



wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 12:24 AM GMT

PDB ID : 2AAF
Title : Structure of H278A arginine deiminase with L-arginine forming a S-alkylthiuronium reaction intermediate
Authors : Galkin, A.; Lu, X.; Dunaway-Mariano, D.; Herzberg, O.
Deposited on : 2005-07-13
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
Mogul : 1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix) : 1.9-1692
EDS : rb-20026688
Percentile statistics : 20151230.v01 (using entries in the PDB archive December 30th 2015)
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) : trunk26865

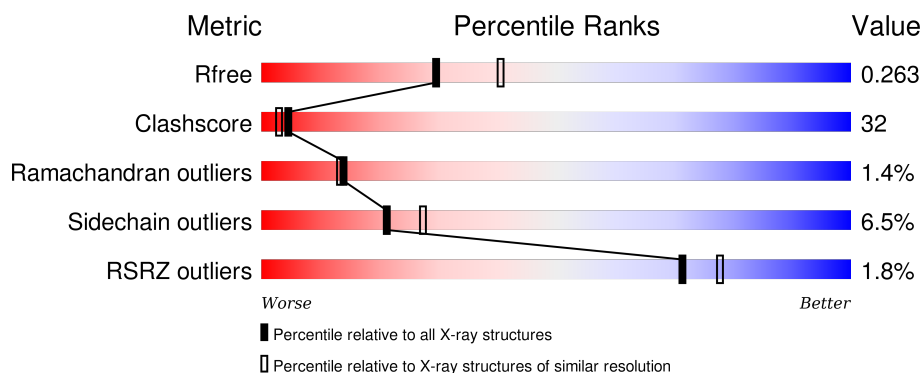
1 Overall quality at a glance

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}	91344	3852 (2.30-2.30)
Clashscore	102246	4452 (2.30-2.30)
Ramachandran outliers	100387	4410 (2.30-2.30)
Sidechain outliers	100360	4409 (2.30-2.30)
RSRZ outliers	91569	3857 (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	418	<div> <div>53%</div> <div>38%</div> <div>5%</div> <div>• •</div> </div>
1	B	418	<div> <div>61%</div> <div>32%</div> <div>• •</div> </div>
1	C	418	<div> <div>2%</div> <div>53%</div> <div>40%</div> <div>• •</div> </div>
1	D	418	<div> <div>4%</div> <div>51%</div> <div>41%</div> <div>• •</div> </div>

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 13542 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 Arginine deiminase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	405	Total	C	N	O	S	0	0	0
			3170	2004	553	596	17			
1	B	409	Total	C	N	O	S	0	0	0
			3202	2024	557	604	17			
1	C	406	Total	C	N	O	S	0	0	0
			3185	2014	554	600	17			
1	D	406	Total	C	N	O	S	0	0	0
			3183	2013	554	599	17			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	278	ALA	HIS	ENGINEERED	UNP P13981
A	406	CYR	CYS	MODIFIED RESIDUE	UNP P13981
B	278	ALA	HIS	ENGINEERED	UNP P13981
B	406	CYR	CYS	MODIFIED RESIDUE	UNP P13981
C	278	ALA	HIS	ENGINEERED	UNP P13981
C	406	CYR	CYS	MODIFIED RESIDUE	UNP P13981
D	278	ALA	HIS	ENGINEERED	UNP P13981
D	406	CYR	CYS	MODIFIED RESIDUE	UNP P13981

