



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 18, 2020 – 03:22 pm BST

PDB ID : 3O3W  
Title : Crystal Structure of BH2092 protein (residues 14-131) from *Bacillus halodurans*, Northeast Structural Genomics Consortium Target BhR228A  
Authors : Forouhar, F.; Neely, H.; Seetharaman, J.; Sahdev, S.; Xiao, R.; Ciccocanti, C.; Lee, D.; Everett, J.K.; Nair, R.; Acton, T.B.; Rost, B.; Montelione, G.T.; Tong, L.; Hunt, J.F.; Northeast Structural Genomics Consortium (NESG)  
Deposited on : 2010-07-26  
Resolution : 2.91 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

<https://www.wwpdb.org/validation/2017/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  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtriage (Phenix) : 1.13  
EDS : 2.11  
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) : 2.11

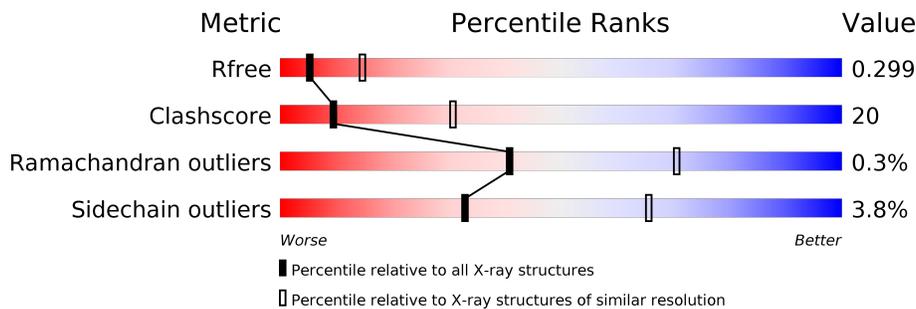
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.91 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	2307 (2.94-2.90)
Clashscore	141614	2531 (2.94-2.90)
Ramachandran outliers	138981	2462 (2.94-2.90)
Sidechain outliers	138945	2464 (2.94-2.90)

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 on the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	126	
1	B	126	
1	C	126	
1	D	126	
1	E	126	
1	F	126	
1	G	126	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	H	126	
1	I	126	

## 2 Entry composition [i](#)

There are 2 unique types of molecules in this entry. The entry contains 8353 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 BH2092 protein.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	N	O	S	Se			
1	A	118	928	591	159	174	3	1	0	0	0
1	B	117	928	591	158	175	3	1	0	0	0
1	C	118	933	594	159	176	3	1	0	0	0
1	D	114	907	578	155	170	3	1	0	0	0
1	E	118	933	594	159	176	3	1	0	0	0
1	F	118	929	591	161	173	3	1	0	0	0
1	G	114	907	578	155	170	3	1	0	0	0
1	H	116	919	586	157	172	3	1	0	0	0
1	I	113	890	567	153	166	3	1	0	0	0

There are 72 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	132	LEU	-	EXPRESSION TAG	UNP Q9KB42
A	133	GLU	-	EXPRESSION TAG	UNP Q9KB42
A	134	HIS	-	EXPRESSION TAG	UNP Q9KB42
A	135	HIS	-	EXPRESSION TAG	UNP Q9KB42
A	136	HIS	-	EXPRESSION TAG	UNP Q9KB42
A	137	HIS	-	EXPRESSION TAG	UNP Q9KB42
A	138	HIS	-	EXPRESSION TAG	UNP Q9KB42
A	139	HIS	-	EXPRESSION TAG	UNP Q9KB42
B	132	LEU	-	EXPRESSION TAG	UNP Q9KB42
B	133	GLU	-	EXPRESSION TAG	UNP Q9KB42
B	134	HIS	-	EXPRESSION TAG	UNP Q9KB42

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
B	135	HIS	-	EXPRESSION TAG	UNP Q9KB42
B	136	HIS	-	EXPRESSION TAG	UNP Q9KB42
B	137	HIS	-	EXPRESSION TAG	UNP Q9KB42
B	138	HIS	-	EXPRESSION TAG	UNP Q9KB42
B	139	HIS	-	EXPRESSION TAG	UNP Q9KB42
C	132	LEU	-	EXPRESSION TAG	UNP Q9KB42
C	133	GLU	-	EXPRESSION TAG	UNP Q9KB42
C	134	HIS	-	EXPRESSION TAG	UNP Q9KB42
C	135	HIS	-	EXPRESSION TAG	UNP Q9KB42
C	136	HIS	-	EXPRESSION TAG	UNP Q9KB42
C	137	HIS	-	EXPRESSION TAG	UNP Q9KB42
C	138	HIS	-	EXPRESSION TAG	UNP Q9KB42
C	139	HIS	-	EXPRESSION TAG	UNP Q9KB42
D	132	LEU	-	EXPRESSION TAG	UNP Q9KB42
D	133	GLU	-	EXPRESSION TAG	UNP Q9KB42
D	134	HIS	-	EXPRESSION TAG	UNP Q9KB42
D	135	HIS	-	EXPRESSION TAG	UNP Q9KB42
D	136	HIS	-	EXPRESSION TAG	UNP Q9KB42
D	137	HIS	-	EXPRESSION TAG	UNP Q9KB42
D	138	HIS	-	EXPRESSION TAG	UNP Q9KB42
D	139	HIS	-	EXPRESSION TAG	UNP Q9KB42
E	132	LEU	-	EXPRESSION TAG	UNP Q9KB42
E	133	GLU	-	EXPRESSION TAG	UNP Q9KB42
E	134	HIS	-	EXPRESSION TAG	UNP Q9KB42
E	135	HIS	-	EXPRESSION TAG	UNP Q9KB42
E	136	HIS	-	EXPRESSION TAG	UNP Q9KB42
E	137	HIS	-	EXPRESSION TAG	UNP Q9KB42
E	138	HIS	-	EXPRESSION TAG	UNP Q9KB42
E	139	HIS	-	EXPRESSION TAG	UNP Q9KB42
F	132	LEU	-	EXPRESSION TAG	UNP Q9KB42
F	133	GLU	-	EXPRESSION TAG	UNP Q9KB42
F	134	HIS	-	EXPRESSION TAG	UNP Q9KB42
F	135	HIS	-	EXPRESSION TAG	UNP Q9KB42
F	136	HIS	-	EXPRESSION TAG	UNP Q9KB42
F	137	HIS	-	EXPRESSION TAG	UNP Q9KB42
F	138	HIS	-	EXPRESSION TAG	UNP Q9KB42
F	139	HIS	-	EXPRESSION TAG	UNP Q9KB42
G	132	LEU	-	EXPRESSION TAG	UNP Q9KB42
G	133	GLU	-	EXPRESSION TAG	UNP Q9KB42
G	134	HIS	-	EXPRESSION TAG	UNP Q9KB42
G	135	HIS	-	EXPRESSION TAG	UNP Q9KB42
G	136	HIS	-	EXPRESSION TAG	UNP Q9KB42

