wwPDB X-ray Structure Validation Summary Report (i)

Aug 8, 2023 – 08:39 PM EDT

PDB ID	:	1P3F
Title	:	Crystallographic Studies of Nucleosome Core Particles containing Histone 'Sin' Mutants
Authors	:	Muthurajan, U.M.; Bao, Y.; Forsberg, L.J.; Edayathumangalam, R.S.; Dyer, P.N.; White, C.L.; Luger, K.
Deposited on	:	2003-04-17
Resolution	:	2.90 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

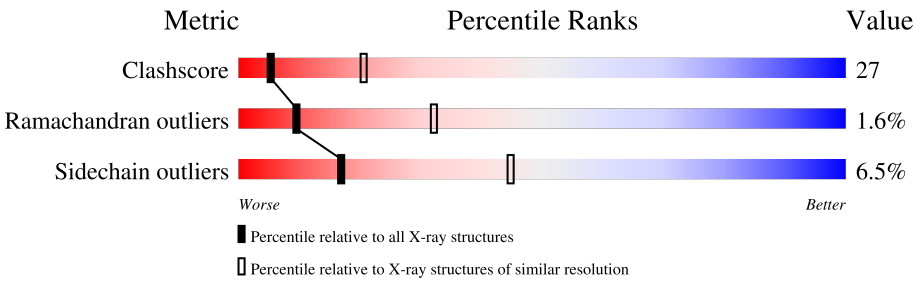
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	NOT EXECUTED
EDS	:	NOT EXECUTED
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.90 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))
Clashscore	141614	2172 (2.90-2.90)
Ramachandran outliers	138981	2115 (2.90-2.90)
Sidechain outliers	138945	2117 (2.90-2.90)

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%

Note EDS was not executed.

Mol	Chain	Length	Quality of chain
1	I	146	21% 63% 14% •
1	J	146	23% 53% 19% 5%
2	A	135	41% 27% • • 27%
2	E	135	47% 19% 6% • 27%
3	B	102	39% 34% 5% 22%
3	F	102	45% 29% 6% 20%
4	C	129	47% 26% 7% 19%
4	G	129	43% 36% • • 17%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
5	D	125	45% 24% • • 27%
5	H	125	41% 32% • 24%

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 12150 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 DNA chain called Palindromic 146bp Human Alpha-Satellite DNA fragment.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
1	I	146	2990	1430	541	874	145	0	0	0
1	J	146	2990	1430	541	874	145	0	0	0

- Molecule 2 is a protein called Histone H3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	A	99	817	515	158	141	3	0	0	0
2	E	99	817	515	158	141	3	0	0	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	434	GLU	GLY	conflict	UNP Q7ZT64
A	435	SER	VAL	conflict	UNP Q7ZT64
A	502	ALA	GLY	conflict	UNP Q7ZT64
E	634	GLU	GLY	conflict	UNP Q7ZT64
E	635	SER	VAL	conflict	UNP Q7ZT64
E	702	ALA	GLY	conflict	UNP Q7ZT64

- Molecule 3 is a protein called Histone H4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	B	80	633	398	122	111	2	0	0	0
3	F	82	648	409	124	113	2	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	45	CYS	ARG	conflict	UNP P62799
F	245	CYS	ARG	conflict	UNP P62799

- Molecule 4 is a protein called Histone H2A.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
			Total	C	N	O			
4	C	104	804	507	157	140	0	0	0
			Total	C	N	O			
4	G	107	827	522	162	143	0	0	0

