wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 02:38 AM GMT

PDB ID	:	2HYI
Title	:	Structure of the human exon junction complex with a trapped DEAD-box helicase bound to RNA
Authors	:	Andersen, C.B.F.; Le Hir, H.; Andersen, G.R.
Deposited on	:	2006-08-06
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 (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

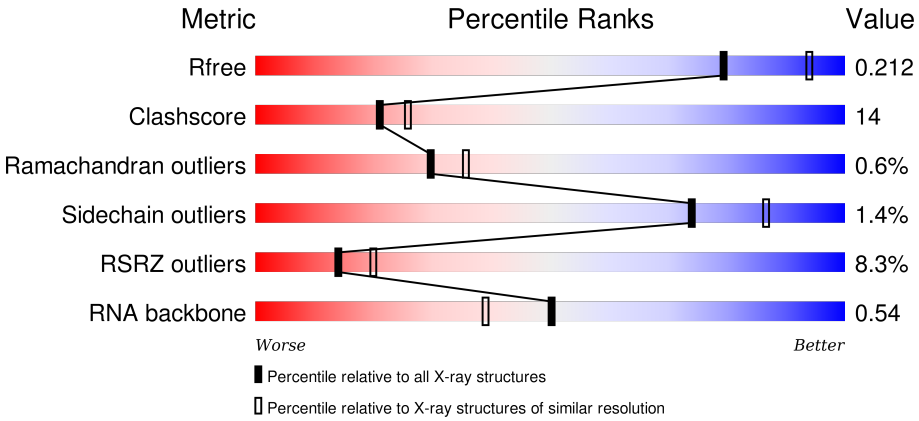
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

1 Overall quality at a glance (i)

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.



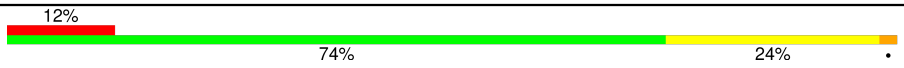
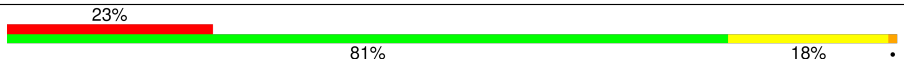
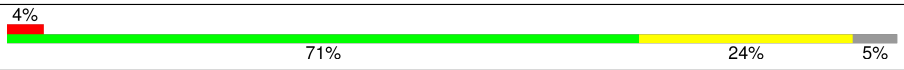
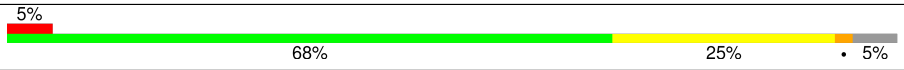
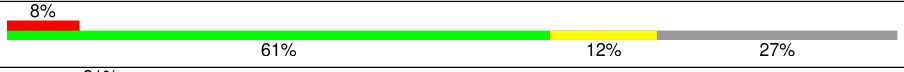
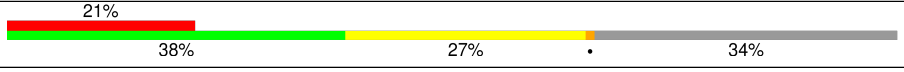
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	3852 (2.30-2.30)
Clashscore	102246	4452 (2.30-2.30)
Ramachandran outliers	100387	4410 (2.30-2.30)
Sidechain outliers	100360	4409 (2.30-2.30)
RSRZ outliers	91569	3857 (2.30-2.30)
RNA backbone	2183	1011 (2.84-1.76)

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 <=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	F	6	100%
1	L	6	17% 83% 17%
2	A	146	8% 79% 18% ••
2	G	146	6% 71% 27% ••

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	B	91	
3	H	91	
4	C	413	
4	I	413	
5	D	77	
5	J	77	

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
6	MG	C	701	-	-	-	X

2 Entry composition

There are 8 unique types of molecules in this entry. The entry contains 12033 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 5'-R(*UP*UP*UP*UP*UP*U)-3'.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	F	6	Total	C	N	O	P	0	0	0
			117	54	12	46	5			
1	L	6	Total	C	N	O	P	0	0	0
			117	54	12	46	5			

