



wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 5, 2026 – 08:23 PM UTC

PDB ID : 4GXP / pdb_00004gxp
Title : Chimeric Family 1 beta-glucosidase made with non-contiguous SCHEMA
Authors : Smith, M.A.; Romero, P.A.; Wu, T.; Brustad, E.M.; Arnold, F.H.
Deposited on : 2012-09-04
Resolution : 3.00 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
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

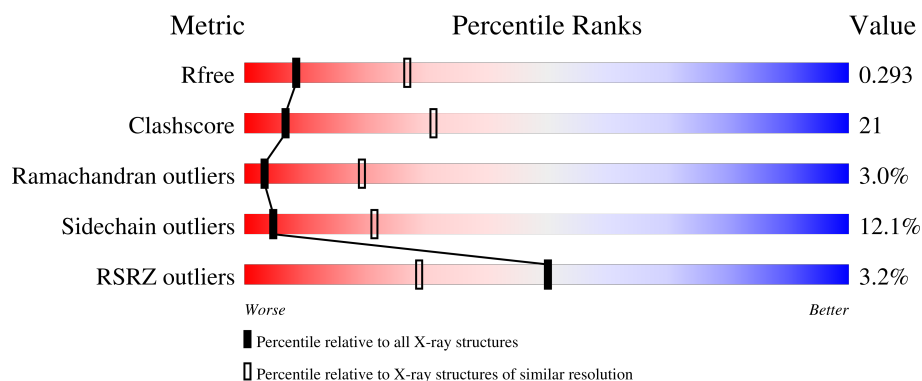
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.00 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}	180053	2672 (3.00-3.00)
Clashscore	190562	2977 (3.00-3.00)
Ramachandran outliers	187476	2877 (3.00-3.00)
Sidechain outliers	187428	2880 (3.00-3.00)
RSRZ outliers	180081	2671 (3.00-3.00)

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 $\leq 5\%$. 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	467	2% 61% 31% 5% .
1	B	467	% 56% 34% 6% . .
1	C	467	7% 55% 34% 8% . .

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 9869 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 Beta-glucosidase Chimeric protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	456	Total	C	N	O	S	0	0	0
			3451	2206	592	644	9			
1	B	456	Total	C	N	O	S	0	0	0
			3288	2084	577	618	9			
1	C	454	Total	C	N	O	S	0	0	0
			3128	1957	554	609	8			

There are 39 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	expression tag	UNP Q08638
A	2	HIS	-	expression tag	UNP Q08638
A	3	HIS	-	expression tag	UNP Q08638
A	4	HIS	-	expression tag	UNP Q08638
A	5	HIS	-	expression tag	UNP Q08638
A	6	HIS	-	expression tag	UNP Q08638
A	7	HIS	-	expression tag	UNP Q08638
A	127	PHE	TYR	engineered mutation	UNP Q08638
A	141	LEU	TRP	engineered mutation	UNP Q08638
A	142	LEU	ALA	engineered mutation	UNP Q08638
A	398	TYR	HIS	engineered mutation	UNP Q08638
A	340	ALA	GLN	engineered mutation	UNP O93785
A	341	MET	SER	engineered mutation	UNP O93785
B	1	MET	-	expression tag	UNP Q08638
B	2	HIS	-	expression tag	UNP Q08638
B	3	HIS	-	expression tag	UNP Q08638
B	4	HIS	-	expression tag	UNP Q08638
B	5	HIS	-	expression tag	UNP Q08638
B	6	HIS	-	expression tag	UNP Q08638
B	7	HIS	-	expression tag	UNP Q08638
B	127	PHE	TYR	engineered mutation	UNP Q08638
B	141	LEU	TRP	engineered mutation	UNP Q08638
B	142	LEU	ALA	engineered mutation	UNP Q08638

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	398	TYR	HIS	engineered mutation	UNP Q08638
B	340	ALA	GLN	engineered mutation	UNP O93785
B	341	MET	SER	engineered mutation	UNP O93785
C	1	MET	-	expression tag	UNP Q08638
C	2	HIS	-	expression tag	UNP Q08638
C	3	HIS	-	expression tag	UNP Q08638
C	4	HIS	-	expression tag	UNP Q08638
C	5	HIS	-	expression tag	UNP Q08638
C	6	HIS	-	expression tag	UNP Q08638
C	7	HIS	-	expression tag	UNP Q08638
C	127	PHE	TYR	engineered mutation	UNP Q08638
C	141	LEU	TRP	engineered mutation	UNP Q08638
C	142	LEU	ALA	engineered mutation	UNP Q08638
C	398	TYR	HIS	engineered mutation	UNP Q08638
C	340	ALA	GLN	engineered mutation	UNP O93785
C	341	MET	SER	engineered mutation	UNP O93785