There are 48 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	814	ALA	SER	conflict	UNP Q7ZT66
C	867	GLY	TRP	conflict	UNP Q7ZT66
C	868	ASN	GLU	conflict	UNP Q7ZT66
C	869	ALA	ARG	conflict	UNP Q7ZT66
C	870	ALA	LEU	conflict	UNP Q7ZT66
C	871	ARG	PRO	conflict	UNP Q7ZT66
C	872	ASP	GLU	conflict	UNP Q7ZT66
C	873	ASN	ILE	conflict	UNP Q7ZT66
C	874	LYS	TRP	conflict	UNP Q7ZT66
C	876	THR	ARG	conflict	UNP Q7ZT66
C	877	ARG	PRO	conflict	UNP Q7ZT66
C	878	ILE	VAL	conflict	UNP Q7ZT66
C	879	ILE	LEU	conflict	UNP Q7ZT66
C	880	PRO	SER	conflict	UNP Q7ZT66
C	881	ARG	PRO	conflict	UNP Q7ZT66
C	882	HIS	GLY	conflict	UNP Q7ZT66
C	883	LEU	TRP	conflict	UNP Q7ZT66
C	884	GLN	CYS	conflict	UNP Q7ZT66
C	885	LEU	ASN	conflict	UNP Q7ZT66
C	886	ALA	SER	conflict	UNP Q7ZT66
C	887	VAL	LEU	conflict	UNP Q7ZT66
C	888	ARG	CYS	conflict	UNP Q7ZT66
C	923	ALA	SER	conflict	UNP Q7ZT66
C	926	ALA	THR	conflict	UNP Q7ZT66
G	1014	ALA	SER	conflict	UNP Q7ZT66
G	1067	GLY	TRP	conflict	UNP Q7ZT66
G	1068	ASN	GLU	conflict	UNP Q7ZT66
G	1069	ALA	ARG	conflict	UNP Q7ZT66
G	1070	ALA	LEU	conflict	UNP Q7ZT66

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
G	1071	ARG	PRO	conflict	UNP Q7ZT66
G	1072	ASP	GLU	conflict	UNP Q7ZT66
G	1073	ASN	ILE	conflict	UNP Q7ZT66
G	1074	LYS	TRP	conflict	UNP Q7ZT66
G	1076	THR	ARG	conflict	UNP Q7ZT66
G	1077	ARG	PRO	conflict	UNP Q7ZT66
G	1078	ILE	VAL	conflict	UNP Q7ZT66
G	1079	ILE	LEU	conflict	UNP Q7ZT66
G	1080	PRO	SER	conflict	UNP Q7ZT66
G	1081	ARG	PRO	conflict	UNP Q7ZT66
G	1082	HIS	GLY	conflict	UNP Q7ZT66
G	1083	LEU	TRP	conflict	UNP Q7ZT66
G	1084	GLN	CYS	conflict	UNP Q7ZT66
G	1085	LEU	ASN	conflict	UNP Q7ZT66
G	1086	ALA	SER	conflict	UNP Q7ZT66
G	1087	VAL	LEU	conflict	UNP Q7ZT66
G	1088	ARG	CYS	conflict	UNP Q7ZT66
G	1123	ALA	SER	conflict	UNP Q7ZT66
G	1126	ALA	THR	conflict	UNP Q7ZT66

- Molecule 5 is a protein called Histone H2B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	D	91	709	447	125	135	2	0	0	0
			Total	C	N	O	S			
5	H	95	744	468	134	140	2	0	0	0

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	1219	GLN	PRO	conflict	UNP P02281
D	1242	LEU	MET	conflict	UNP P02281
D	1257	SER	GLY	conflict	UNP P02281
D	1266	VAL	ILE	conflict	UNP P02281
H	1419	GLN	PRO	conflict	UNP P02281
H	1442	LEU	MET	conflict	UNP P02281
H	1457	SER	GLY	conflict	UNP P02281
H	1466	VAL	ILE	conflict	UNP P02281

- Molecule 6 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	I	39	39	39	0	0
			Total	O		
6	J	39	39	39	0	0
			Total	O		
6	A	9	9	9	0	0
			Total	O		
6	B	8	8	8	0	0
			Total	O		
6	C	16	16	16	0	0
			Total	O		
6	D	9	9	9	0	0
			Total	O		
6	E	17	17	17	0	0
			Total	O		
6	F	16	16	16	0	0
			Total	O		
6	G	13	13	13	0	0
			Total	O		
6	H	5	5	5	0	0

