		
3	B	193	193	193	0	0
			Total	O		
3	C	136	136	136	0	0

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.

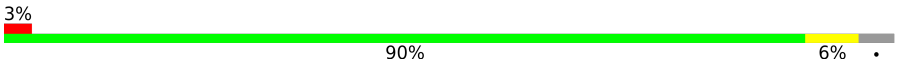
• Molecule 1: GENERAL CONTROL PROTEIN GCN4, PUTATIVE INNER MEMBRANE PROTEIN, GENERAL CONTROL PROTEIN GCN4

Chain A:



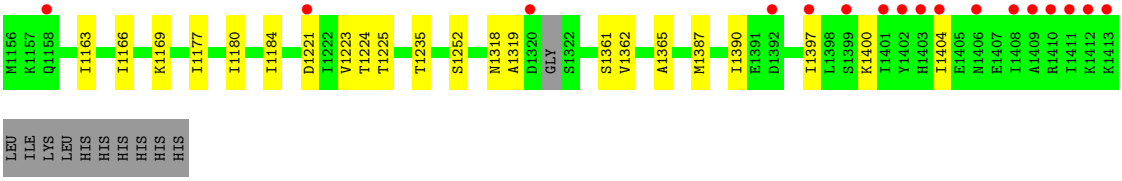
• Molecule 1: GENERAL CONTROL PROTEIN GCN4, PUTATIVE INNER MEMBRANE PROTEIN, GENERAL CONTROL PROTEIN GCN4

Chain B:



• Molecule 1: GENERAL CONTROL PROTEIN GCN4, PUTATIVE INNER MEMBRANE PROTEIN, GENERAL CONTROL PROTEIN GCN4

Chain C:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	189.32Å 46.17Å 103.76Å 90.00° 98.03° 90.00°	Depositor
Resolution (Å)	37.31 – 2.00 37.32 – 2.00	Depositor EDS
% Data completeness (in resolution range)	98.3 (37.31-2.00) 98.3 (37.32-2.00)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.69 (at 2.00Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.200 , 0.253 0.204 , 0.252	Depositor DCC
R_{free} test set	2984 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	28.4	Xtriage
Anisotropy	0.192	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 40.7	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	0.95	EDS
Total number of atoms	6246	wwPDB-VP
Average B, all atoms (Å^2)	42.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 7.66% 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: CL

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	0.53	0/1921	0.59	1/2598 (0.0%)
1	B	0.52	0/1924	0.61	1/2606 (0.0%)
1	C	0.51	0/1866	0.61	0/2529
All	All	0.52	0/5711	0.60	2/7733 (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})
1	B	1286	ARG	NE-CZ-NH1	5.11	122.86	120.30
1	A	1339	ARG	NE-CZ-NH2	-5.02	117.79	120.30

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	1909	0	1888	23	0
1	B	1906	0	1862	11	0
1	C	1857	0	1804	26	0
2	B	1	0	0	0	0
3	A	244	0	0	2	0
3	B	193	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	C	136	0	0	0	0
All	All	6246	0	5554	34	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 3.

All (34) 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:1387:MET:CE	1:C:1390:ILE:HD12	2.08	0.84
1:A:1394:ILE:CD1	1:C:1390:ILE:HG23	2.12	0.80
1:A:1387:MET:HE2	1:C:1390:ILE:HD12	1.63	0.79
1:A:1219:ILE:HG21	1:B:1219:ILE:HD11	1.64	0.77
3:A:2014:HOH:O	1:C:1169:LYS:NZ	2.23	0.70
1:A:1169:LYS:HE2	1:B:1174:GLU:OE2	1.99	0.61
1:A:1387:MET:HE3	1:C:1390:ILE:HD12	1.81	0.60
1:A:1394:ILE:HD11	1:C:1390:ILE:HG23	1.84	0.60
1:A:1387:MET:HE1	1:C:1387:MET:HG3	1.86	0.57
1:A:1169:LYS:CE	1:B:1174:GLU:OE2	2.59	0.51
1:A:1177:ILE:HG23	1:C:1180:ILE:HD11	1.93	0.51
1:A:1208:LEU:HD21	1:B:1209:ASP:HA	1.93	0.51
1:A:1383:VAL:HG13	1:B:1387:MET:CE	2.41	0.51
1:A:1387:MET:HE3	1:C:1390:ILE:CD1	2.42	0.50
1:B:1361[B]:SER:HB3	1:C:1362:VAL:CG2	2.41	0.50
1:A:1408:ILE:CD1	1:C:1404:ILE:HG23	2.43	0.48
1:C:1223:VAL:HG23	1:C:1224:THR:HG23	1.95	0.47
1:A:1167:LEU:HD13	1:C:1166:ILE:CG1	2.45	0.46
1:B:1346:ALA:HB1	1:B:1351:ASP:HB2	1.98	0.46
1:B:1173:ILE:HG23	1:C:1177:ILE:HD11	1.98	0.45
1:C:1235:THR:HG21	1:C:1252:SER:HB3	1.98	0.45
1:B:1180:ILE:HG23	1:C:1184:ILE:HD11	1.99	0.44
1:C:1163[B]:ILE:HD13	1:C:1163[B]:ILE:HA	1.80	0.44
1:A:1401:ILE:CD1	1:C:1397:ILE:HG23	2.48	0.44
1:C:1221:ASP:O	1:C:1225:THR:OG1	2.31	0.43
1:A:1408:ILE:HD11	1:C:1404:ILE:HG23	2.02	0.42
1:A:1366:ASN:HD21	1:C:1365:ALA:HA	1.84	0.42
1:A:1401:ILE:HD11	1:C:1397:ILE:HG23	2.01	0.41
1:A:1394:ILE:HD13	1:C:1390:ILE:HG23	1.99	0.41
1:A:1401:ILE:HG21	1:C:1400:LYS:HE2	2.03	0.41
1:B:1346:ALA:HB2	1:B:1352:ALA:O	2.21	0.41
1:A:1167:LEU:HD13	1:C:1166:ILE:HG12	2.02	0.40

