wwPDB X-ray Structure Validation Summary Report (i)

Sep 3, 2023 – 01:28 PM EDT

PDB ID	:	3S6P
Title	:	Crystal Structure of Helicoverpa Armigera Stunt Virus
Authors	:	Speir, J.A.; Chen, Z.; Taylor, D.J.; Johnson, J.E.
Deposited on	:	2011-05-25
Resolution	:	2.50 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:

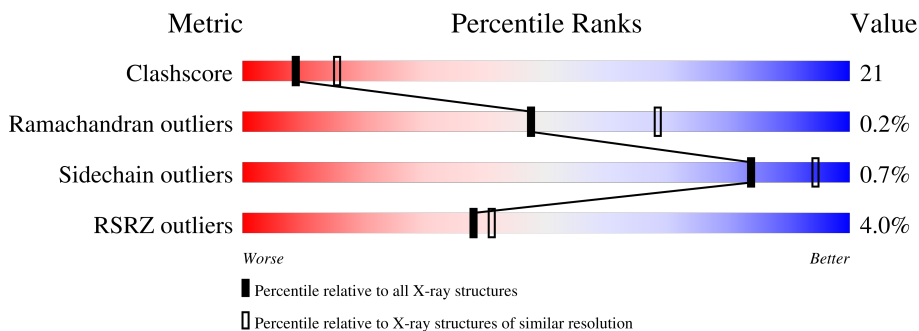|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

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 2.50 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.
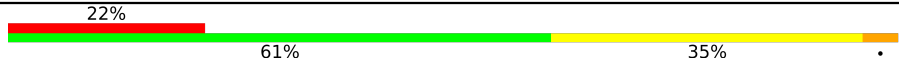
Metric	Percentile Ranks	Value
Clashscore		21
Ramachandran outliers		0.2%
Sidechain outliers		0.7%
RSRZ outliers		4.0%

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))
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	575	4% — 65% · 27% · 8%
1	B	575	3% — 65% · 26% · 8%
1	C	575	.% — 65% · 28% · 7%
1	D	575	.% — 70% · 23% · 7%
2	E	72	8% — 32% · 14% · 54%
2	F	72	6% — 22% · 11% · 67%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	G	72	22% · · · 61% · · · 35% ·
2	H	72	19% · · · 54% · · · 33% · · · 12%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 19383 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 Capsid protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	529	Total	C	N	O	S	0	4	0
			4145	2629	695	809	12			
1	B	530	Total	C	N	O	S	0	3	0
			4143	2628	694	809	12			
1	C	536	Total	C	N	O	S	0	2	0
			4182	2652	704	814	12			
1	D	532	Total	C	N	O	S	0	6	0
			4177	2652	701	811	13			

- Molecule 2 is a protein called Capsid protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	E	33	Total	C	N	O	S	0	0	0
			227	143	36	47	1			
2	F	24	Total	C	N	O	S	0	0	0
			164	103	26	34	1			
2	G	72	Total	C	N	O	S	0	5	0
			547	339	112	93	3			
2	H	63	Total	C	N	O	S	0	5	0
			485	300	99	83	3			

- Molecule 3 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	1	Total	Ca	0	0
			1	1		
3	C	1	Total	Ca	0	0
			1	1		

- Molecule 4 is CHLORIDE ION (three-letter code: CL) (formula: Cl).