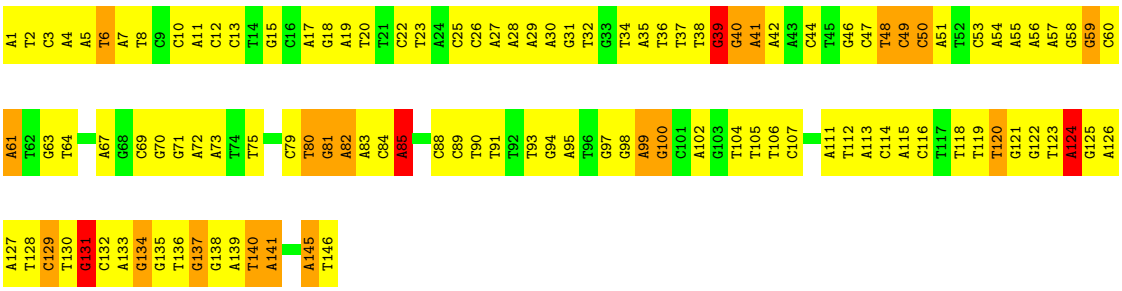
3 Residue-property plots (i)

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. 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. 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.

Note EDS was not executed.

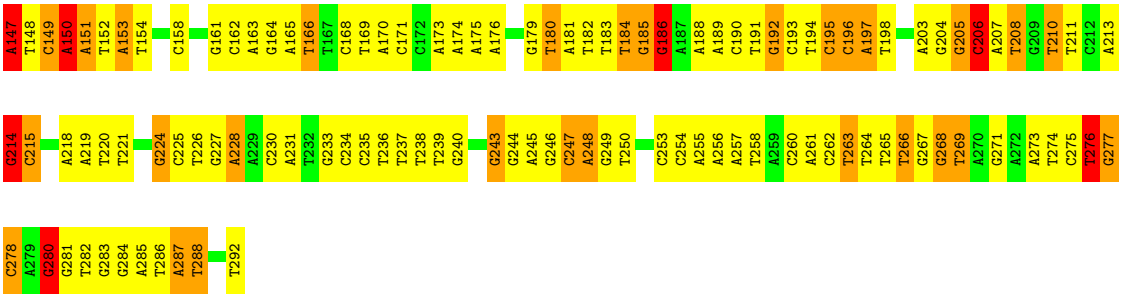
• Molecule 1: Palindromic 146bp Human Alpha-Satellite DNA fragment

Chain I: 21% 63% 14%



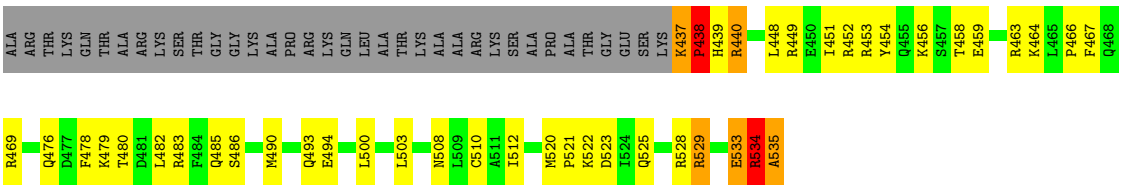
• Molecule 1: Palindromic 146bp Human Alpha-Satellite DNA fragment