- Molecule 2 is a protein called Protein mago nashi homolog.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	144	Total	C	N	O	S	0	0	0
			1196	772	200	221	3			
2	G	144	Total	C	N	O	S	0	0	0
			1196	772	200	221	3			

- Molecule 3 is a protein called RNA-binding protein 8A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	B	91	Total	C	N	O	S	0	0	0
			731	463	122	143	3			
3	H	91	Total	C	N	O	S	0	0	0
			731	463	122	143	3			

- Molecule 4 is a protein called Probable ATP-dependent RNA helicase DDX48.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	C	392	Total	C	N	O	S	0	0	0
			3148	1987	548	594	19			
4	I	392	Total	C	N	O	S	0	0	0
			3148	1987	548	594	19			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	412	LEU	-	CLONING ARTIFACT	UNP P38919
C	413	GLU	-	CLONING ARTIFACT	UNP P38919
I	412	LEU	-	CLONING ARTIFACT	UNP P38919
I	413	GLU	-	CLONING ARTIFACT	UNP P38919

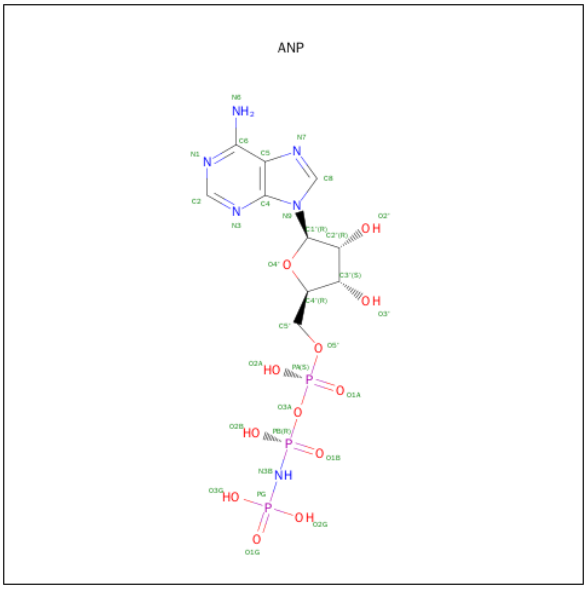
- Molecule 5 is a protein called Protein CASC3.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
5	D	56	Total	C	N	O	0	0	0
			485	303	89	93			
5	J	51	Total	C	N	O	0	0	0
			440	278	84	78			

- Molecule 6 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	I	1	Total	Mg	0	0
			1	1		
6	C	1	Total	Mg	0	0
			1	1		

- Molecule 7 is PHOSPHOAMINOPHOSPHONIC ACID-ADENYLATE ESTER (three-letter code: ANP) (formula: $C_{10}H_{17}N_6O_{12}P_3$).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	C	1	Total	C	N	O	P	0	0
			31	10	6	12	3		

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
7	I	1	Total	C	N	O	P	0	0
			31	10	6	12	3		

- Molecule 8 is water.

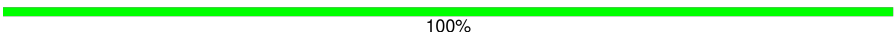
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
8	A	55	Total	O	0	0
			55	55		
8	B	46	Total	O	0	0
			46	46		
8	C	229	Total	O	0	0
			229	229		
8	D	26	Total	O	0	0
			26	26		
8	F	12	Total	O	0	0
			12	12		
8	G	65	Total	O	0	0
			65	65		
8	H	20	Total	O	0	0
			20	20		
8	I	190	Total	O	0	0
			190	190		
8	J	13	Total	O	0	0
			13	13		
8	L	4	Total	O	0	0
			4	4		

3 Residue-property plots (i)

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: 5'-R(*UP*UP*UP*UP*UP*U)-3'

Chain F:



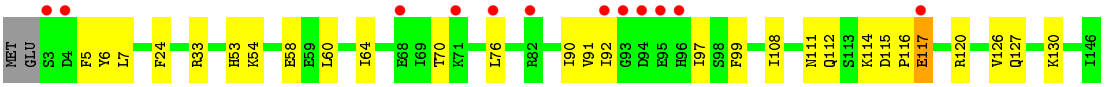
There are no outlier residues recorded for this chain.

