



wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 22, 2023 – 12:35 PM EDT

PDB ID : 2RKJ
Title : Cocrystal structure of a tyrosyl-tRNA synthetase splicing factor with a group I intron RNA
Authors : Paukstelis, P.J.; Chen, J.-H.; Chase, E.; Lambowitz, A.M.; Golden, B.L.
Deposited on : 2007-10-16
Resolution : 4.50 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) : 1.13
EDS : 2.35
Percentile statistics : 20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac : 5.8.0158
CCP4 : 7.0.044 (Gargrove)
Ideal geometry (proteins) : Engh & Huber (2001)
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP) : 2.35

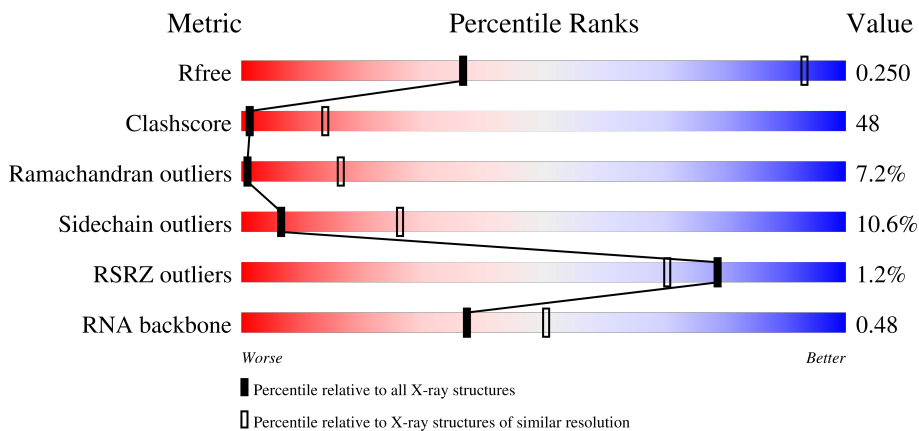
1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

The reported resolution of this entry is 4.50 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.



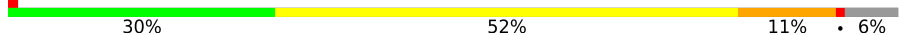
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1055 (5.20-3.80)
Clashscore	141614	1123 (5.20-3.80)
Ramachandran outliers	138981	1069 (5.20-3.80)
Sidechain outliers	138945	1050 (5.20-3.80)
RSRZ outliers	127900	1101 (5.30-3.70)
RNA backbone	3102	1063 (6.00-3.00)

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	C	246	
1	G	246	
1	K	246	
1	O	246	

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	4	 75% 25%
2	H	4	 75% 25%
2	L	4	 25% 50% 25%
2	P	4	 75% 25%
3	A	392	 27% 57% 9% • 6%
3	B	392	 30% 52% 11% • 6%
3	E	392	 25% 58% 10% • 6%
3	F	392	 30% 52% 11% • 6%
3	I	392	 26% 57% 10% • 6%
3	J	392	 30% 52% 11% • 6%
3	M	392	 26% 58% 9% • 6%
3	N	392	 30% 51% 11% • 6%

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 44328 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 RNA chain called RNA (238-MER).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	C	238	5096	2282	938	1638	238	0	0	0
1	G	238	5096	2282	938	1638	238	0	0	0
1	K	238	5096	2282	938	1638	238	0	0	0
1	O	238	5096	2282	938	1638	238	0	0	0

- Molecule 2 is a RNA chain called RNA (5'-R(P*GP*CP*UP*U)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	4	84	37	12	31	4	0	0	0
2	H	4	84	37	12	31	4	0	0	0
2	L	4	84	37	12	31	4	0	0	0
2	P	4	84	37	12	31	4	0	0	0

- Molecule 3 is a protein called Tyrosyl-tRNA synthetase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	A	367	2951	1876	515	544	16	0	0	0
3	B	367	2951	1876	515	544	16	0	0	0
3	E	367	2951	1876	515	544	16	0	0	0
3	F	367	2951	1876	515	544	16	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	I	367	Total	C	N	O	S	0	0	0
			2951	1876	515	544	16			
3	J	367	Total	C	N	O	S	0	0	0
			2951	1876	515	544	16			
3	M	367	Total	C	N	O	S	0	0	0
			2951	1876	515	544	16			
3	N	367	Total	C	N	O	S	0	0	0
			2951	1876	515	544	16			

