



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 03:04 am GMT

PDB ID : 2B43  
Title : Crystal structure of the Norwalk virus RNA dependent RNA polymerase from strain Hu/NLV/Dresden174/1997/GE  
Authors : Hogbom, M.; Rohayem, J.; Unge, T.; Jones, T.A.  
Deposited on : 2005-09-22  
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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

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

MolProbity : 4.02b-467  
Xtriage (Phenix) : 1.9-1692  
EDS : trunk28620  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : recalc28949

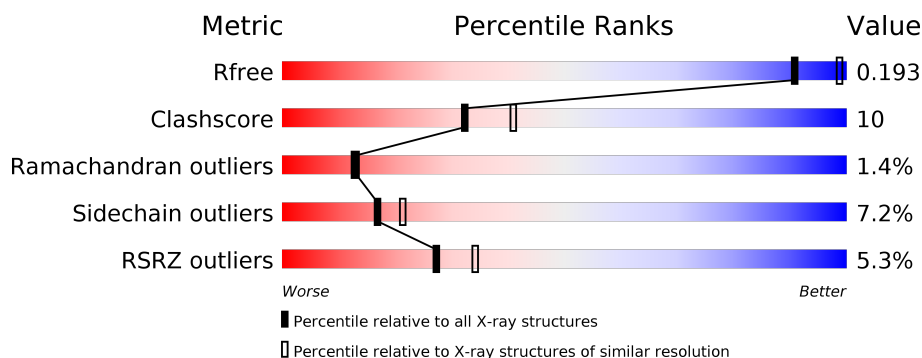
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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))
$R_{free}$	100719	4130 (2.30-2.30)
Clashscore	112137	4751 (2.30-2.30)
Ramachandran outliers	110173	4705 (2.30-2.30)
Sidechain outliers	110143	4704 (2.30-2.30)
RSRZ outliers	101464	4156 (2.30-2.30)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	526	<div> <div>4%</div> <div> <div></div> <div>76%</div> <div>15%</div> <div>5%</div> </div> </div>
1	B	526	<div> <div>6%</div> <div> <div></div> <div>76%</div> <div>15%</div> <div>5%</div> </div> </div>
1	C	526	<div> <div>5%</div> <div> <div></div> <div>73%</div> <div>18%</div> <div>5%</div> </div> </div>
1	D	526	<div> <div>6%</div> <div> <div></div> <div>77%</div> <div>14%</div> <div>5%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 16838 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 non-structural polyprotein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	501	Total	C	N	O	S	0	0	0
			3929	2498	673	736	22			
1	B	501	Total	C	N	O	S	0	0	0
			3929	2498	673	736	22			
1	C	501	Total	C	N	O	S	0	0	0
			3929	2498	673	736	22			
1	D	501	Total	C	N	O	S	0	0	0
			3929	2498	673	736	22			

There are 92 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	CLONING ARTIFACT	UNP Q6REV3
A	34	THR	ALA	SEE REMARK 999	UNP Q6REV3
A	107	THR	ALA	SEE REMARK 999	UNP Q6REV3
A	153	GLY	GLU	SEE REMARK 999	UNP Q6REV3
A	236	LYS	ARG	SEE REMARK 999	UNP Q6REV3
A	257	ALA	SER	SEE REMARK 999	UNP Q6REV3
A	291	LYS	THR	SEE REMARK 999	UNP Q6REV3
A	363	LYS	ARG	SEE REMARK 999	UNP Q6REV3
A	511	VAL	-	CLONING ARTIFACT	UNP Q6REV3
A	512	ASP	-	CLONING ARTIFACT	UNP Q6REV3
A	513	LYS	-	CLONING ARTIFACT	UNP Q6REV3
A	514	LEU	-	CLONING ARTIFACT	UNP Q6REV3
A	515	ALA	-	CLONING ARTIFACT	UNP Q6REV3
A	516	ALA	-	CLONING ARTIFACT	UNP Q6REV3
A	517	ALA	-	CLONING ARTIFACT	UNP Q6REV3
A	518	LEU	-	CLONING ARTIFACT	UNP Q6REV3
A	519	GLU	-	CLONING ARTIFACT	UNP Q6REV3
A	520	HIS	-	EXPRESSION TAG	UNP Q6REV3
A	521	HIS	-	EXPRESSION TAG	UNP Q6REV3
A	522	HIS	-	EXPRESSION TAG	UNP Q6REV3
A	523	HIS	-	EXPRESSION TAG	UNP Q6REV3

