Full wwPDB X-ray Structure Validation Report ⓘ

Oct 27, 2025 – 09:09 am GMT

PDB ID	:	9H5R / pdb_00009h5r
Title	:	X-ray structure of Trichomonas vaginalis inactive mutant hydrogenosomal processing peptidase heterodimer (HPPin)
Authors	:	Samad, A.; Cianci, M.; Motlova, L.; Barinka, C.
Deposited on	:	2024-10-23
Resolution	:	2.84 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

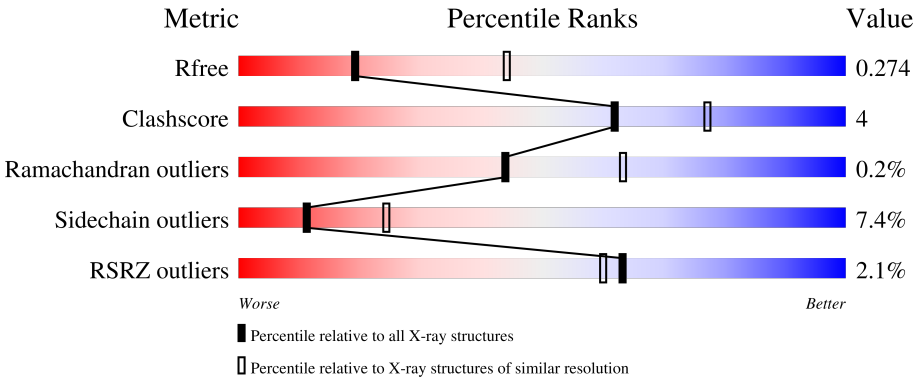
MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.010 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.46

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 2.84 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}	164625	1367 (2.86-2.82)
Clashscore	180529	1455 (2.86-2.82)
Ramachandran outliers	177936	1422 (2.86-2.82)
Sidechain outliers	177891	1423 (2.86-2.82)
RSRZ outliers	164620	1368 (2.86-2.82)

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	431	2% 77% 14% · 8%
2	B	419	2% 81% 16% ··

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6462 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 Alpha-MPP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	397	3092	1978	519	581	14	0	1	0

There are 14 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	240	GLY	GLU	conflict	UNP A2D7B7
A	419	GLY	-	expression tag	UNP A2D7B7
A	420	SER	-	expression tag	UNP A2D7B7
A	421	ALA	-	expression tag	UNP A2D7B7
A	422	LEU	-	expression tag	UNP A2D7B7
A	423	GLU	-	expression tag	UNP A2D7B7
A	424	HIS	-	expression tag	UNP A2D7B7
A	425	HIS	-	expression tag	UNP A2D7B7
A	426	HIS	-	expression tag	UNP A2D7B7
A	427	HIS	-	expression tag	UNP A2D7B7
A	428	HIS	-	expression tag	UNP A2D7B7
A	429	HIS	-	expression tag	UNP A2D7B7
A	430	HIS	-	expression tag	UNP A2D7B7
A	431	HIS	-	expression tag	UNP A2D7B7

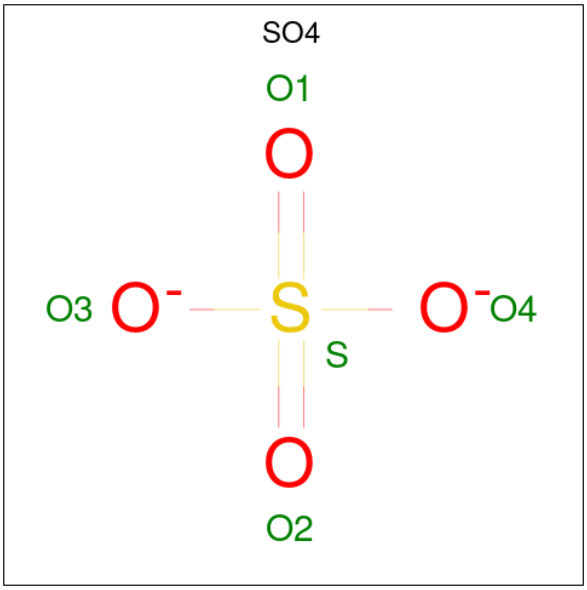
- Molecule 2 is a protein called Clan ME, family M16, insulinase-like metallopeptidase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	415	3272	2077	552	637	6	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	56	GLN	GLU	engineered mutation	UNP A2ES04
B	260	SER	PRO	conflict	UNP A2ES04

- Molecule 3 is SULFATE ION (CCD ID: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	A	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0
3	B	1	Total O S 5 4 1	0	0

- Molecule 4 is ZINC ION (CCD ID: ZN) (formula: Zn).