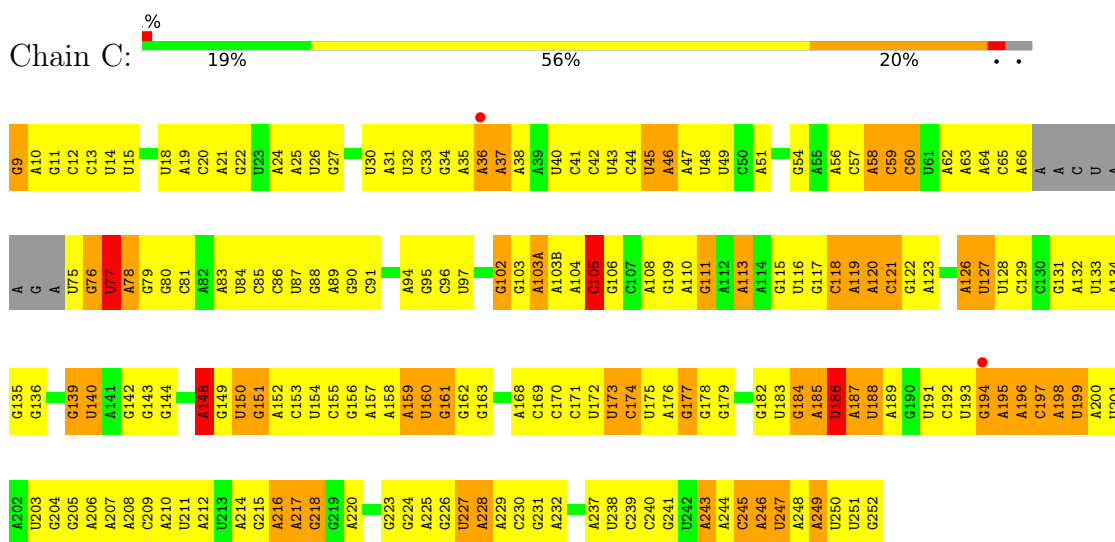
There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	32	MET	-	initiating methionine	UNP P12063
B	32	MET	-	initiating methionine	UNP P12063
E	32	MET	-	initiating methionine	UNP P12063
F	32	MET	-	initiating methionine	UNP P12063
I	32	MET	-	initiating methionine	UNP P12063
J	32	MET	-	initiating methionine	UNP P12063
M	32	MET	-	initiating methionine	UNP P12063
N	32	MET	-	initiating methionine	UNP P12063

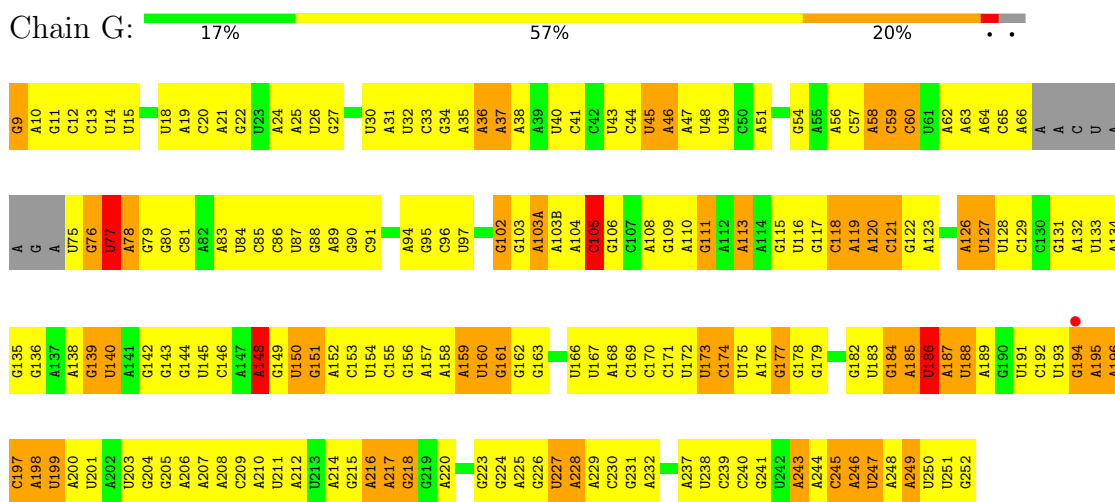
3 Residue-property plots [i](#)

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: RNA (238-MER)

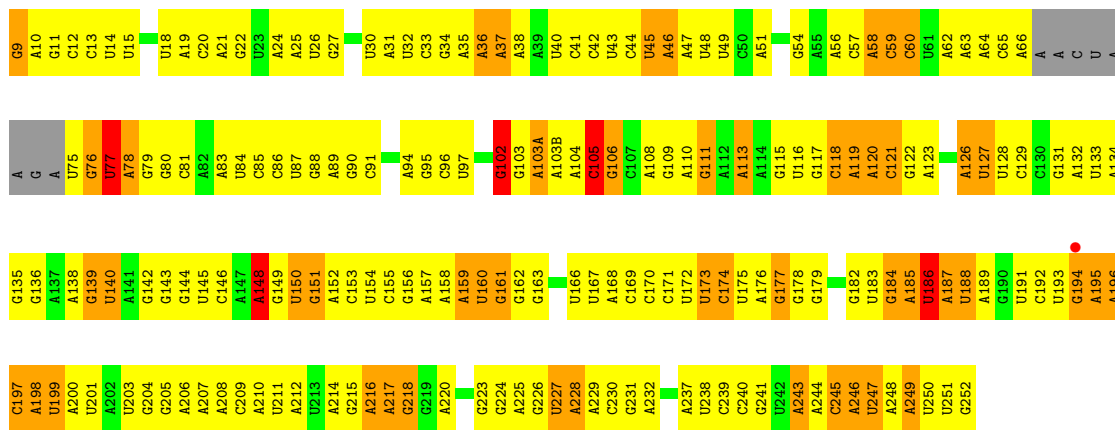


- Molecule 1: RNA (238-MER)

