



# wwPDB X-ray Structure Validation Summary Report

Feb 27, 2014 – 04:35 PM GMT

PDB ID : 3D23  
Title : Main protease of HCoV-HKU1  
Authors : Zhao, Q.; Chen, C.; Li, S.; Zou, Y.  
Deposited on : 2008-05-07  
Resolution : 2.50 Å(reported)

This is a wwPDB 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/ValidationPDFNotes.html>

---

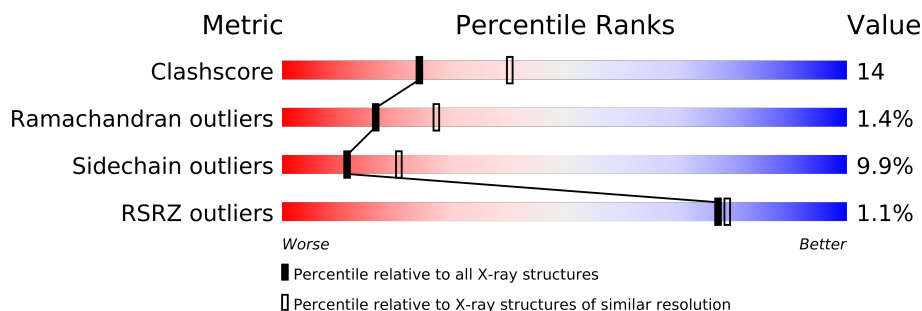
The following versions of software and data (see [references](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.15 2013  
Xtriage (Phenix) : dev-1323  
EDS : stable22639  
Percentile statistics : 21963  
Refmac : 5.8.0049  
CCP4 : 6.3.0 (Settle)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : stable22683

# 1 Overall quality at a glance

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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
Clashscore	79885	3562 (2.50-2.50)
Ramachandran outliers	78287	3480 (2.50-2.50)
Sidechain outliers	78261	3482 (2.50-2.50)
RSRZ outliers	66119	2785 (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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 2, 1 and 0 types of geometric quality criteria. The upper red bar (where present) indicates the fraction of residues that have poor fit to the electron density.

Mol	Chain	Length	Quality of chain
1	A	302	
1	B	302	
1	C	302	
1	D	302	

The following table lists non-polymeric compounds that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Geometry	Electron density
2	3IH	C	1145	-	X

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9579 atoms, of which 0 are hydrogen and 0 are deuterium.

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.

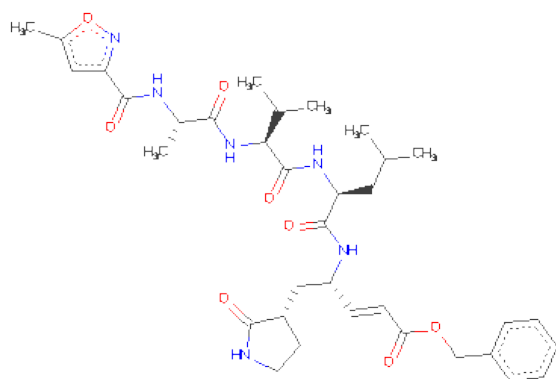
- Molecule 1 is a protein called 3C-like proteinase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	301	Total	C	N	O	S	0	1	0
			2312	1472	377	441	22			
1	A	299	Total	C	N	O	S	0	0	0
			2288	1461	372	433	22			
1	C	298	Total	C	N	O	S	0	0	0
			2285	1457	372	434	22			
1	D	299	Total	C	N	O	S	0	0	0
			2290	1460	372	436	22			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	-1	ALA	-	EXPRESSION TAG	UNP Q5MQD2
B	0	SER	-	EXPRESSION TAG	UNP Q5MQD2
A	-1	ALA	-	EXPRESSION TAG	UNP Q5MQD2
A	0	SER	-	EXPRESSION TAG	UNP Q5MQD2
C	-1	ALA	-	EXPRESSION TAG	UNP Q5MQD2
C	0	SER	-	EXPRESSION TAG	UNP Q5MQD2
D	-1	ALA	-	EXPRESSION TAG	UNP Q5MQD2
D	0	SER	-	EXPRESSION TAG	UNP Q5MQD2

- Molecule 2 is N-[(5-METHYLISOXAZOL-3-YL)CARBONYL]ALANYL-L-VALYL-N 1 -((1R,2Z)-4-(BENZYLOXY)-4-OXO-1-[(3R)-2-OXOPYRROLIDIN-3-YL]METHYL}BUT-2-ENYL)-L-LEUCINAMIDE (three-letter code: 3IH) (formula: C<sub>35</sub>H<sub>48</sub>N<sub>6</sub>O<sub>8</sub>).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	B	1	Total	C	N	O	0	0
			49	35	6	8		
2	A	1	Total	C	N	O	0	0
			49	35	6	8		
2	C	1	Total	C	N	O	0	0
			49	35	6	8		
2	D	1	Total	C	N	O	0	0
			49	35	6	8		

