



wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 1, 2016 – 02:07 PM EDT

PDB ID : 5SY5
Title : Crystal Structure of the Heterodimeric NPAS1-ARNT Complex
Authors : Wu, D.; Su, X.; Potluri, N.; Kim, Y.; Rastinejad, F.
Deposited on : 2016-08-10
Resolution : 3.20 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	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20028320
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)	:	rb-20028320

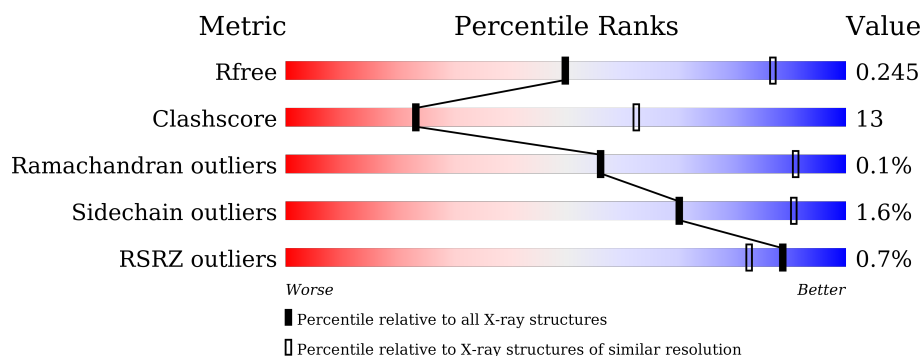
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 3.20 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	1124 (3.24-3.16)
Clashscore	102246	1024 (3.22-3.18)
Ramachandran outliers	100387	1004 (3.22-3.18)
Sidechain outliers	100360	1003 (3.22-3.18)
RSRZ outliers	91569	1129 (3.24-3.16)

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	384	
1	C	384	
1	E	384	
2	B	391	
2	D	391	
2	F	391	

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 13303 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 Aryl hydrocarbon receptor nuclear translocator.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	278	Total	C	N	O	S	0	0	0
			2240	1418	395	408	19			
1	C	269	Total	C	N	O	S	0	0	0
			2161	1374	374	397	16			
1	E	279	Total	C	N	O	S	0	0	0
			2258	1431	399	409	19			

There are 3 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	81	MET	-	initiating methionine	UNP P53762
C	81	MET	-	initiating methionine	UNP P53762
E	81	MET	-	initiating methionine	UNP P53762

- Molecule 2 is a protein called Neuronal PAS domain-containing protein 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	278	Total	C	N	O	S	0	0	0
			2191	1407	391	387	6			
2	D	283	Total	C	N	O	S	0	0	0
			2228	1433	396	393	6			
2	F	282	Total	C	N	O	S	0	0	0
			2225	1428	398	393	6			

There are 30 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	41	MET	-	initiating methionine	UNP P97459
B	42	ILE	-	expression tag	UNP P97459
B	424	LEU	-	expression tag	UNP P97459
B	425	GLU	-	expression tag	UNP P97459
B	426	HIS	-	expression tag	UNP P97459

Continued on next page...