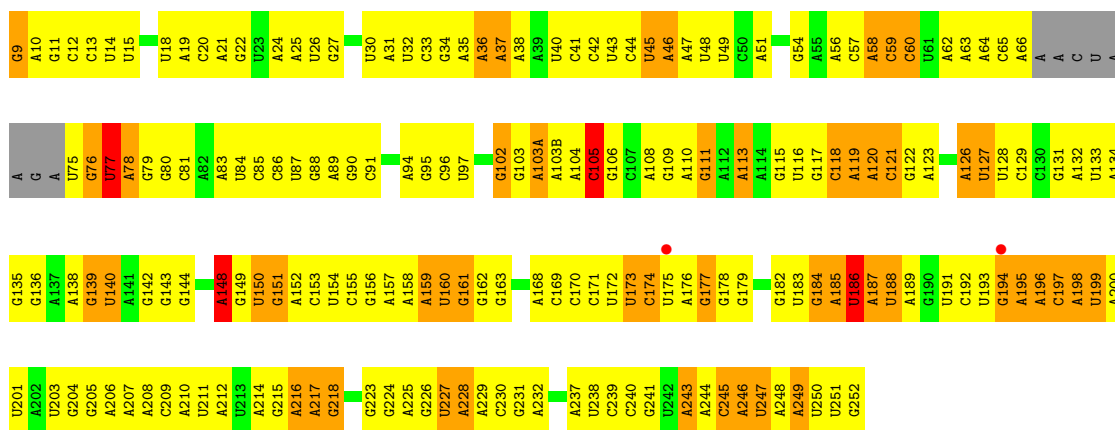
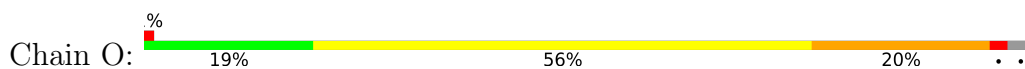


- Molecule 1: RNA (238-MER)

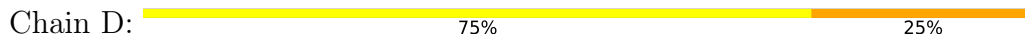




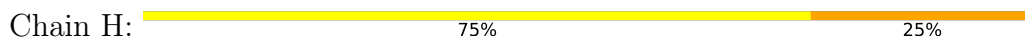
- Molecule 1: RNA (238-MER)



- Molecule 2: RNA (5'-R(P*GP*CP*UP*U)-3')



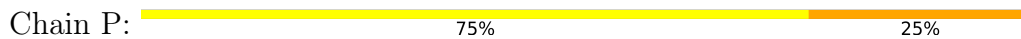
- Molecule 2: RNA (5'-R(P*GP*CP*UP*U)-3')



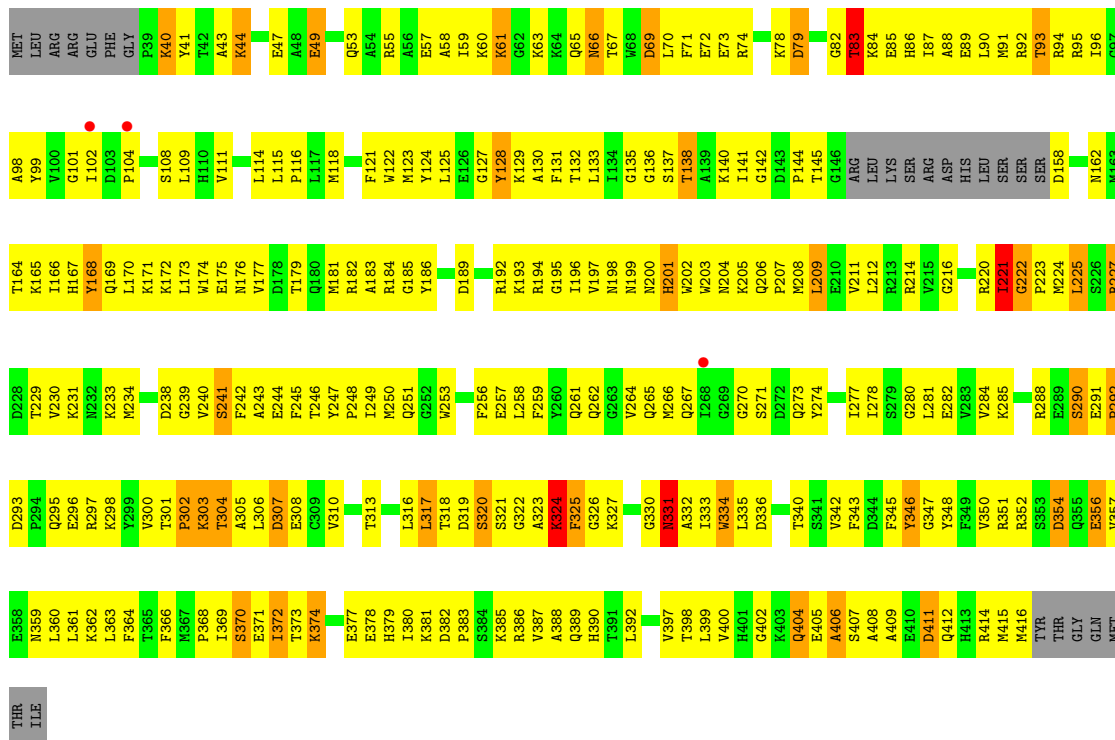
- Molecule 2: RNA (5'-R(P*GP*CP*UP*U)-3')



