



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 6, 2026 – 08:53 PM UTC

PDB ID : 3H09 / pdb_00003h09
Title : The structure of Haemophilus influenzae IgA1 protease
Authors : Johnson, T.A.; Qiu, J.; Plaut, A.G.; Holyoak, T.
Deposited on : 2009-04-08
Resolution : 1.75 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

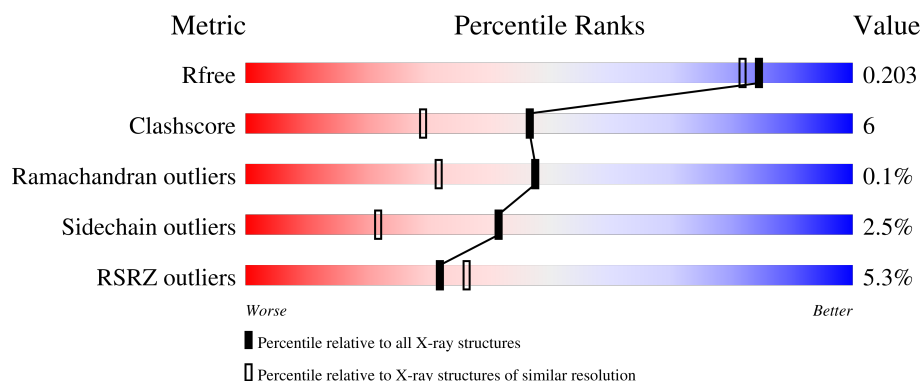
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 1.75 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	3183 (1.76-1.76)
Clashscore	190562	3299 (1.76-1.76)
Ramachandran outliers	187476	3274 (1.76-1.76)
Sidechain outliers	187428	3274 (1.76-1.76)
RSRZ outliers	180081	3183 (1.76-1.76)

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	989	5% 87% 10% ..
1	B	989	6% 87% 10% ..

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 17310 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 Immunoglobulin A1 protease.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	961	Total	C	N	O	S	0	58	0
			7718	4809	1353	1545	11			
1	B	964	Total	C	N	O	S	0	57	0
			7725	4814	1351	1549	11			

There are 12 discrepancies between the modelled and reference sequences:

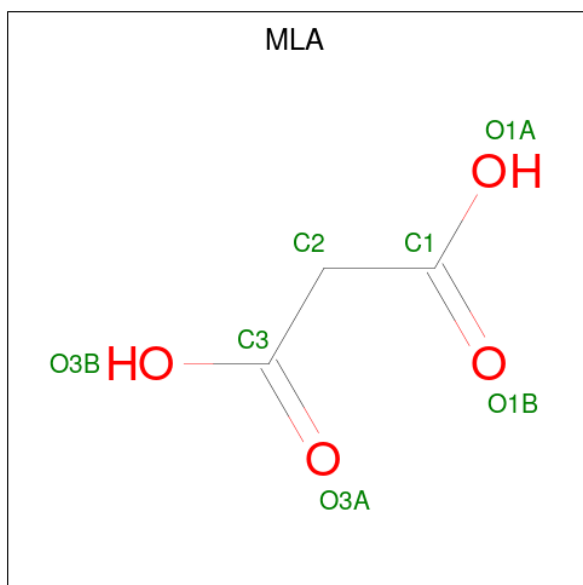
Chain	Residue	Modelled	Actual	Comment	Reference
A	1007	HIS	ASN	engineered mutation	UNP P44969
A	1008	HIS	ASN	engineered mutation	UNP P44969
A	1009	HIS	ILE	engineered mutation	UNP P44969
A	1010	HIS	GLN	engineered mutation	UNP P44969
A	1011	HIS	ALA	engineered mutation	UNP P44969
A	1012	HIS	ASP	engineered mutation	UNP P44969
B	1007	HIS	ASN	engineered mutation	UNP P44969
B	1008	HIS	ASN	engineered mutation	UNP P44969
B	1009	HIS	ILE	engineered mutation	UNP P44969
B	1010	HIS	GLN	engineered mutation	UNP P44969
B	1011	HIS	ALA	engineered mutation	UNP P44969
B	1012	HIS	ASP	engineered mutation	UNP P44969

