wwPDB X-ray Structure Validation Summary Report (i)

May 21, 2020 – 08:45 pm BST

PDB ID	:	5NJ8
Title	:	Structural basis for aryl hydrocarbon receptor mediated gene activation
Authors	:	Daumke, O.; Schulte, K.W.
Deposited on	:	2017-03-28
Resolution	:	3.30 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
https://www.wwpdb.org/validation/2017/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:

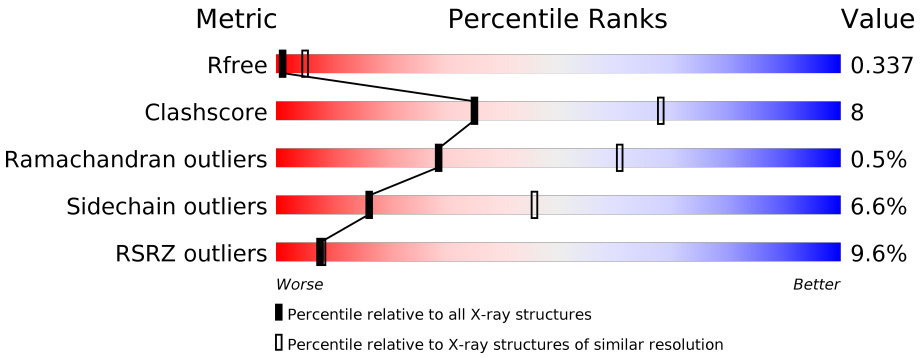
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 3.30 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}	130704	1149 (3.34-3.26)
Clashscore	141614	1205 (3.34-3.26)
Ramachandran outliers	138981	1183 (3.34-3.26)
Sidechain outliers	138945	1182 (3.34-3.26)
RSRZ outliers	127900	1115 (3.34-3.26)

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 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	254	3% 55% 13% • 31%
1	C	254	9% 54% 14% • 31%
2	B	239	11% 55% 16% • 27%
2	D	239	5% 50% 5% • 44%
3	E	12	33% 67%
3	G	12	58% 42%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	F	12	67% 33%
4	H	12	50% 50%

2 Entry composition

There are 6 unique types of molecules in this entry. The entry contains 10871 atoms, of which 5023 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.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	176	Total	C	H	N	O	S	0	0	0
			2787	902	1391	241	248	5			
1	C	176	Total	C	H	N	O	S	0	0	0
			2332	786	1090	219	235	2			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	20	GLY	-	expression tag	UNP P35869
A	21	PRO	-	expression tag	UNP P35869
A	22	MET	-	expression tag	UNP P35869
C	20	GLY	-	expression tag	UNP P35869
C	21	PRO	-	expression tag	UNP P35869
C	22	MET	-	expression tag	UNP P35869

- Molecule 2 is a protein called Aryl hydrocarbon receptor nuclear translocator.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
2	B	174	Total	C	H	N	O	S	0	0	0
			2711	857	1343	246	254	11			
2	D	134	Total	C	H	N	O	S	0	0	0
			1522	518	664	165	169	6			

There are 70 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	79	GLY	-	expression tag	UNP P53762
B	80	PRO	-	expression tag	UNP P53762
B	81	GLY	-	expression tag	UNP P53762
B	82	SER	-	expression tag	UNP P53762
B	83	ASP	-	expression tag	UNP P53762
B	84	ALA	-	expression tag	UNP P53762

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
B	256	SER	CYS	engineered mutation	UNP P53762
B	?	-	VAL	deletion	UNP P53762
B	?	-	ASP	deletion	UNP P53762
B	?	-	PRO	deletion	UNP P53762
B	?	-	VAL	deletion	UNP P53762
B	?	-	SER	deletion	UNP P53762
B	?	-	MET	deletion	UNP P53762
B	?	-	ASN	deletion	UNP P53762
B	?	-	ARG	deletion	UNP P53762
B	?	-	LEU	deletion	UNP P53762
B	?	-	SER	deletion	UNP P53762
B	?	-	PHE	deletion	UNP P53762
B	?	-	LEU	deletion	UNP P53762
B	?	-	ARG	deletion	UNP P53762
B	?	-	ASN	deletion	UNP P53762
B	?	-	ARG	deletion	UNP P53762
B	?	-	CYS	deletion	UNP P53762
B	?	-	ARG	deletion	UNP P53762
B	?	-	ASN	deletion	UNP P53762
B	?	-	GLY	deletion	UNP P53762
B	?	-	LEU	deletion	UNP P53762
B	?	-	GLY	deletion	UNP P53762
B	?	-	SER	deletion	UNP P53762
B	?	-	VAL	deletion	UNP P53762
B	?	-	LYS	deletion	UNP P53762
B	?	-	GLU	deletion	UNP P53762
B	?	-	GLY	deletion	UNP P53762
B	?	-	GLU	deletion	UNP P53762
B	?	-	PRO	deletion	UNP P53762
D	79	GLY	-	expression tag	UNP P53762
D	80	PRO	-	expression tag	UNP P53762
D	81	GLY	-	expression tag	UNP P53762
D	82	SER	-	expression tag	UNP P53762
D	83	ASP	-	expression tag	UNP P53762
D	84	ALA	-	expression tag	UNP P53762
D	256	SER	CYS	engineered mutation	UNP P53762
D	?	-	VAL	deletion	UNP P53762
D	?	-	ASP	deletion	UNP P53762
D	?	-	PRO	deletion	UNP P53762
D	?	-	VAL	deletion	UNP P53762
D	?	-	SER	deletion	UNP P53762
D	?	-	MET	deletion	UNP P53762