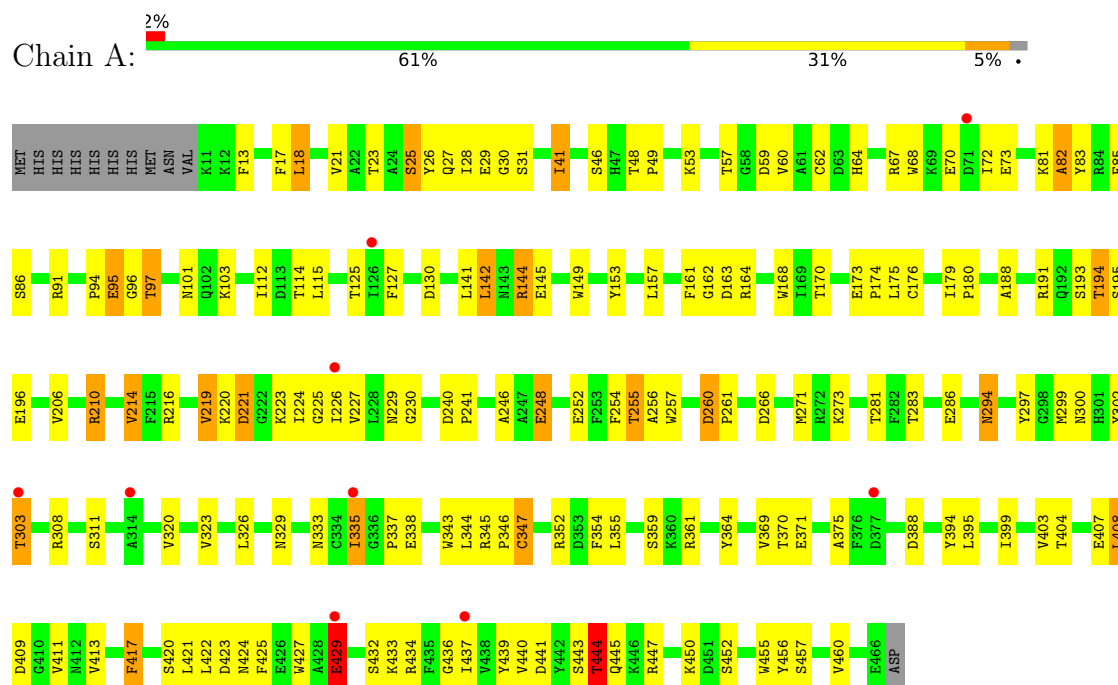
- Molecule 2 is water.