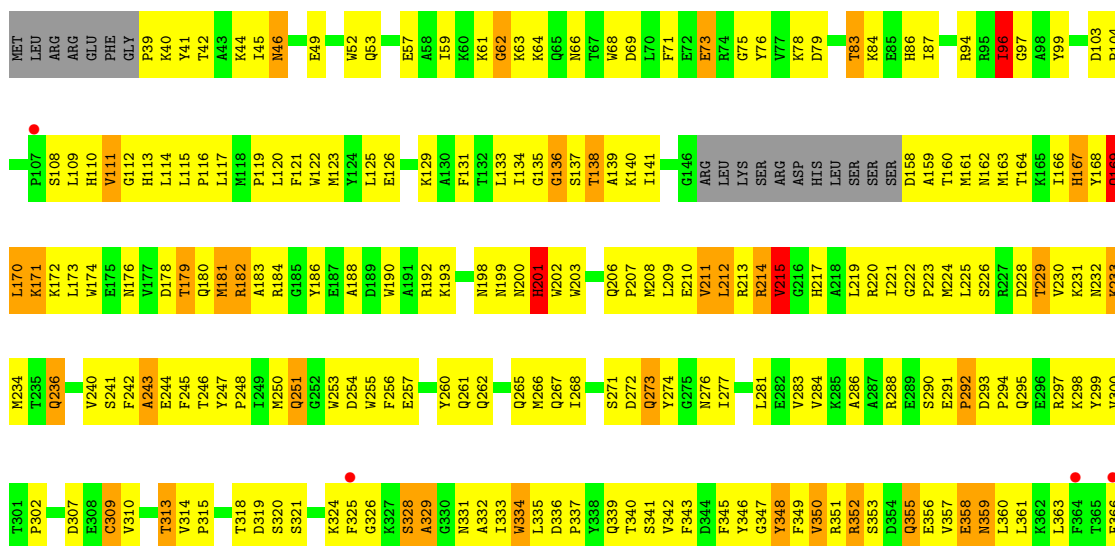
- Molecule 2: RNA (5'-R(P*GP*CP*UP*U)-3')



