wwPDB X-ray Structure Validation Summary Report (i)

Aug 27, 2023 – 05:52 AM EDT

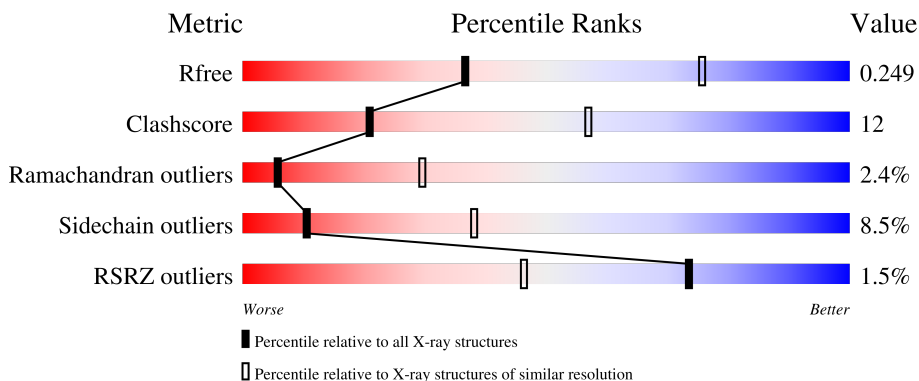|		
PDB ID	:	3HZR
Title	:	Tryptophanyl-tRNA synthetase homolog from Entamoeba histolytica
Authors	:	Arakaki, T.; Merritt, E.A.; Medical Structural Genomics of Pathogenic Protozoa (MSGPP)
Deposited on	:	2009-06-24
Resolution	:	3.00 Å(reported)

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	Percentile Ranks	Value
Rfree		0.249
Clashscore		12
Ramachandran outliers		2.4%
Sidechain outliers		8.5%
RSRZ outliers		1.5%

Worse *Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2092 (3.00-3.00)
Clashscore	141614	2416 (3.00-3.00)
Ramachandran outliers	138981	2333 (3.00-3.00)
Sidechain outliers	138945	2336 (3.00-3.00)
RSRZ outliers	127900	1990 (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 <=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	386	2% / 71% 20% 5% · ·
1	B	386	.% / 73% 18% 6% · ·
1	C	386	.% / 70% 21% 6% · ·
1	D	386	72% 19% 5% · ·
1	E	386	6% / 73% 18% 5% ·

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	386	72% 19% 5% ··

2 Entry composition ⓘ

There is only 1 type of molecule in this entry. The entry contains 17738 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 Tryptophanyl-tRNA synthetase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	374	Total	C	N	O	S	0	0	0
			2958	1905	479	565	9			
1	B	374	Total	C	N	O	S	0	0	0
			2954	1903	479	563	9			
1	C	374	Total	C	N	O	S	0	0	0
			2957	1906	479	563	9			
1	D	374	Total	C	N	O	S	0	0	0
			2958	1905	479	565	9			
1	E	374	Total	C	N	O	S	0	0	0
			2954	1903	479	563	9			
1	F	374	Total	C	N	O	S	0	0	0
			2957	1906	479	563	9			

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-8	MET	-	expression tag	PDB 3HZR
A	-7	ALA	-	expression tag	PDB 3HZR
A	-6	HIS	-	expression tag	PDB 3HZR
A	-5	HIS	-	expression tag	PDB 3HZR
A	-4	HIS	-	expression tag	PDB 3HZR
A	-3	HIS	-	expression tag	PDB 3HZR
A	-2	HIS	-	expression tag	PDB 3HZR
A	-1	HIS	-	expression tag	PDB 3HZR
B	-8	MET	-	expression tag	PDB 3HZR
B	-7	ALA	-	expression tag	PDB 3HZR
B	-6	HIS	-	expression tag	PDB 3HZR
B	-5	HIS	-	expression tag	PDB 3HZR
B	-4	HIS	-	expression tag	PDB 3HZR
B	-3	HIS	-	expression tag	PDB 3HZR
B	-2	HIS	-	expression tag	PDB 3HZR
B	-1	HIS	-	expression tag	PDB 3HZR
C	-8	MET	-	expression tag	PDB 3HZR

Continued on next page...

