



wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 23, 2024 – 12:13 AM EDT

PDB ID : 4NIY
Title : Crystal structure of trypsiligase (K60E/N143H/Y151H/D189K trypsin)
complexed to YRH-ecotin (M84Y/M85R/A86H ecotin)
Authors : Schoepfel, M.; Parthier, C.; Stubbs, M.T.
Deposited on : 2013-11-08
Resolution : 2.84 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	:	4.02b-467
Xtriage (Phenix)	:	1.20.1
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.39

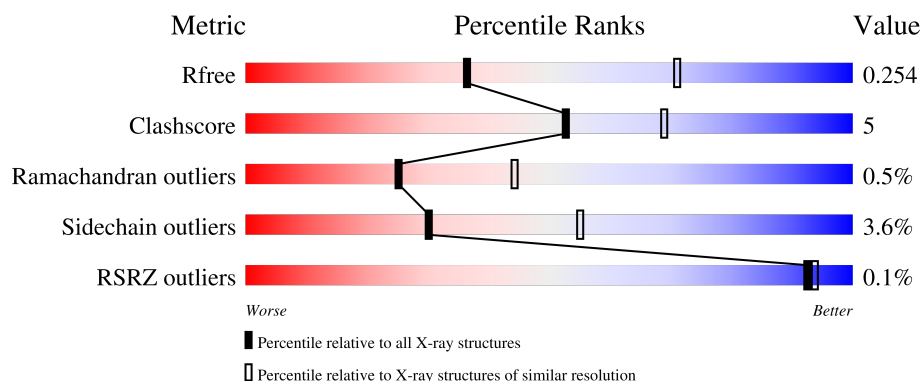
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.84 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 $\leq 5\%$. 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	223	<div> <div>83%</div> <div>13%</div> <div>••</div> </div>
1	B	223	<div> <div>81%</div> <div>13%</div> <div>5%</div> </div>
1	C	223	<div> <div>79%</div> <div>15%</div> <div>5%</div> </div>
1	D	223	<div> <div>%</div> <div>68%</div> <div>18%</div> <div>•</div> <div>13%</div> </div>
2	E	142	<div> <div>85%</div> <div>9%</div> <div>••</div> </div>

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	F	142	 87% 10% •
2	G	142	 83% 11% • 5%
2	H	142	 80% 15% •

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 10574 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 Cationic trypsin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	216	Total	C	N	O	S	0	0	0
			1585	987	274	310	14			
1	B	211	Total	C	N	O	S	0	0	0
			1546	963	266	303	14			
1	C	212	Total	C	N	O	S	0	0	0
			1552	966	266	306	14			
1	D	195	Total	C	N	O	S	0	0	0
			1435	894	244	284	13			

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	60	GLU	LYS	engineered mutation	UNP P00760
A	143	HIS	ASN	engineered mutation	UNP P00760
A	151	HIS	TYR	engineered mutation	UNP P00760
A	189	LYS	ASP	engineered mutation	UNP P00760
B	60	GLU	LYS	engineered mutation	UNP P00760
B	143	HIS	ASN	engineered mutation	UNP P00760
B	151	HIS	TYR	engineered mutation	UNP P00760
B	189	LYS	ASP	engineered mutation	UNP P00760
C	60	GLU	LYS	engineered mutation	UNP P00760
C	143	HIS	ASN	engineered mutation	UNP P00760
C	151	HIS	TYR	engineered mutation	UNP P00760
C	189	LYS	ASP	engineered mutation	UNP P00760
D	60	GLU	LYS	engineered mutation	UNP P00760
D	143	HIS	ASN	engineered mutation	UNP P00760
D	151	HIS	TYR	engineered mutation	UNP P00760
D	189	LYS	ASP	engineered mutation	UNP P00760

- Molecule 2 is a protein called Ecotin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	137	Total	C	N	O	S	0	0	0
			1107	708	186	209	4			
2	F	137	Total	C	N	O	S	0	0	0
			1112	711	187	210	4			
2	G	135	Total	C	N	O	S	0	0	0
			1096	701	184	207	4			
2	H	136	Total	C	N	O	S	0	0	0
			1103	705	185	209	4			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
E	84	TYR	MET	engineered mutation	UNP P23827
E	85	ARG	MET	engineered mutation	UNP P23827
E	86	HIS	ALA	engineered mutation	UNP P23827
F	84	TYR	MET	engineered mutation	UNP P23827
F	85	ARG	MET	engineered mutation	UNP P23827
F	86	HIS	ALA	engineered mutation	UNP P23827
G	84	TYR	MET	engineered mutation	UNP P23827
G	85	ARG	MET	engineered mutation	UNP P23827
G	86	HIS	ALA	engineered mutation	UNP P23827
H	84	TYR	MET	engineered mutation	UNP P23827
H	85	ARG	MET	engineered mutation	UNP P23827
H	86	HIS	ALA	engineered mutation	UNP P23827