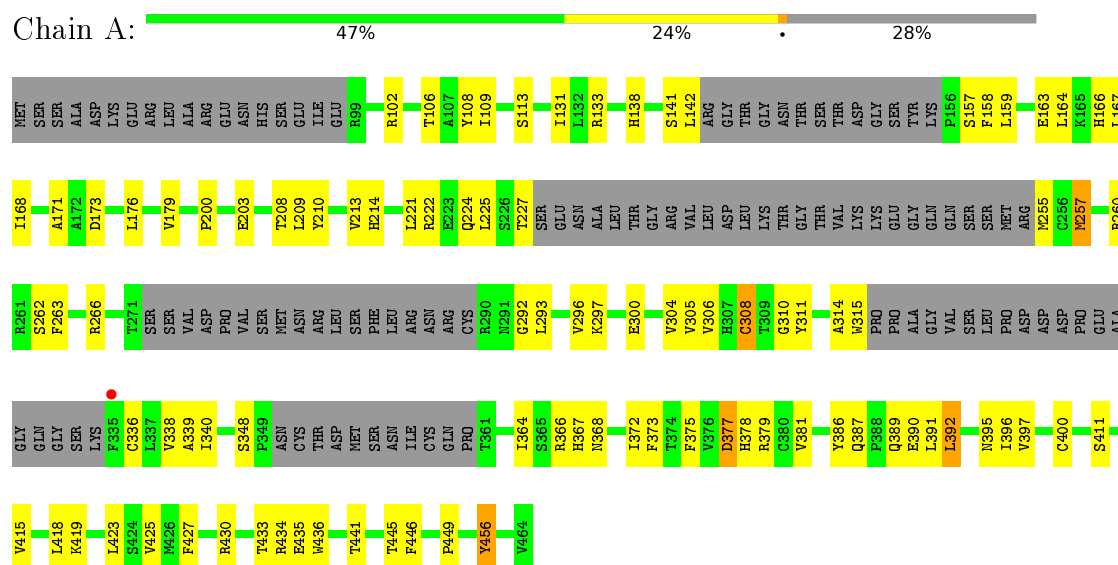
Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	427	HIS	-	expression tag	UNP P97459
B	428	HIS	-	expression tag	UNP P97459
B	429	HIS	-	expression tag	UNP P97459
B	430	HIS	-	expression tag	UNP P97459
B	431	HIS	-	expression tag	UNP P97459
D	41	MET	-	initiating methionine	UNP P97459
D	42	ILE	-	expression tag	UNP P97459
D	424	LEU	-	expression tag	UNP P97459
D	425	GLU	-	expression tag	UNP P97459
D	426	HIS	-	expression tag	UNP P97459
D	427	HIS	-	expression tag	UNP P97459
D	428	HIS	-	expression tag	UNP P97459
D	429	HIS	-	expression tag	UNP P97459
D	430	HIS	-	expression tag	UNP P97459
D	431	HIS	-	expression tag	UNP P97459
F	41	MET	-	initiating methionine	UNP P97459
F	42	ILE	-	expression tag	UNP P97459
F	424	LEU	-	expression tag	UNP P97459
F	425	GLU	-	expression tag	UNP P97459
F	426	HIS	-	expression tag	UNP P97459
F	427	HIS	-	expression tag	UNP P97459
F	428	HIS	-	expression tag	UNP P97459
F	429	HIS	-	expression tag	UNP P97459
F	430	HIS	-	expression tag	UNP P97459
F	431	HIS	-	expression tag	UNP P97459

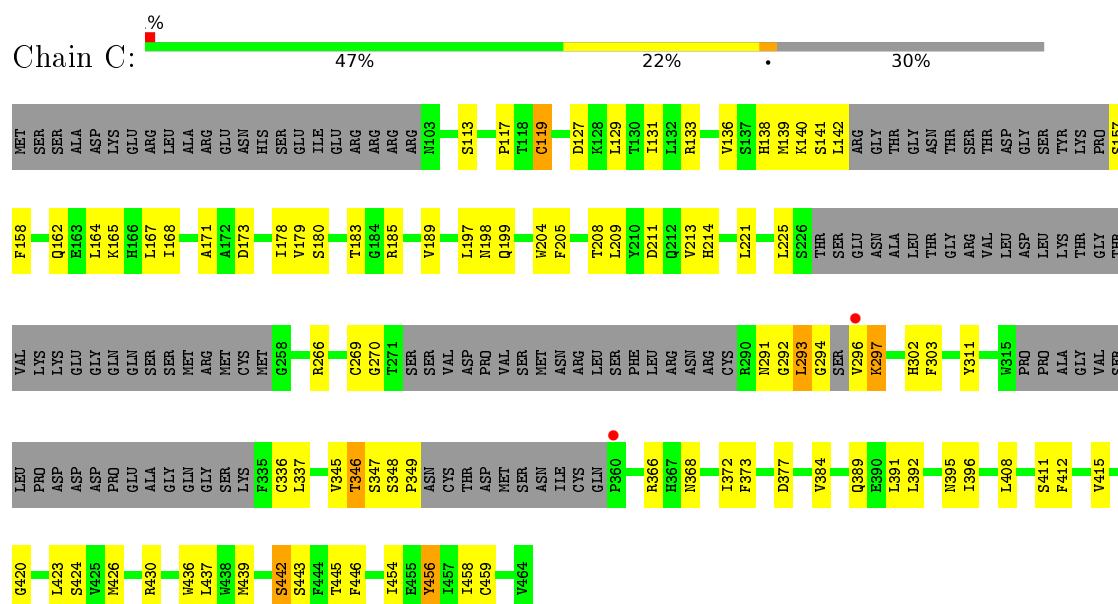
3 Residue-property plots


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: Aryl hydrocarbon receptor nuclear translocator