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

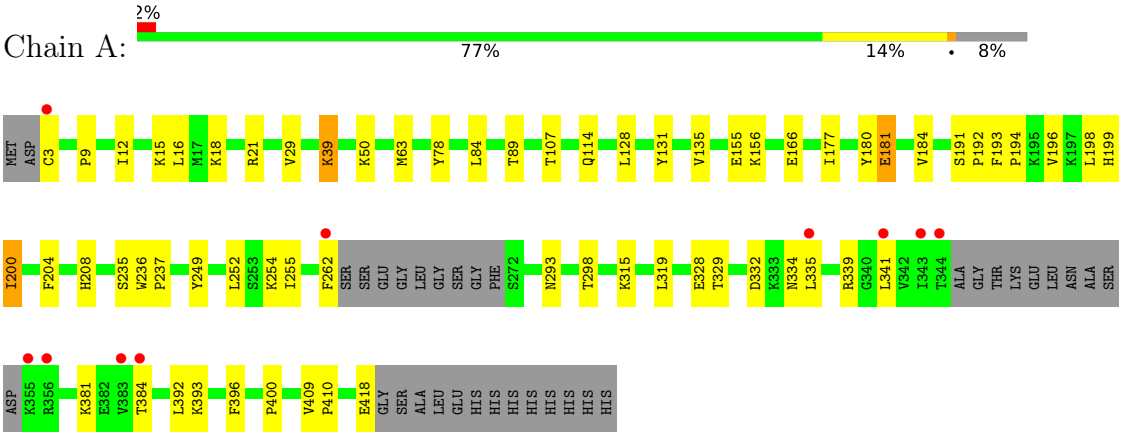
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	30	Total O 30 30	0	0
5	B	37	Total O 37 37	0	0

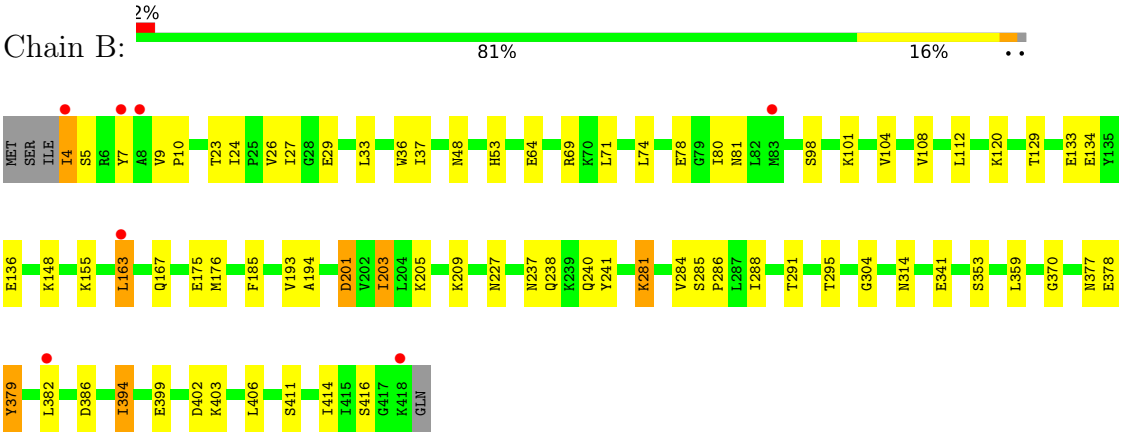
3 Residue-property plots (i)

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: Alpha-MPP



• Molecule 2: Clan ME, family M16, insulinase-like metallopeptidase



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	87.91Å 115.10Å 125.11Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.00 – 2.84 48.00 – 2.84	Depositor EDS
% Data completeness (in resolution range)	99.7 (48.00-2.84) 99.7 (48.00-2.84)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.86 (at 2.86Å)	Xtriage
Refinement program	REFMAC 5.8.0135	Depositor
R, R_{free}	0.203 , 0.273 0.203 , 0.274	Depositor DCC
R_{free} test set	1456 reflections (4.76%)	wwPDB-VP
Wilson B-factor (Å^2)	53.8	Xtriage
Anisotropy	0.081	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 43.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	6462	wwPDB-VP
Average B, all atoms (Å^2)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.57% 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: SO4, 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	0.67	0/3155	1.04	2/4264 (0.0%)
2	B	0.71	0/3335	1.06	3/4524 (0.1%)
All	All	0.69	0/6490	1.05	5/8788 (0.1%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	24	ILE	CB-CA-C	5.71	115.67	110.13
2	B	379	TYR	N-CA-C	5.50	117.39	110.24
1	A	39	LYS	N-CA-C	5.46	117.37	109.07
1	A	155	GLU	N-CA-C	5.42	116.91	110.19
2	B	81	ASN	N-CA-C	5.34	117.22	108.52

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	3092	0	3131	25	0
2	B	3272	0	3260	29	0
3	A	20	0	0	0	0
3	B	10	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	1	0	0	0	0
5	A	30	0	0	0	0
5	B	37	0	0	0	0
All	All	6462	0	6391	51	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 4.