Continued from previous page...

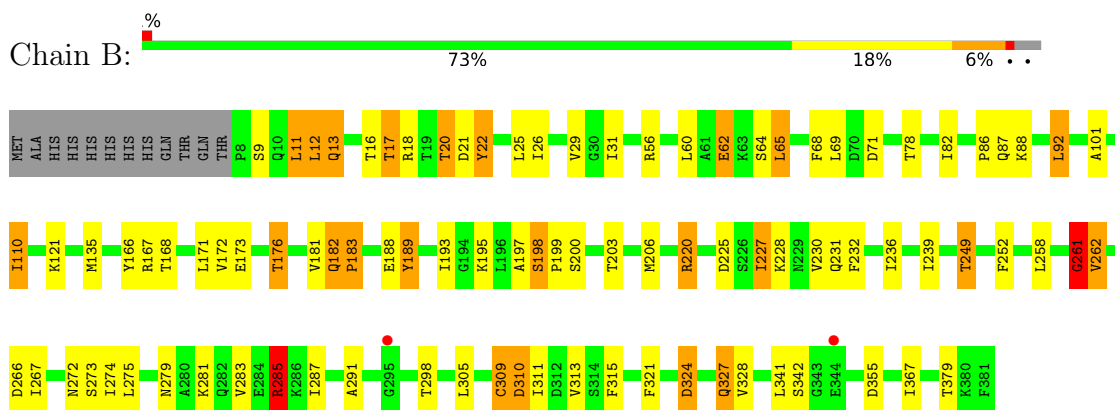| Chain	Residue	Modelled	Actual	Comment	Reference
C	-7	ALA	-	expression tag	PDB 3HZR
C	-6	HIS	-	expression tag	PDB 3HZR
C	-5	HIS	-	expression tag	PDB 3HZR
C	-4	HIS	-	expression tag	PDB 3HZR
C	-3	HIS	-	expression tag	PDB 3HZR
C	-2	HIS	-	expression tag	PDB 3HZR
C	-1	HIS	-	expression tag	PDB 3HZR
D	-8	MET	-	expression tag	PDB 3HZR
D	-7	ALA	-	expression tag	PDB 3HZR
D	-6	HIS	-	expression tag	PDB 3HZR
D	-5	HIS	-	expression tag	PDB 3HZR
D	-4	HIS	-	expression tag	PDB 3HZR
D	-3	HIS	-	expression tag	PDB 3HZR
D	-2	HIS	-	expression tag	PDB 3HZR
D	-1	HIS	-	expression tag	PDB 3HZR
E	-8	MET	-	expression tag	PDB 3HZR
E	-7	ALA	-	expression tag	PDB 3HZR
E	-6	HIS	-	expression tag	PDB 3HZR
E	-5	HIS	-	expression tag	PDB 3HZR
E	-4	HIS	-	expression tag	PDB 3HZR
E	-3	HIS	-	expression tag	PDB 3HZR
E	-2	HIS	-	expression tag	PDB 3HZR
E	-1	HIS	-	expression tag	PDB 3HZR
F	-8	MET	-	expression tag	PDB 3HZR
F	-7	ALA	-	expression tag	PDB 3HZR
F	-6	HIS	-	expression tag	PDB 3HZR
F	-5	HIS	-	expression tag	PDB 3HZR
F	-4	HIS	-	expression tag	PDB 3HZR
F	-3	HIS	-	expression tag	PDB 3HZR
F	-2	HIS	-	expression tag	PDB 3HZR
F	-1	HIS	-	expression tag	PDB 3HZR