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
A	524	HIS	-	EXPRESSION TAG	UNP Q6REV3
A	525	HIS	-	EXPRESSION TAG	UNP Q6REV3
B	0	MET	-	CLONING ARTIFACT	UNP Q6REV3
B	34	THR	ALA	SEE REMARK 999	UNP Q6REV3
B	107	THR	ALA	SEE REMARK 999	UNP Q6REV3
B	153	GLY	GLU	SEE REMARK 999	UNP Q6REV3
B	236	LYS	ARG	SEE REMARK 999	UNP Q6REV3
B	257	ALA	SER	SEE REMARK 999	UNP Q6REV3
B	291	LYS	THR	SEE REMARK 999	UNP Q6REV3
B	363	LYS	ARG	SEE REMARK 999	UNP Q6REV3
B	511	VAL	-	CLONING ARTIFACT	UNP Q6REV3
B	512	ASP	-	CLONING ARTIFACT	UNP Q6REV3
B	513	LYS	-	CLONING ARTIFACT	UNP Q6REV3
B	514	LEU	-	CLONING ARTIFACT	UNP Q6REV3
B	515	ALA	-	CLONING ARTIFACT	UNP Q6REV3
B	516	ALA	-	CLONING ARTIFACT	UNP Q6REV3
B	517	ALA	-	CLONING ARTIFACT	UNP Q6REV3
B	518	LEU	-	CLONING ARTIFACT	UNP Q6REV3
B	519	GLU	-	CLONING ARTIFACT	UNP Q6REV3
B	520	HIS	-	EXPRESSION TAG	UNP Q6REV3
B	521	HIS	-	EXPRESSION TAG	UNP Q6REV3
B	522	HIS	-	EXPRESSION TAG	UNP Q6REV3
B	523	HIS	-	EXPRESSION TAG	UNP Q6REV3
B	524	HIS	-	EXPRESSION TAG	UNP Q6REV3
B	525	HIS	-	EXPRESSION TAG	UNP Q6REV3
C	0	MET	-	CLONING ARTIFACT	UNP Q6REV3
C	34	THR	ALA	SEE REMARK 999	UNP Q6REV3
C	107	THR	ALA	SEE REMARK 999	UNP Q6REV3
C	153	GLY	GLU	SEE REMARK 999	UNP Q6REV3
C	236	LYS	ARG	SEE REMARK 999	UNP Q6REV3
C	257	ALA	SER	SEE REMARK 999	UNP Q6REV3
C	291	LYS	THR	SEE REMARK 999	UNP Q6REV3
C	363	LYS	ARG	SEE REMARK 999	UNP Q6REV3
C	511	VAL	-	CLONING ARTIFACT	UNP Q6REV3
C	512	ASP	-	CLONING ARTIFACT	UNP Q6REV3
C	513	LYS	-	CLONING ARTIFACT	UNP Q6REV3
C	514	LEU	-	CLONING ARTIFACT	UNP Q6REV3
C	515	ALA	-	CLONING ARTIFACT	UNP Q6REV3
C	516	ALA	-	CLONING ARTIFACT	UNP Q6REV3
C	517	ALA	-	CLONING ARTIFACT	UNP Q6REV3
C	518	LEU	-	CLONING ARTIFACT	UNP Q6REV3
C	519	GLU	-	CLONING ARTIFACT	UNP Q6REV3

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
C	520	HIS	-	EXPRESSION TAG	UNP Q6REV3
C	521	HIS	-	EXPRESSION TAG	UNP Q6REV3
C	522	HIS	-	EXPRESSION TAG	UNP Q6REV3
C	523	HIS	-	EXPRESSION TAG	UNP Q6REV3
C	524	HIS	-	EXPRESSION TAG	UNP Q6REV3
C	525	HIS	-	EXPRESSION TAG	UNP Q6REV3
D	0	MET	-	CLONING ARTIFACT	UNP Q6REV3
D	34	THR	ALA	SEE REMARK 999	UNP Q6REV3
D	107	THR	ALA	SEE REMARK 999	UNP Q6REV3
D	153	GLY	GLU	SEE REMARK 999	UNP Q6REV3
D	236	LYS	ARG	SEE REMARK 999	UNP Q6REV3
D	257	ALA	SER	SEE REMARK 999	UNP Q6REV3
D	291	LYS	THR	SEE REMARK 999	UNP Q6REV3
D	363	LYS	ARG	SEE REMARK 999	UNP Q6REV3
D	511	VAL	-	CLONING ARTIFACT	UNP Q6REV3
D	512	ASP	-	CLONING ARTIFACT	UNP Q6REV3
D	513	LYS	-	CLONING ARTIFACT	UNP Q6REV3
D	514	LEU	-	CLONING ARTIFACT	UNP Q6REV3
D	515	ALA	-	CLONING ARTIFACT	UNP Q6REV3
D	516	ALA	-	CLONING ARTIFACT	UNP Q6REV3
D	517	ALA	-	CLONING ARTIFACT	UNP Q6REV3
D	518	LEU	-	CLONING ARTIFACT	UNP Q6REV3
D	519	GLU	-	CLONING ARTIFACT	UNP Q6REV3
D	520	HIS	-	EXPRESSION TAG	UNP Q6REV3
D	521	HIS	-	EXPRESSION TAG	UNP Q6REV3
D	522	HIS	-	EXPRESSION TAG	UNP Q6REV3
D	523	HIS	-	EXPRESSION TAG	UNP Q6REV3
D	524	HIS	-	EXPRESSION TAG	UNP Q6REV3
D	525	HIS	-	EXPRESSION TAG	UNP Q6REV3