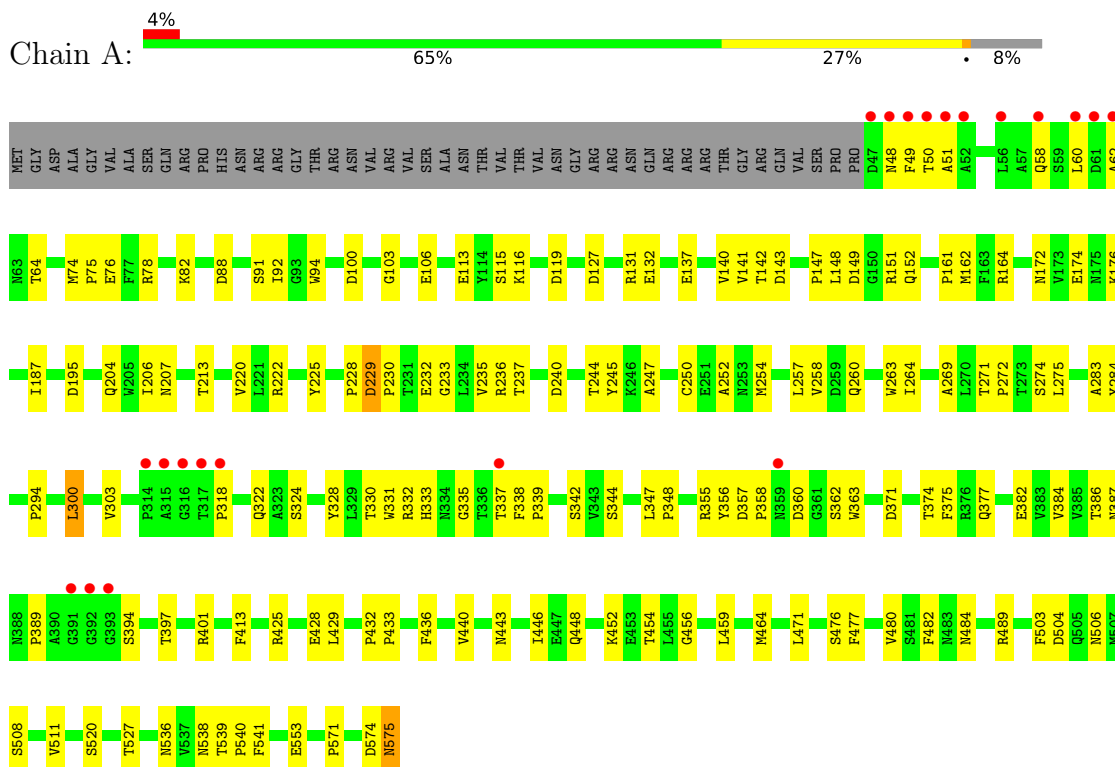| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	C	1	Total 1	Cl 1	0	0
4	D	1	Total 1	Cl 1	0	0

• Molecule 5 is water.

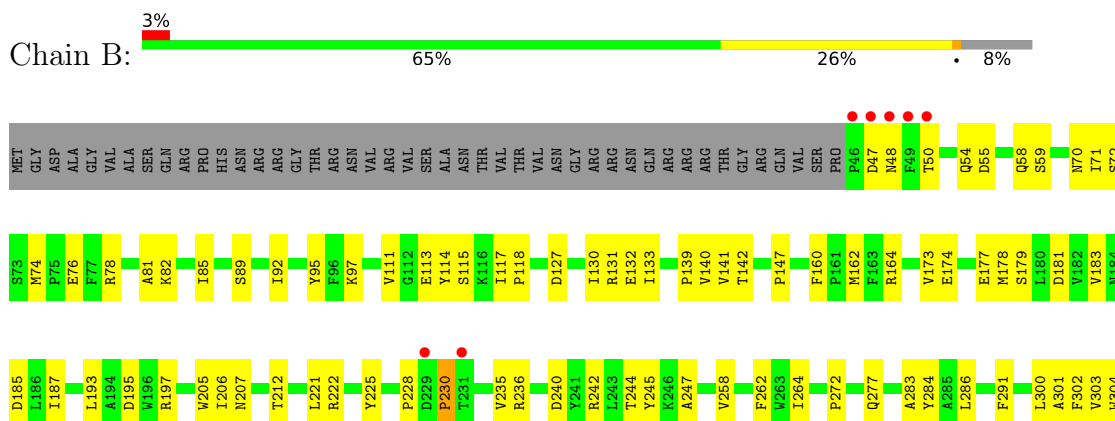| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	330	Total 330	O 330	0	0
5	E	7	Total 7	O 7	0	0
5	B	313	Total 313	O 313	0	0
5	F	4	Total 4	O 4	0	0
5	C	290	Total 290	O 290	0	0
5	G	43	Total 43	O 43	0	0
5	D	287	Total 287	O 287	0	0
5	H	35	Total 35	O 35	0	0

