



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 02:13 am GMT

PDB ID : 3SQ5
Title : Crystal Structure Analysis of the Yeast Tyrosyl-DNA Phosphodiesterase H432N Mutant
Authors : Gajewski, S.; White, S.W.
Deposited on : 2011-07-04
Resolution : 2.30 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

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

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

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	trunk28620
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	recalc28949

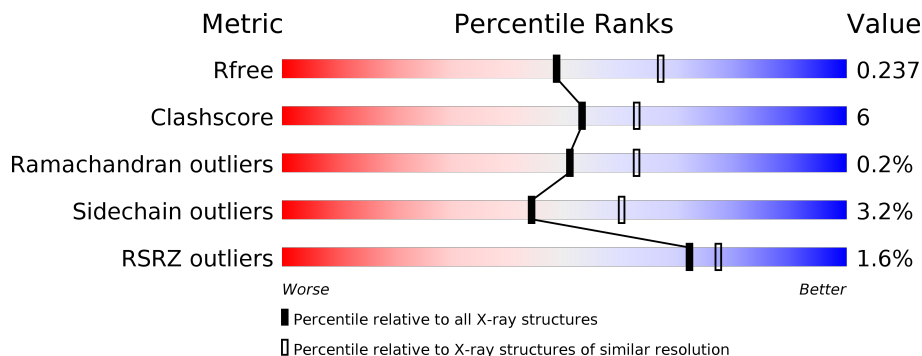
1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 2.30 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}	100719	4130 (2.30-2.30)
Clashscore	112137	4751 (2.30-2.30)
Ramachandran outliers	110173	4705 (2.30-2.30)
Sidechain outliers	110143	4704 (2.30-2.30)
RSRZ outliers	101464	4156 (2.30-2.30)

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 on the lower bar indicate the fraction of residues that contain outliers for ≥ 3 , 2, 1 and 0 types of geometric quality criteria. 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	470	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, green 75%, yellow 13%, grey 11%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> % 75% 13% • 11% </div> </div>
1	B	470	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, green 77%, yellow 11%, grey 11%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> % 77% 11% • 11% </div> </div>
1	C	470	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 3%, green 78%, yellow 11%, grey 10%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> 3% 78% 11% 10% </div> </div>
1	D	470	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, green 76%, yellow 11%, grey 11%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> % 76% 11% • 11% </div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 13812 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 Tyrosyl-DNA phosphodiesterase 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	419	Total	C	N	O	S	0	0	0
			3415	2215	560	620	20			
1	B	420	Total	C	N	O	S	0	1	0
			3426	2219	567	620	20			
1	C	422	Total	C	N	O	S	0	1	0
			3430	2225	563	622	20			
1	D	418	Total	C	N	O	S	0	1	0
			3416	2216	562	618	20			

There are 40 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	78	MET	-	INITIATING METHIONINE	UNP P38319
A	432	ASN	HIS	ENGINEERED MUTATION	UNP P38319
A	540	LEU	-	EXPRESSION TAG	UNP P38319
A	541	HIS	-	EXPRESSION TAG	UNP P38319
A	542	HIS	-	EXPRESSION TAG	UNP P38319
A	543	HIS	-	EXPRESSION TAG	UNP P38319
A	544	HIS	-	EXPRESSION TAG	UNP P38319
A	545	HIS	-	EXPRESSION TAG	UNP P38319
A	546	HIS	-	EXPRESSION TAG	UNP P38319
A	547	HIS	-	EXPRESSION TAG	UNP P38319
B	78	MET	-	INITIATING METHIONINE	UNP P38319
B	432	ASN	HIS	ENGINEERED MUTATION	UNP P38319
B	540	LEU	-	EXPRESSION TAG	UNP P38319
B	541	HIS	-	EXPRESSION TAG	UNP P38319
B	542	HIS	-	EXPRESSION TAG	UNP P38319
B	543	HIS	-	EXPRESSION TAG	UNP P38319
B	544	HIS	-	EXPRESSION TAG	UNP P38319
B	545	HIS	-	EXPRESSION TAG	UNP P38319
B	546	HIS	-	EXPRESSION TAG	UNP P38319
B	547	HIS	-	EXPRESSION TAG	UNP P38319
C	78	MET	-	INITIATING METHIONINE	UNP P38319