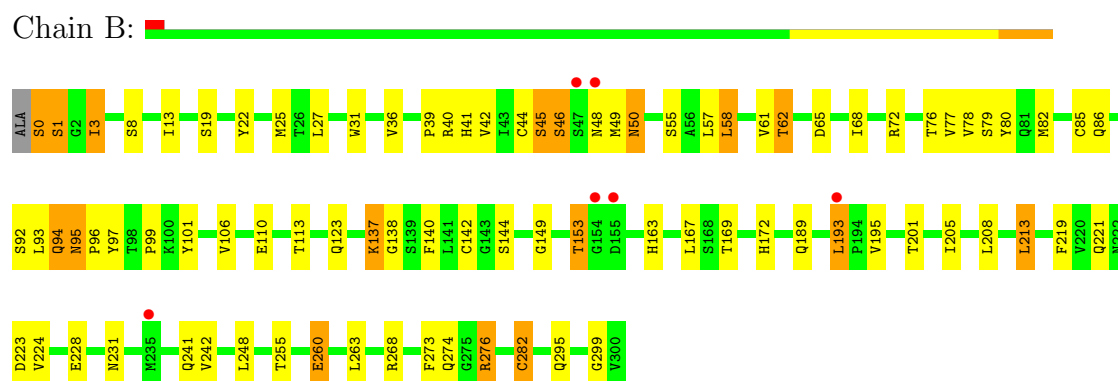
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	56	Total	O	0	0
			56	56		
3	B	56	Total	O	0	0
			56	56		
3	C	41	Total	O	0	0
			41	41		
3	D	55	Total	O	0	0
			55	55		

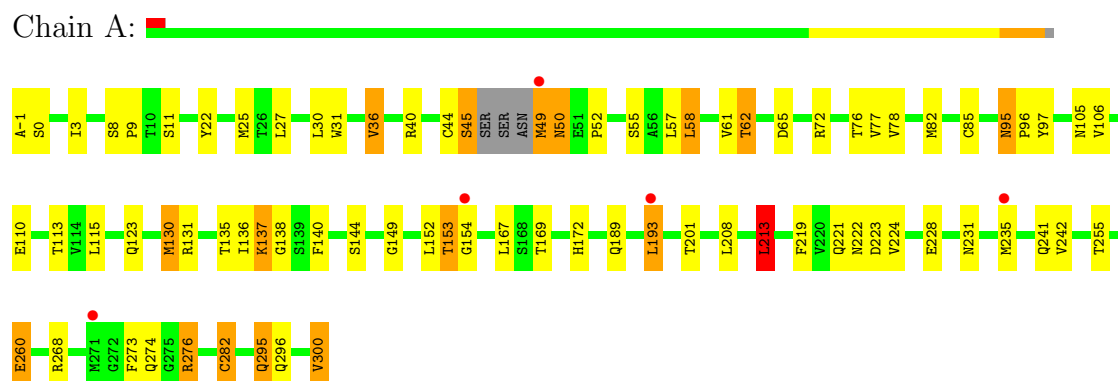
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