- Molecule 1: Aryl hydrocarbon receptor nuclear translocator



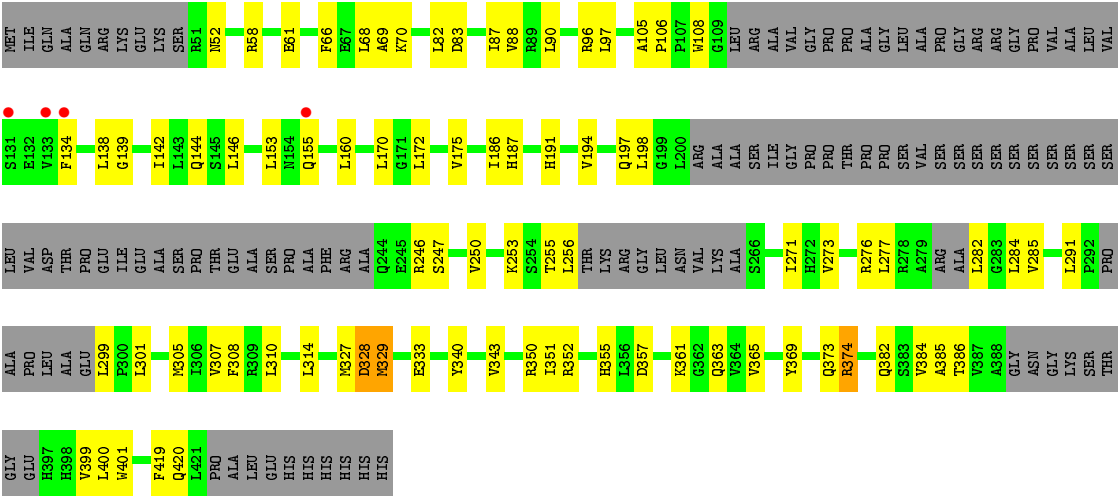
- Chain E: 
- 53% 18% 27%
- Metabolites (Amino Acids): MET, SER, SER, ALA, ASP, LYS, GLU, ARG, LEU, ALA, ARG, GLU, ASN, HIS, SER, GLU, ILE, E98, R99, R100, R101, R102, M105, E111, L112, C119, S120, A121, L122, K125, P126, D127, I131, V136, M139, R140, S141, L142, R143, GLY, THR, GLY, GLY, ASN, THR, THR, THR, ASP, GLY, S153, S157, F158
- Metabolites (Sugars): D161, K165, H166, L167, L168, L169, E170, A171, A172, D173, G174, V179, V189, S192, P195, M198, Q201, W204, F205, L221, L225, S226, THR, SER, GLU, ASN, ALA, THR, GLY, ARG, VAL, LEU, ASP, LEU, LYS, THR, GLY, THR, VAL, LYS, LYS, GLU, GLY, GLN, SER, SER, MET, ASP
- Metabolites (Nucleotides): W255, C256, R260, C269, GLY, THR, SER, SER, VAL, ASP, PRO, VAL, SER, MET, ASN, ARG, LEU, SER, PHE, LEU, ARG, ASN, ARG, CYS, N291, G292, L293, G294, S295, V296, LYS, GLU, G299, H302, V305, W315, PRO, PRO, ALA, GLY, VAL, SER, PRO, ASP, ASP, PRO, ALA, GLY, GLN, GLY, SER
- Metabolites (Lipids): E455, Y456, M461, T462, M463, Y464, LYS, F336, C336, G341, R342, L343, Q344, S347, S348, F349, ASN, CYS, THR, ASP, MET, SER, ASN, ILE, CYS, GLN, P360, T361, E362, N368, I372, D377, H378, R379, T383, V384, D404, S411, V415, G420, S424, F427, R430, W436, T441, S442, T445, F446, T454

- [illegible]

- [illegible]



● Molecule 2: Neuronal PAS domain-containing protein 1



4 Data and refinement statistics

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α , β , γ	69.95Å 81.19Å 138.06Å 90.38° 95.08° 107.39°	Depositor
Resolution (Å)	38.93 – 3.20 38.93 – 3.20	Depositor EDS
% Data completeness (in resolution range)	82.3 (38.93-3.20) 82.6 (38.93-3.20)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.60 (at 3.18Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.176 , 0.247 0.178 , 0.245	Depositor DCC
R_{free} test set	1997 reflections (5.11%)	DCC
Wilson B-factor (Å ²)	55.8	Xtriage
Anisotropy	0.138	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.28 , 31.0	EDS
L-test for twinning ²	$\langle L \rangle = 0.49$, $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	13303	wwPDB-VP
Average B, all atoms (Å ²)	49.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.86% 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	0.59	1/2284 (0.0%)	0.77	3/3079 (0.1%)
1	C	0.58	0/2204	0.73	2/2974 (0.1%)
1	E	0.54	2/2303 (0.1%)	0.68	1/3103 (0.0%)
2	B	0.56	0/2239	0.76	1/3026 (0.0%)
2	D	0.58	2/2276 (0.1%)	0.78	0/3076
2	F	0.52	0/2273	0.73	2/3070 (0.1%)
All	All	0.56	5/13579 (0.0%)	0.74	9/18328 (0.0%)