- Molecule 1: 5'-R(*UP*UP*UP*UP*UP*U)-3'

Chain L:

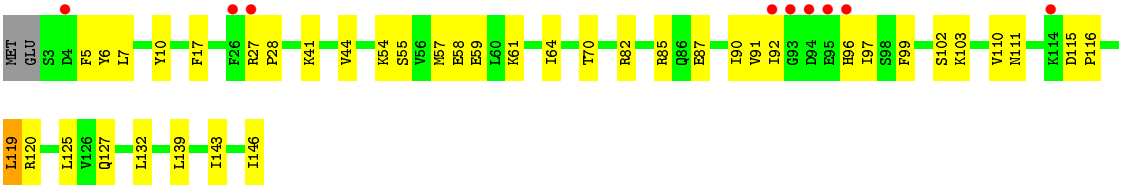
- Molecule 2: Protein mago nashi homolog

Chain A:



- Molecule 2: Protein mago nashi homolog

Chain G:

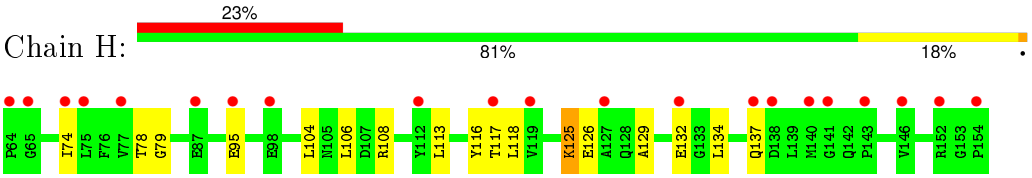


- Molecule 3: RNA-binding protein 8A

Chain B:

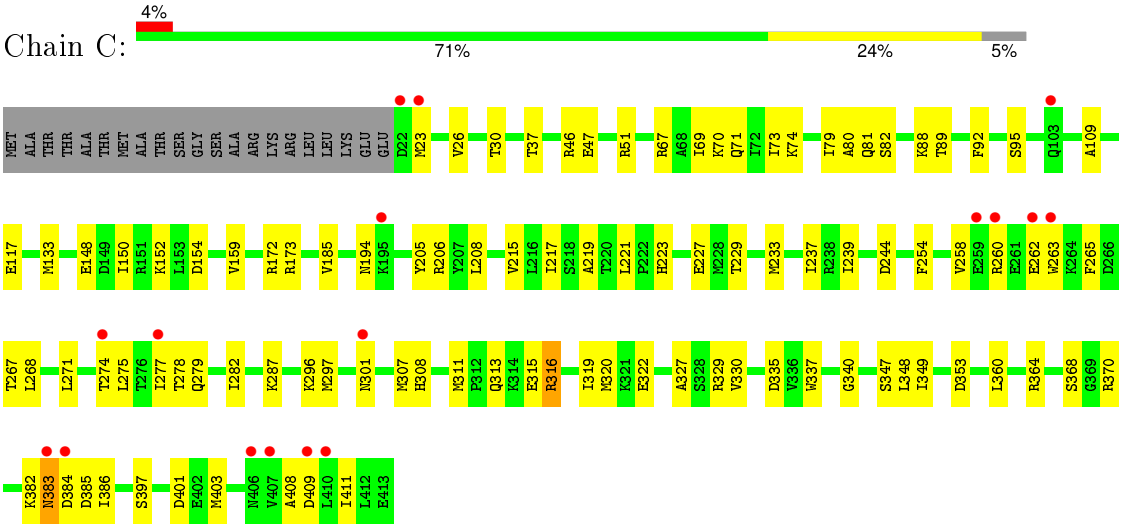
- Molecule 3: RNA-binding protein 8A

Chain H:



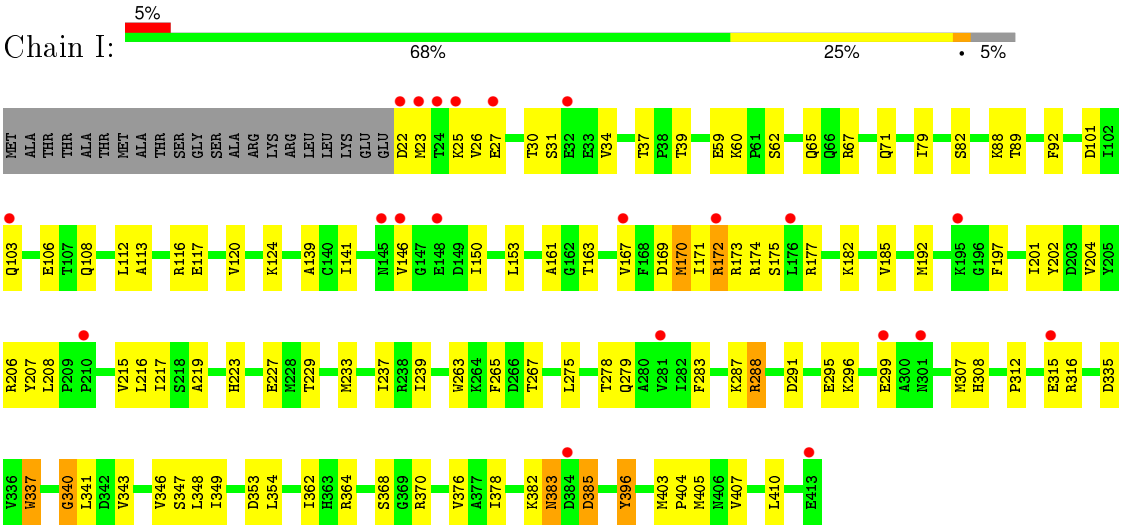
• Molecule 4: Probable ATP-dependent RNA helicase DDX48

Chain C:



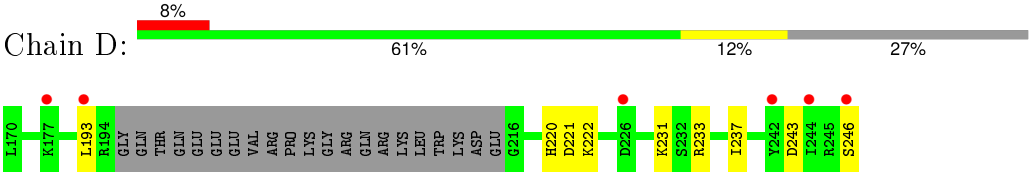
• Molecule 4: Probable ATP-dependent RNA helicase DDX48

Chain I:



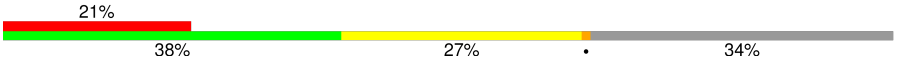
• Molecule 5: Protein CASC3

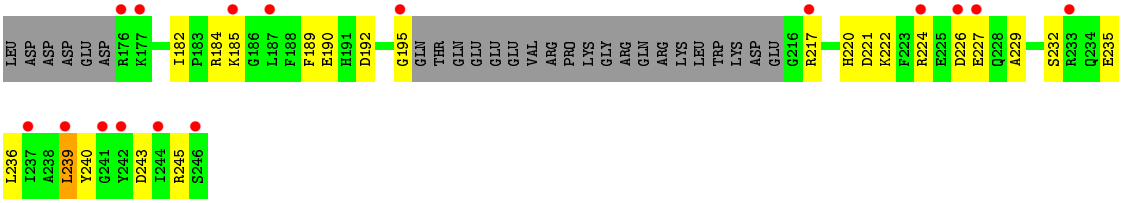
Chain D:



• Molecule 5: Protein CASC3

Chain J:




LEU ASP ASP ASP GLU ASP R176 K177 I182 P183 R184 K185 G186 L187 F188 F189 E190 H191 D192 G195 GLN THR GLN GLU GLU GLU VAL ARG PRO LYS GLY ARG GLN ARG LYS LEU TRP LYS ASP GLU G216 R217 H220 D221 K222 F223 R224 E225 D226 E227 Q228 A229 S232 R233 Q234 E235
L236 I237 A238 L239 Y240 G241 Y242 D243 I244 R245 S246

