



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 10, 2026 – 10:30 AM UTC

PDB ID : 1TO9 / pdb_00001to9
Title : Crystal structure of THI-4 protein from *Bacillus subtilis*
Authors : Rajan, S.S.; Shuvalova, L.; Yang, X.; Anderson, W.F.; Midwest Center for Structural Genomics (MCSG)
Deposited on : 2004-06-14
Resolution : 2.40 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 ⓘ 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	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
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)	:	2.49

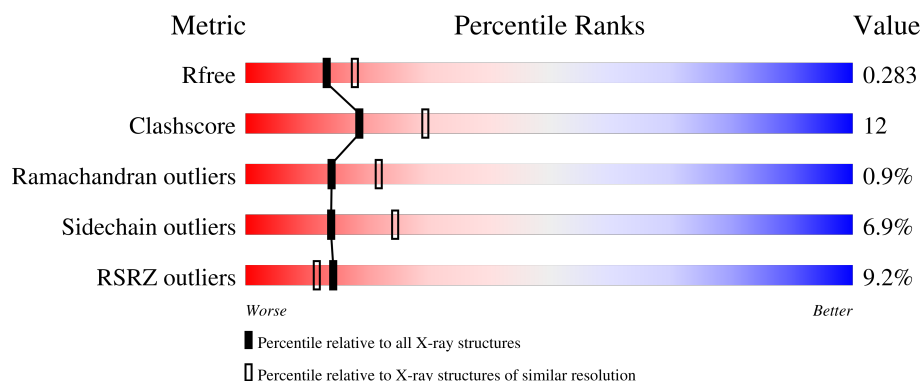
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 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}	180053	4912 (2.40-2.40)
Clashscore	190562	5391 (2.40-2.40)
Ramachandran outliers	187476	5320 (2.40-2.40)
Sidechain outliers	187428	5321 (2.40-2.40)
RSRZ outliers	180081	4916 (2.40-2.40)

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 $\leq 5\%$. 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	260	13% 55% 28% •• 13%
2	B	260	3% 71% 14% • 14%

2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 3758 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 THI-4 protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	225	1860	1201	301	349	4	5	0	1	0

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-23	MSE	-	cloning artifact	UNP P25052
A	-22	HIS	-	cloning artifact	UNP P25052
A	-21	HIS	-	cloning artifact	UNP P25052
A	-20	HIS	-	cloning artifact	UNP P25052
A	-19	HIS	-	cloning artifact	UNP P25052
A	-18	HIS	-	cloning artifact	UNP P25052
A	-17	HIS	-	cloning artifact	UNP P25052
A	-16	SER	-	cloning artifact	UNP P25052
A	-15	SER	-	cloning artifact	UNP P25052
A	-14	GLY	-	cloning artifact	UNP P25052
A	-13	VAL	-	cloning artifact	UNP P25052
A	-12	ASP	-	cloning artifact	UNP P25052
A	-11	LEU	-	cloning artifact	UNP P25052
A	-10	GLY	-	cloning artifact	UNP P25052
A	-9	THR	-	cloning artifact	UNP P25052
A	-8	GLU	-	cloning artifact	UNP P25052
A	-7	ASN	-	cloning artifact	UNP P25052
A	-6	LEU	-	cloning artifact	UNP P25052
A	-5	TYR	-	cloning artifact	UNP P25052
A	-4	PHE	-	cloning artifact	UNP P25052
A	-3	GLN	-	cloning artifact	UNP P25052
A	-2	SER	-	cloning artifact	UNP P25052
A	-1	ASN	-	cloning artifact	UNP P25052
A	0	ALA	-	cloning artifact	UNP P25052
A	1	MSE	MET	modified residue	UNP P25052
A	71	MSE	MET	modified residue	UNP P25052
A	83	MSE	MET	modified residue	UNP P25052

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	116	MSE	MET	modified residue	UNP P25052
A	194	MSE	MET	modified residue	UNP P25052
A	211	MSE	MET	modified residue	UNP P25052

- Molecule 2 is a protein called THI-4 protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	223	Total	C	N	O	S	Se	0	1	0
			1845	1188	299	348	4	6			

There are 31 discrepancies between the modelled and reference sequences:

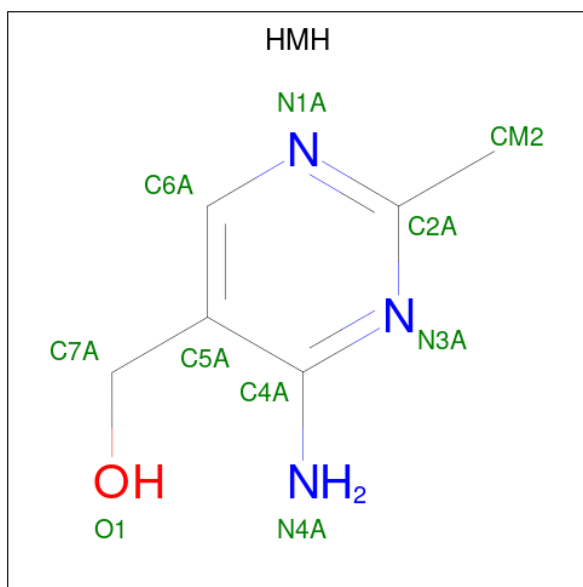
Chain	Residue	Modelled	Actual	Comment	Reference
B	-23	MSE	-	cloning artifact	UNP P25052
B	-22	HIS	-	cloning artifact	UNP P25052
B	-21	HIS	-	cloning artifact	UNP P25052
B	-20	HIS	-	cloning artifact	UNP P25052
B	-19	HIS	-	cloning artifact	UNP P25052
B	-18	HIS	-	cloning artifact	UNP P25052
B	-17	HIS	-	cloning artifact	UNP P25052
B	-16	SER	-	cloning artifact	UNP P25052
B	-15	SER	-	cloning artifact	UNP P25052
B	-14	GLY	-	cloning artifact	UNP P25052
B	-13	VAL	-	cloning artifact	UNP P25052
B	-12	ASP	-	cloning artifact	UNP P25052
B	-11	LEU	-	cloning artifact	UNP P25052
B	-10	GLY	-	cloning artifact	UNP P25052
B	-9	THR	-	cloning artifact	UNP P25052
B	-8	GLU	-	cloning artifact	UNP P25052
B	-7	ASN	-	cloning artifact	UNP P25052
B	-6	LEU	-	cloning artifact	UNP P25052
B	-5	TYR	-	cloning artifact	UNP P25052
B	-4	PHE	-	cloning artifact	UNP P25052
B	-3	GLN	-	cloning artifact	UNP P25052
B	-2	SER	-	cloning artifact	UNP P25052
B	-1	ASN	-	cloning artifact	UNP P25052
B	0	ALA	-	cloning artifact	UNP P25052
B	1	MSE	MET	modified residue	UNP P25052
B	71	MSE	MET	modified residue	UNP P25052
B	83	MSE	MET	modified residue	UNP P25052
B	116	MSE	MET	modified residue	UNP P25052
B	149	CSD	CYS	SEE REMARK 999	UNP P25052

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	194	MSE	MET	modified residue	UNP P25052
B	211	MSE	MET	modified residue	UNP P25052

- Molecule 3 is 4-AMINO-5-HYDROXYMETHYL-2-METHYLPYRIMIDINE (CCD ID: HMH) (formula: C₆H₉N₃O).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
3	A	1	Total	C	N	O	0	1
			12	7	3	2		
3	B	1	Total	C	N	O	0	1
			12	7	3	2		

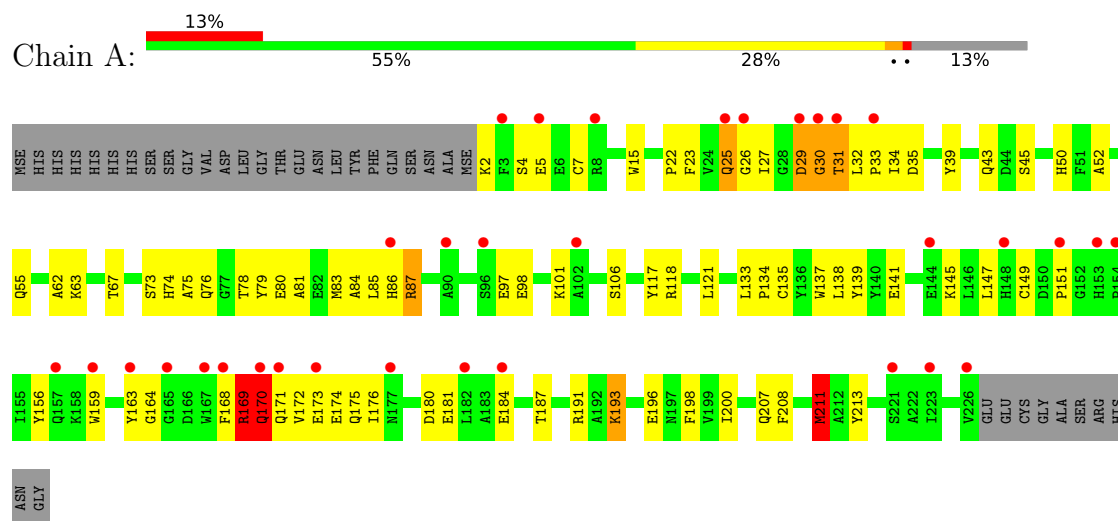
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	9	Total	O	0	0
			9	9		
4	B	20	Total	O	0	0
			20	20		

