wwPDB X-ray Structure Validation Summary Report (i)

Oct 21, 2023 – 01:47 PM EDT

PDB ID	:	8GB2
Title	:	Crystal structure of Apo-SAMHD1
Authors	:	Egleston, M.; Dong, L.; Howlader, A.H.; Bhat, S.; Orris, B.; Bianchet, M.A.; Greenberg, M.M.; Stivers, J.T.
Deposited on	:	2023-02-24
Resolution	:	3.07 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:

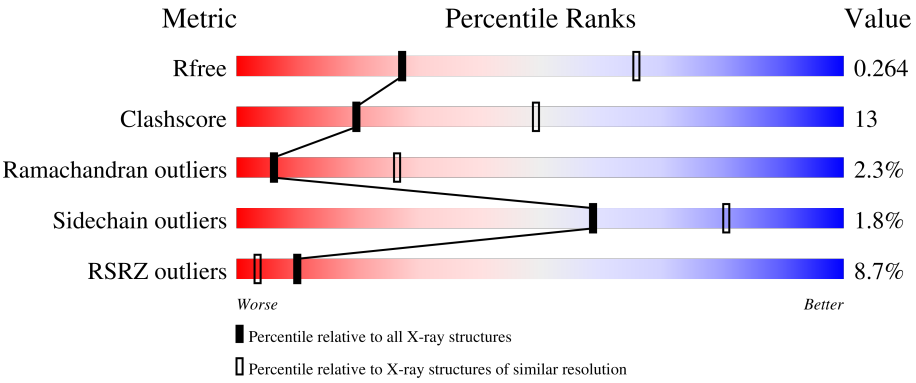
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.07 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	1447 (3.10-3.06)
Clashscore	141614	1546 (3.10-3.06)
Ramachandran outliers	138981	1487 (3.10-3.06)
Sidechain outliers	138945	1486 (3.10-3.06)
RSRZ outliers	127900	1416 (3.10-3.06)

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	516	65% 20% · 14%
1	B	516	11% 59% 25% · 14%
1	C	516	18% 49% 34% · 13%
1	D	516	% 67% 17% · 14%

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
2	FE	D	701	-	-	-	X

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 28956 atoms, of which 14422 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 Deoxynucleoside triphosphate triphosphohydrolase SAMHD1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	D	445	Total	C	H	N	O	S	0	0	0
			7243	2326	3607	633	658	19			
1	A	443	Total	C	H	N	O	S	0	0	0
			7224	2323	3601	631	650	19			
1	B	442	Total	C	H	N	O	S	0	0	0
			7203	2314	3589	630	651	19			
1	C	447	Total	C	H	N	O	S	0	0	0
			7281	2340	3625	636	660	20			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	111	SER	-	expression tag	UNP Q9Y3Z3
D	112	MET	-	expression tag	UNP Q9Y3Z3
A	111	SER	-	expression tag	UNP Q9Y3Z3
A	112	MET	-	expression tag	UNP Q9Y3Z3
B	111	SER	-	expression tag	UNP Q9Y3Z3
B	112	MET	-	expression tag	UNP Q9Y3Z3
C	111	SER	-	expression tag	UNP Q9Y3Z3
C	112	MET	-	expression tag	UNP Q9Y3Z3

- Molecule 2 is FE (III) ION (three-letter code: FE) (formula: Fe).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	D	1	Total	Fe	0	0
			1	1		
2	A	1	Total	Fe	0	0
			1	1		
2	B	1	Total	Fe	0	0
			1	1		
2	C	1	Total	Fe	0	0
			1	1		

- Molecule 3 is water.