• Molecule 3: Tyrosyl-tRNA synthetase

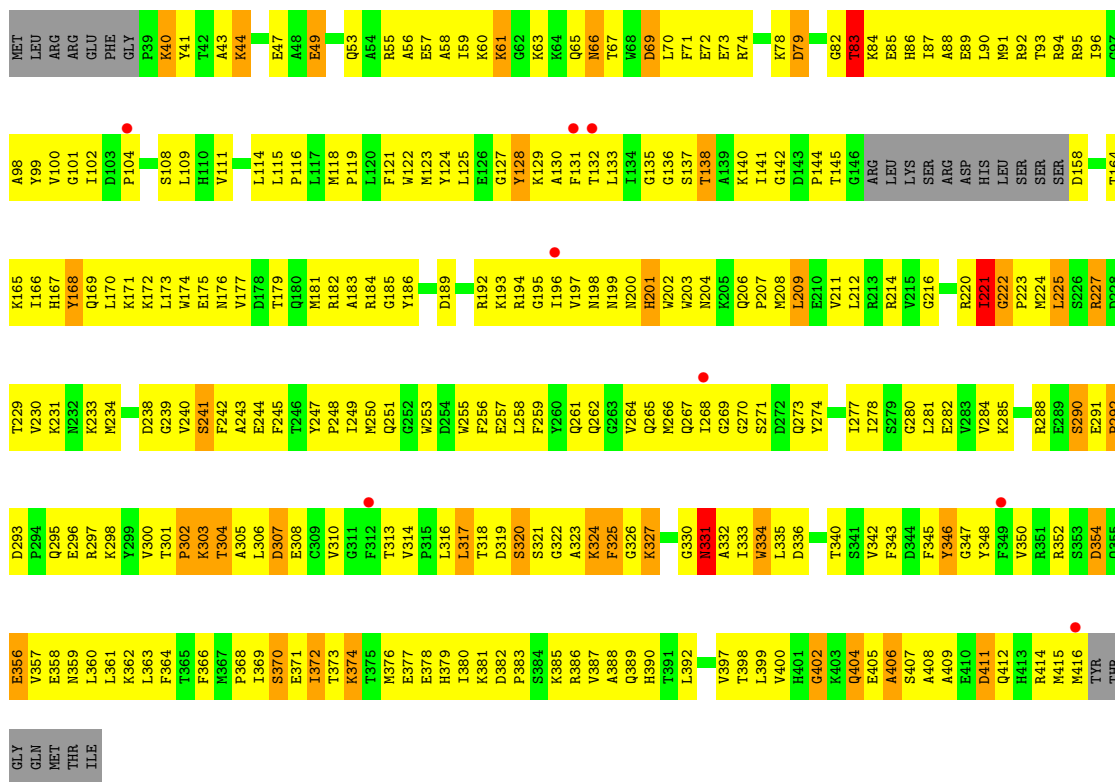


• Molecule 3: Tyrosyl-tRNA synthetase

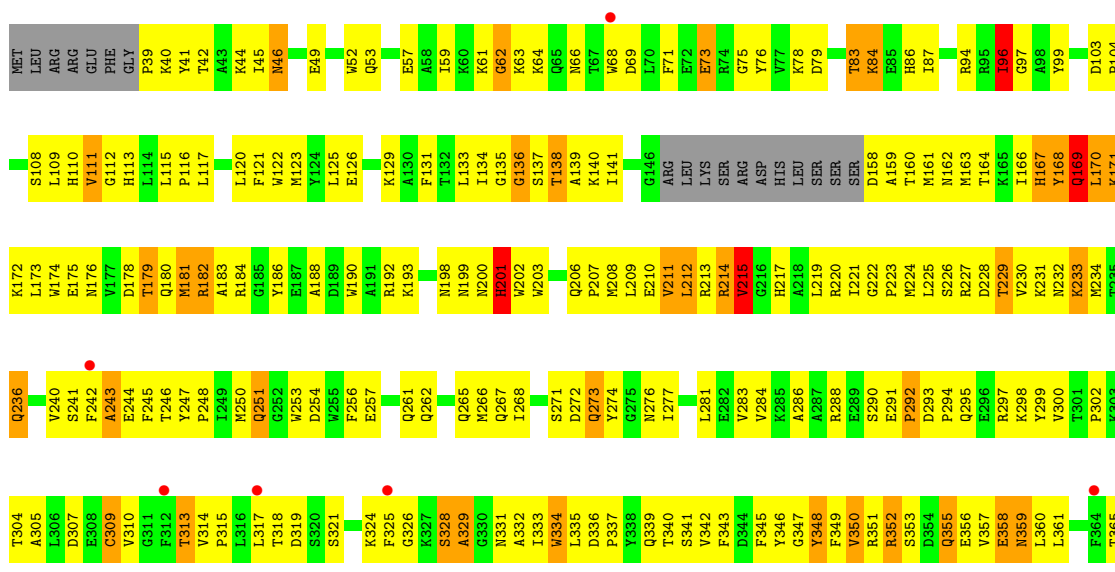


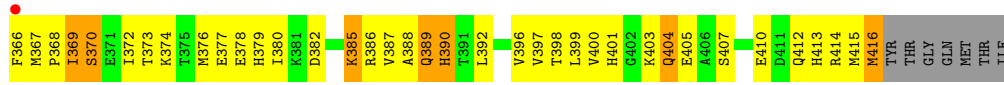


• Molecule 3: Tyrosyl-tRNA synthetase

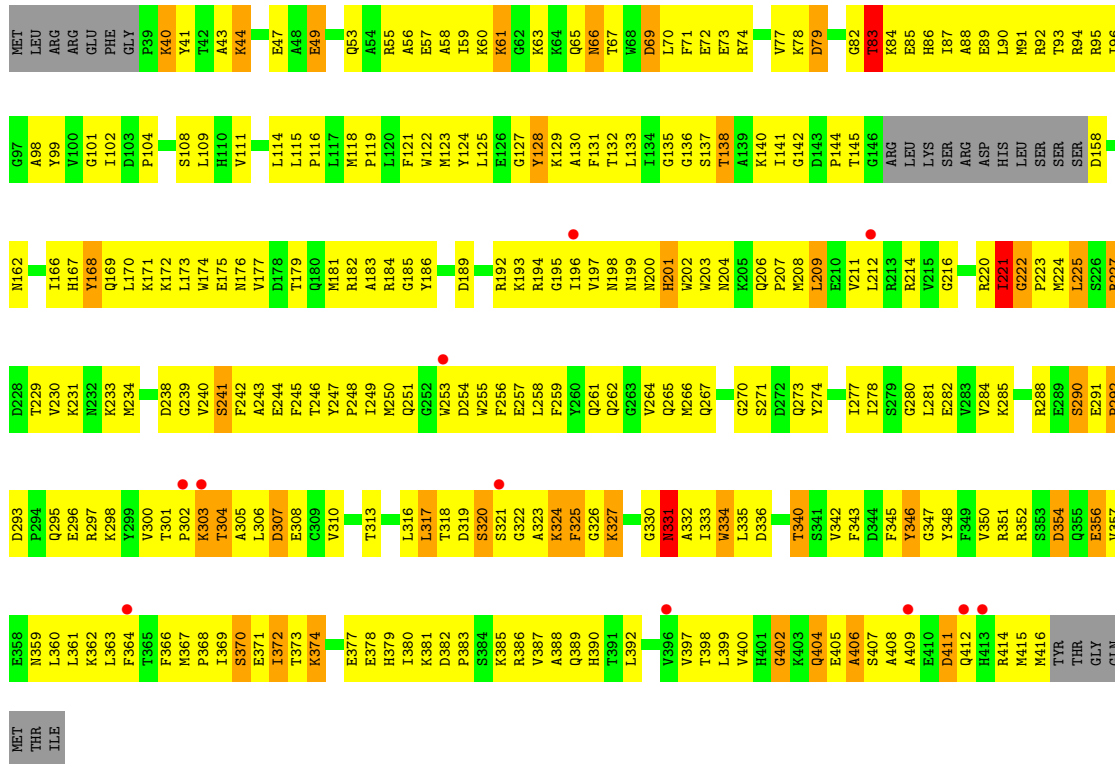


• Molecule 3: Tyrosyl-tRNA synthetase

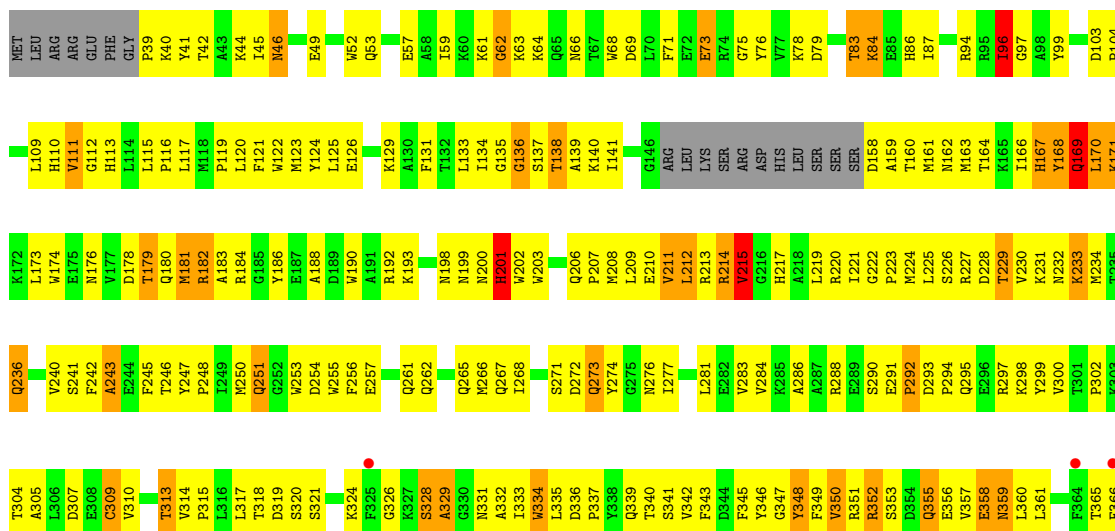


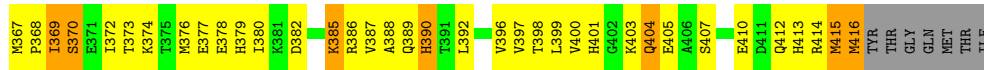