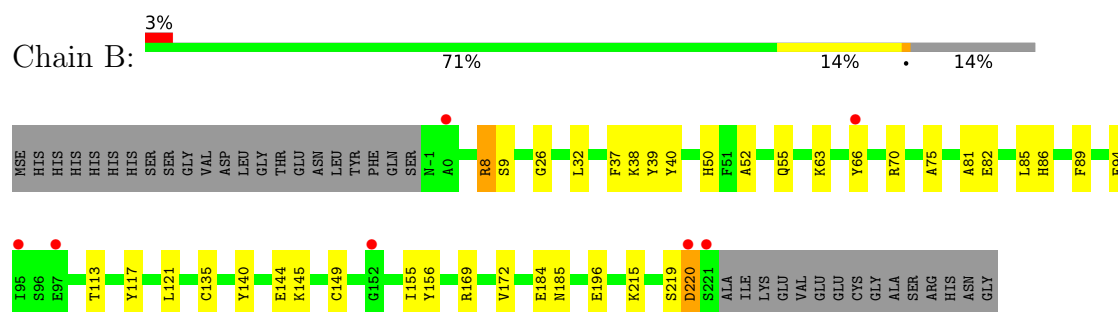
3 Residue-property plots [i](#)

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: THI-4 protein



• Molecule 2: THI-4 protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 41 21 2	Depositor
Cell constants a, b, c, α , β , γ	58.38Å 58.38Å 296.24Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	30.00 – 2.40 30.00 – 2.40	Depositor EDS
% Data completeness (in resolution range)	99.7 (30.00-2.40) 99.6 (30.00-2.40)	Depositor EDS
R_{merge}	0.04	Depositor
R_{sym}	0.04	Depositor
$\langle I/\sigma(I) \rangle$ ¹	9.51 (at 2.39Å)	Xtriage
Refinement program	REFMAC 5.1.24	Depositor
R, R_{free}	0.210 , 0.257 (Not available) , 0.283	Depositor DCC
R_{free} test set	1084 reflections (5.13%)	wwPDB-VP
Wilson B-factor (Å ²)	41.6	Xtriage
Anisotropy	0.039	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.33 , 60.8	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.92	EDS
Total number of atoms	3758	wwPDB-VP
Average B, all atoms (Å ²)	43.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.14% of the height of the origin peak. No significant pseudotranslation is detected.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ 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: CSD, HMM

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	0.86	8/1918 (0.4%)	0.92	2/2588 (0.1%)
2	B	0.77	0/1894	0.85	0/2554
All	All	0.82	8/3812 (0.2%)	0.88	2/5142 (0.0%)

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	29	ASP	C-N	11.05	1.48	1.33
1	A	30	GLY	N-CA	10.21	1.60	1.45
1	A	151	PRO	C-O	9.86	1.34	1.23
1	A	29	ASP	C-O	7.17	1.33	1.24
1	A	31	THR	N-CA	6.56	1.54	1.46

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	30	GLY	N-CA-C	-9.06	101.11	114.61
1	A	31	THR	N-CA-C	-6.86	104.91	113.28

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	1860	0	1732	64	0
2	B	1845	0	1711	23	0
3	A	12	0	6	0	0
3	B	12	0	6	0	0
4	A	9	0	0	1	0
4	B	20	0	0	0	0
All	All	3758	0	3455	86	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 12.