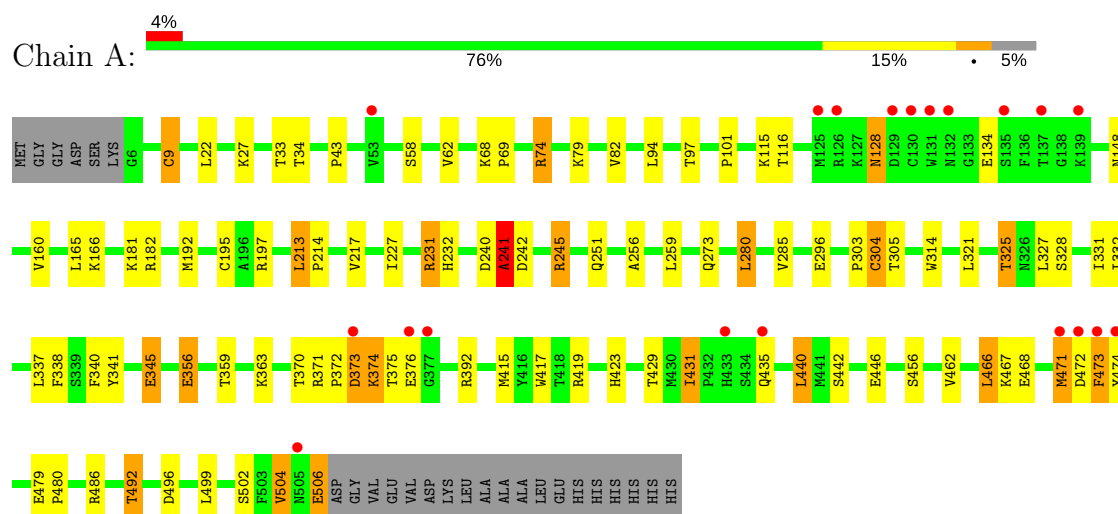
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	274	Total O 274 274	0	0
2	B	251	Total O 251 251	0	0
2	C	334	Total O 334 334	0	0
2	D	263	Total O 263 263	0	0

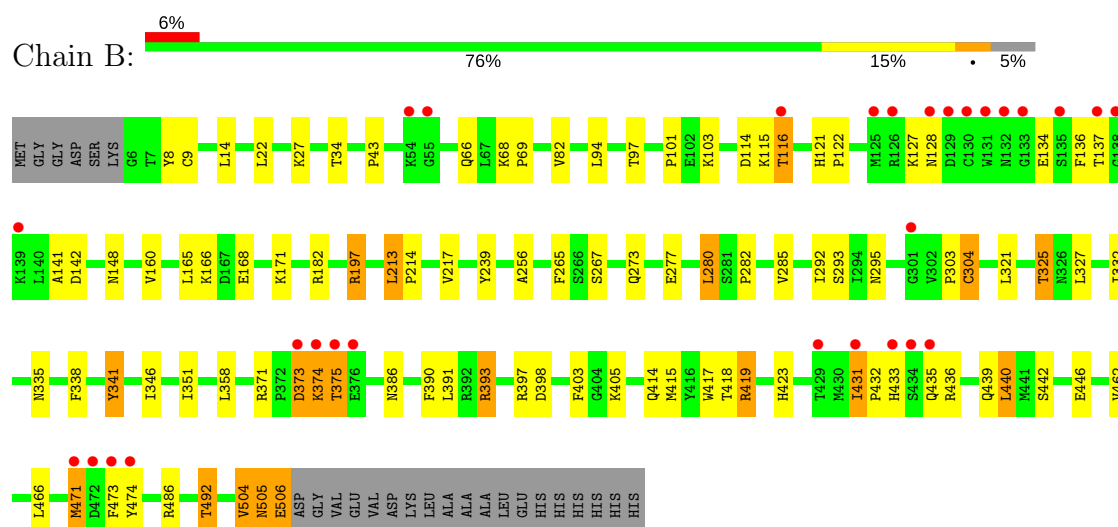
### 3 Residue-property plots

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.