- Molecule 3: Tyrosyl-tRNA synthetase

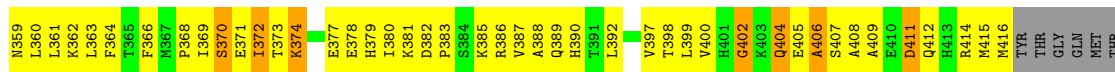
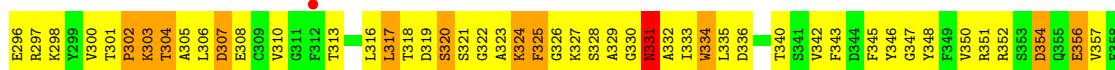
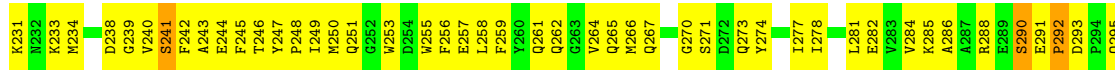
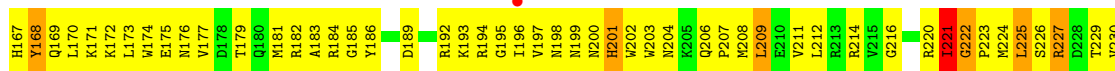
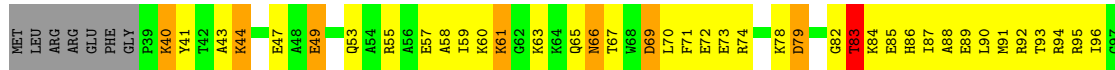


- Molecule 3: Tyrosyl-tRNA synthetase



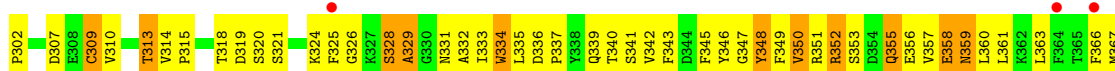
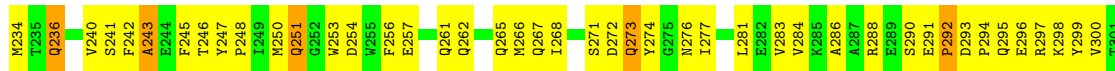
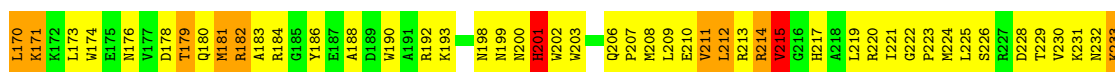
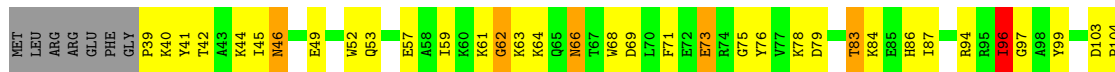