All (51) 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)
2:B:71:LEU:HD21	2:B:112:LEU:HD11	1.77	0.66
1:A:181:GLU:H	1:A:181:GLU:CD	2.07	0.62
2:B:341:GLU:HG3	2:B:394:ILE:HD11	1.80	0.62
1:A:339:ARG:HH21	1:A:384:THR:HA	1.64	0.62
2:B:201:ASP:O	2:B:205:LYS:HG2	2.03	0.59
2:B:4:ILE:HA	2:B:359:LEU:O	2.02	0.59
1:A:29:VAL:HG23	1:A:78:TYR:HB3	1.86	0.58
2:B:23:THR:HG22	2:B:194:ALA:HB3	1.86	0.58
2:B:284:VAL:HG11	2:B:291:THR:HG22	1.86	0.57
1:A:204:PHE:CG	1:A:237:PRO:HG3	2.39	0.57
2:B:23:THR:HG21	2:B:203:ILE:HG13	1.87	0.55
1:A:84:LEU:HD11	2:B:353:SER:HA	1.88	0.55
1:A:339:ARG:NH2	1:A:384:THR:HA	2.22	0.55
2:B:48:ASN:HA	2:B:176:MET:HE3	1.88	0.54
1:A:339:ARG:HD3	1:A:381:LYS:HA	1.89	0.54
1:A:315:LYS:O	1:A:319:LEU:HD12	2.09	0.53
2:B:240:GLN:HG2	2:B:414:ILE:HB	1.91	0.52
2:B:53:HIS:CE1	2:B:133:GLU:OE1	2.62	0.52
2:B:80:ILE:HD11	2:B:101:LYS:HB3	1.91	0.51
1:A:254:LYS:HD2	1:A:341:LEU:HB3	1.90	0.51
1:A:392:LEU:HD12	1:A:396:PHE:CD2	2.47	0.50
1:A:180:TYR:O	1:A:184:VAL:HG23	2.13	0.49
1:A:252:LEU:HA	1:A:255:ILE:HD12	1.94	0.49
2:B:237:ASN:O	2:B:411:SER:HA	2.13	0.48
1:A:200:ILE:HG22	1:A:298:THR:HG21	1.96	0.47
2:B:33:LEU:HA	2:B:193:VAL:O	2.14	0.47
2:B:227:ASN:ND2	2:B:304:GLY:HA2	2.29	0.47
2:B:104:VAL:O	2:B:108:VAL:HG23	2.15	0.46
1:A:12:ILE:HD13	1:A:194:PRO:HG3	1.98	0.46
2:B:74:LEU:HD23	2:B:78:GLU:HG3	1.97	0.46

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:37:ILE:HD13	2:B:185:PHE:CE2	2.51	0.46
2:B:295:THR:CG2	2:B:314:ASN:H	2.28	0.45
1:A:236:TRP:CE2	1:A:400:PRO:HB3	2.52	0.44
2:B:148:LYS:HD2	2:B:241:TYR:CZ	2.53	0.44
1:A:262:PHE:HB3	2:B:136:GLU:HG3	2.00	0.44
1:A:392:LEU:HD12	1:A:396:PHE:HD2	1.82	0.44
2:B:285:SER:HB3	2:B:288:ILE:HG12	2.00	0.44
1:A:192:PRO:HA	1:A:193:PHE:HA	1.85	0.43
2:B:53:HIS:NE2	2:B:133:GLU:OE1	2.52	0.42
1:A:249:TYR:HB3	1:A:293:ASN:ND2	2.34	0.42
2:B:36:TRP:CD1	2:B:370:GLY:HA3	2.55	0.42
2:B:53:HIS:CE1	2:B:163:LEU:HD12	2.54	0.42
2:B:281:LYS:HE3	2:B:281:LYS:HB2	1.89	0.42
1:A:3:CYS:HA	1:A:16:LEU:O	2.20	0.41
1:A:50:LYS:HD2	1:A:50:LYS:HA	1.88	0.41
1:A:63:MET:HE2	2:B:286:PRO:HB3	2.02	0.41
2:B:10:PRO:HD3	2:B:379:TYR:CE2	2.56	0.41
2:B:7:TYR:CD1	2:B:382:LEU:HD23	2.56	0.41
1:A:131:TYR:O	1:A:135:VAL:HG23	2.21	0.40
1:A:39:LYS:HE2	1:A:156:LYS:O	2.21	0.40
1:A:208:HIS:HB3	1:A:410:PRO:HB2	2.02	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	392/431 (91%)	366 (93%)	25 (6%)	1 (0%)	37	55
2	B	413/419 (99%)	390 (94%)	22 (5%)	1 (0%)	44	63
All	All	805/850 (95%)	756 (94%)	47 (6%)	2 (0%)	44	63