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.
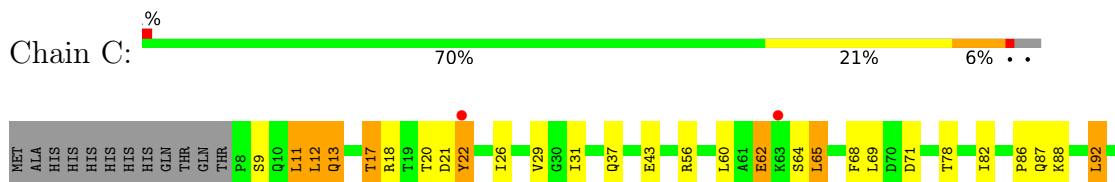
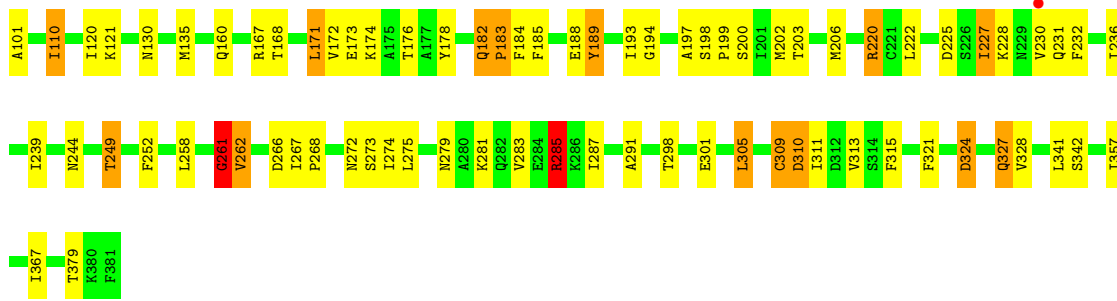• Molecule 1: Tryptophanyl-tRNA synthetase
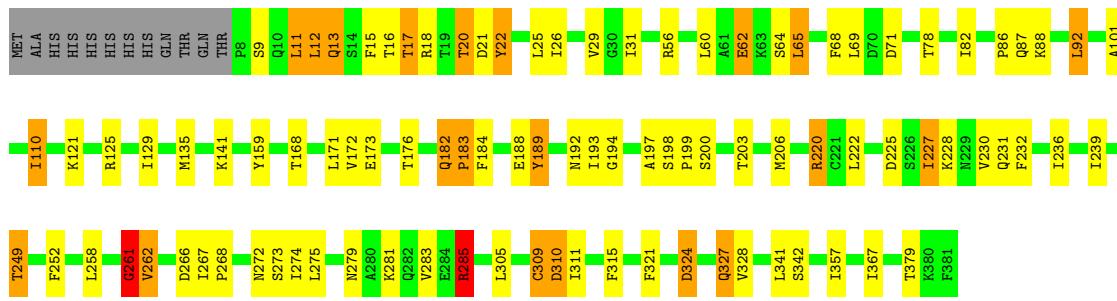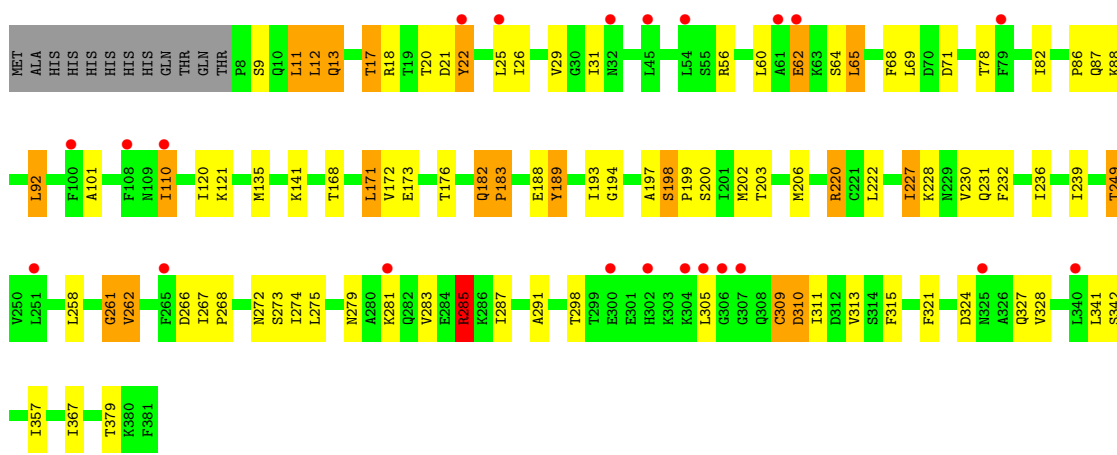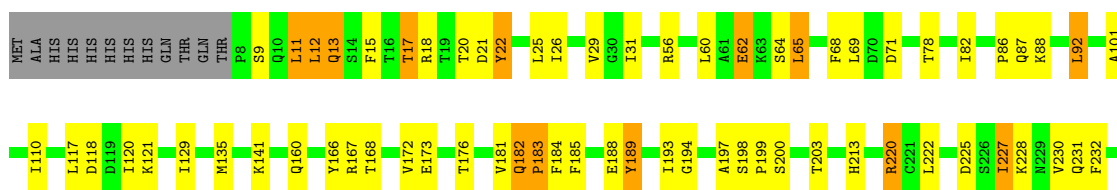
Chain A:

• Molecule 1: Tryptophanyl-tRNA synthetase

Chain B:

• Molecule 1: Tryptophanyl-tRNA synthetase

Chain C:

A101 I110 I120 K121 N130 M135 Q160 R167 T168 L171 V172 E173 K174 A175 T176 A177 Y178 Q182 P183 F184 F185 E188 Y189 I193 G194 A197 S198 P199 S200 I201 M202 T203 M206 R220 C221 L222 D225 S226 I227 K228 N229 V230 Q231 F232 I236

I239 N244 T249 F252 L258 G261 V262 D266 I267 P268 N272 S273 I274 L275 N279 A280 K281 Q282 V283 E284 R285 K286 I287 A291 T298 E301 L305 C309 D310 I311 D312 V313 S314 F315 F321 D324 Q327 V328 L341 S342 I357

I367 T379 K380 F381

• Molecule 1: Tryptophanyl-tRNA synthetase

Chain D: 72% 19% 5% • •

MET ALA HIS HIS HIS HIS HIS GLN THR GLN THR P8 S9 Q10 L11 L12 Q13 S14 F15 T16 T17 R18 T19 T20 D21 Y22 L25 I26 V29 G30 I31 R56 L60 A61 E62 K63 S64 L65 F68 L69 D70 D71 T78 I82 P86 Q87 K88 L92 A101

I110 K121 R125 I129 M135 K141 Y159 T168 L171 V172 E173 T176 Q182 P183 F184 E188 Y189 N192 I193 G194 A197 S198 P199 S200 T203 M206 R220 C221 L222 D225 S226 I227 K228 N229 V230 Q231 F232 I236 I239

T249 F252 L258 G261 V262 D266 I267 P268 N272 S273 I274 L275 N279 A280 K281 Q282 V283 E284 R285 L305 C309 D310 I311 F315 F321 D324 Q327 V328 L341 S342 I357 I367 T379 K380 F381

• Molecule 1: Tryptophanyl-tRNA synthetase

Chain E: 6% 73% 18% 5% •

MET ALA HIS HIS HIS HIS HIS GLN THR GLN THR P8 S9 Q10 L11 L12 Q13 T17 R18 T19 T20 D21 Y22 L25 I26 V29 G30 I31 N32 L45 L54 S55 R56 L60 A61 E62 K63 S64 L65 F68 L69 D70 D71 T78 F79 I82 P86 Q87 K88

L92 F100 A101 F108 N109 I110 I120 K121 M135 K141 T168 L171 V172 E173 T176 Q182 P183 E188 Y189 I193 G194 A197 S198 P199 S200 I201 M202 T203 M206 R220 C221 L222 I227 K228 N229 V230 Q231 F232 I236 I239 T249