• Molecule 3: Tyrosyl-tRNA synthetase



ILE

• Molecule 3: Tyrosyl-tRNA synthetase



P368	I369	S370	E371	I372	T373	K374	I376	M376	E377	E378	H379	I380	K381	D382	K385	R386	V387	A388	Q389	H390	T391	L392	V396	V397	I398	L399	V400	H401	G402	K403	Q404	E405	A406	S407	E410	D411	Q412	H413	R414	M415	M416	TYR	THR	THR	GLY	GLN	MET	THR	ILE
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----	-----	-----	-----	-----	-----	-----

4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 2 1	Depositor
Cell constants a, b, c, α , β , γ	231.06Å 123.53Å 235.23Å 90.00° 90.26° 90.00°	Depositor
Resolution (Å)	47.88 – 4.50 49.47 – 4.47	Depositor EDS
% Data completeness (in resolution range)	83.3 (47.88-4.50) 82.1 (49.47-4.47)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	2.10 (at 4.45Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.235 , 0.249 0.235 , 0.250	Depositor DCC
R_{free} test set	3334 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å ²)	167.2	Xtriage
Anisotropy	0.414	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.27 , 232.6	EDS
L-test for twinning ²	$\langle L \rangle = 0.48$, $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.247 for l,k,-h 0.388 for h,-k,-l 0.257 for l,-k,h	Xtriage
F_o, F_c correlation	0.91	EDS
Total number of atoms	44328	wwPDB-VP
Average B, all atoms (Å ²)	211.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.69% 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	C	0.62	1/5710 (0.0%)	0.81	2/8894 (0.0%)
1	G	0.62	1/5710 (0.0%)	0.81	2/8894 (0.0%)
1	K	0.68	5/5710 (0.1%)	0.83	6/8894 (0.1%)
1	O	0.62	1/5710 (0.0%)	0.80	2/8894 (0.0%)
2	D	1.05	1/92 (1.1%)	0.86	0/139
2	H	1.02	1/92 (1.1%)	0.86	1/139 (0.7%)
2	L	1.03	1/92 (1.1%)	0.87	1/139 (0.7%)
2	P	0.99	1/92 (1.1%)	0.84	0/139
3	A	0.48	0/3023	0.64	0/4083
3	B	0.48	0/3023	0.64	1/4083 (0.0%)
3	E	0.50	0/3023	0.64	0/4083
3	F	0.49	0/3023	0.65	1/4083 (0.0%)
3	I	0.49	0/3023	0.64	0/4083
3	J	0.49	0/3023	0.65	1/4083 (0.0%)
3	M	0.47	0/3023	0.64	0/4083
3	N	0.49	0/3023	0.64	1/4083 (0.0%)
All	All	0.57	12/47392 (0.0%)	0.74	18/68796 (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	C	0	9
1	G	0	9
1	K	0	9
1	O	0	9
All	All	0	36

The worst 5 of 12 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	K	105	C	P-OP2	8.61	1.63	1.49
1	K	9	G	OP3-P	-7.41	1.52	1.61
1	G	9	G	OP3-P	-7.30	1.52	1.61
1	O	9	G	OP3-P	-7.26	1.52	1.61
2	D	1	G	OP3-P	-7.17	1.52	1.61

The worst 5 of 18 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	G	186	U	N1-C1'-C2'	6.62	122.60	114.00
1	O	186	U	N1-C1'-C2'	6.62	122.60	114.00
1	C	186	U	N1-C1'-C2'	6.45	122.38	114.00
1	K	186	U	N1-C1'-C2'	6.37	122.28	114.00
1	G	66	A	N9-C1'-C2'	6.21	122.07	114.00

There are no chirality outliers.

5 of 36 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	102	G	Sidechain
1	C	45	U	Sidechain
1	C	75	U	Sidechain
1	C	76	G	Sidechain
1	C	77	U	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	C	5096	0	2564	305	0
1	G	5096	0	2564	303	0
1	K	5096	0	2564	315	0
1	O	5096	0	2564	307	0
2	D	84	0	43	9	0
2	H	84	0	43	9	0
2	L	84	0	43	8	0
2	P	84	0	43	9	0
3	A	2951	0	2903	341	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	2951	0	2903	312	0
3	E	2951	0	2903	337	0
3	F	2951	0	2903	320	0
3	I	2951	0	2903	342	0
3	J	2951	0	2903	312	0
3	M	2951	0	2903	343	0
3	N	2951	0	2903	318	0
All	All	44328	0	33652	3740	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 48.