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	432	ASN	HIS	ENGINEERED MUTATION	UNP P38319
C	540	LEU	-	EXPRESSION TAG	UNP P38319
C	541	HIS	-	EXPRESSION TAG	UNP P38319
C	542	HIS	-	EXPRESSION TAG	UNP P38319
C	543	HIS	-	EXPRESSION TAG	UNP P38319
C	544	HIS	-	EXPRESSION TAG	UNP P38319
C	545	HIS	-	EXPRESSION TAG	UNP P38319
C	546	HIS	-	EXPRESSION TAG	UNP P38319
C	547	HIS	-	EXPRESSION TAG	UNP P38319
D	78	MET	-	INITIATING METHIONINE	UNP P38319
D	432	ASN	HIS	ENGINEERED MUTATION	UNP P38319
D	540	LEU	-	EXPRESSION TAG	UNP P38319
D	541	HIS	-	EXPRESSION TAG	UNP P38319
D	542	HIS	-	EXPRESSION TAG	UNP P38319
D	543	HIS	-	EXPRESSION TAG	UNP P38319
D	544	HIS	-	EXPRESSION TAG	UNP P38319
D	545	HIS	-	EXPRESSION TAG	UNP P38319
D	546	HIS	-	EXPRESSION TAG	UNP P38319
D	547	HIS	-	EXPRESSION TAG	UNP P38319

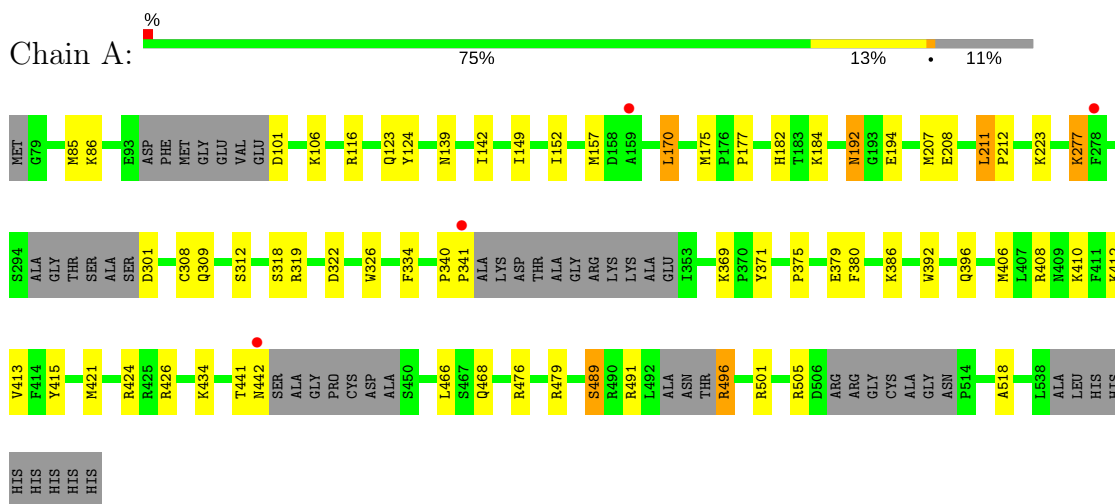
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	34	Total O 34 34	0	0
2	B	36	Total O 36 36	0	0
2	C	28	Total O 28 28	0	0
2	D	27	Total O 27 27	0	0

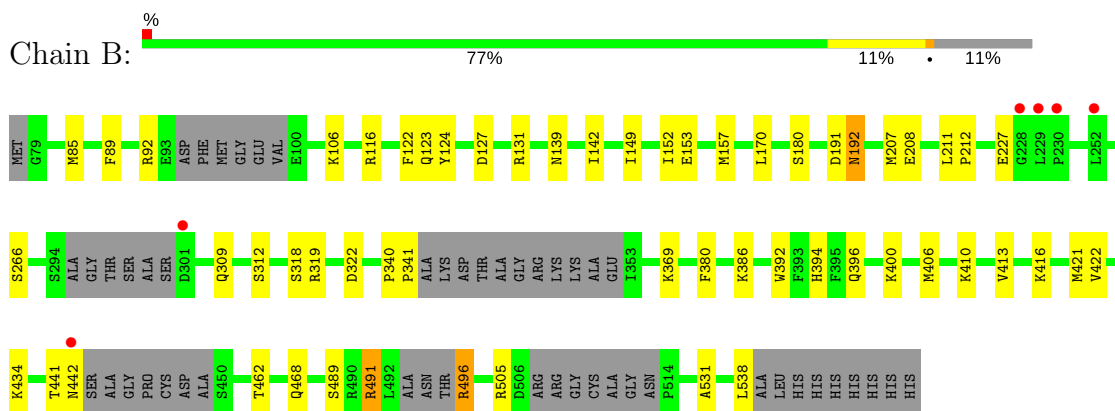
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA 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.