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	1	Total	Ca	0	0
			1	1		
3	B	1	Total	Ca	0	0
			1	1		
3	C	1	Total	Ca	0	0
			1	1		
3	D	1	Total	Ca	0	0
			1	1		

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn).

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

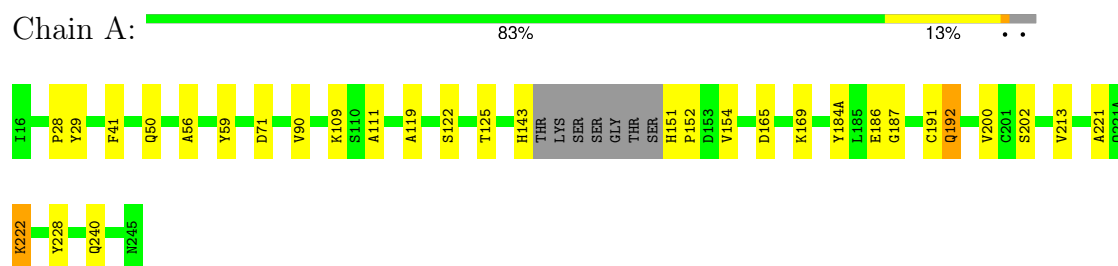
- Molecule 5 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	A	5	Total O 5 5	0	0
5	B	7	Total O 7 7	0	0
5	C	9	Total O 9 9	0	0
5	D	2	Total O 2 2	0	0
5	E	2	Total O 2 2	0	0
5	F	5	Total O 5 5	0	0
5	G	2	Total O 2 2	0	0
5	H	1	Total O 1 1	0	0

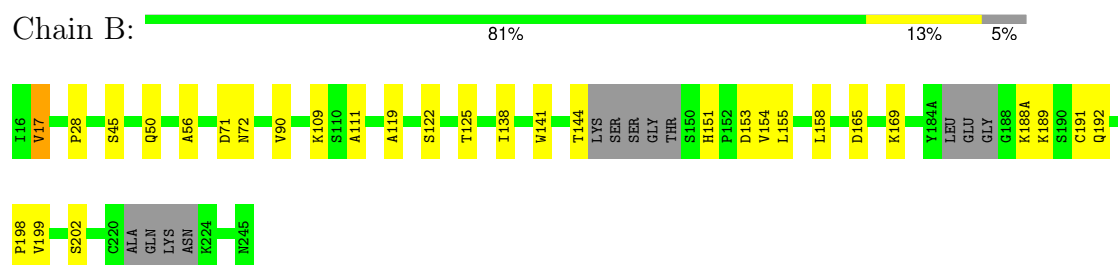
3 Residue-property plots [i](#)

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.