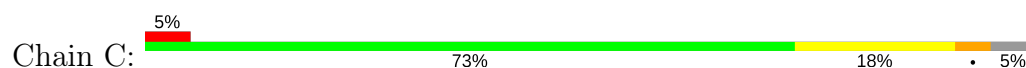
#### • Molecule 1: non-structural polyprotein

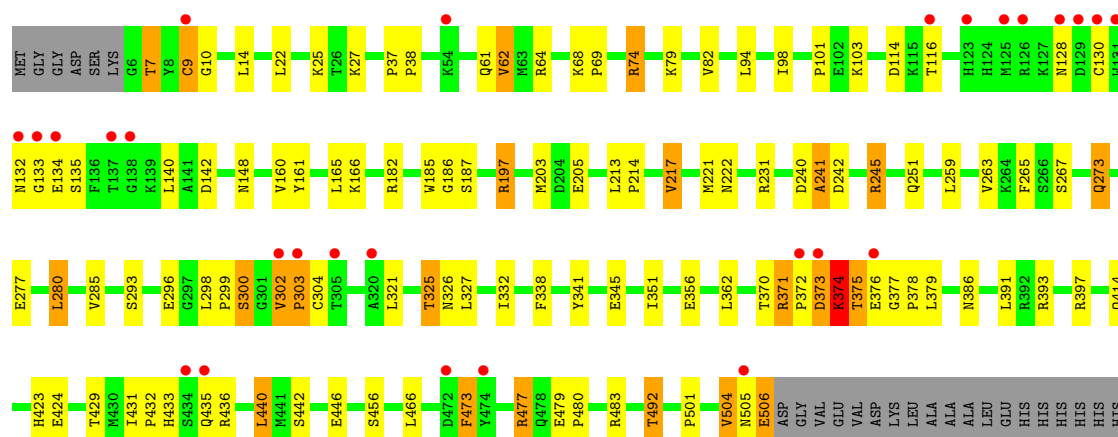


#### • Molecule 1: non-structural polyprotein

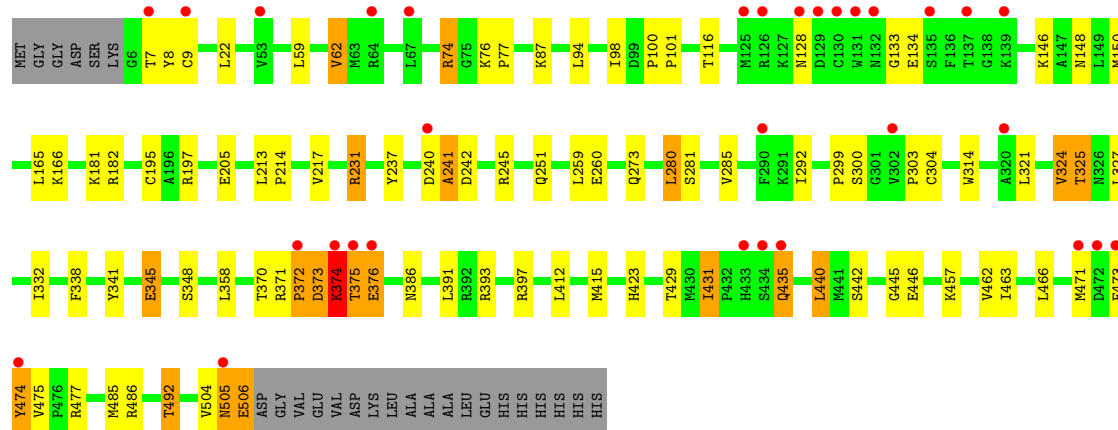
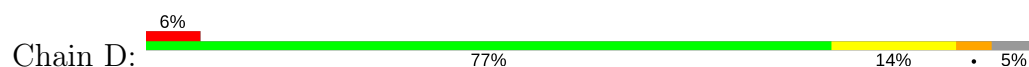


#### • Molecule 1: non-structural polyprotein





• Molecule 1: non-structural polypeptide



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.47Å 125.56Å 218.13Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.30 29.72 – 2.30	Depositor EDS
% Data completeness (in resolution range)	97.2 (20.00-2.30) 97.2 (29.72-2.30)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.09	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.72 (at 2.31Å)	Xtriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.192 , 0.257 0.195 , 0.193	Depositor DCC
$R_{free}$ test set	5401 reflections (5.28%)	DCC
Wilson B-factor (Å <sup>2</sup> )	29.2	Xtriage
Anisotropy	0.393	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 46.9	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	16838	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	31.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 40.63 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 2.6575e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<sup>1</sup> Intensities estimated from amplitudes.