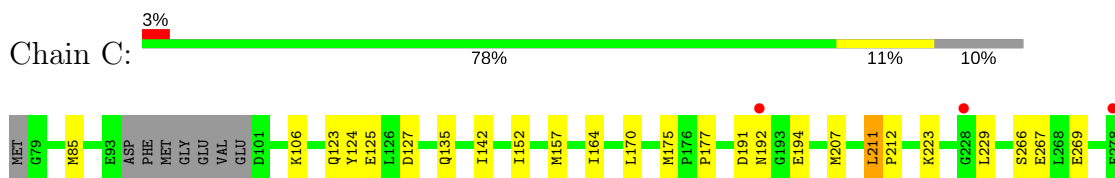
• Molecule 1: Tyrosyl-DNA phosphodiesterase 1

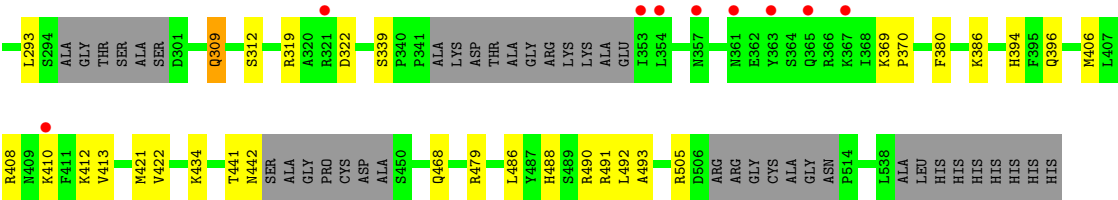


• Molecule 1: Tyrosyl-DNA phosphodiesterase 1

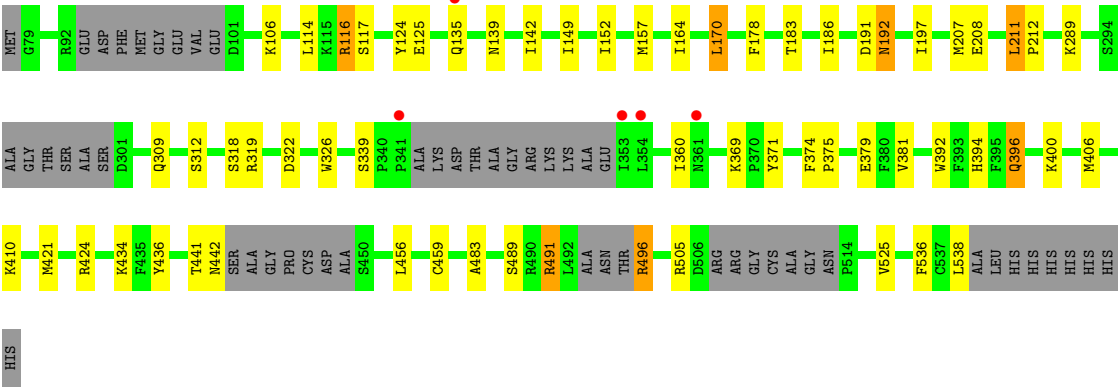
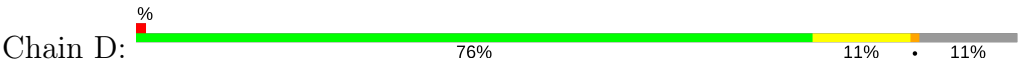


• Molecule 1: Tyrosyl-DNA phosphodiesterase 1





● Molecule 1: Tyrosyl-DNA phosphodiesterase 1



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	64.45Å 82.04Å 98.68Å 89.80° 94.33° 112.78°	Depositor
Resolution (Å)	50.00 – 2.30 49.38 – 1.90	Depositor EDS
% Data completeness (in resolution range)	96.1 (50.00-2.30) 69.1 (49.38-1.90)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.02 (at 1.90Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.203 , 0.236 0.204 , 0.237	Depositor DCC
R_{free} test set	3872 reflections (5.13%)	DCC
Wilson B-factor (Å ²)	27.3	Xtriage
Anisotropy	0.181	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.34 , 42.3	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	0.004 for -h,h+k,-l	Xtriage
F_o, F_c correlation	0.94	EDS
Total number of atoms	13812	wwPDB-VP
Average B, all atoms (Å ²)	53.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.16% 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](#)