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.
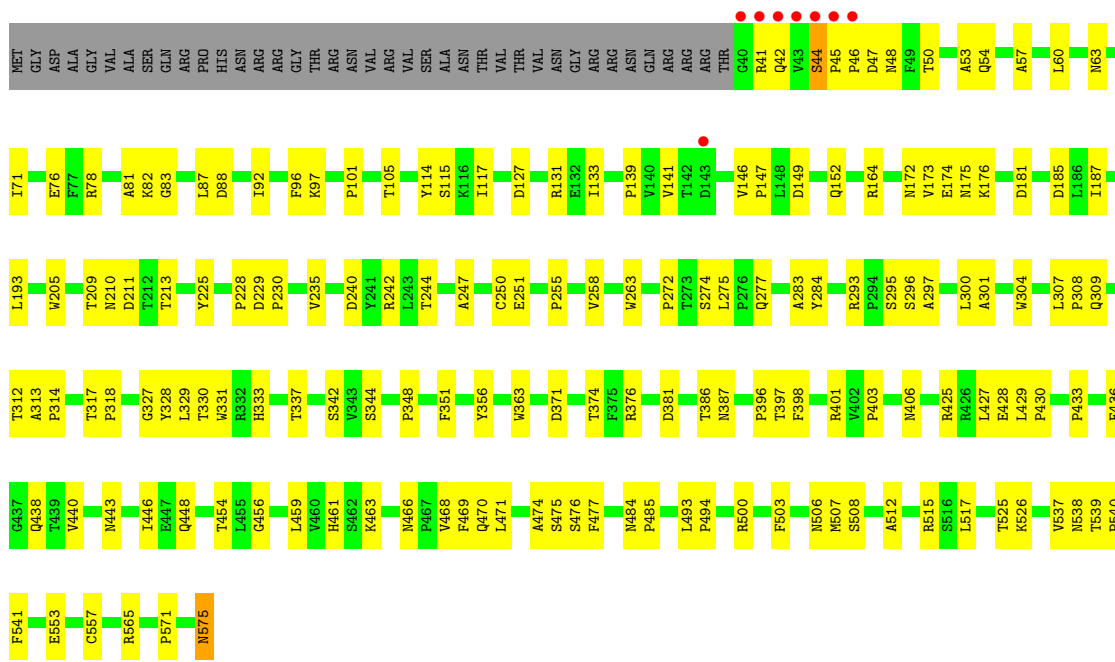
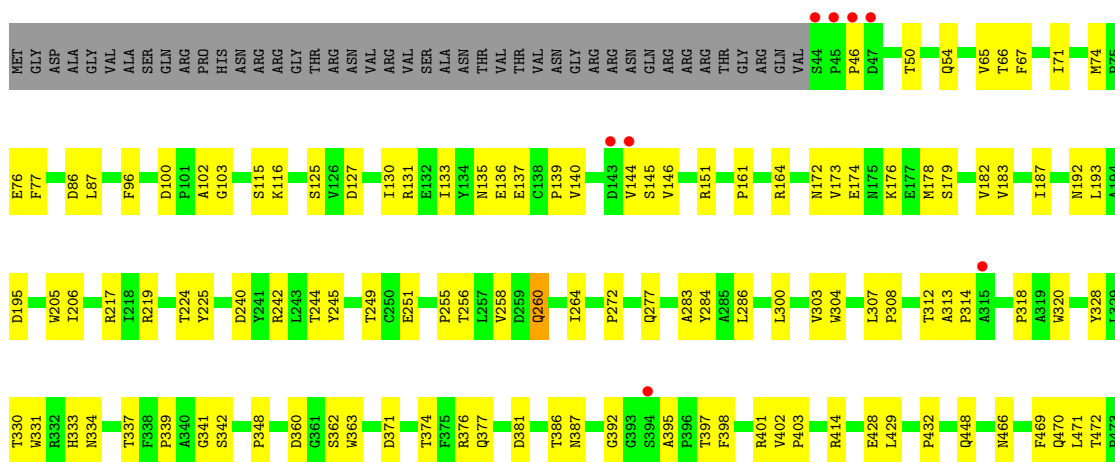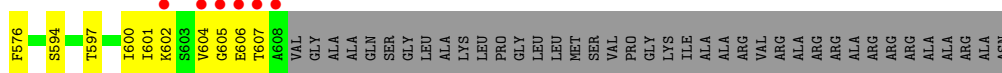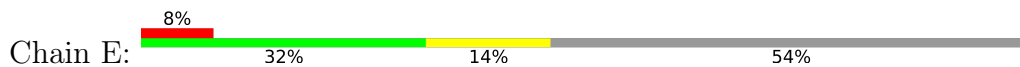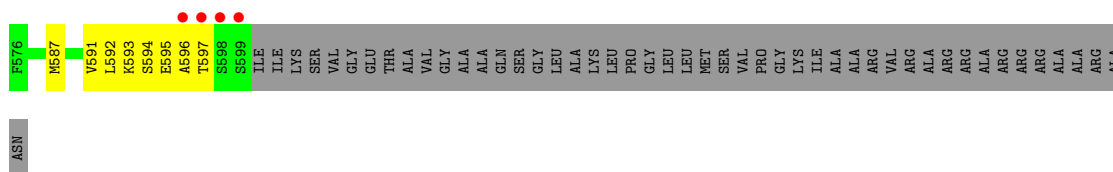• Molecule 1: Capsid protein