Chain J: 23% 53% 19% 5%



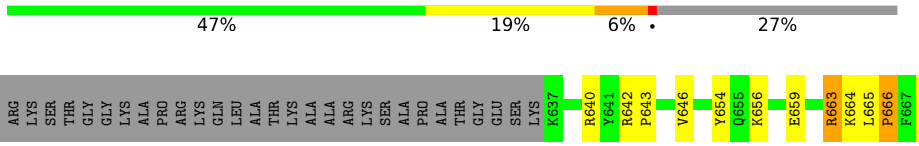
• Molecule 2: Histone H3

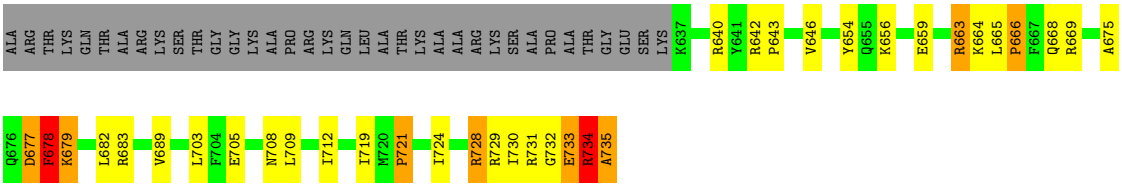
Chain A: 41% 27% 27%



• Molecule 2: Histone H3

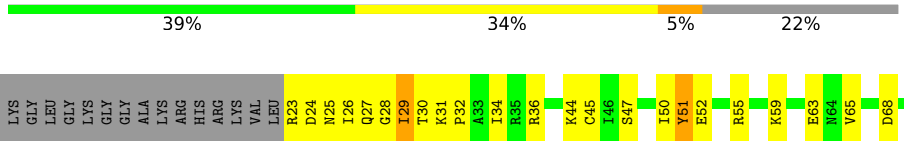
Chain E:

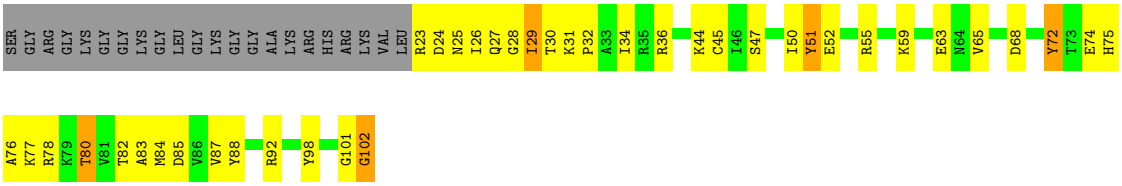




- Molecule 3: Histone H4

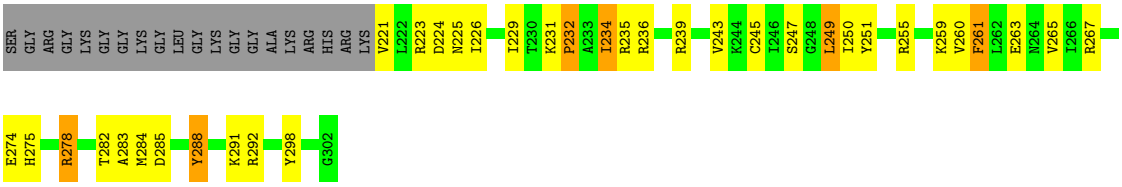
Chain B:





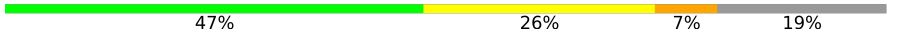
- Molecule 3: Histone H4

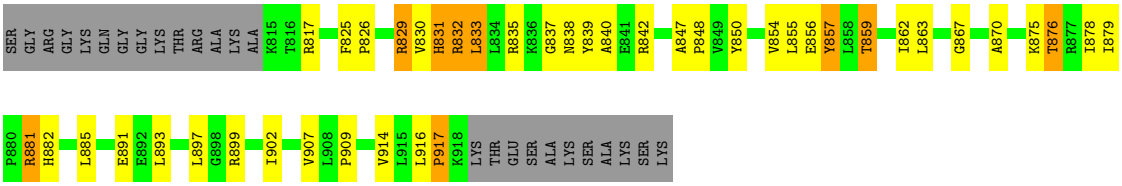
Chain F:



- Molecule 4: Histone H2A

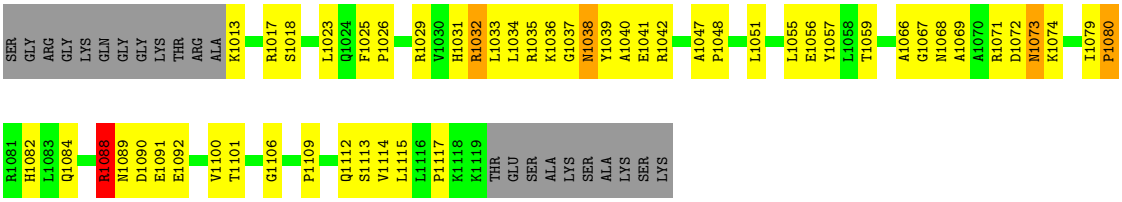
Chain C:





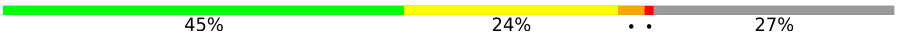
- Molecule 4: Histone H2A

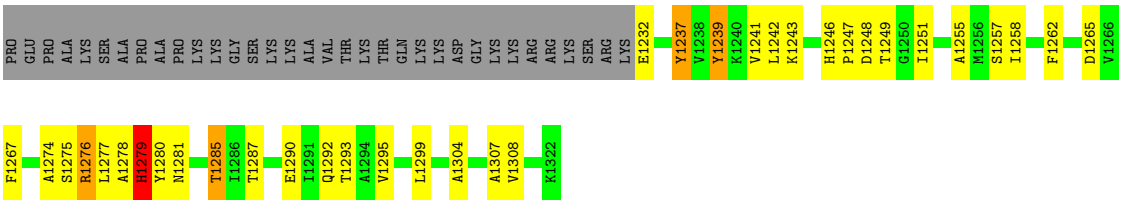
Chain G:



- Molecule 5: Histone H2B

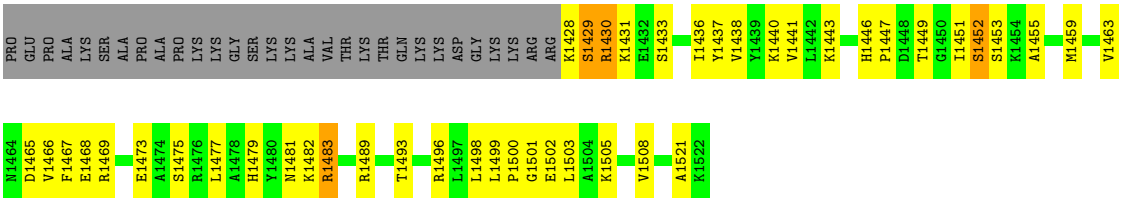
Chain D:





- Molecule 5: Histone H2B

Chain H:



4 Data and refinement statistics (i)

Xtriage (Phenix) and EDS were not executed - this section is therefore incomplete.

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	105.74Å 109.50Å 181.65Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	40.00 – 2.90	Depositor
% Data completeness (in resolution range)	94.1 (40.00-2.90)	Depositor
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
Refinement program	CNS	Depositor
R, R_{free}	0.214 , 0.272	Depositor
Estimated twinning fraction	No twinning to report.	Xtriage
Total number of atoms	12150	wwPDB-VP
Average B, all atoms (Å^2)	55.0	wwPDB-VP

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	I	1.40	11/3354 (0.3%)	1.22	17/5175 (0.3%)
1	J	1.37	14/3354 (0.4%)	1.25	19/5175 (0.4%)
2	A	1.52	7/829 (0.8%)	1.42	5/1111 (0.5%)
2	E	1.82	10/829 (1.2%)	1.57	19/1111 (1.7%)
3	B	1.46	2/640 (0.3%)	1.31	3/856 (0.4%)
3	F	1.70	8/655 (1.2%)	1.51	9/877 (1.0%)
4	C	1.37	3/814 (0.4%)	1.36	5/1099 (0.5%)
4	G	1.22	1/837 (0.1%)	1.25	5/1128 (0.4%)
5	D	1.40	3/720 (0.4%)	1.37	6/969 (0.6%)
5	H	1.43	6/755 (0.8%)	1.22	3/1013 (0.3%)
All	All	1.44	65/12787 (0.5%)	1.30	91/18514 (0.5%)

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	I	0	18
1	J	0	25
2	A	0	1
2	E	0	1
3	B	0	3
3	F	0	1
4	C	0	1
4	G	0	1
5	D	0	3
All	All	0	54

The worst 5 of 65 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	735	ALA	C-O	15.70	1.53	1.23

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	E	735	ALA	CA-CB	15.32	1.84	1.52
2	E	677	ASP	CB-CG	14.08	1.81	1.51
2	E	733	GLU	CG-CD	12.22	1.70	1.51
2	A	533	GLU	CG-CD	10.66	1.68	1.51

The worst 5 of 91 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	A	534	ARG	N-CA-C	12.64	145.14	111.00
1	J	206	DC	O5'-P-OP2	-12.29	94.64	105.70
2	E	677	ASP	CB-CG-OD1	11.81	128.93	118.30
3	F	278	ARG	NE-CZ-NH2	-10.47	115.06	120.30
4	C	862	ILE	CG1-CB-CG2	-10.24	88.87	111.40

There are no chirality outliers.