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	1.14	4/3501 (0.1%)	0.97	6/4734 (0.1%)
1	B	1.11	3/3512 (0.1%)	0.91	3/4749 (0.1%)
1	C	1.00	2/3518 (0.1%)	0.91	3/4761 (0.1%)
1	D	1.02	6/3503 (0.2%)	0.89	4/4737 (0.1%)
All	All	1.07	15/14034 (0.1%)	0.92	16/18981 (0.1%)

The worst 5 of 15 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	308	CYS	CB-SG	-9.31	1.66	1.82
1	A	424	ARG	CA-CB	6.61	1.68	1.53
1	D	381	VAL	CB-CG1	6.26	1.66	1.52
1	D	125	GLU	CG-CD	5.70	1.60	1.51
1	C	125	GLU	CG-CD	5.66	1.60	1.51

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	85	MET	CG-SD-CE	-8.67	86.33	100.20
1	A	85	MET	CG-SD-CE	-8.52	86.57	100.20
1	C	85	MET	CG-SD-CE	-8.41	86.74	100.20
1	C	175	MET	CA-CB-CG	-6.80	101.74	113.30
1	A	175	MET	CA-CB-CG	-6.67	101.97	113.30

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	3415	0	3409	41	0
1	B	3426	0	3417	45	0
1	C	3430	0	3410	44	0
1	D	3416	0	3409	42	0
2	A	34	0	0	6	0
2	B	36	0	0	4	0
2	C	28	0	0	1	0
2	D	27	0	0	1	0
All	All	13812	0	13645	168	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 168 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:89:PHE:HD2	1:C:229:LEU:HD11	1.13	1.07
1:B:89:PHE:CD2	1:C:229:LEU:HD11	1.91	1.05
1:C:191:ASP:O	1:C:192:ASN:HB2	1.66	0.95
1:B:191:ASP:O	1:B:192:ASN:HB2	1.63	0.95
1:D:322:ASP:HB2	1:D:406:MET:CE	1.97	0.94

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	405/470 (86%)	396 (98%)	8 (2%)	1 (0%)	51	63
1	B	407/470 (87%)	396 (97%)	11 (3%)	0	100	100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	C	411/470 (87%)	400 (97%)	10 (2%)	1 (0%)	51 63
1	D	405/470 (86%)	395 (98%)	9 (2%)	1 (0%)	51 63
All	All	1628/1880 (87%)	1587 (98%)	38 (2%)	3 (0%)	51 63

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	211	LEU
1	A	211	LEU
1	C	211	LEU

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	386/423 (91%)	369 (96%)	17 (4%)	33 45
1	B	386/423 (91%)	375 (97%)	11 (3%)	49 65
1	C	385/423 (91%)	377 (98%)	8 (2%)	59 76
1	D	386/423 (91%)	372 (96%)	14 (4%)	40 55
All	All	1543/1692 (91%)	1493 (97%)	50 (3%)	44 60

5 of 50 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	369	LYS
1	B	496	ARG
1	D	489	SER
1	B	413	VAL
1	B	489	SER

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 26 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	288	ASN
1	C	123	GLN
1	D	288	ASN
1	B	309	GLN
1	B	432	ASN

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 carbohydrates 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	419/470 (89%)	-0.37	4 (0%) 82 86	32, 48, 74, 91	0
1	B	420/470 (89%)	-0.47	6 (1%) 75 80	32, 48, 79, 96	0
1	C	422/470 (89%)	-0.12	12 (2%) 53 61	32, 54, 93, 115	0
1	D	418/470 (88%)	-0.33	5 (1%) 79 82	35, 53, 84, 99	0
All	All	1679/1880 (89%)	-0.32	27 (1%) 72 77	32, 51, 83, 115	0

The worst 5 of 27 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	135	GLN	5.1
1	C	321	ARG	4.7
1	C	361	ASN	4.3
1	A	442	ASN	4.0
1	C	365	GLN	3.7

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 carbohydrates 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.