Chain A:

• Molecule 1: Capsid protein

Chain B:

• Molecule 1: Capsid protein

Chain C:

• Molecule 1: Capsid protein

Chain D:

A474 N484 L493 P494 R500 F503 D504 Q505 N506 M507 S508 V511 T525 K526 N552 E553 E554 N575

• Molecule 2: Capsid protein

Chain E:

8%

32% 14% 54%

F576 S594 T597 I600 I601 K602 S603 V604 G605 E606 T607 A608 VAL GLY ALA ALA GLN SER GLY LEU ALA LYS LEU PRO GLY LEU MET SER VAL PRO GLY LYS ILE ALA ALA ARG VAL ALA ARG ALA ARG ARG ALA ARG ALA ASN

• Molecule 2: Capsid protein

Chain F:

6%

22% 11% 67%

F576 M587 V591 L592 K593 S594 E595 A596 T597 S598 S599 ILE ILE LYS SER VAL GLY GLU THR ALA VAL GLY ALA ALA GLN SER GLY LEU ALA LYS LEU PRO GLY LEU LEU MET SER VAL PRO GLY LYS ILE ALA ALA ARG VAL ARG ALA ARG ALA ARG ARG ARG ALA ALA ARG ALA

ASN

• Molecule 2: Capsid protein

Chain G:

22%

61% 35% •

F576 S589 K593 S594 E595 A596 T597 S598 S599 I600 I601 K602 S603 V604 G605 E606 T607 A608 V609 G610 A611 A612 Q613 S614 G615 L623 M624 A631 A632 R633 V634 R637 R638 R641 R642 R645 A646 N647

• Molecule 2: Capsid protein

Chain H:

19%

54% 33% 12%

F576 L592 K593 S594 E595 A596 T597 S598 S599 I600 I601 K602 S603 V604 G605 E606 T607 A608 V609 G610 ALA ALA GLN SER GLY LEU ALA LYS LEU P620 M624 R633 V634 R635 A636 R637 R638 A639 R640 R641 R642 A643 A644 R645 A646 N647

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1	Depositor
Cell constants a, b, c, α, β, γ	404.12Å 405.57Å 406.04Å 119.20° 114.44° 94.79°	Depositor
Resolution (Å)	40.00 – 2.50 39.85 – 2.50	Depositor EDS
% Data completeness (in resolution range)	27.3 (40.00-2.50) 27.3 (39.85-2.50)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.64 (at 2.51Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.237 , (Not available) 0.240 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	wwPDB-VP
Wilson B-factor (Å^2)	32.9	Xtriage
Anisotropy	0.132	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.31 , 60.1	EDS
L-test for twinning[2]	$<|L|>$ = 0.48, $<L^2>$ = 0.31	Xtriage
Estimated twinning fraction	0.002 for k,h,-h-k-l 0.002 for l,-h-k-l,h 0.002 for -h-k-l,l,k	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	19383	wwPDB-VP
Average B, all atoms (Å^2)	40.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.55% 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 (i)

5.1 Standard geometry (i)

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

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	0.56	0/4270	0.74	0/5846
1	B	0.55	0/4266	0.73	0/5842
1	C	0.55	0/4303	0.74	1/5892 (0.0%)
1	D	0.56	0/4311	0.74	0/5902
2	E	0.52	0/228	0.57	0/307
2	F	0.62	0/165	0.56	0/222
2	G	0.54	0/565	0.73	0/753
2	H	0.57	0/502	0.69	0/664
All	All	0.55	0/18610	0.73	1/25428 (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
2	G	0	2

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	C	44	SER	C-N-CD	-5.05	109.49	120.60

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	G	645[A]	ARG	Sidechain

Continued on next page...

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Continued from previous page...