The worst 5 of 86 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:30:GLY:HA2	1:A:156:TYR:CD2	1.87	1.09
1:A:30:GLY:CA	1:A:156:TYR:CD2	2.63	0.81
2:B:82:GLU:O	2:B:86:HIS:HB2	1.85	0.76
1:A:85:LEU:HD13	1:A:163:TYR:OH	1.90	0.71
1:A:171:GLN:O	1:A:175:GLN:HG2	1.91	0.70

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	224/260 (86%)	204 (91%)	17 (8%)	3 (1%)	9	14
2	B	221/260 (85%)	212 (96%)	8 (4%)	1 (0%)	24	37
All	All	445/520 (86%)	416 (94%)	25 (6%)	4 (1%)	14	22

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	169	ARG
1	A	170	GLN
2	B	220	ASP
1	A	168	PHE

5.3.2 Protein sidechains ⓘ

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	190/211 (90%)	173 (91%)	17 (9%)	9	15
2	B	187/210 (89%)	178 (95%)	9 (5%)	23	40
All	All	377/421 (90%)	351 (93%)	26 (7%)	14	24

5 of 26 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	193	LYS
2	B	8	ARG
2	B	185	ASN
1	A	211	MSE
2	B	9	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 15 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	197	ASN
2	B	157	GLN
1	A	207	GLN
2	B	197	ASN
2	B	76	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

1 non-standard protein/DNA/RNA residue is modelled in this entry.

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	$\# Z > 2$
2	CSD	B	149	2	4,7,8	4.29	2 (50%)	1,8,10	3.24	1 (100%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	CSD	B	149	2	-	1/2/6/8	-

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	149	CSD	OD1-SG	-7.30	1.41	1.47
2	B	149	CSD	O-C	4.07	1.35	1.20

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	149	CSD	OD1-SG-CB	-3.24	99.63	105.60

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	149	CSD	CA-CB-SG-OD1

There are no ring outliers.

No monomer is involved in short contacts.

5.5 Carbohydrates [i](#)

There are no oligosaccharides in this entry.

5.6 Ligand geometry [i](#)

4 ligands are modelled in this entry.

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	HMH	A	301[A]	-	10,10,10	0.92	0	12,13,13	2.21	6 (50%)
3	HMH	A	301[B]	-	10,10,10	0.91	0	12,13,13	2.25	6 (50%)
3	HMH	B	302[B]	-	10,10,10	1.68	1 (10%)	12,13,13	2.46	8 (66%)
3	HMH	B	302[A]	-	10,10,10	1.70	2 (20%)	12,13,13	2.33	7 (58%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
3	HMH	A	301[A]	-	-	2/2/2/2	0/1/1/1
3	HMH	A	301[B]	-	-	1/2/2/2	0/1/1/1
3	HMH	B	302[B]	-	-	0/2/2/2	0/1/1/1
3	HMH	B	302[A]	-	-	1/2/2/2	0/1/1/1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	302[A]	HMH	C4A-N3A	3.50	1.39	1.35
3	B	302[B]	HMH	C4A-N3A	3.50	1.39	1.35
3	B	302[A]	HMH	O1-C7A	2.18	1.50	1.41

The worst 5 of 27 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	302[A]	HMH	N1A-C2A-N3A	-4.38	118.24	125.53
3	B	302[B]	HMH	N1A-C2A-N3A	-4.38	118.24	125.53
3	A	301[A]	HMH	N1A-C2A-N3A	-3.69	119.39	125.53
3	A	301[B]	HMH	N1A-C2A-N3A	-3.69	119.39	125.53
3	A	301[A]	HMH	C6A-N1A-C2A	3.57	121.94	116.07

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	301[A]	HMH	C6A-C5A-C7A-O1
3	B	302[A]	HMH	C6A-C5A-C7A-O1
3	A	301[A]	HMH	C4A-C5A-C7A-O1
3	A	301[B]	HMH	C6A-C5A-C7A-O1

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	220/260 (84%)	0.96	33 (15%) 5 4	24, 47, 89, 114	1 (0%)
2	B	216/260 (83%)	0.13	7 (3%) 50 46	16, 29, 74, 111	1 (0%)
All	All	436/520 (83%)	0.55	40 (9%) 14 12	16, 39, 80, 114	2 (0%)

The worst 5 of 40 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	30	GLY	6.9
1	A	177	ASN	4.7
1	A	168	PHE	4.4
1	A	148	HIS	4.1
1	A	184	GLU	3.9

6.2 Non-standard residues in protein, DNA, RNA chains [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	CSD	B	149	8/9	0.85	0.12	32,35,40,48	0

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, 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	HMH	A	301[A]	10/10	0.53	0.26	38,60,71,75	10
3	HMH	A	301[B]	10/10	0.53	0.26	38,58,70,75	10
3	HMH	B	302[A]	10/10	0.83	0.17	31,43,48,50	2
3	HMH	B	302[B]	10/10	0.83	0.17	5,37,47,48	2

6.5 Other polymers [i](#)

There are no such residues in this entry.