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	?	-	ASN	deletion	UNP P53762
D	?	-	ARG	deletion	UNP P53762
D	?	-	LEU	deletion	UNP P53762
D	?	-	SER	deletion	UNP P53762
D	?	-	PHE	deletion	UNP P53762
D	?	-	LEU	deletion	UNP P53762
D	?	-	ARG	deletion	UNP P53762
D	?	-	ASN	deletion	UNP P53762
D	?	-	ARG	deletion	UNP P53762
D	?	-	CYS	deletion	UNP P53762
D	?	-	ARG	deletion	UNP P53762
D	?	-	ASN	deletion	UNP P53762
D	?	-	GLY	deletion	UNP P53762
D	?	-	LEU	deletion	UNP P53762
D	?	-	GLY	deletion	UNP P53762
D	?	-	SER	deletion	UNP P53762
D	?	-	VAL	deletion	UNP P53762
D	?	-	LYS	deletion	UNP P53762
D	?	-	GLU	deletion	UNP P53762
D	?	-	GLY	deletion	UNP P53762
D	?	-	GLU	deletion	UNP P53762
D	?	-	PRO	deletion	UNP P53762

- Molecule 3 is a DNA chain called DNA (5'-D(*GP*GP*TP*CP*AP*CP*GP*CP*AP*AP*CP*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
3	E	12	Total	C	H	N	O	P	0	0	0
			372	115	132	47	67	11			
3	G	12	Total	C	H	N	O	P	0	0	0
			372	115	132	47	67	11			

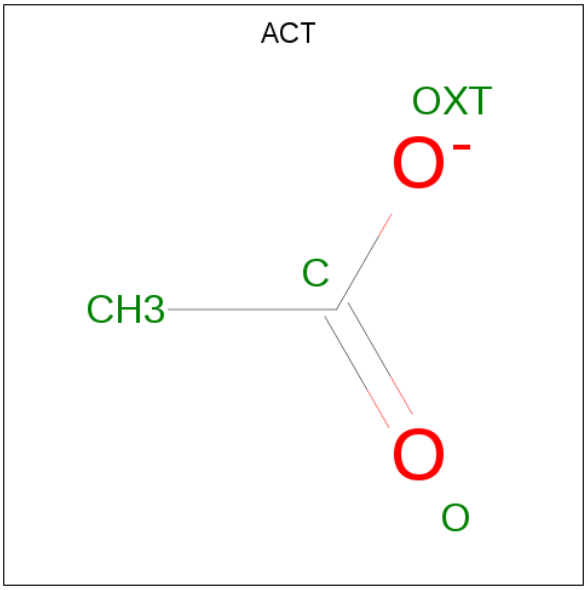
- Molecule 4 is a DNA chain called DNA (5'-D(*GP*GP*TP*TP*GP*CP*GP*TP*GP*AP*CP*C)-3').

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
4	F	12	Total	C	H	N	O	P	0	0	0
			378	117	134	45	71	11			
4	H	12	Total	C	H	N	O	P	0	0	0
			378	117	134	45	71	11			

- Molecule 5 is ERBIUM (III) ION (three-letter code: ER3) (formula: Er).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	H	1	Total Er 1 1	0	0
5	B	4	Total Er 4 4	0	0
5	A	3	Total Er 3 3	0	0
5	D	2	Total Er 2 2	0	0
5	C	2	Total Er 2 2	0	0

- Molecule 6 is ACETATE ION (three-letter code: ACT) (formula: $C_2H_3O_2$).