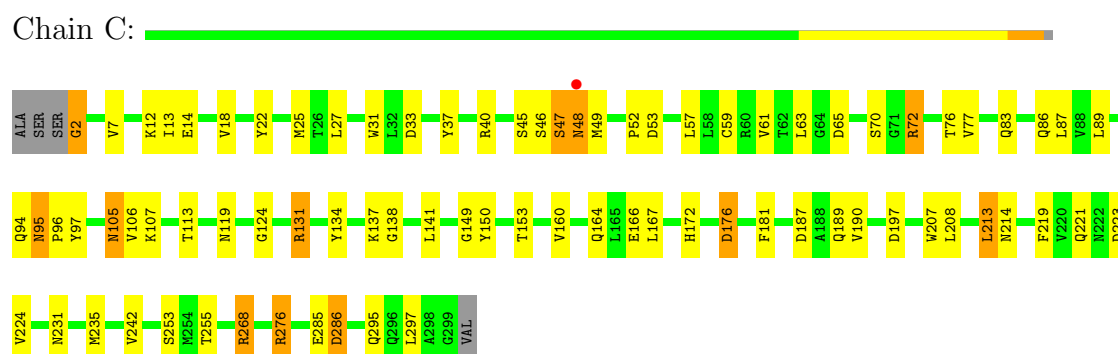
- Molecule 1: 3C-like proteinase



- Molecule 1: 3C-like proteinase

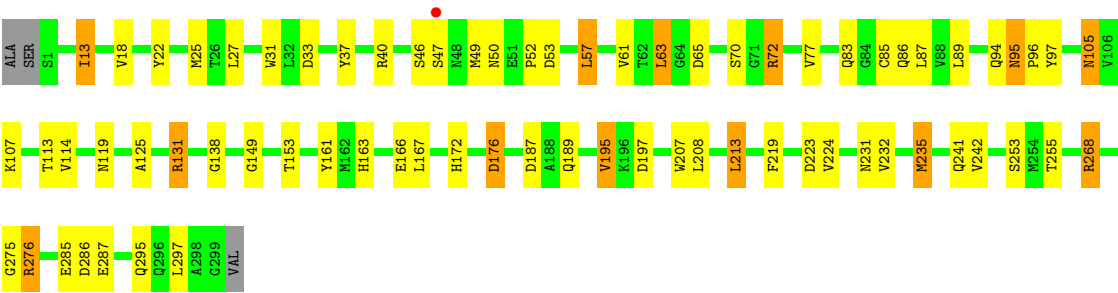


- Molecule 1: 3C-like proteinase



● Molecule 1: 3C-like proteinase

Chain D: 



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 41	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	91.77Å 91.77Å 187.91Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 2.50 41.04 – 1.97	Depositor EDS
% Data completeness (in resolution range)	98.6 (50.00-2.50) 63.2 (41.04-1.97)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.62 (at 1.97Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.229 , 0.285 0.251 , (Not available)	Depositor DCC
$R_{free}$ test set	No test flags present.	DCC
Wilson B-factor (Å <sup>2</sup> )	30.1	Xtriage
Anisotropy	0.269	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.35 , 25.9	EDS
Estimated twinning fraction	0.063 for h,-k,-l	Xtriage
L-test for twinning	$\langle  L  \rangle = 0.45$ , $\langle L^2 \rangle = 0.29$	Xtriage
Outliers	9 of 106783 reflections (0.008%)	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	9579	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	38.0	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 99.57 % 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 2.5557e-13. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

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

## 5 Model quality ⓘ

### 5.1 Standard geometry ⓘ

Bond lengths and bond angles in the following residue types are not validated in this section: 3IH

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	0.99	2/2338 (0.1%)	0.99	4/3181 (0.1%)
1	B	1.02	3/2363 (0.1%)	0.94	0/3216
1	C	1.07	6/2336 (0.3%)	1.01	7/3179 (0.2%)
1	D	1.03	0/2341	0.99	5/3185 (0.2%)
All	All	1.03	11/9378 (0.1%)	0.98	16/12761 (0.1%)

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	1	1

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	C	150	TYR	CD1-CE1	7.23	1.50	1.39
1	A	44	CYS	CB-SG	-6.63	1.71	1.82
1	C	166	GLU	CB-CG	6.38	1.64	1.52
1	B	44	CYS	CB-SG	-5.58	1.72	1.81
1	B	137	LYS	CE-NZ	5.56	1.62	1.49

The worst 5 of 16 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	274	GLN	CB-CA-C	18.50	147.41	110.40
1	A	50	ASN	N-CA-C	6.52	128.60	111.00
1	D	268	ARG	NE-CZ-NH1	6.34	123.47	120.30
1	C	164	GLN	N-CA-C	6.09	127.44	111.00

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	268	ARG	NE-CZ-NH2	-5.81	117.39	120.30

All (1) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	A	274	GLN	CA

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	49	MET	Peptide

## 5.2 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 hydrogens added by MolProbity. The Clashes column lists the number of clashes within the asymmetric unit, and the number in parentheses is this value normalized per 1000 atoms of the molecule in the chain. The Symm-Clashes column gives symmetry related clashes, in the same way as for the Clashes column.

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	A	2288	0	2249	67	0
1	B	2312	0	2269	75	0
1	C	2285	0	2242	70	0
1	D	2290	0	2246	58	0
2	A	49	0	47	5	0
2	B	49	0	47	5	0
2	C	49	0	47	10	0
2	D	49	0	47	9	0
3	A	56	0	0	3	0
3	B	56	0	0	2	0
3	C	41	0	0	5	0
3	D	55	0	0	5	0
All	All	9579	0	9194	262	0

Clashscore is defined as the number of clashes calculated for the entry per 1000 atoms (including hydrogens) of the entry. The overall clashscore for this entry is 14.

The worst 5 of 262 close contacts within the same asymmetric unit are listed below.