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

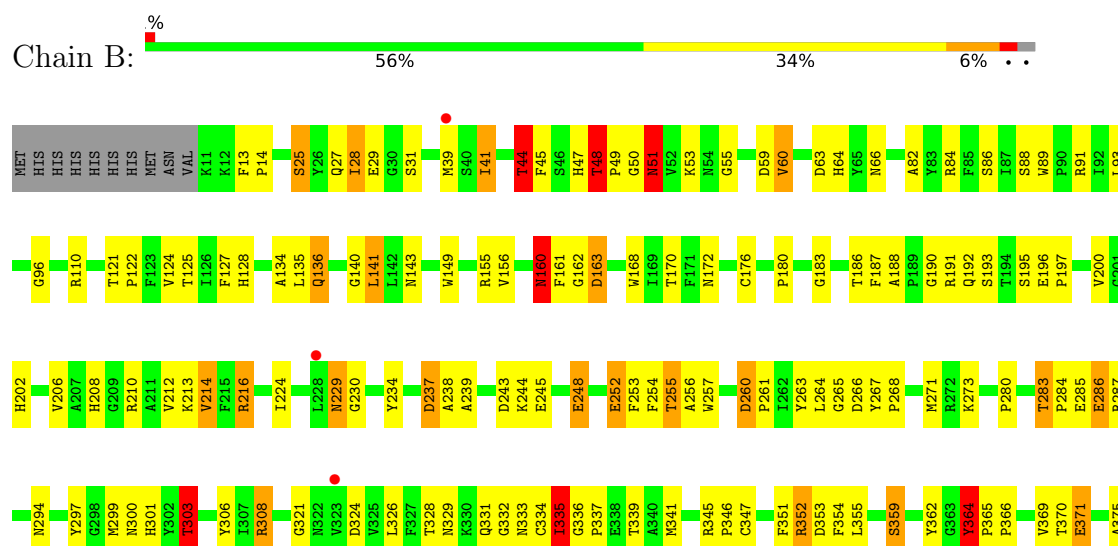
3 Residue-property plots

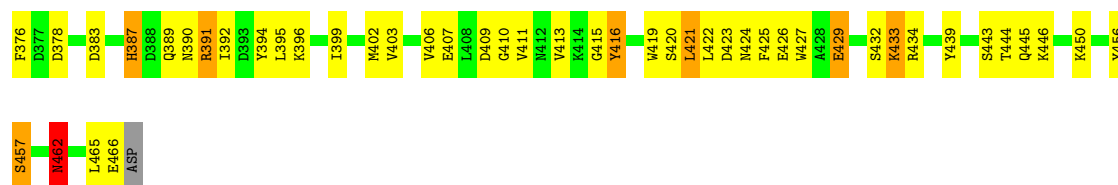
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: Beta-glucosidase Chimeric protein