- Molecule 2 is ACETATE ION (CCD ID: ACT) (formula: C₂H₃O₂).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			4	2	2		
2	A	1	Total	C	O	0	0
			4	2	2		
2	B	1	Total	C	O	0	0
			4	2	2		
2	B	1	Total	C	O	0	0
			4	2	2		

- Molecule 3 is MALONIC ACID (CCD ID: MLA) (formula: C₃H₄O₄).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 7 3 4	0	0
3	A	1	Total C O 7 3 4	0	0
3	A	1	Total C O 7 3 4	0	0
3	B	1	Total C O 7 3 4	0	0
3	B	1	Total C O 7 3 4	0	0

- Molecule 4 is SODIUM ION (CCD ID: NA) (formula: Na).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	1	Total Na 1 1	0	0
4	B	3	Total Na 3 3	0	0

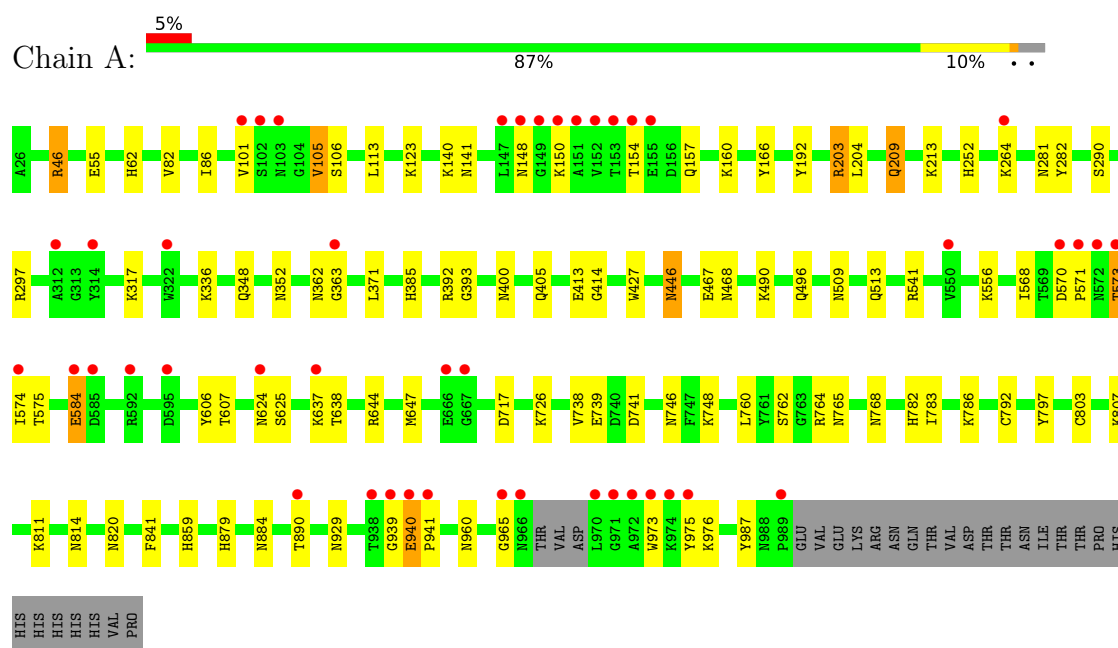
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	919	Total O 919 919	0	0
5	B	893	Total O 893 893	0	0