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	292	PRO	N-CD	-7.00	1.38	1.47
1	E	99	ARG	CZ-NH1	-5.95	1.25	1.33
1	E	99	ARG	CZ-NH2	-5.67	1.25	1.33
2	D	285	VAL	CB-CG1	-5.25	1.41	1.52
1	A	308	CYS	CB-SG	-5.11	1.73	1.81

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

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	F	299	LEU	CA-CB-CG	8.69	135.28	115.30
2	B	299	LEU	CA-CB-CG	6.59	130.45	115.30
1	E	198	ASN	C-N-CA	6.40	137.70	121.70
1	A	225	LEU	CA-CB-CG	6.27	129.71	115.30
1	C	164	LEU	CB-CG-CD2	-5.95	100.89	111.00

There are no chirality outliers.

There are no planarity outliers.

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	2240	0	2228	77	1
1	C	2161	0	2140	75	0
1	E	2258	0	2244	56	1
2	B	2191	0	2189	66	0
2	D	2228	0	2230	65	0
2	F	2225	0	2223	66	0
All	All	13303	0	13254	355	1

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 13.

The worst 5 of 355 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:293:LEU:HD13	1:A:305:VAL:HG21	1.53	0.90
2:B:93:THR:HG21	2:B:179:GLY:HA2	1.66	0.77
2:F:308:PHE:HB2	2:F:401:TRP:HB3	1.65	0.76
2:D:154:ASN:HB3	2:D:160:LEU:HD11	1.65	0.76
2:D:319:CYS:SG	2:D:323:VAL:HG21	2.26	0.74

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:133:ARG:NH2	1:E:404:ASP:OD1[1_445]	2.18	0.02

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	266/384 (69%)	239 (90%)	27 (10%)	0	100	100
1	C	255/384 (66%)	234 (92%)	20 (8%)	1 (0%)	39	80
1	E	265/384 (69%)	240 (91%)	25 (9%)	0	100	100
2	B	264/391 (68%)	249 (94%)	15 (6%)	0	100	100
2	D	269/391 (69%)	252 (94%)	17 (6%)	0	100	100
2	F	268/391 (68%)	255 (95%)	13 (5%)	0	100	100
All	All	1587/2325 (68%)	1469 (93%)	117 (7%)	1 (0%)	56	91

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	346	THR

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	255/347 (74%)	251 (98%)	4 (2%)	70	91
1	C	246/347 (71%)	240 (98%)	6 (2%)	57	86
1	E	257/347 (74%)	251 (98%)	6 (2%)	58	87
2	B	234/322 (73%)	233 (100%)	1 (0%)	93	98
2	D	238/322 (74%)	236 (99%)	2 (1%)	86	96
2	F	238/322 (74%)	234 (98%)	4 (2%)	68	90
All	All	1468/2007 (73%)	1445 (98%)	23 (2%)	70	91

5 of 23 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	456	TYR
2	D	134	PHE

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	F	329	MET
2	D	52	ASN
1	E	99	ARG

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	E	166	HIS
2	F	404	HIS
1	E	302	HIS
2	D	373	GLN
2	F	373	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 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	278/384 (72%)	-0.45	1 (0%) 93 90	20, 43, 90, 116	0
1	C	269/384 (70%)	-0.32	2 (0%) 89 83	22, 48, 88, 124	0
1	E	279/384 (72%)	-0.29	5 (1%) 71 58	29, 54, 94, 126	0
2	B	278/391 (71%)	-0.51	0 100 100	19, 40, 76, 105	0
2	D	283/391 (72%)	-0.44	0 100 100	20, 40, 83, 109	0
2	F	282/391 (72%)	-0.35	4 (1%) 78 65	29, 46, 83, 124	0
All	All	1669/2325 (71%)	-0.39	12 (0%) 89 83	19, 45, 88, 126	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	295	SER	3.9
2	F	134	PHE	3.3
1	C	360	PRO	3.1
2	F	155	GLN	2.8
1	C	296	VAL	2.6

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no carbohydrates in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.