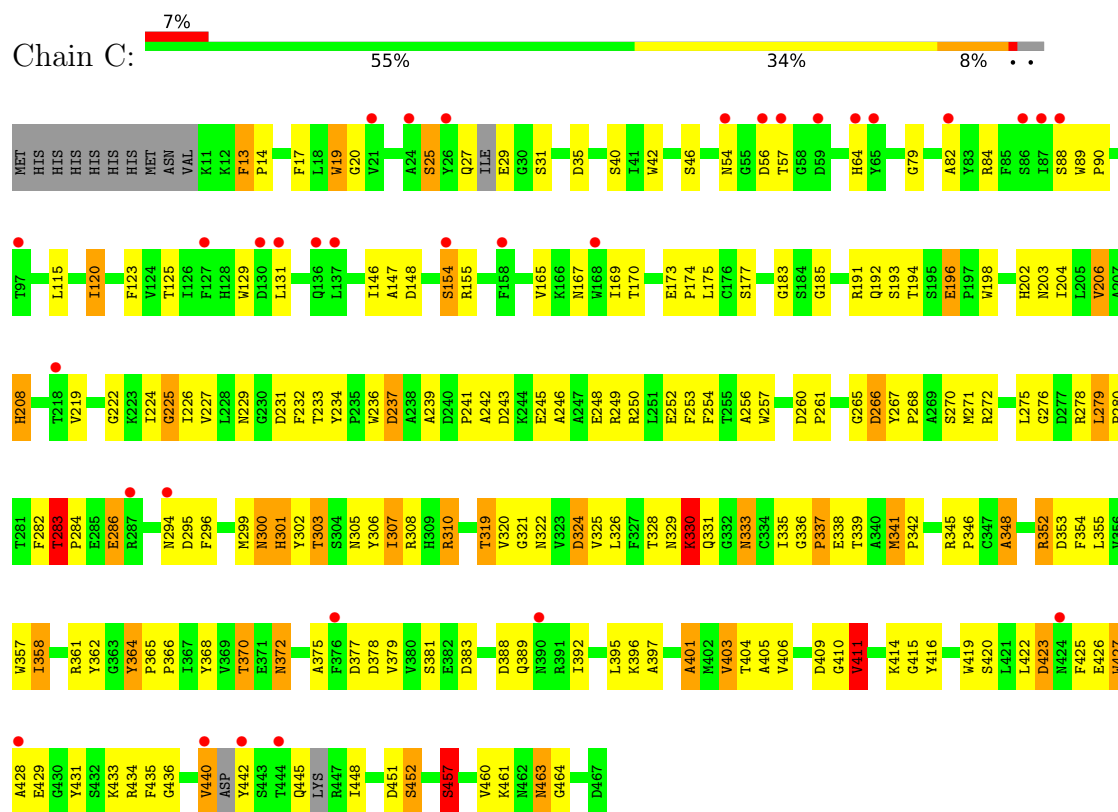


• Molecule 1: Beta-glucosidase Chimeric protein





• Molecule 1: Beta-glucosidase Chimeric protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 31 1 2	Depositor
Cell constants a, b, c, α , β , γ	115.38Å 115.38Å 282.54Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	37.43 – 3.00 37.43 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.5 (37.43-3.00) 99.5 (37.43-3.00)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.05	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.90 (at 3.01Å)	Xtriage
Refinement program	REFMAC 5.6.0117	Depositor
R, R_{free}	0.239 , 0.294 0.238 , 0.293	Depositor DCC
R_{free} test set	2188 reflections (5.03%)	wwPDB-VP
Wilson B-factor (Å ²)	92.2	Xtriage
Anisotropy	0.075	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.31 , 125.4	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	0.037 for -h,-k,l	Xtriage
F_o, F_c correlation	0.93	EDS
Total number of atoms	9869	wwPDB-VP
Average B, all atoms (Å ²)	130.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.87% 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

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	0.96	0/3563	1.19	18/4871 (0.4%)
1	B	1.15	10/3398 (0.3%)	1.28	35/4665 (0.8%)
1	C	1.17	12/3220 (0.4%)	1.20	34/4418 (0.8%)
All	All	1.09	22/10181 (0.2%)	1.23	87/13954 (0.6%)

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	B	0	1

The worst 5 of 22 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	401	ALA	C-O	15.82	1.42	1.24
1	C	404	THR	CB-OG1	11.64	1.62	1.43
1	C	301	HIS	ND1-CE1	9.87	1.42	1.32
1	C	348	ALA	C-O	7.40	1.32	1.24
1	C	457	SER	CB-OG	7.35	1.56	1.42

The worst 5 of 87 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	267	TYR	CA-C-N	8.45	128.95	119.83
1	B	267	TYR	C-N-CA	8.45	128.95	119.83
1	B	89	TRP	CA-C-N	-7.92	112.75	120.83
1	B	89	TRP	C-N-CA	-7.92	112.75	120.83
1	B	48	THR	CA-C-N	7.69	129.45	119.84

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	364	TYR	Sidechain

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	3451	0	3027	106	0
1	B	3288	0	2645	140	0
1	C	3128	0	2393	141	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
All	All	9869	0	8065	384	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 21.

The worst 5 of 384 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:B:299:MET:HE1	1:B:355:LEU:HD21	1.11	1.06
1:B:299:MET:HE1	1:B:355:LEU:CD2	1.84	1.06
1:B:303:THR:HG21	1:B:345:ARG:H	1.26	1.00
1:B:303:THR:HG22	1:B:345:ARG:O	1.66	0.95
1:B:283:THR:HG22	1:B:286:GLU:H	1.32	0.93

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	454/467 (97%)	393 (87%)	56 (12%)	5 (1%)	11	43
1	B	454/467 (97%)	390 (86%)	48 (11%)	16 (4%)	3	16
1	C	446/467 (96%)	338 (76%)	88 (20%)	20 (4%)	2	12
All	All	1354/1401 (97%)	1121 (83%)	192 (14%)	41 (3%)	3	19

5 of 41 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	220	LYS
1	B	445	GLN
1	B	462	ASN
1	C	35	ASP
1	C	266	ASP

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	321/391 (82%)	287 (89%)	34 (11%)	6	27
1	B	268/391 (68%)	233 (87%)	35 (13%)	4	19
1	C	232/391 (59%)	202 (87%)	30 (13%)	4	19
All	All	821/1173 (70%)	722 (88%)	99 (12%)	5	21

5 of 99 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	283	THR
1	C	125	THR
1	B	308	ARG
1	B	411	VAL
1	C	208	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	372	ASN
1	B	462	ASN
1	C	372	ASN
1	B	424	ASN
1	C	43	HIS

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](#)

There are no ligands in this entry.

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	456/467 (97%)	0.34	9 (1%) 65 41	93, 112, 124, 136	6 (1%)
1	B	456/467 (97%)	0.15	3 (0%) 84 66	94, 112, 129, 150	2 (0%)
1	C	454/467 (97%)	0.52	32 (7%) 22 12	94, 171, 225, 249	1 (0%)
All	All	1366/1401 (97%)	0.33	44 (3%) 50 29	93, 116, 208, 249	9 (0%)

The worst 5 of 44 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	130	ASP	6.6
1	C	26	TYR	5.3
1	C	82	ALA	3.8
1	B	323	VAL	3.3
1	A	437	ILE	3.1

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.