<sup>2</sup> Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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](#)

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	0.60	0/4031	0.71	2/5467 (0.0%)
1	B	0.58	0/4031	0.70	4/5467 (0.1%)
1	C	0.64	0/4031	0.74	2/5467 (0.0%)
1	D	0.58	0/4031	0.70	2/5467 (0.0%)
All	All	0.60	0/16124	0.71	10/21868 (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	3
1	B	0	1
1	D	0	1
All	All	0	5

There are no bond length outliers.

The worst 5 of 10 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	393	ARG	NE-CZ-NH1	6.55	123.58	120.30
1	D	486	ARG	NE-CZ-NH2	-5.99	117.31	120.30
1	B	486	ARG	NE-CZ-NH2	-5.82	117.39	120.30
1	B	393	ARG	NE-CZ-NH2	-5.70	117.45	120.30
1	A	241	ALA	N-CA-C	5.66	126.27	111.00

There are no chirality outliers.

All (5) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	240	ASP	Peptide

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Group
1	A	345	GLU	Peptide
1	A	504	VAL	Peptide
1	B	504	VAL	Peptide
1	D	345	GLU	Peptide

## 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	3929	0	3909	77	0
1	B	3929	0	3909	78	0
1	C	3929	0	3909	95	0
1	D	3929	0	3909	69	0
2	A	274	0	0	8	0
2	B	251	0	0	5	0
2	C	334	0	0	11	0
2	D	263	0	0	11	0
All	All	16838	0	15636	313	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 10.

The worst 5 of 313 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:442:SER:HB3	1:B:504:VAL:HG12	1.30	1.10
1:B:492:THR:HG21	2:C:540:HOH:O	1.55	1.07
1:C:386:ASN:ND2	1:C:397:ARG:H	1.64	0.95
1:A:492:THR:HG21	2:A:573:HOH:O	1.66	0.93
1:D:474:TYR:HD1	1:D:474:TYR:H	1.15	0.91

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	499/526 (95%)	478 (96%)	16 (3%)	5 (1%)	18	20
1	B	499/526 (95%)	479 (96%)	17 (3%)	3 (1%)	28	34
1	C	499/526 (95%)	473 (95%)	15 (3%)	11 (2%)	8	6
1	D	499/526 (95%)	472 (95%)	18 (4%)	9 (2%)	10	9
All	All	1996/2104 (95%)	1902 (95%)	66 (3%)	28 (1%)	13	13

5 of 28 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	473	PHE
1	B	473	PHE
1	C	241	ALA
1	C	303	PRO
1	C	473	PHE

### 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	432/451 (96%)	395 (91%)	37 (9%)	12	14
1	B	432/451 (96%)	404 (94%)	28 (6%)	20	26
1	C	432/451 (96%)	402 (93%)	30 (7%)	18	23
1	D	432/451 (96%)	402 (93%)	30 (7%)	18	23
All	All	1728/1804 (96%)	1603 (93%)	125 (7%)	17	21

5 of 125 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	375	THR
1	C	62	VAL
1	D	374	LYS
1	B	423	HIS
1	B	505	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 46 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	386	ASN
1	C	143	GLN
1	D	414	GLN
1	B	414	GLN
1	B	433	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 carbohydrates 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 ⓘ

There are no chain breaks in this entry.

## 6 Fit of model and data [i](#)

### 6.1 Protein, DNA and RNA chains [i](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	501/526 (95%)	-0.00	20 (3%)	39	46	19, 30, 49, 58	0
1	B	501/526 (95%)	0.02	29 (5%)	24	30	19, 30, 49, 57	0
1	C	501/526 (95%)	0.00	27 (5%)	26	33	19, 30, 49, 58	0
1	D	501/526 (95%)	0.07	31 (6%)	21	28	19, 30, 49, 58	0
All	All	2004/2104 (95%)	0.02	107 (5%)	27	34	19, 30, 49, 58	0

The worst 5 of 107 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	472	ASP	5.6
1	B	472	ASP	5.2
1	B	135	SER	4.6
1	C	125	MET	4.4
1	D	434	SER	4.4

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

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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.