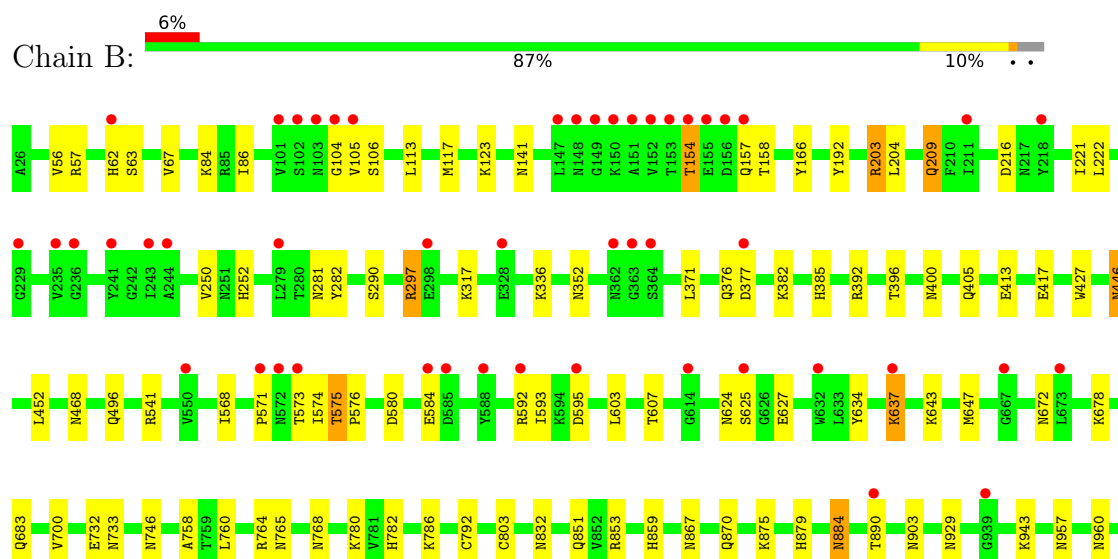
3 Residue-property plots [i](#)

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: Immunoglobulin A1 protease



• Molecule 1: Immunoglobulin A1 protease





WORLD WIDE
PDB
PROTEIN DATA BANK

4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	94.39Å 131.87Å 111.81Å 90.00° 113.11° 90.00°	Depositor
Resolution (Å)	33.18 – 1.75 33.18 – 1.75	Depositor EDS
% Data completeness (in resolution range)	99.6 (33.18-1.75) 99.6 (33.18-1.75)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.03 (at 1.75Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.161 , 0.190 0.180 , 0.203	Depositor DCC
R_{free} test set	12657 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	15.6	Xtriage
Anisotropy	0.572	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.40 , 36.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.012 for h,-k,-h-l	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	17310	wwPDB-VP
Average B, all atoms (Å ²)	9.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.65% 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 [i](#)

5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: MLA, ACT, NA

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	0.75	0/8030	0.84	1/10869 (0.0%)
1	B	0.71	1/8036 (0.0%)	0.81	1/10887 (0.0%)
All	All	0.73	1/16066 (0.0%)	0.82	2/21756 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	250	VAL	CA-CB	5.13	1.59	1.54

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	513	GLN	N-CA-C	5.12	117.65	111.40
1	B	966	ASN	N-CA-C	-5.10	105.36	112.30

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	7718	0	7610	88	0
1	B	7725	0	7605	84	0
2	A	8	0	6	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	8	0	6	0	0
3	A	21	0	6	1	0
3	B	14	0	4	1	0
4	A	1	0	0	0	0
4	B	3	0	0	0	0
5	A	919	0	0	14	0
5	B	893	0	0	17	0
All	All	17310	0	15237	174	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 6.

The worst 5 of 174 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:571:PRO:HA	1:A:647[B]:MET:SD	1.81	1.19
1:A:352[B]:ASN:ND2	5:A:1740:HOH:O	1.79	1.12
1:A:46[A]:ARG:HD3	5:A:1467:HOH:O	1.49	1.10
1:A:939:GLY:HA2	1:A:975:TYR:OH	1.58	1.03
1:B:571:PRO:HA	1:B:647[B]:MET:SD	1.99	1.03

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	1015/989 (103%)	984 (97%)	30 (3%)	1 (0%)	48	32
1	B	1019/989 (103%)	988 (97%)	31 (3%)	0	100	100
All	All	2034/1978 (103%)	1972 (97%)	61 (3%)	1 (0%)	48	100

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	363	GLY

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	873/843 (104%)	850 (97%)	23 (3%)	40	20
1	B	875/843 (104%)	854 (98%)	21 (2%)	43	23
All	All	1748/1686 (104%)	1704 (98%)	44 (2%)	42	22