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	64	GLU
1	A	9	PRO

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	346/371 (93%)	322 (93%)	24 (7%)	13	26
2	B	371/375 (99%)	342 (92%)	29 (8%)	10	22
All	All	717/746 (96%)	664 (93%)	53 (7%)	11	24

All (53) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	15	LYS
1	A	18	LYS
1	A	21	ARG
1	A	89	THR
1	A	107	THR
1	A	114	GLN
1	A	128	LEU
1	A	166	GLU
1	A	177	ILE
1	A	181	GLU
1	A	191	SER
1	A	196	VAL
1	A	198	LEU
1	A	199	HIS
1	A	200	ILE
1	A	235	SER
1	A	328	GLU
1	A	329	THR
1	A	332	ASP
1	A	334	ASN
1	A	335	LEU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	393	LYS
1	A	409	VAL
1	A	418	GLU
2	B	4	ILE
2	B	5	SER
2	B	9	VAL
2	B	26	VAL
2	B	27	ILE
2	B	29	GLU
2	B	69	ARG
2	B	98	SER
2	B	120	LYS
2	B	129	THR
2	B	134	GLU
2	B	155	LYS
2	B	163	LEU
2	B	167	GLN
2	B	175	GLU
2	B	201	ASP
2	B	203	ILE
2	B	209	LYS
2	B	238	GLN
2	B	281	LYS
2	B	377	ASN
2	B	378	GLU
2	B	386	ASP
2	B	394	ILE
2	B	399	GLU
2	B	402	ASP
2	B	403	LYS
2	B	406	LEU
2	B	416	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (14) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	44	GLN
1	A	71	GLN
1	A	73	ASN
1	A	126	ASN
1	A	147	HIS
1	A	208	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	358	ASN
1	A	411	ASN
2	B	63	ASN
2	B	94	ASN
2	B	142	ASN
2	B	180	GLN
2	B	189	ASN
2	B	238	GLN

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 oligosaccharides in this entry.

5.6 Ligand geometry (i)

Of 7 ligands modelled in this entry, 1 is monoatomic - leaving 6 for Mogul analysis.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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 $|Z| > 2$ is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
3	SO4	B	503	-	4,4,4	0.38	0	6,6,6	0.12	0
3	SO4	A	504	-	4,4,4	0.37	0	6,6,6	0.29	0
3	SO4	A	503	-	4,4,4	0.35	0	6,6,6	0.10	0
3	SO4	B	502	-	4,4,4	0.38	0	6,6,6	0.17	0
3	SO4	A	501	-	4,4,4	0.37	0	6,6,6	0.13	0
3	SO4	A	502	-	4,4,4	0.34	0	6,6,6	0.17	0

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	397/431 (92%)	-0.17	10 (2%) 58 55	28, 57, 99, 125	1 (0%)
2	B	415/419 (99%)	-0.34	7 (1%) 69 65	31, 50, 87, 126	0
All	All	812/850 (95%)	-0.26	17 (2%) 63 60	28, 55, 93, 126	1 (0%)

All (17) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	262	PHE	5.6
1	A	344	THR	4.6
2	B	7	TYR	3.6
1	A	343	ILE	3.2
1	A	383	VAL	3.1
2	B	163	LEU	3.0
2	B	382	LEU	3.0
2	B	418	LYS	2.6
1	A	384	THR	2.6
2	B	4	ILE	2.4
1	A	335	LEU	2.3
1	A	341	LEU	2.3
2	B	8	ALA	2.1
2	B	83	MET	2.1
1	A	355	LYS	2.1
1	A	3	CYS	2.1
1	A	356	ARG	2.1

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 (i)

There are no oligosaccharides 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
3	SO4	A	501	5/5	0.82	0.08	118,120,122,123	0
3	SO4	A	503	5/5	0.83	0.12	104,115,119,120	0
3	SO4	B	502	5/5	0.84	0.13	103,108,112,114	0
3	SO4	A	504	5/5	0.89	0.14	44,44,46,47	5
3	SO4	B	503	5/5	0.92	0.13	104,104,105,105	0
3	SO4	A	502	5/5	0.95	0.12	69,69,74,77	0
4	ZN	B	501	1/1	0.99	0.02	51,51,51,51	0

6.5 Other polymers (i)

There are no such residues in this entry.