V250 L251 L258 G261 V262 F265 D266 I267 P268 N272 S273 I274 L275 N279 A280 K281 Q282 V283 E284 R285 K286 I287 A291 T298 T299 E300 E301 H302 K303 K304 L305 G306 G307 Q308 C309 D310 I311 D312 V313 S314 F315 F321 D324 N325 A326 Q327 V328 L340 L341 S342

I357 I367 T379 K380 F381

• Molecule 1: Tryptophanyl-tRNA synthetase

Chain F: 72% 19% 5% • •

MET ALA HIS HIS HIS HIS HIS GLN THR GLN THR P8 S9 Q10 L11 L12 Q13 S14 F15 T16 T17 R18 T19 T20 D21 Y22 L25 I26 V29 G30 I31 R56 L60 A61 E62 K63 S64 L65 F68 L69 D70 D71 T78 I82 P86 Q87 K88 L92 A101

I110 L117 D118 D119 I120 K121 I129 M135 K141 Q160 Y166 R167 T168 V172 E173 T176 V181 Q182 P183 F184 F185 E188 Y189 I193 G194 A197 S198 P199 S200 T203 H213 R220 C221 L222 D225 S226 I227 K228 N229 V230 Q231 F232

I239 T249 F252 L258 G261 V262 D266 I267 N272 S273 I274 L275 N279 A280 K281 Q282 V283 E284 R285 L305 C309 D310 I311 F321 D324 N325 A326 Q327 V328 L341 S342 I367 T379 K380 F381

4   Data and refinement statistics   ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	93.20Å 158.88Å 100.28Å 90.00° 107.33° 90.00°	Depositor
Resolution (Å)	47.89 – 3.00 47.86 – 3.00	Depositor EDS
% Data completeness (in resolution range)	97.6 (47.89-3.00) 97.6 (47.86-3.00)	Depositor EDS
R_{merge}	0.12	Depositor
R_{sym}	0.12	Depositor
$< I/\sigma(I) >$ [1]	1.64 (at 3.01Å)	Xtriage
Refinement program	REFMAC refmac_5.5.0089	Depositor
R, R_{free}	0.228 , 0.250 0.226 , 0.249	Depositor DCC
R_{free} test set	2753 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	80.9	Xtriage
Anisotropy	0.042	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 43.7	EDS
L-test for twinning[2]	$< |L| > = 0.47, < L^2 > = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	17738	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.51% of the height of the origin peak. No significant pseudotranslation is detected.*

---

[1]Intensities estimated from amplitudes.
[2]Theoretical values of $< |L| >$, $< L^2 >$ 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	0.61	0/3019	0.70	2/4084 (0.0%)
1	B	0.61	0/3015	0.70	2/4079 (0.0%)
1	C	0.56	1/3018 (0.0%)	0.68	2/4083 (0.0%)
1	D	0.60	0/3019	0.69	2/4084 (0.0%)
1	E	0.50	0/3015	0.65	1/4079 (0.0%)
1	F	0.59	1/3018 (0.0%)	0.70	2/4083 (0.0%)
All	All	0.58	2/18104 (0.0%)	0.69	11/24492 (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	A	0	2
1	B	0	2
1	C	0	2
1	D	0	2
1	E	0	2
1	F	0	2
All	All	0	12

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	F	160	GLN	CG-CD	5.97	1.64	1.51
1	C	160	GLN	CG-CD	5.30	1.63	1.51

The worst 5 of 11 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	285	ARG	NE-CZ-NH1	8.18	124.39	120.30
1	F	285	ARG	NE-CZ-NH1	7.97	124.28	120.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	285	ARG	NE-CZ-NH1	7.93	124.27	120.30
1	A	285	ARG	NE-CZ-NH1	7.81	124.21	120.30
1	D	285	ARG	NE-CZ-NH1	7.68	124.14	120.30

There are no chirality outliers.