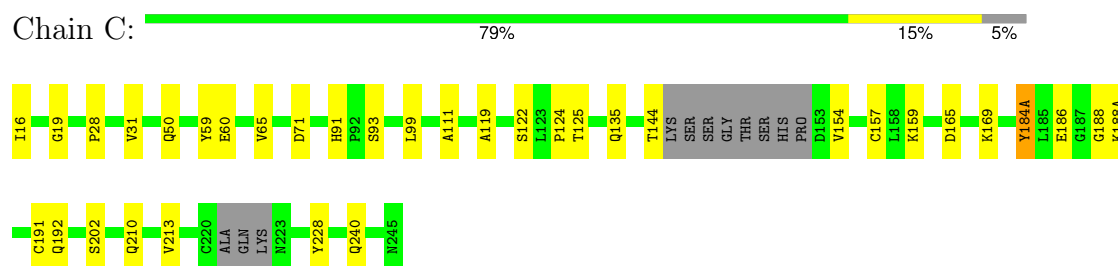
- Molecule 1: Cationic trypsin



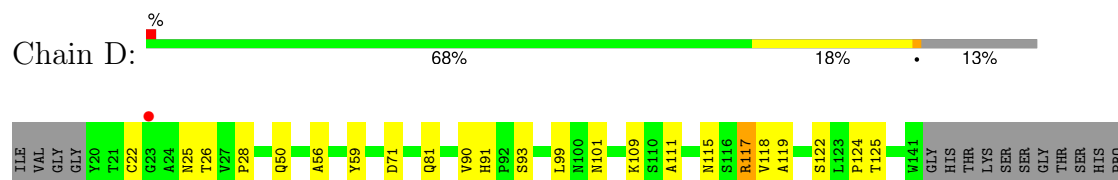
- Molecule 1: Cationic trypsin

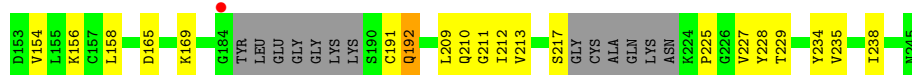


- Molecule 1: Cationic trypsin

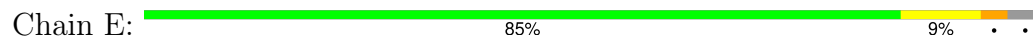


- Molecule 1: Cationic trypsin

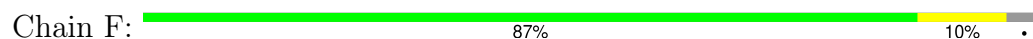




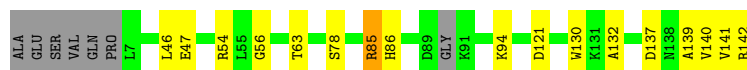
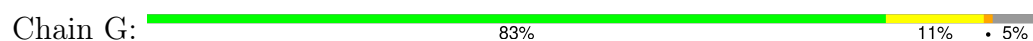
- Molecule 2: Ecotin



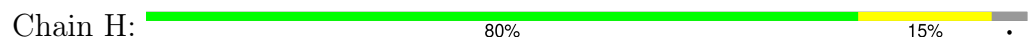
- Molecule 2: Ecotin



- Molecule 2: Ecotin



- Molecule 2: Ecotin



4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α , β , γ	94.24Å 78.61Å 98.08Å 90.00° 96.62° 90.00°	Depositor
Resolution (Å)	49.60 – 2.84 49.60 – 2.84	Depositor EDS
% Data completeness (in resolution range)	99.6 (49.60-2.84) 99.7 (49.60-2.84)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.21 (at 2.86Å)	Xtriage
Refinement program	PHENIX 1.8_1069	Depositor
R, R_{free}	0.192 , 0.252 0.194 , 0.254	Depositor DCC
R_{free} test set	1682 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	49.7	Xtriage
Anisotropy	0.717	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 30.2	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.018 for l,-k,h	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	10574	wwPDB-VP
Average B, all atoms (Å ²)	26.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 9.25% 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: CA, 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	0.48	0/1616	0.58	0/2189
1	B	0.47	0/1575	0.58	0/2132
1	C	0.50	0/1580	0.58	0/2139
1	D	0.46	0/1460	0.59	0/1978
2	E	0.52	0/1131	0.93	3/1528 (0.2%)
2	F	0.55	0/1135	0.66	2/1533 (0.1%)
2	G	0.50	0/1118	0.71	2/1509 (0.1%)
2	H	0.52	0/1126	0.63	0/1522
All	All	0.50	0/10741	0.65	7/14530 (0.0%)

Chiral center outliers are detected by calculating the chiral volume of a chiral center and verifying if the center is modelled as a planar moiety or with the opposite hand. A planarity outlier is detected by checking planarity of atoms in a peptide group, atoms in a mainchain group or atoms of a sidechain that are expected to be planar.

Mol	Chain	#Chirality outliers	#Planarity outliers
1	A	0	2
1	C	0	1
1	D	0	1
2	E	0	1
All	All	0	5

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	85	ARG	NE-CZ-NH1	-16.91	111.84	120.30
2	E	85	ARG	NE-CZ-NH2	15.35	127.97	120.30
2	G	85	ARG	NE-CZ-NH2	-9.54	115.53	120.30
2	G	85	ARG	NE-CZ-NH1	7.80	124.20	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	E	85	ARG	CD-NE-CZ	7.54	134.16	123.60

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	152	PRO	Peptide
1	A	59	TYR	Peptide
1	C	59	TYR	Peptide
1	D	59	TYR	Peptide
2	E	91	LYS	Peptide