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
3:A:2033:HOH:O	1:B:1184:ILE:HG23	2.20	0.40
1:A:1177:ILE:HG23	1:C:1180:ILE:CD1	2.51	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	256/268 (96%)	254 (99%)	2 (1%)	0	100	100
1	B	259/268 (97%)	254 (98%)	4 (2%)	1 (0%)	34	30
1	C	254/268 (95%)	249 (98%)	3 (1%)	2 (1%)	19	13
All	All	769/804 (96%)	757 (98%)	9 (1%)	3 (0%)	34	30

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	1319	ALA
1	C	1318	ASN
1	B	1335	GLY

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	208/223 (93%)	200 (96%)	8 (4%)	33	31
1	B	205/223 (92%)	202 (98%)	3 (2%)	65	69
1	C	193/223 (86%)	192 (100%)	1 (0%)	88	92
All	All	606/669 (91%)	594 (98%)	12 (2%)	55	58

All (12) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	1162	LYS
1	A	1167	LEU
1	A	1259	GLU
1	A	1372	LYS
1	A	1395	GLU
1	A	1400	LYS
1	A	1410	ARG
1	A	1412	LYS
1	B	1221	ASP
1	B	1309	GLU
1	B	1316	GLU
1	C	1361	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (9) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	1296	ASN
1	A	1366	ASN
1	B	1243	GLN
1	B	1329	ASN
1	B	1366	ASN
1	C	1158	GLN
1	C	1267	ASN
1	C	1296	ASN
1	C	1366	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 1 ligands modelled in this entry, 1 is 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(Å^2)	Q<0.9
1	A	257/268 (95%)	-0.19	10 (3%)	39	38	18, 35, 69, 124	0
1	B	258/268 (96%)	-0.26	9 (3%)	44	43	19, 34, 76, 97	0
1	C	257/268 (95%)	0.02	17 (6%)	18	17	19, 36, 106, 182	0
All	All	772/804 (96%)	-0.14	36 (4%)	31	30	18, 35, 83, 182	0

All (36) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	1410	ARG	7.2
1	C	1411	ILE	7.0
1	A	1219	ILE	6.7
1	C	1408	ILE	6.2
1	C	1404	ILE	5.6
1	C	1409	ALA	5.0
1	C	1406	ASN	4.5
1	C	1399	SER	4.2
1	C	1402	TYR	3.9
1	A	1223	VAL	3.8
1	A	1410	ARG	3.6
1	C	1158	GLN	3.3
1	B	1406	ASN	3.1
1	A	1411	ILE	3.0
1	B	1402	TYR	2.9
1	C	1412	LYS	2.9
1	A	1408	ILE	2.7
1	A	1159	ILE	2.7
1	B	1159	ILE	2.6
1	B	1156	MET	2.5
1	C	1320	ASP	2.5
1	B	1223	VAL	2.5
1	A	1405	GLU	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	1413	LYS	2.4
1	C	1221	ASP	2.4
1	B	1221	ASP	2.4
1	B	1220	GLY	2.4
1	C	1392	ASP	2.3
1	A	1402	TYR	2.3
1	B	1411	ILE	2.2
1	A	1412	LYS	2.2
1	A	1215	ILE	2.1
1	C	1401	ILE	2.1
1	C	1403	HIS	2.1
1	B	1396	GLU	2.1
1	C	1397	ILE	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 monosaccharides 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
2	CL	B	2387	1/1	0.99	0.07	26,26,26,26	0

6.5 Other polymers (i)

There are no such residues in this entry.