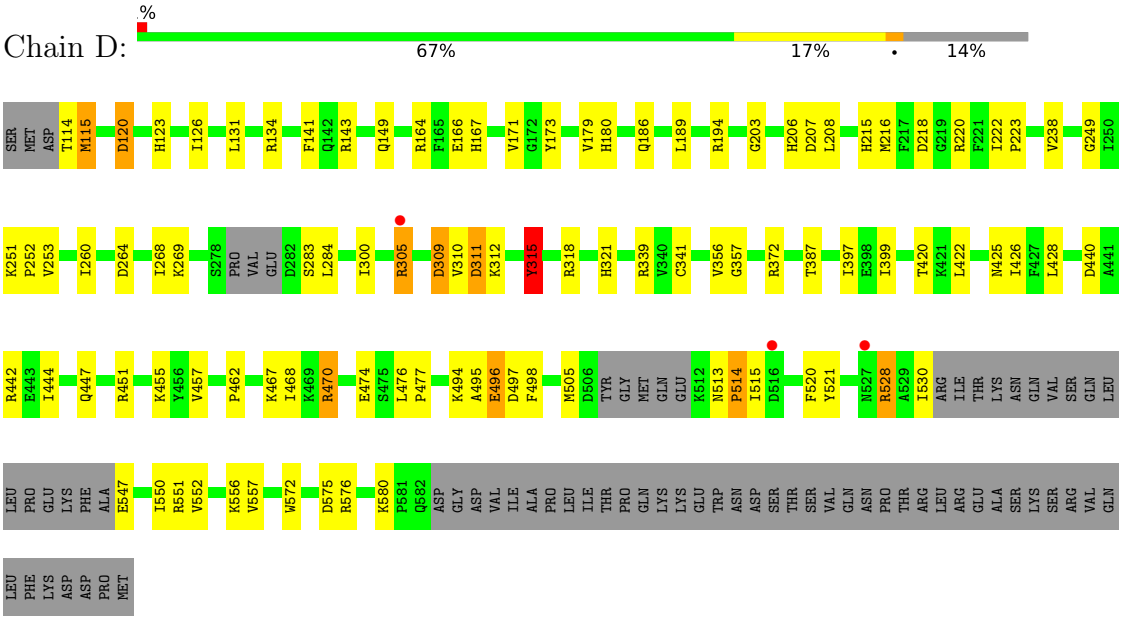
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total O 1 1	0	0

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 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.

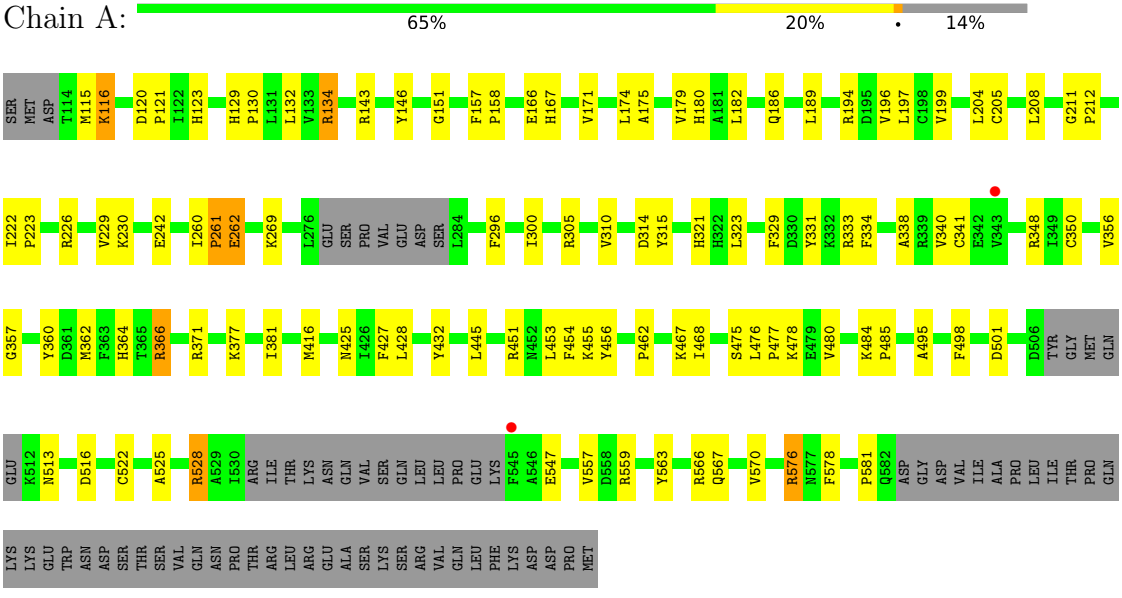
• Molecule 1: Deoxynucleoside triphosphate triphosphohydrolase SAMHD1

Chain D:



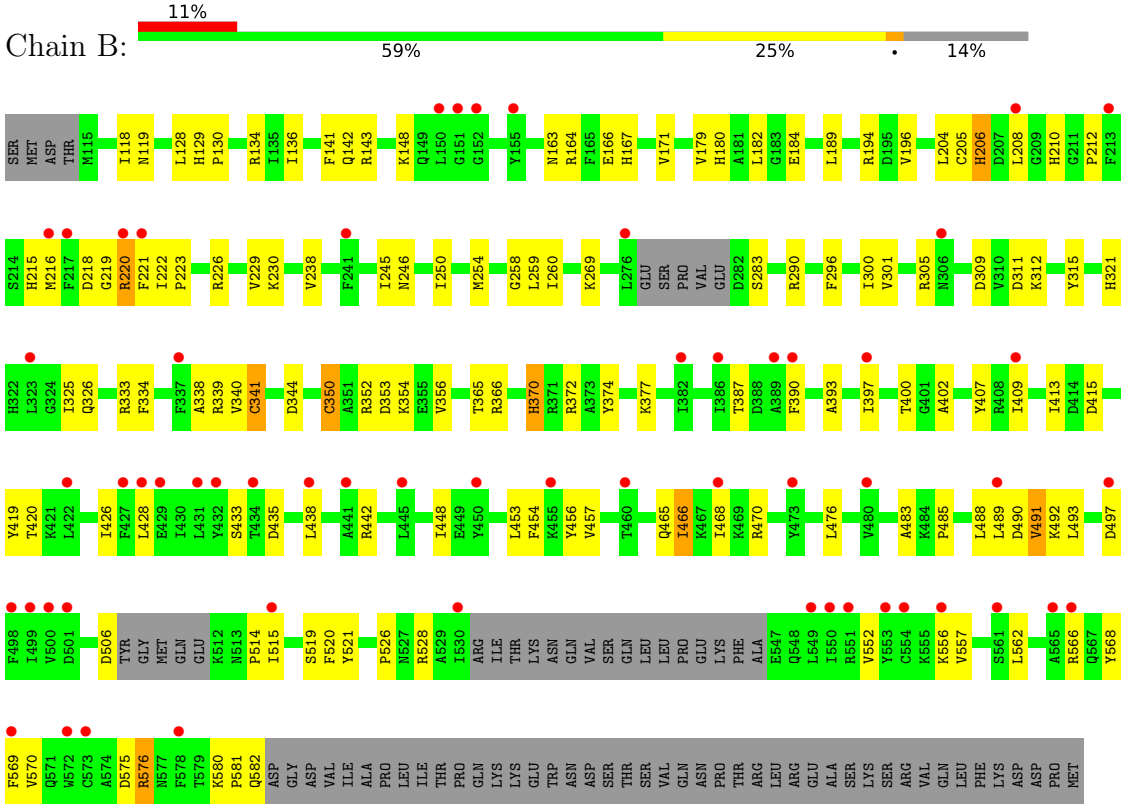
• Molecule 1: Deoxynucleoside triphosphate triphosphohydrolase SAMHD1

Chain A:



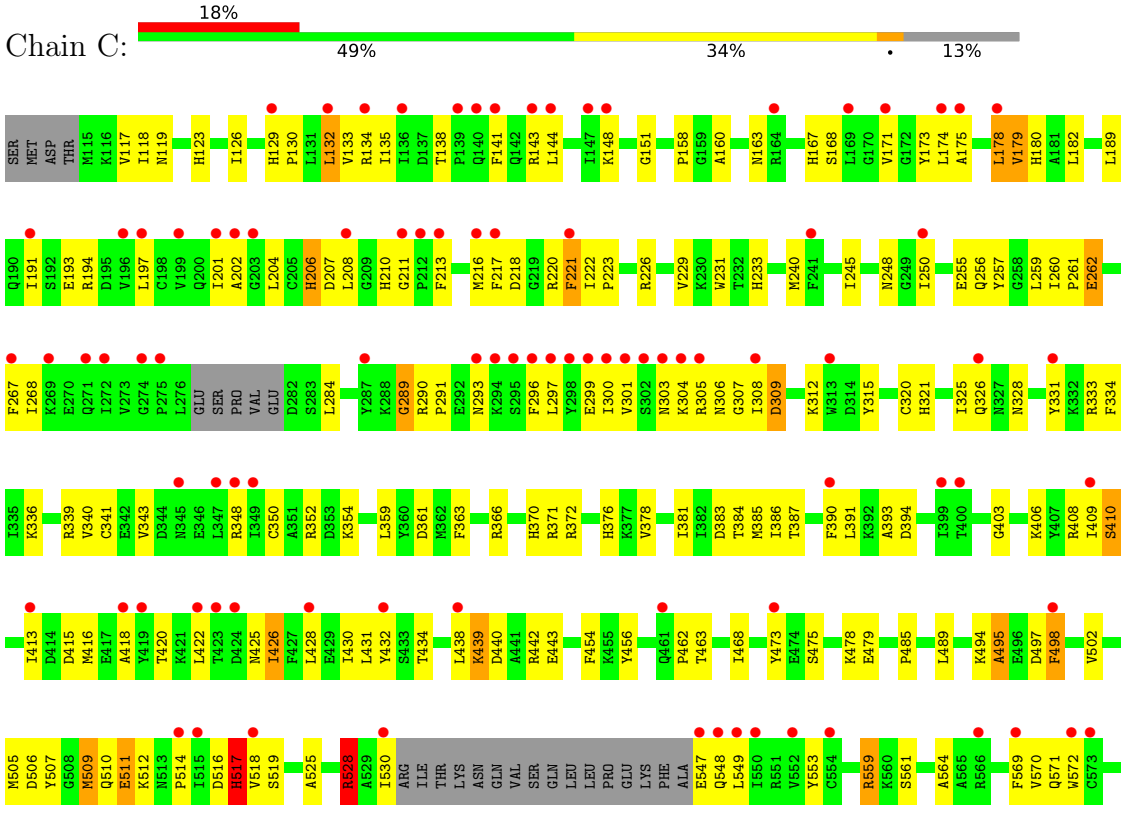
• Molecule 1: Deoxynucleoside triphosphate triphosphohydrolase SAMHD1

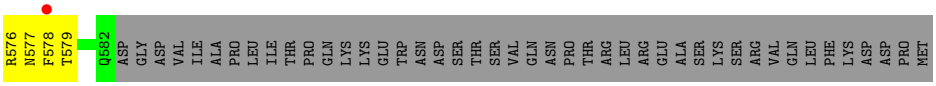
Chain B:



• Molecule 1: Deoxynucleoside triphosphate triphosphohydrolase SAMHD1

Chain C:




R576
N577
F578
T579
Q582
ASP
GLY
ASP
VAL
ILE
ALA
PRO
LEU
ILE
THR
PRO
GLN
LYS
LYS
GLU
TRP
ASN
ASP
SER
THR
SER
VAL
GLN
ASN
PRO
THR
ARG
LEU
ARG
GLU
ALA
SER
LYS
SER
ARG
VAL
GLN
LEU
PHE
LYS
ASP
ASP
PRO
MET

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	202.57Å 81.88Å 154.03Å 90.00° 118.52° 90.00°	Depositor
Resolution (Å)	27.61 – 3.07 135.34 – 2.78	Depositor EDS
% Data completeness (in resolution range)	99.4 (27.61-3.07) 99.6 (135.34-2.78)	Depositor EDS
R_{merge}	0.13	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.53 (at 2.77Å)	Xtriage
Refinement program	PHENIX (1.20.1_4487: ???)	Depositor
R, R_{free}	0.197 , 0.250 0.212 , 0.264	Depositor DCC
R_{free} test set	2864 reflections (5.10%)	wwPDB-VP
Wilson B-factor (Å^2)	79.2	Xtriage
Anisotropy	0.041	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 86.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	28956	wwPDB-VP
Average B, all atoms (Å^2)	135.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.71% 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: FE

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	0.67	0/3708	0.84	3/5001 (0.1%)
1	B	0.49	1/3698 (0.0%)	0.70	0/4987
1	C	0.46	1/3742 (0.0%)	0.69	0/5047
1	D	0.65	0/3720	0.85	4/5017 (0.1%)
All	All	0.58	2/14868 (0.0%)	0.77	7/20052 (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	10
1	B	0	9
1	C	0	9
1	D	0	10
All	All	0	38

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	B	341	CYS	CB-SG	-6.65	1.71	1.82
1	C	320	CYS	CB-SG	-6.01	1.72	1.82

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	311	ASP	CB-CG-OD2	10.57	127.82	118.30
1	A	314	ASP	CB-CG-OD1	8.51	125.96	118.30
1	D	120	ASP	CB-CG-OD2	-7.31	111.72	118.30

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	D	315	TYR	CA-CB-CG	6.00	124.80	113.40
1	D	311	ASP	OD1-CG-OD2	-5.75	112.37	123.30

There are no chirality outliers.