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	1585	0	1543	14	0
1	B	1546	0	1501	14	0
1	C	1552	0	1509	18	0
1	D	1435	0	1395	24	0
2	E	1107	0	1115	9	0
2	F	1112	0	1118	7	0
2	G	1096	0	1103	14	0
2	H	1103	0	1105	18	0
3	A	1	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
4	B	1	0	0	0	0
5	A	5	0	0	0	0
5	B	7	0	0	0	0
5	C	9	0	0	0	0
5	D	2	0	0	0	0
5	E	2	0	0	0	0
5	F	5	0	0	0	0
5	G	2	0	0	0	0
5	H	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
All	All	10574	0	10389	103	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 5.

The worst 5 of 103 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:E:89:ASP:HB2	2:E:90:GLY:HA3	1.44	0.95
1:D:115:ASN:OD1	1:D:117:ARG:HG3	1.85	0.76
2:G:139:ALA:HB1	2:H:23:GLN:HB3	1.70	0.72
1:B:138:ILE:HG12	1:B:199:VAL:HG22	1.74	0.70
2:E:89:ASP:HB2	2:E:90:GLY:CA	2.21	0.67

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	212/223 (95%)	199 (94%)	12 (6%)	1 (0%)	25	44
1	B	203/223 (91%)	198 (98%)	5 (2%)	0	100	100
1	C	206/223 (92%)	199 (97%)	7 (3%)	0	100	100
1	D	187/223 (84%)	174 (93%)	11 (6%)	2 (1%)	12	24
2	E	135/142 (95%)	127 (94%)	6 (4%)	2 (2%)	8	18
2	F	133/142 (94%)	130 (98%)	2 (2%)	1 (1%)	16	32
2	G	131/142 (92%)	128 (98%)	3 (2%)	0	100	100
2	H	132/142 (93%)	126 (96%)	5 (4%)	1 (1%)	16	32
All	All	1339/1460 (92%)	1281 (96%)	51 (4%)	7 (0%)	25	44

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	192	GLN
2	E	17	GLU
2	H	17	GLU
1	D	225	PRO
2	F	17	GLU

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	178/184 (97%)	171 (96%)	7 (4%)	27	52
1	B	175/184 (95%)	168 (96%)	7 (4%)	27	51
1	C	175/184 (95%)	168 (96%)	7 (4%)	27	51
1	D	164/184 (89%)	158 (96%)	6 (4%)	29	54
2	E	122/126 (97%)	117 (96%)	5 (4%)	26	50
2	F	123/126 (98%)	120 (98%)	3 (2%)	44	69
2	G	121/126 (96%)	117 (97%)	4 (3%)	33	58
2	H	122/126 (97%)	119 (98%)	3 (2%)	42	68
All	All	1180/1240 (95%)	1138 (96%)	42 (4%)	30	55

5 of 42 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	85	ARG
2	G	54	ARG
2	E	86	HIS
2	F	54	ARG
2	G	86	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	135	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 5 ligands modelled in this entry, 5 are 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(Å²)	Q<0.9	
1	A	216/223 (96%)	-0.75	0	100	100	4, 21, 53, 67	0
1	B	211/223 (94%)	-0.77	0	100	100	9, 22, 42, 62	0
1	C	212/223 (95%)	-0.84	0	100	100	5, 17, 47, 76	0
1	D	195/223 (87%)	-0.01	2 (1%)	79	77	18, 44, 61, 67	0
2	E	137/142 (96%)	-0.66	0	100	100	9, 23, 53, 58	0
2	F	137/142 (96%)	-0.78	0	100	100	4, 17, 39, 46	0
2	G	135/142 (95%)	-0.43	0	100	100	13, 26, 48, 59	0
2	H	136/142 (95%)	-0.62	0	100	100	10, 24, 55, 73	0
All	All	1379/1460 (94%)	-0.61	2 (0%)	92	94	4, 24, 54, 76	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	184	GLY	2.6
1	D	23	GLY	2.3

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](#)

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
4	ZN	B	302	1/1	0.92	0.06	81,81,81,81	0
3	CA	D	301	1/1	0.94	0.07	56,56,56,56	0
3	CA	A	301	1/1	0.95	0.05	43,43,43,43	0
3	CA	C	301	1/1	0.95	0.04	32,32,32,32	0
3	CA	B	301	1/1	0.98	0.03	20,20,20,20	0

6.5 Other polymers [i](#)

There are no such residues in this entry.