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	185.83Å 88.26Å 145.77Å 90.00° 110.77° 90.00°	Depositor
Resolution (Å)	20.00 – 2.30 19.93 – 2.30	Depositor EDS
% Data completeness (in resolution range)	(Not available) (20.00-2.30) 100.0 (19.93-2.30)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.06	Depositor
$< I/\sigma(I) >$ [1]	3.04 (at 2.30Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.212 , 0.231 0.207 , 0.212	Depositor DCC
R_{free} test set	1948 reflections (1.99%)	DCC
Wilson B-factor (Å^2)	44.6	Xtriage
Anisotropy	0.443	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 43.3	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Outliers	0 of 97913 reflections	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	12033	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 3.85% 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, 0.375 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: MG, ANP

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	F	0.71	0/128	0.80	0/196
1	L	0.86	0/128	0.80	0/196
2	A	0.38	0/1225	0.60	0/1648
2	G	0.39	0/1225	0.61	0/1648
3	B	0.36	0/749	0.71	3/1012 (0.3%)
3	H	0.31	0/749	0.57	0/1012
4	C	0.38	0/3197	0.65	0/4314
4	I	0.37	0/3197	0.64	2/4314 (0.0%)
5	D	0.39	0/496	0.61	0/662
5	J	0.38	0/451	0.53	0/600
All	All	0.39	0/11545	0.64	5/15602 (0.0%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	B	114	LYS	N-CA-C	-6.42	93.67	111.00
3	B	118	LEU	CA-CB-CG	6.29	129.77	115.30
4	I	340	GLY	N-CA-C	5.33	126.42	113.10
4	I	396	TYR	N-CA-C	-5.22	96.89	111.00
3	B	115	GLY	N-CA-C	5.22	126.14	113.10

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	F	117	0	62	0	0
1	L	117	0	62	1	0
2	A	1196	0	1182	33	0
2	G	1196	0	1182	45	0
3	B	731	0	690	25	0
3	H	731	0	690	15	0
4	C	3148	0	3190	88	0
4	I	3148	0	3190	111	0
5	D	485	0	452	11	0
5	J	440	0	422	24	0
6	C	1	0	0	0	0
6	I	1	0	0	0	0
7	C	31	0	13	4	0
7	I	31	0	13	3	0
8	A	55	0	0	1	0
8	B	46	0	0	1	0
8	C	229	0	0	11	0
8	D	26	0	0	0	0
8	F	12	0	0	0	0
8	G	65	0	0	2	0
8	H	20	0	0	1	0
8	I	190	0	0	8	0
8	J	13	0	0	1	0
8	L	4	0	0	0	0
All	All	12033	0	11148	323	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 14.

The worst 5 of 323 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)
2:G:58:GLU:HA	2:G:61:LYS:HE2	1.23	1.20
2:A:70:THR:HG21	2:A:92:ILE:HD11	1.29	1.08
2:G:58:GLU:HA	2:G:61:LYS:CE	1.82	1.08
4:I:172:ARG:HH11	4:I:172:ARG:HG3	0.95	1.05
5:J:224:ARG:HG3	5:J:227:GLU:HG3	1.32	1.05

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
2	A	142/146 (97%)	138 (97%)	4 (3%)	0	100 100
2	G	142/146 (97%)	139 (98%)	3 (2%)	0	100 100
3	B	89/91 (98%)	87 (98%)	1 (1%)	1 (1%)	17 18
3	H	89/91 (98%)	86 (97%)	3 (3%)	0	100 100
4	C	390/413 (94%)	378 (97%)	9 (2%)	3 (1%)	24 27
4	I	390/413 (94%)	374 (96%)	13 (3%)	3 (1%)	24 27
5	D	52/77 (68%)	50 (96%)	2 (4%)	0	100 100
5	J	47/77 (61%)	42 (89%)	4 (8%)	1 (2%)	9 7
All	All	1341/1454 (92%)	1294 (96%)	39 (3%)	8 (1%)	30 36

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	C	383	ASN
4	I	383	ASN
3	B	115	GLY
4	C	340	GLY
4	C	385	ASP

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
2	A	132/134 (98%)	131 (99%)	1 (1%)	86 94