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	261	GLY	Peptide
1	A	309	CYS	Peptide
1	B	261	GLY	Peptide
1	B	309	CYS	Peptide
1	C	261	GLY	Peptide

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	2958	0	2910	76	4
1	B	2954	0	2906	70	2
1	C	2957	0	2915	88	3
1	D	2958	0	2910	83	1
1	E	2954	0	2906	73	0
1	F	2957	0	2915	75	2
All	All	17738	0	17462	429	6

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 12.

The worst 5 of 429 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:29:VAL:HG11	1:B:230:VAL:HG11	1.24	1.16
1:F:29:VAL:HG11	1:F:230:VAL:HG11	1.34	1.08
1:E:29:VAL:HG11	1:E:230:VAL:HG11	1.32	1.07

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:29:VAL:HG11	1:C:230:VAL:HG11	1.34	1.05
1:A:29:VAL:HG11	1:A:230:VAL:HG11	1.33	1.03

The worst 5 of 6 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:344:GLU:OE2	1:C:244:ASN:ND2[2_546]	1.90	0.30
1:B:189:TYR:CB	1:C:37:GLN:NE2[1_655]	1.97	0.23
1:A:189:TYR:OH	1:C:43:GLU:OE1[1_655]	2.00	0.20
1:B:355:ASP:OD2	1:F:88:LYS:NZ[2_555]	2.09	0.11
1:A:40:GLN:NE2	1:F:18:ARG:NE[1_556]	2.14	0.06

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	372/386 (96%)	345 (93%)	19 (5%)	8 (2%)	6	31
1	B	372/386 (96%)	347 (93%)	16 (4%)	9 (2%)	6	29
1	C	372/386 (96%)	346 (93%)	17 (5%)	9 (2%)	6	29
1	D	372/386 (96%)	347 (93%)	16 (4%)	9 (2%)	6	29
1	E	372/386 (96%)	347 (93%)	16 (4%)	9 (2%)	6	29
1	F	372/386 (96%)	346 (93%)	17 (5%)	9 (2%)	6	29
All	All	2232/2316 (96%)	2078 (93%)	101 (4%)	53 (2%)	6	29

5 of 53 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	12	LEU
1	A	22	TYR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	182	GLN
1	A	228	LYS
1	A	310	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	323/348 (93%)	296 (92%)	27 (8%)	11	38
1	B	322/348 (92%)	295 (92%)	27 (8%)	11	38
1	C	323/348 (93%)	295 (91%)	28 (9%)	10	37
1	D	323/348 (93%)	296 (92%)	27 (8%)	11	38
1	E	322/348 (92%)	295 (92%)	27 (8%)	11	38
1	F	323/348 (93%)	295 (91%)	28 (9%)	10	37
All	All	1936/2088 (93%)	1772 (92%)	164 (8%)	10	38

5 of 164 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	65	LEU
1	F	65	LEU
1	E	171	LEU
1	E	311	ILE
1	F	220	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	272	ASN
1	D	279	ASN
1	F	279	ASN
1	E	279	ASN
1	F	272	ASN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

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

5.5 Carbohydrates ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

There are no such residues in this entry.

5.8 Polymer linkage issues ⓘ

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(Å²)	Q<0.9
1	A	374/386 (96%)	-0.10	6 (1%)	72	44	20, 20, 20, 50	0
1	B	374/386 (96%)	-0.13	2 (0%)	91	75	20, 20, 20, 50	0
1	C	374/386 (96%)	-0.11	3 (0%)	86	65	20, 20, 20, 50	0
1	D	374/386 (96%)	-0.25	0	100	100	20, 20, 20, 50	0
1	E	374/386 (96%)	0.26	22 (5%)	22	7	20, 20, 20, 50	0
1	F	374/386 (96%)	-0.23	1 (0%)	94	84	20, 20, 20, 50	0
All	All	2244/2316 (96%)	-0.09	34 (1%)	73	46	20, 20, 20, 50	0

The worst 5 of 34 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	61	ALA	4.5
1	E	54	LEU	4.1
1	E	22	TYR	3.5
1	A	308	GLN	3.5
1	E	45	LEU	3.3

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 monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers ⓘ

There are no such residues in this entry.