Atom-1	Atom-2	Distance(Å)	Clash(Å)
1:A:276:ARG:HG2	1:A:276:ARG:HH11	1.06	1.15
1:B:276:ARG:HH11	1:B:276:ARG:HG2	1.13	1.07
1:D:189:GLN:HE21	2:D:1145:3IH:HN4	1.05	0.99
1:B:0:SER:O	1:B:1:SER:HB3	1.64	0.97
1:A:130:MET:HA	1:A:130:MET:HE3	1.46	0.97

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	295/302 (98%)	280 (95%)	12 (4%)	3 (1%)	22	38
1	B	300/302 (99%)	276 (92%)	15 (5%)	9 (3%)	7	9
1	C	296/302 (98%)	272 (92%)	22 (7%)	2 (1%)	30	50
1	D	297/302 (98%)	272 (92%)	21 (7%)	4 (1%)	18	29
All	All	1188/1208 (98%)	1100 (93%)	70 (6%)	18 (2%)	16	25

5 of 18 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	46	SER
1	B	50[A]	ASN
1	B	50[B]	ASN
1	B	299	GLY
1	C	47	SER

### 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	254/258 (98%)	226 (89%)	28 (11%)	9	16
1	B	259/258 (100%)	232 (90%)	27 (10%)	10	18
1	C	255/258 (99%)	232 (91%)	23 (9%)	14	25
1	D	255/258 (99%)	232 (91%)	23 (9%)	14	25
All	All	1023/1032 (99%)	922 (90%)	101 (10%)	11	21

5 of 101 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	167	LEU
1	C	48	ASN
1	D	197	ASP
1	A	193	LEU
1	A	276	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 53 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	237	ASN
1	C	86	GLN
1	D	221	GLN
1	A	241	GLN
1	A	295	GLN

### 5.3.3 RNA ⓘ

There are no RNA chains 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 carbohydrates in this entry.

## 5.6 Ligand geometry ⓘ

4 ligands are modelled in this entry.

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$
2	3IH	A	1145	-	51,51,51	1.67	6 (11%)	69,69,69	1.67	12 (17%)
2	3IH	B	1145	-	51,51,51	1.82	5 (9%)	69,69,69	2.17	18 (26%)
2	3IH	C	1145	-	51,51,51	2.54	12 (23%)	69,69,69	3.27	22 (31%)
2	3IH	D	1145	-	51,51,51	1.89	8 (15%)	69,69,69	2.05	19 (27%)

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the chemical component dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	3IH	A	1145	-	-	1/53/64/64	0/3/3/3
2	3IH	B	1145	-	-	1/53/64/64	0/3/3/3
2	3IH	C	1145	-	-	1/53/64/64	0/3/3/3
2	3IH	D	1145	-	-	1/53/64/64	0/3/3/3

The worst 5 of 31 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	1145	3IH	C21-C20	8.63	1.55	1.32
2	C	1145	3IH	C21-C20	8.31	1.55	1.32
2	C	1145	3IH	C28-N6	8.07	1.63	1.46
2	B	1145	3IH	C21-C20	7.32	1.52	1.32
2	A	1145	3IH	C21-C20	7.20	1.52	1.32

The worst 5 of 71 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	1145	3IH	C27-C28-N6	-15.38	89.40	102.80
2	C	1145	3IH	C28-N6-C29	12.65	121.37	114.17

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	B	1145	3IH	C27-C28-N6	-10.15	93.96	102.80
2	D	1145	3IH	C3-O1-N1	-6.32	105.61	107.66
2	B	1145	3IH	C19-N5-C18	-6.15	113.44	123.36

There are no chirality outliers.

All (4) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	1145	3IH	C2-C1-C4-N2
2	C	1145	3IH	O2-C4-C1-N1
2	B	1145	3IH	C2-C1-C4-N2
2	A	1145	3IH	C2-C1-C4-N2

There are no ring outliers.

## 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<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	299/302 (99%)	-0.21	5 (1%) 67 69	22, 36, 60, 82	0
1	B	301/302 (99%)	-0.18	6 (1%) 62 64	21, 36, 61, 84	0
1	C	298/302 (98%)	-0.42	1 (0%) 91 93	19, 36, 61, 72	0
1	D	299/302 (99%)	-0.43	1 (0%) 91 93	21, 36, 62, 72	0
All	All	1197/1208 (99%)	-0.31	13 (1%) 77 79	19, 36, 61, 84	0

The worst 5 of 13 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	48	ASN	4.4
1	B	47	SER	4.1
1	A	193	LEU	3.7
1	B	193	LEU	3.1
1	A	271	MET	2.9

### 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. The B-factors column lists the minimum, median, 95<sup>th</sup> 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	RSR	LLDF	B-factors( $\text{\AA}^2$ )	Q<0.9
2	3IH	C	1145	49/49	0.20	2.37	23,36,69,70	0
2	3IH	D	1145	49/49	0.17	1.39	19,29,74,75	0
2	3IH	B	1145	49/49	0.17	0.89	30,40,79,79	0
2	3IH	A	1145	49/49	0.17	0.43	30,39,77,78	0

## 6.5 Other polymers ⓘ

There are no such residues in this entry.