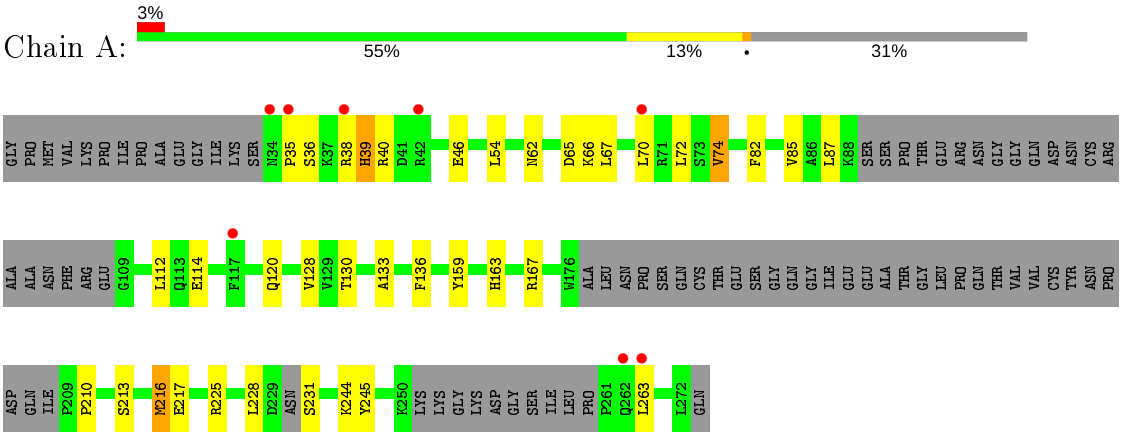
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
6	A	1	Total C H O 7 2 3 2	0	0

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 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.

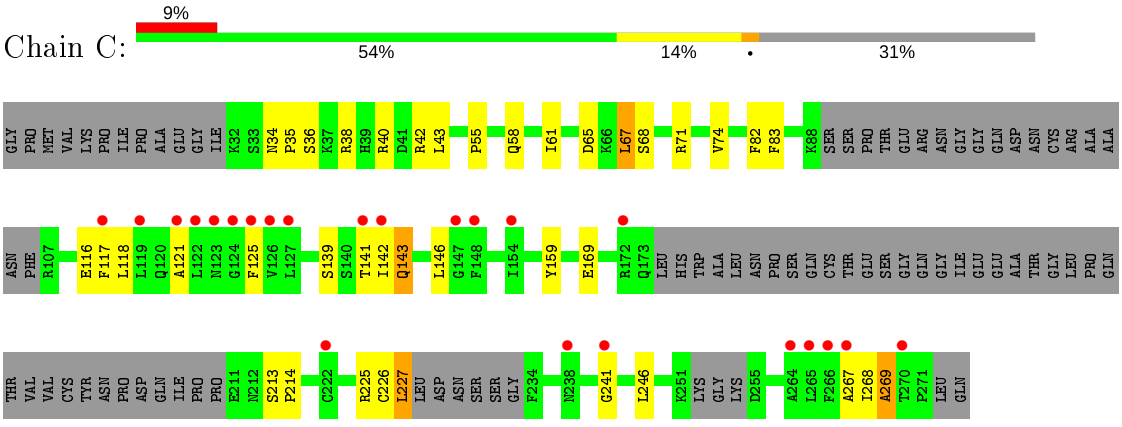
• Molecule 1: Aryl hydrocarbon receptor

Chain A:



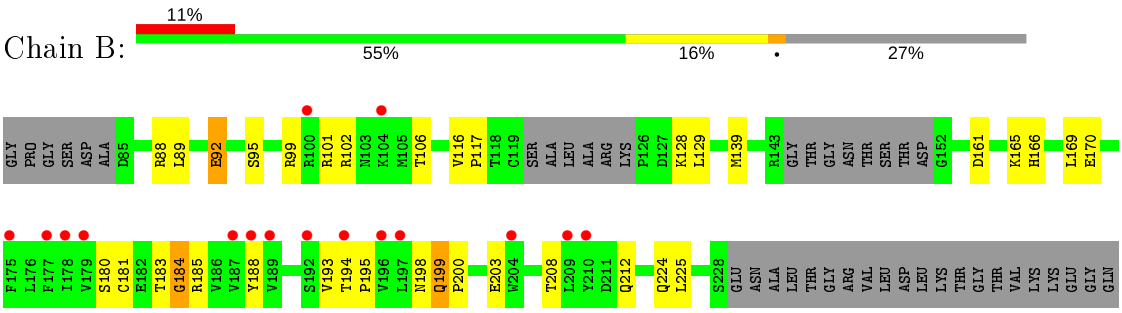
• Molecule 1: Aryl hydrocarbon receptor

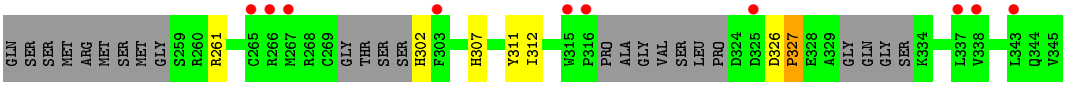
Chain C:



• Molecule 2: Aryl hydrocarbon receptor nuclear translocator

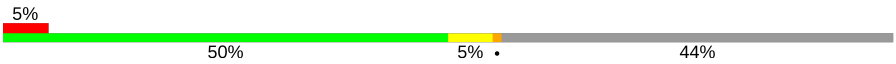
Chain B:

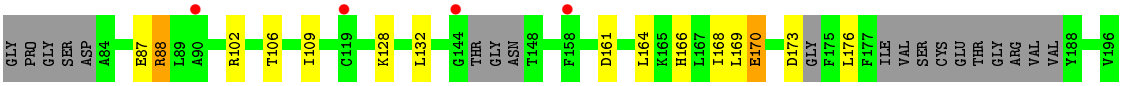


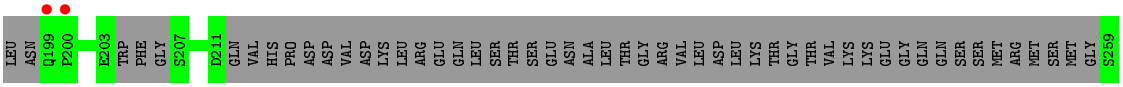


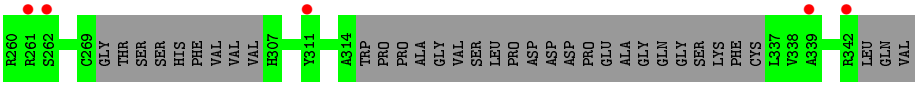
• Molecule 2: Aryl hydrocarbon receptor nuclear translocator

Chain D:



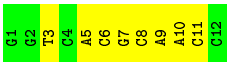






• Molecule 3: DNA (5'-D(*GP*GP*TP*CP*AP*CP*GP*CP*AP*AP*CP*C)-3')

Chain E:

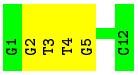


• Molecule 3: DNA (5'-D(*GP*GP*TP*CP*AP*CP*GP*CP*AP*AP*CP*C)-3')

Chain G:

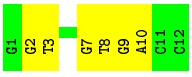
• Molecule 4: DNA (5'-D(*GP*GP*TP*TP*GP*CP*GP*TP*GP*AP*CP*C)-3')

Chain F:



• Molecule 4: DNA (5'-D(*GP*GP*TP*TP*GP*CP*GP*TP*GP*AP*CP*C)-3')

Chain H:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65 2 2	Depositor
Cell constants a, b, c, α, β, γ	91.32Å 91.32Å 464.91Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.83 – 3.30 46.83 – 3.30	Depositor EDS
% Data completeness (in resolution range)	99.5 (46.83-3.30) 99.8 (46.83-3.30)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	0.10	Depositor
$< I/\sigma(I) >$ [1]	5.90 (at 3.32Å)	Xtriage
Refinement program	PHENIX (1.11.1_2575: ???)	Depositor
R, R_{free}	0.292 , 0.333 0.293 , 0.337	Depositor DCC
R_{free} test set	919 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å^2)	86.6	Xtriage
Anisotropy	0.557	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 99.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.43$, $< L^2 > = 0.25$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.85	EDS
Total number of atoms	10871	wwPDB-VP
Average B, all atoms (Å^2)	115.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.13% 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

Bond lengths and bond angles in the following residue types are not validated in this section: ACT, ER3

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	0.28	0/1423	0.45	0/1922
1	C	0.28	0/1259	0.48	0/1715
2	B	0.26	0/1389	0.46	0/1871
2	D	0.24	0/858	0.43	0/1158
3	E	0.55	0/269	0.84	0/413
3	G	0.62	0/269	0.88	0/413
4	F	0.62	0/273	0.98	0/421
4	H	0.70	0/273	0.94	0/421
All	All	0.36	0/6013	0.58	0/8334

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	2

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	C	268	ILE	Peptide
1	C	269	ALA	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	1396	1391	1390	26	0
1	C	1242	1090	1090	24	0
2	B	1368	1343	1338	27	0
2	D	858	664	662	9	2
3	E	240	132	132	7	0
3	G	240	132	132	4	2
4	F	244	134	134	5	0
4	H	244	134	134	7	0
5	A	3	0	0	0	0
5	B	4	0	0	0	0
5	C	2	0	0	0	0
5	D	2	0	0	0	0
5	H	1	0	0	0	0
6	A	4	3	3	0	0
All	All	5848	5023	5015	89	2

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 8.