5 of 38 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	134	ARG	Sidechain
1	D	305	ARG	Sidechain
1	D	318	ARG	Sidechain
1	D	372	ARG	Sidechain
1	D	442	ARG	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	A	3623	3601	3601	68	1
1	B	3614	3589	3589	96	1
1	C	3656	3625	3625	151	1
1	D	3636	3607	3607	64	0
2	A	1	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	1	0	0	1	0
All	All	14534	14422	14422	368	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 13.

The worst 5 of 368 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:A:242:GLU:OE1	1:A:269:LYS:NZ	2.00	0.94

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:513:ASN:ND2	1:A:516:ASP:OD2	2.01	0.94
1:B:428:LEU:HD13	1:C:425:ASN:HB3	1.52	0.89
1:C:134:ARG:NH2	1:C:250:ILE:HD13	1.88	0.88
1:D:220:ARG:HG2	1:D:387:THR:HG21	1.55	0.87

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)
1:B:344:ASP:O	1:C:577:ASN:ND2[1_545]	1.93	0.27
1:A:262:GLU:OE2	1:A:262:GLU:OE2[2_555]	2.13	0.07

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	435/516 (84%)	397 (91%)	33 (8%)	5 (1%)	14 44
1	B	434/516 (84%)	385 (89%)	38 (9%)	11 (2%)	5 25
1	C	441/516 (86%)	351 (80%)	70 (16%)	20 (4%)	2 13
1	D	437/516 (85%)	409 (94%)	24 (6%)	4 (1%)	17 49
All	All	1747/2064 (85%)	1542 (88%)	165 (9%)	40 (2%)	6 26

5 of 40 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	491	VAL
1	C	439	LYS
1	C	498	PHE
1	C	517	HIS
1	D	496	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
1	A	392/461 (85%)	386 (98%)	6 (2%)	65 84
1	B	392/461 (85%)	385 (98%)	7 (2%)	59 80
1	C	396/461 (86%)	385 (97%)	11 (3%)	43 71
1	D	395/461 (86%)	390 (99%)	5 (1%)	69 86
All	All	1575/1844 (85%)	1546 (98%)	29 (2%)	59 80

5 of 29 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	419	TYR
1	C	528	ARG
1	B	506	ASP
1	C	383	ASP
1	B	497	ASP

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

Mol	Chain	Res	Type
1	C	248	ASN
1	C	322	HIS
1	A	527	ASN
1	B	322	HIS
1	B	447	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 4 ligands modelled in this entry, 4 are monoatomic - leaving 0 for Mogul analysis.

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 (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	A	443/516 (85%)	0.01	2 (0%) 91 80	57, 82, 145, 229	0
1	B	442/516 (85%)	0.54	58 (13%) 3 1	83, 139, 206, 241	0
1	C	447/516 (86%)	0.93	92 (20%) 1 0	96, 169, 221, 269	0
1	D	445/516 (86%)	0.06	3 (0%) 87 74	59, 85, 160, 233	0
All	All	1777/2064 (86%)	0.38	155 (8%) 10 3	57, 113, 205, 269	0

The worst 5 of 155 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	498	PHE	8.4
1	C	296	PHE	7.7
1	B	480	VAL	6.8
1	C	297	LEU	6.8
1	C	300	ILE	6.0

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

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
2	FE	D	701	1/1	0.01	0.52	201,201,201,201	0
2	FE	C	701	1/1	0.75	0.16	163,163,163,163	0
2	FE	A	701	1/1	0.87	0.27	76,76,76,76	0
2	FE	B	701	1/1	0.95	0.15	111,111,111,111	0

6.5 Other polymers (i)

There are no such residues in this entry.