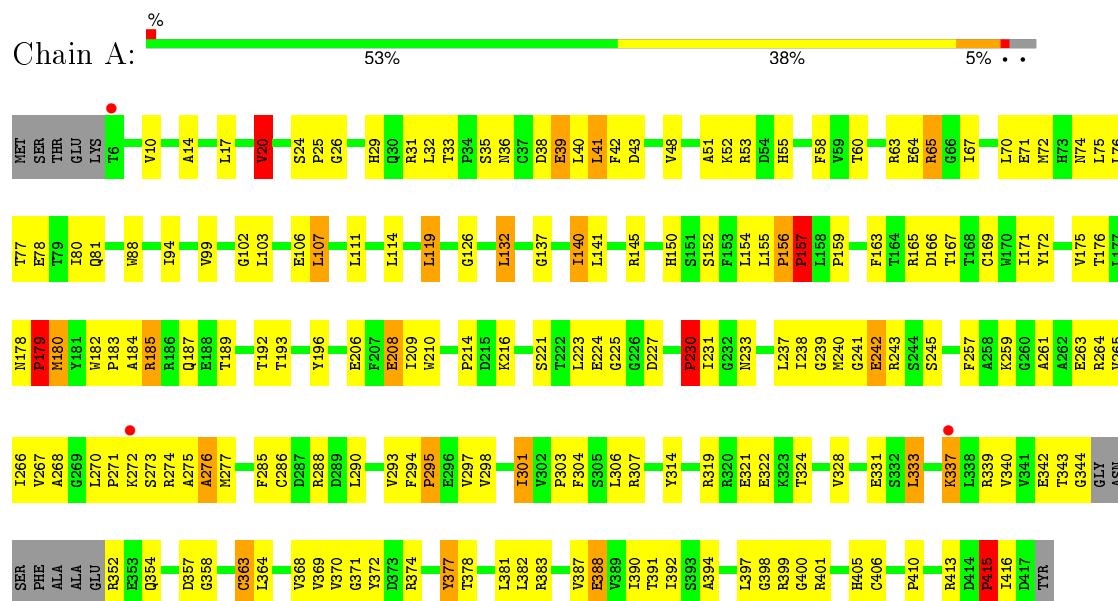
- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	227	Total	O	0	0
			227	227		
2	B	209	Total	O	0	0
			209	209		
2	C	187	Total	O	0	0
			187	187		
2	D	179	Total	O	0	0
			179	179		

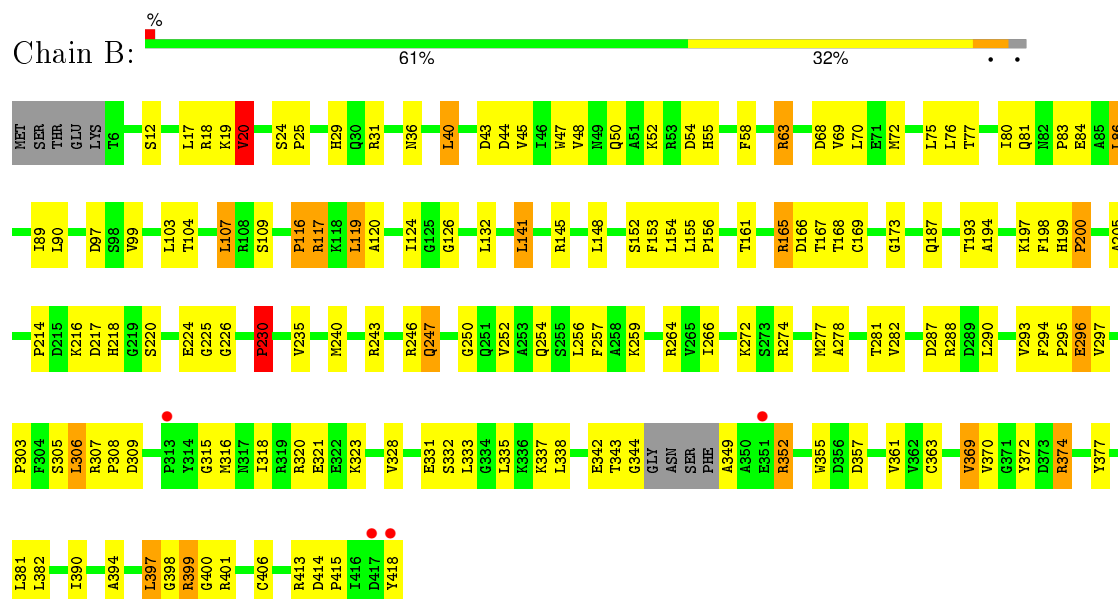
3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: Arginine deiminase



• Molecule 1: Arginine deiminase



[illegible]

Chain D:

Sequence logo for Chain D showing amino acid conservation across 400 positions. The y-axis represents the log-odds of an amino acid being at a position given the background frequency. The x-axis shows positions from 1 to 400. Amino acids are color-coded: red for high conservation, orange for medium, yellow for low, and grey for very low. Key conserved residues include MET, SER, THR, GLU, LYS, T6, K7, V10, K16, L17, R18, K19, V20, C23, S24, P25, G26, L27, A28, R31, N36, C37, D38, E39, L40, L41, F42, D43, E45, F56, D57, F58, V59, T60, K61, H62, R63, E64, R65, V69, L70, E71, N74, L75, L76, L80, E84, A85, L86, L90, D91, R92, K93, I94, D97, S98, V99, G100, I101, H102, L103, T104, S105, E106, L107, R108, S109, W110, L111, L114, E115, L119, I124, G125, L132, F135, S135, A138, H139, I140, L141, K142, T143, Y144, R145, G149, E150, S151, S152, F153, L154, L155, L158, P159, H160, T161, Q162, R165, T167, T168, C169, W170, K171, I172, V175, T176, L177, N178, P179, H180, Y181, W182, P183, A184, R185, R186, T189, W190, L191, I195, F198, H199, P200, E201, D213, P214, D217, H218, G219, S220, S221, T222, L223, E224, G225, G226, D227, P230, T231, T238, G239, N240, G241, E242, R243, R246, A253, T257, V265, I266, W267, G268, G269, L270, L271, K272, S273, R274, A275, A276, A277, A278, S284, F285, G286, D287, R288, D289, L290, V291, T292, V293, P294, P295, E296, V297, R298, R299, E300, F304, S305, L306, R307, P308, D309, P310, S311, S312, P313, Y314, G315, G316, R317, L318, R319, K323, T324, F325, L326, E327, V328, V329, A330, L333, L338, R339, V340, G344, GLY, ASN, SER, PHE, ALA, GLU, R352, R353, R354, Q354, R355, D356, D357, G358, R359, R360, E365, P366, G367, R374, T376, Y377, T378, R383, V387, I392, S393, A394, S395, E396, L397, G398, R399, G400, R401, H405, C406, P410, F411, V412, I416, D417, W419.

4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	90.80Å 121.20Å 151.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.30 19.91 – 2.30	Depositor EDS
% Data completeness (in resolution range)	(Not available) (20.00-2.30) 98.8 (19.91-2.30)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.33 (at 2.30Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.200 , 0.264 0.200 , 0.263	Depositor DCC
R_{free} test set	3759 reflections (5.09%)	DCC
Wilson B-factor (Å ²)	28.2	Xtriage
Anisotropy	0.290	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.32 , 60.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 73810 reflections	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	13542	wwPDB-VP
Average B, all atoms (Å ²)	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 23.50 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 4.6062e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical values of $\langle |L| \rangle$, $\langle L^2 \rangle$ for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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: CYR

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	0.72	1/3218 (0.0%)	1.01	8/4363 (0.2%)
1	B	0.77	0/3251	0.98	4/4407 (0.1%)
1	C	0.70	1/3234 (0.0%)	0.94	3/4383 (0.1%)
1	D	0.72	1/3232 (0.0%)	0.97	4/4381 (0.1%)
All	All	0.73	3/12935 (0.0%)	0.97	19/17534 (0.1%)

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	1

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	169	CYS	CB-SG	-5.37	1.73	1.81
1	C	37	CYS	CB-SG	-5.19	1.73	1.81
1	A	415	PRO	N-CD	5.11	1.55	1.47

The worst 5 of 19 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	277	MET	C-N-CA	12.39	152.68	121.70
1	D	278	ALA	N-CA-CB	9.03	122.74	110.10
1	B	230	PRO	CA-N-CD	-8.89	99.06	111.50
1	A	157	PRO	CA-N-CD	-7.63	100.81	111.50
1	A	20	VAL	CB-CA-C	-7.32	97.50	111.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	196	TYR	Sidechain

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	3170	0	3160	224	0
1	B	3202	0	3185	187	0
1	C	3185	0	3168	209	0
1	D	3183	0	3168	226	0
2	A	227	0	0	81	0
2	B	209	0	0	53	0
2	C	187	0	0	79	0
2	D	179	0	0	73	0
All	All	13542	0	12681	820	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 32.

The worst 5 of 820 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:C:364:LEU:HG	2:C:604:HOH:O	1.29	1.27
1:A:364:LEU:HB3	2:A:634:HOH:O	1.30	1.27
1:B:187:GLN:HG2	2:B:423:HOH:O	1.33	1.26
1:A:192:THR:HB	2:A:623:HOH:O	1.26	1.24
1:A:74:ASN:HB3	2:A:620:HOH:O	1.36	1.22

There are no symmetry-related clashes.