5 of 44 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	216	ASP
1	B	595	ASP
1	B	297	ARG
1	B	573	THR
1	B	678	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 60 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	953	ASN
1	B	872	ASN
1	B	252	HIS
1	B	870	GLN
1	B	966	ASN

5.3.3 RNA ⓘ

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 13 ligands modelled in this entry, 4 are monoatomic - leaving 9 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	MLA	A	2000	-	6,6,6	1.08	0	7,7,7	1.13	0
2	ACT	B	1900	-	3,3,3	0.82	0	3,3,3	1.34	0
2	ACT	A	1800	-	3,3,3	0.91	0	3,3,3	1.00	0
3	MLA	A	2100	-	6,6,6	1.25	0	7,7,7	0.87	0
3	MLA	A	2200	-	6,6,6	1.26	0	7,7,7	0.78	0
3	MLA	B	2000	-	6,6,6	1.24	0	7,7,7	0.49	0
3	MLA	B	1800	-	6,6,6	1.00	0	7,7,7	1.65	2 (28%)
2	ACT	A	1900	-	3,3,3	0.90	0	3,3,3	1.47	0
2	ACT	B	1700	-	3,3,3	0.96	0	3,3,3	1.38	0

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	MLA	A	2000	-	-	2/4/4/4	-
3	MLA	A	2100	-	-	4/4/4/4	-
3	MLA	A	2200	-	-	0/4/4/4	-
3	MLA	B	2000	-	-	2/4/4/4	-
3	MLA	B	1800	-	-	0/4/4/4	-

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed($^{\circ}$)	Ideal($^{\circ}$)
3	B	1800	MLA	O1B-C1-C2	-2.44	115.16	122.11
3	B	1800	MLA	C3-C2-C1	-2.00	105.89	112.95

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	A	2100	MLA	O1A-C1-C2-C3
3	A	2100	MLA	O1B-C1-C2-C3
3	A	2100	MLA	C1-C2-C3-O3B
3	B	2000	MLA	O1A-C1-C2-C3
3	A	2100	MLA	C1-C2-C3-O3A

There are no ring outliers.

2 monomers are involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	2200	MLA	1	0
3	B	2000	MLA	1	0

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	961/989 (97%)	-0.11	45 (4%)	36	42	1, 5, 22, 49	58 (6%)
1	B	964/989 (97%)	0.28	57 (5%)	28	32	2, 6, 22, 44	57 (5%)
All	All	1925/1978 (97%)	0.08	102 (5%)	32	36	1, 5, 22, 49	115 (5%)

The worst 5 of 102 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	152	VAL	8.7
1	A	152	VAL	7.2
1	B	149	GLY	6.7
1	A	149	GLY	6.2
1	A	973	TRP	5.8

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 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($\text{\AA}^2$)	Q<0.9
2	ACT	B	1900	4/4	0.65	0.23	46,47,47,47	0
2	ACT	A	1900	4/4	0.76	0.21	37,38,38,38	0
3	MLA	A	2000	7/7	0.81	0.15	38,41,49,51	0
3	MLA	B	1800	7/7	0.84	0.14	33,37,43,46	0
3	MLA	B	2000	7/7	0.88	0.11	28,34,44,45	0
2	ACT	A	1800	4/4	0.93	0.08	14,14,15,15	0
3	MLA	A	2200	7/7	0.93	0.09	19,22,35,37	0
3	MLA	A	2100	7/7	0.94	0.10	19,22,27,27	0
2	ACT	B	1700	4/4	0.96	0.07	15,15,17,17	0
4	NA	A	2300	1/1	0.96	0.06	20,20,20,20	0
4	NA	B	2300	1/1	0.97	0.09	25,25,25,25	0
4	NA	B	2200	1/1	0.98	0.05	30,30,30,30	0
4	NA	B	2100	1/1	0.98	0.05	28,28,28,28	0

6.5 Other polymers [i](#)

There are no such residues in this entry.