The worst 5 of 89 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:D:169:LEU:O	2:D:173:ASP:N	2.05	0.88
1:A:112:LEU:O	2:B:161:ASP:OD2	1.90	0.87
1:A:62:ASN:OD1	1:C:143:GLN:NE2	2.11	0.83
1:C:139:SER:OG	1:C:141:THR:OG1	1.99	0.80
2:B:128:LYS:NZ	4:F:5:DG:OP2	2.15	0.75

All (2) 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)
2:D:88:ARG:HH12	3:G:9:DA:OP1[8_556]	1.53	0.07
2:D:88:ARG:NH1	3:G:9:DA:OP1[8_556]	2.16	0.04

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
1	A	166/254 (65%)	145 (87%)	21 (13%)	0	100 100
1	C	166/254 (65%)	155 (93%)	10 (6%)	1 (1%)	25 57
2	B	160/239 (67%)	147 (92%)	11 (7%)	2 (1%)	12 40
2	D	116/239 (48%)	107 (92%)	9 (8%)	0	100 100
All	All	608/986 (62%)	554 (91%)	51 (8%)	3 (0%)	29 61

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	327	PRO
2	B	184	GLY
1	C	55	PRO

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
1	A	151/222 (68%)	143 (95%)	8 (5%)	22 53
1	C	110/222 (50%)	99 (90%)	11 (10%)	7 27
2	B	152/208 (73%)	143 (94%)	9 (6%)	19 49
2	D	60/208 (29%)	57 (95%)	3 (5%)	24 55
All	All	473/860 (55%)	442 (93%)	31 (7%)	16 46

5 of 31 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	302	HIS
1	C	36	SER
2	D	87	GLU
2	B	307	HIS
1	C	43	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	62	ASN
1	C	143	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)

Of 13 ligands modelled in this entry, 12 are monoatomic - leaving 1 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	$\#\|Z\| > 2$
6	ACT	A	303	5	1,3,3	1.64	0	0,3,3	0.00	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

No monomer is involved in short contacts.

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, 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	A	176/254 (69%)	0.37	8 (4%) 33 32	48, 96, 153, 198	0
1	C	176/254 (69%)	0.70	23 (13%) 3 3	52, 120, 185, 216	0
2	B	174/239 (72%)	0.77	26 (14%) 2 2	66, 116, 160, 196	0
2	D	134/239 (56%)	0.49	11 (8%) 11 11	48, 107, 189, 222	0
3	E	12/12 (100%)	0.19	0 100 100	70, 88, 123, 132	0
3	G	12/12 (100%)	0.52	0 100 100	51, 62, 77, 110	0
4	F	12/12 (100%)	0.44	0 100 100	76, 95, 124, 158	0
4	H	12/12 (100%)	0.22	0 100 100	55, 63, 74, 83	0
All	All	708/1034 (68%)	0.57	68 (9%) 8 8	48, 108, 174, 222	0

The worst 5 of 68 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	267	ALA	8.3
1	C	124	GLY	8.0
1	C	123	ASN	6.1
1	C	127	LEU	5.7
2	B	265	CYS	5.4

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

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. 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	B-factors(Å^2)	Q<0.9
5	ER3	D	401	1/1	0.74	0.09	190,190,190,190	0
6	ACT	A	303	4/4	0.90	0.23	31,37,86,95	0
5	ER3	A	304	1/1	0.93	0.12	132,132,132,132	0
5	ER3	C	301	1/1	0.95	0.18	137,137,137,137	0
5	ER3	D	402	1/1	0.96	0.51	159,159,159,159	0
5	ER3	B	401	1/1	0.96	0.03	113,113,113,113	0
5	ER3	B	404	1/1	0.98	0.19	111,111,111,111	0
5	ER3	B	402	1/1	0.98	0.30	121,121,121,121	0
5	ER3	A	302	1/1	0.98	0.15	57,57,57,57	0
5	ER3	A	301	1/1	0.98	0.17	95,95,95,95	0
5	ER3	B	403	1/1	0.98	0.04	103,103,103,103	0
5	ER3	H	101	1/1	0.99	0.15	67,67,67,67	1
5	ER3	C	302	1/1	0.99	0.05	97,97,97,97	0

6.5 Other polymers

There are no such residues in this entry.