The worst 5 of 3740 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)
3:E:108:SER:HA	3:E:169:GLN:HE22	1.15	1.11
3:M:108:SER:HA	3:M:169:GLN:HE22	1.16	1.09
3:A:108:SER:HA	3:A:169:GLN:HE22	1.17	1.07
3:I:108:SER:HA	3:I:169:GLN:HE22	1.14	1.05
1:K:102:G:N2	1:K:105:C:N3	2.05	1.02

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	
3	A	363/392 (93%)	257 (71%)	81 (22%)	25 (7%)	1	17
3	B	363/392 (93%)	256 (70%)	80 (22%)	27 (7%)	1	16
3	E	363/392 (93%)	257 (71%)	81 (22%)	25 (7%)	1	17
3	F	363/392 (93%)	256 (70%)	80 (22%)	27 (7%)	1	16

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	I	363/392 (93%)	258 (71%)	79 (22%)	26 (7%)	1	16
3	J	363/392 (93%)	256 (70%)	79 (22%)	28 (8%)	1	15
3	M	363/392 (93%)	258 (71%)	81 (22%)	24 (7%)	1	18
3	N	363/392 (93%)	257 (71%)	79 (22%)	27 (7%)	1	16
All	All	2904/3136 (93%)	2055 (71%)	640 (22%)	209 (7%)	1	16

5 of 209 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	A	320	SER
3	A	334	TRP
3	B	96	ILE
3	B	169	GLN
3	B	358	GLU

5.3.2 Protein sidechains [i](#)

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	
3	A	310/333 (93%)	278 (90%)	32 (10%)	7	27
3	B	310/333 (93%)	276 (89%)	34 (11%)	6	25
3	E	310/333 (93%)	278 (90%)	32 (10%)	7	27
3	F	310/333 (93%)	277 (89%)	33 (11%)	6	26
3	I	310/333 (93%)	278 (90%)	32 (10%)	7	27
3	J	310/333 (93%)	277 (89%)	33 (11%)	6	26
3	M	310/333 (93%)	278 (90%)	32 (10%)	7	27
3	N	310/333 (93%)	275 (89%)	35 (11%)	6	24
All	All	2480/2664 (93%)	2217 (89%)	263 (11%)	6	26

5 of 263 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	N	83	THR
3	N	182	ARG
3	N	414	ARG
3	E	411	ASP
3	E	374	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 123 such sidechains are listed below:

Mol	Chain	Res	Type
3	F	331	ASN
3	N	162	ASN
3	I	236	GLN
3	N	110	HIS
3	N	331	ASN

5.3.3 RNA [i](#)

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	C	236/246 (95%)	48 (20%)	14 (5%)
1	G	236/246 (95%)	48 (20%)	14 (5%)
1	K	236/246 (95%)	48 (20%)	14 (5%)
1	O	236/246 (95%)	48 (20%)	14 (5%)
2	D	3/4 (75%)	0	0
2	H	3/4 (75%)	0	0
2	L	3/4 (75%)	0	0
2	P	3/4 (75%)	0	0
All	All	956/1000 (95%)	192 (20%)	56 (5%)

5 of 192 RNA backbone outliers are listed below:

Mol	Chain	Res	Type
1	C	36	A
1	C	37	A
1	C	46	A
1	C	58	A
1	C	59	C

5 of 56 RNA pucker outliers are listed below:

Mol	Chain	Res	Type
1	K	58	A

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	O	246	A
1	K	173	U
1	O	217	A
1	O	173	U

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 monosaccharides 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	C	238/246 (96%)	-0.22	2 (0%) 86 79	185, 211, 244, 250	0
1	G	238/246 (96%)	-0.24	1 (0%) 92 87	185, 211, 244, 250	0
1	K	238/246 (96%)	-0.25	1 (0%) 92 87	185, 211, 244, 250	0
1	O	238/246 (96%)	-0.20	2 (0%) 86 79	185, 211, 244, 250	0
2	D	4/4 (100%)	0.17	0 100 100	198, 208, 211, 236	0
2	H	4/4 (100%)	-0.00	0 100 100	198, 208, 211, 236	0
2	L	4/4 (100%)	-0.15	0 100 100	198, 208, 211, 236	0
2	P	4/4 (100%)	-0.24	0 100 100	198, 208, 211, 236	0
3	A	367/392 (93%)	-0.08	3 (0%) 86 79	191, 213, 229, 240	0
3	B	367/392 (93%)	-0.14	4 (1%) 80 72	187, 209, 225, 233	0
3	E	367/392 (93%)	-0.04	8 (2%) 62 52	191, 213, 229, 240	0
3	F	367/392 (93%)	-0.15	7 (1%) 66 58	187, 209, 225, 233	0
3	I	367/392 (93%)	0.04	11 (2%) 50 39	191, 213, 229, 240	0
3	J	367/392 (93%)	-0.11	3 (0%) 86 79	187, 209, 225, 233	0
3	M	367/392 (93%)	-0.04	2 (0%) 91 85	191, 213, 229, 240	0
3	N	367/392 (93%)	-0.16	4 (1%) 80 72	187, 209, 225, 233	0
All	All	3904/4136 (94%)	-0.12	48 (1%) 79 70	185, 211, 232, 250	0

The worst 5 of 48 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	I	413	HIS	5.1
1	O	194	G	4.1
3	N	364	PHE	3.6
3	F	364	PHE	3.4
1	C	194	G	3.4

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](#)

There are no ligands in this entry.

6.5 Other polymers [i](#)

There are no such residues in this entry.