5 of 54 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	I	39	DG	Sidechain
1	I	41	DA	Sidechain
1	I	44	DC	Sidechain
1	I	48	DT	Sidechain
1	I	49	DC	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	I	2990	0	1651	166	0
1	J	2990	0	1651	154	0
2	A	817	0	858	41	0
2	E	817	0	858	39	0
3	B	633	0	668	39	0
3	F	648	0	688	25	0
4	C	804	0	859	48	0
4	G	827	0	890	56	0
5	D	709	0	727	36	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
5	H	744	0	771	46	0
6	A	9	0	0	1	0
6	B	8	0	0	0	0
6	C	16	0	0	0	0
6	D	9	0	0	1	0
6	E	17	0	0	4	0
6	F	16	0	0	0	0
6	G	13	0	0	1	0
6	H	5	0	0	2	0
6	I	39	0	0	12	0
6	J	39	0	0	5	0
All	All	12150	0	9621	567	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 27.

The worst 5 of 567 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:E:735:ALA:CB	2:E:735:ALA:CA	1.84	1.52
3:B:84:MET:SD	3:B:84:MET:CE	2.02	1.46
2:E:677:ASP:CB	2:E:677:ASP:CG	1.81	1.46
4:C:850:TYR:OH	5:D:1292:GLN:HG3	1.45	1.16
2:E:677:ASP:OD2	6:E:97:HOH:O	1.70	1.10

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	97/135 (72%)	91 (94%)	4 (4%)	2 (2%)	7 26

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
2	E	97/135 (72%)	93 (96%)	0	4 (4%)	3	11
3	B	78/102 (76%)	75 (96%)	3 (4%)	0	100	100
3	F	80/102 (78%)	76 (95%)	3 (4%)	1 (1%)	12	37
4	C	102/129 (79%)	93 (91%)	8 (8%)	1 (1%)	15	45
4	G	105/129 (81%)	97 (92%)	7 (7%)	1 (1%)	15	45
5	D	89/125 (71%)	84 (94%)	5 (6%)	0	100	100
5	H	93/125 (74%)	83 (89%)	7 (8%)	3 (3%)	4	16
All	All	741/982 (76%)	692 (93%)	37 (5%)	12 (2%)	9	32

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
3	F	224	ASP
2	E	678	PHE
2	E	733	GLU
5	H	1430	ARG
5	H	1501	GLY

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	86/111 (78%)	80 (93%)	6 (7%)	15	41
2	E	86/111 (78%)	78 (91%)	8 (9%)	9	27
3	B	65/78 (83%)	64 (98%)	1 (2%)	65	87
3	F	67/78 (86%)	65 (97%)	2 (3%)	41	75
4	C	83/100 (83%)	76 (92%)	7 (8%)	11	31
4	G	85/100 (85%)	78 (92%)	7 (8%)	11	32
5	D	77/105 (73%)	73 (95%)	4 (5%)	23	55
5	H	81/105 (77%)	75 (93%)	6 (7%)	13	38
All	All	630/788 (80%)	589 (94%)	41 (6%)	17	45

5 of 41 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	G	1038	ASN
5	H	1429	SER
4	G	1041	GLU
4	G	1088	ARG
5	H	1433	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 9 such sidechains are listed below:

Mol	Chain	Res	Type
5	H	1479	HIS
5	H	1492	GLN
5	D	1292	GLN
3	F	293	GLN
4	G	1031	HIS

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 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)

EDS was not executed - this section is therefore empty.

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

EDS was not executed - this section is therefore empty.

6.3 Carbohydrates (i)

EDS was not executed - this section is therefore empty.

6.4 Ligands (i)

EDS was not executed - this section is therefore empty.

6.5 Other polymers (i)

EDS was not executed - this section is therefore empty.