5.3 Torsion angles

5.3.1 Protein backbone

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	400/418 (96%)	363 (91%)	32 (8%)	5 (1%)	15	15
1	B	404/418 (97%)	374 (93%)	26 (6%)	4 (1%)	19	21
1	C	401/418 (96%)	371 (92%)	28 (7%)	2 (0%)	34	41
1	D	401/418 (96%)	349 (87%)	41 (10%)	11 (3%)	6	4
All	All	1606/1672 (96%)	1457 (91%)	127 (8%)	22 (1%)	14	13

5 of 22 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	275	ALA
1	D	277	MET
1	B	399	ARG
1	D	40	LEU
1	D	273	SER

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	343/353 (97%)	319 (93%)	24 (7%)	19	23
1	B	345/353 (98%)	323 (94%)	22 (6%)	22	28
1	C	344/353 (98%)	320 (93%)	24 (7%)	19	23
1	D	344/353 (98%)	325 (94%)	19 (6%)	27	36
All	All	1376/1412 (98%)	1287 (94%)	89 (6%)	21	27

5 of 89 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	306	LEU
1	C	86	LEU
1	D	243	ARG
1	B	369	VAL
1	B	415	PRO

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	49	ASN
1	D	199	HIS
1	C	160	ASN
1	B	317	ASN
1	C	199	HIS

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

4 non-standard protein/DNA/RNA residues 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	$\# Z > 2$
1	CYR	A	406	1	11,16,17	2.27	2 (18%)	6,19,21	3.31	2 (33%)
1	CYR	B	406	1	11,16,17	2.21	2 (18%)	6,19,21	2.74	3 (50%)
1	CYR	C	406	1	11,16,17	2.21	2 (18%)	6,19,21	3.48	2 (33%)
1	CYR	D	406	1	11,16,17	2.16	2 (18%)	6,19,21	3.15	2 (33%)

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
1	CYR	A	406	1	-	0/10/18/20	0/0/0/0
1	CYR	B	406	1	-	0/10/18/20	0/0/0/0
1	CYR	C	406	1	-	0/10/18/20	0/0/0/0
1	CYR	D	406	1	-	0/10/18/20	0/0/0/0

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	406	CYR	C7-N7	4.47	1.45	1.28
1	C	406	CYR	C7-N7	4.69	1.46	1.28
1	B	406	CYR	C7-N7	4.78	1.46	1.28
1	A	406	CYR	C7-N7	5.03	1.47	1.28
1	B	406	CYR	C7-N6	5.28	1.47	1.36

The worst 5 of 9 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	406	CYR	C5-N6-C7	-7.84	107.44	123.44
1	D	406	CYR	C5-N6-C7	-7.01	109.14	123.44
1	A	406	CYR	C5-N6-C7	-6.46	110.26	123.44
1	B	406	CYR	C5-N6-C7	-5.63	111.94	123.44
1	A	406	CYR	N6-C7-N7	-4.37	109.51	120.88

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

4 monomers are involved in 9 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
1	A	406	CYR	1	0
1	B	406	CYR	2	0
1	C	406	CYR	4	0
1	D	406	CYR	2	0

5.5 Carbohydrates

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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	404/418 (96%)	-0.43	3 (0%) 89 92	14, 31, 57, 74	0
1	B	408/418 (97%)	-0.51	4 (0%) 84 88	12, 26, 55, 77	0
1	C	405/418 (96%)	-0.43	7 (1%) 73 79	13, 34, 59, 76	0
1	D	405/418 (96%)	-0.30	16 (3%) 42 51	16, 34, 63, 79	0
All	All	1622/1672 (97%)	-0.42	30 (1%) 71 78	12, 31, 59, 79	0

The worst 5 of 30 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	418	TYR	8.3
1	D	184	ALA	7.0
1	D	275	ALA	4.8
1	C	313	PRO	4.3
1	D	276	ALA	3.7

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

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å ²)	Q<0.9
1	CYR	C	406	17/18	0.96	0.11	-	18,51,67,68	0
1	CYR	D	406	17/18	0.97	0.09	-	20,36,65,68	0
1	CYR	A	406	17/18	0.95	0.13	-	15,37,68,68	0
1	CYR	B	406	17/18	0.96	0.10	-	18,34,61,63	0

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.