Mol	Chain	Res	Type	Group
2	G	645[B]	ARG	Sidechain

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	A	4145	0	3962	172	0
1	B	4143	0	3957	176	0
1	C	4182	0	4002	169	0
1	D	4177	0	4004	137	0
2	E	227	0	233	25	0
2	F	164	0	163	10	0
2	G	547	0	608	61	0
2	H	485	0	535	41	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
4	C	1	0	0	0	0
4	D	1	0	0	0	0
5	A	330	0	0	18	0
5	B	313	0	0	5	0
5	C	290	0	0	13	0
5	D	287	0	0	8	0
5	E	7	0	0	0	0
5	F	4	0	0	0	0
5	G	43	0	0	9	0
5	H	35	0	0	8	0
All	All	19383	0	17464	730	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 21.

The worst 5 of 730 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:181:ASP:CB	1:D:334:ASN:HD22	1.44	1.28
1:C:44:SER:OG	1:C:45:PRO:HA	1.36	1.23

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:575:ASN:C	1:A:575:ASN:HD22	1.46	1.15
2:G:593:LYS:HD3	2:G:601:ILE:HD13	1.25	1.13
1:B:575:ASN:C	1:B:575:ASN:HD22	1.49	1.12

There are no symmetry-related clashes.

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	531/575 (92%)	509 (96%)	21 (4%)	1 (0%)	47	68
1	B	531/575 (92%)	509 (96%)	21 (4%)	1 (0%)	47	68
1	C	536/575 (93%)	512 (96%)	23 (4%)	1 (0%)	47	68
1	D	536/575 (93%)	517 (96%)	19 (4%)	0	100	100
2	E	31/72 (43%)	31 (100%)	0	0	100	100
2	F	22/72 (31%)	22 (100%)	0	0	100	100
2	G	75/72 (104%)	71 (95%)	3 (4%)	1 (1%)	12	21
2	H	63/72 (88%)	62 (98%)	1 (2%)	0	100	100
All	All	2325/2588 (90%)	2233 (96%)	88 (4%)	4 (0%)	47	68

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	296	SER
2	G	609	VAL
1	A	229	ASP
1	B	230	PRO

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	449/483 (93%)	446 (99%)	3 (1%)	84	94
1	B	449/483 (93%)	446 (99%)	3 (1%)	84	94
1	C	453/483 (94%)	450 (99%)	3 (1%)	84	94
1	D	454/483 (94%)	452 (100%)	2 (0%)	91	97
2	E	24/49 (49%)	24 (100%)	0	100	100
2	F	17/49 (35%)	17 (100%)	0	100	100
2	G	54/49 (110%)	52 (96%)	2 (4%)	34	60
2	H	49/49 (100%)	48 (98%)	1 (2%)	55	79
All	All	1949/2128 (92%)	1935 (99%)	14 (1%)	84	94

5 of 14 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	229	ASP
1	C	575	ASN
2	H	595	GLU
1	D	260	GLN
1	D	575	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 47 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	484	ASN
1	D	152	GLN
1	C	506	ASN
2	G	613	GLN
1	D	260	GLN

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 ⓘ

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 ⓘ

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(Å^2)	Q<0.9
1	A	529/575 (92%)	-0.60	21 (3%)	38	41	19, 33, 84, 133	0
1	B	530/575 (92%)	-0.62	15 (2%)	53	56	18, 31, 86, 125	0
1	C	536/575 (93%)	-0.81	8 (1%)	73	75	19, 31, 73, 139	0
1	D	532/575 (92%)	-0.81	8 (1%)	73	75	19, 30, 69, 124	1 (0%)
2	E	33/72 (45%)	0.17	6 (18%)	1	1	24, 42, 133, 140	0
2	F	24/72 (33%)	0.10	4 (16%)	1	1	21, 30, 119, 131	0
2	G	72/72 (100%)	0.40	16 (22%)	0	0	25, 48, 123, 127	0
2	H	63/72 (87%)	0.51	14 (22%)	0	0	23, 49, 125, 130	0
All	All	2319/2588 (89%)	-0.62	92 (3%)	38	41	18, 32, 87, 140	1 (0%)

The worst 5 of 92 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	H	598	SER	7.6
1	C	43	VAL	7.3
2	F	599	SER	7.1
1	A	60	LEU	6.5
2	G	646	ALA	6.4

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 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, 95th 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
3	CA	C	576	1/1	0.37	0.30	95,95,95,95	0
4	CL	D	576	1/1	0.94	0.07	56,56,56,56	0
4	CL	C	577	1/1	0.96	0.07	64,64,64,64	0
3	CA	B	576	1/1	0.98	0.05	62,62,62,62	0

6.5 Other polymers ⓘ

There are no such residues in this entry.