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles
2	G	132/134 (98%)	129 (98%)	3 (2%)	58 75
3	B	76/76 (100%)	75 (99%)	1 (1%)	76 87
3	H	76/76 (100%)	75 (99%)	1 (1%)	76 87
4	C	347/363 (96%)	342 (99%)	5 (1%)	74 86
4	I	347/363 (96%)	342 (99%)	5 (1%)	74 86
5	D	50/69 (72%)	50 (100%)	0	100 100
5	J	44/69 (64%)	43 (98%)	1 (2%)	58 75
All	All	1204/1284 (94%)	1187 (99%)	17 (1%)	74 86

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	G	115	ASP
2	G	119	LEU
4	I	263	TRP
4	C	360	LEU
4	I	288	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
2	G	140	HIS
4	I	345	GLN
3	H	67	GLN
4	C	394	GLN
3	H	81	HIS

5.3.3 RNA

Mol	Chain	Analysed	Backbone Outliers	Pucker Outliers
1	F	5/6 (83%)	0	0
1	L	5/6 (83%)	0	0
All	All	10/12 (83%)	0	0

There are no RNA backbone outliers to report.

There are no RNA pucker outliers to report.

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)

Of 4 ligands modelled in this entry, 2 are monoatomic - leaving 2 for Mogul analysis.

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	$\#\|Z\| > 2$
7	ANP	C	801	6	27,33,33	1.23	3 (11%)	30,52,52	2.05	5 (16%)
7	ANP	I	802	6	27,33,33	1.10	1 (3%)	30,52,52	2.05	6 (20%)

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
7	ANP	C	801	6	-	0/12/38/38	0/3/3/3
7	ANP	I	802	6	-	0/12/38/38	0/3/3/3

All (4) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
7	C	801	ANP	O4'-C1'	2.18	1.44	1.41
7	I	802	ANP	O4'-C1'	2.34	1.44	1.41
7	C	801	ANP	PG-O1G	2.86	1.49	1.46
7	C	801	ANP	PB-O1B	3.15	1.49	1.46

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
7	I	802	ANP	N3-C2-N1	-9.15	121.89	128.89
7	C	801	ANP	N3-C2-N1	-8.64	122.28	128.89
7	C	801	ANP	C4-C5-N7	-3.82	105.97	109.48
7	I	802	ANP	PA-O3A-PB	-2.98	122.67	132.67
7	C	801	ANP	PA-O3A-PB	-2.75	123.46	132.67

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 7 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
7	C	801	ANP	4	0
7	I	802	ANP	3	0

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, 95^{th} 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	F	6/6 (100%)	0.19	0 100 100	39, 44, 59, 67	0
1	L	6/6 (100%)	0.34	1 (16%) 2 4	46, 49, 62, 79	0
2	A	144/146 (98%)	0.46	12 (8%) 14 20	35, 55, 80, 98	0
2	G	144/146 (98%)	0.21	9 (6%) 23 31	37, 48, 74, 99	0
3	B	91/91 (100%)	0.54	11 (12%) 6 9	38, 52, 70, 80	0
3	H	91/91 (100%)	1.09	21 (23%) 1 1	46, 74, 86, 92	0
4	C	392/413 (94%)	0.19	17 (4%) 39 48	30, 44, 75, 98	0
4	I	392/413 (94%)	0.25	21 (5%) 29 38	31, 50, 72, 99	0
5	D	56/77 (72%)	0.49	6 (10%) 8 12	44, 55, 77, 81	0
5	J	51/77 (66%)	1.37	16 (31%) 1 1	39, 69, 83, 84	0
All	All	1373/1466 (93%)	0.38	114 (8%) 14 20	30, 50, 81, 99	0

The worst 5 of 114 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	94	ASP	8.3
2	G	94	ASP	8.1
4	I	23	MET	6.4
5	D	246	SER	5.5
4	I	22	ASP	5.3

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 (i)

There are no carbohydrates in this entry.

6.4 Ligands (i)

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, 95^{th} 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
6	MG	C	701	1/1	0.96	0.27	6.73	35,35,35,35	0
7	ANP	I	802	31/31	0.97	0.12	-0.16	33,40,43,44	0
7	ANP	C	801	31/31	0.98	0.12	-0.18	31,34,35,37	0
6	MG	I	702	1/1	0.97	0.19	-	39,39,39,39	0

6.5 Other polymers (i)

There are no such residues in this entry.