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
G	137	HIS	-	EXPRESSION TAG	UNP Q9KB42
G	138	HIS	-	EXPRESSION TAG	UNP Q9KB42
G	139	HIS	-	EXPRESSION TAG	UNP Q9KB42
H	132	LEU	-	EXPRESSION TAG	UNP Q9KB42
H	133	GLU	-	EXPRESSION TAG	UNP Q9KB42
H	134	HIS	-	EXPRESSION TAG	UNP Q9KB42
H	135	HIS	-	EXPRESSION TAG	UNP Q9KB42
H	136	HIS	-	EXPRESSION TAG	UNP Q9KB42
H	137	HIS	-	EXPRESSION TAG	UNP Q9KB42
H	138	HIS	-	EXPRESSION TAG	UNP Q9KB42
H	139	HIS	-	EXPRESSION TAG	UNP Q9KB42
I	132	LEU	-	EXPRESSION TAG	UNP Q9KB42
I	133	GLU	-	EXPRESSION TAG	UNP Q9KB42
I	134	HIS	-	EXPRESSION TAG	UNP Q9KB42
I	135	HIS	-	EXPRESSION TAG	UNP Q9KB42
I	136	HIS	-	EXPRESSION TAG	UNP Q9KB42
I	137	HIS	-	EXPRESSION TAG	UNP Q9KB42
I	138	HIS	-	EXPRESSION TAG	UNP Q9KB42
I	139	HIS	-	EXPRESSION TAG	UNP Q9KB42

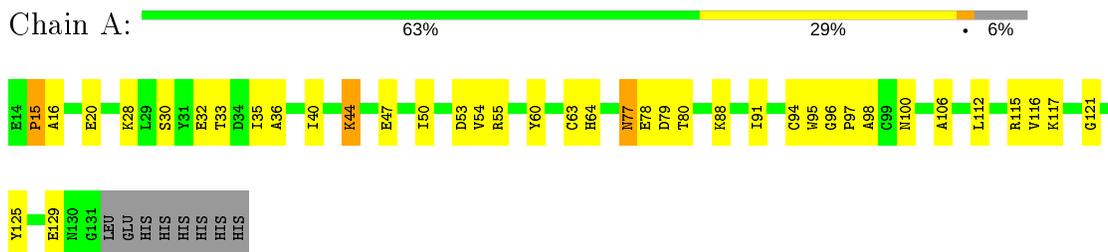
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	14	Total O 14 14	0	0
2	B	9	Total O 9 9	0	0
2	C	6	Total O 6 6	0	0
2	D	10	Total O 10 10	0	0
2	E	11	Total O 11 11	0	0
2	F	7	Total O 7 7	0	0
2	G	10	Total O 10 10	0	0
2	H	2	Total O 2 2	0	0
2	I	10	Total O 10 10	0	0

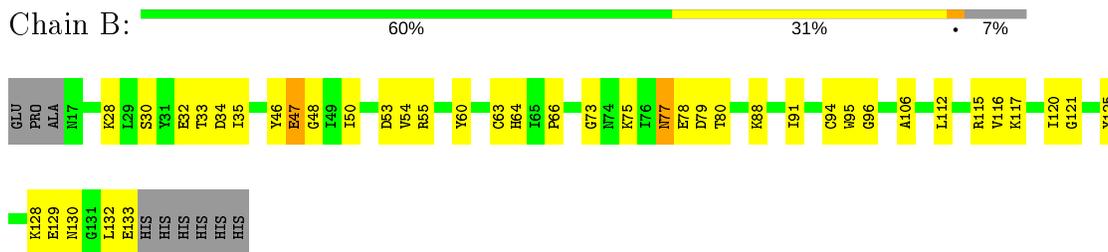
### 3 Residue-property plots [i](#)

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. 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.

- Molecule 1: BH2092 protein



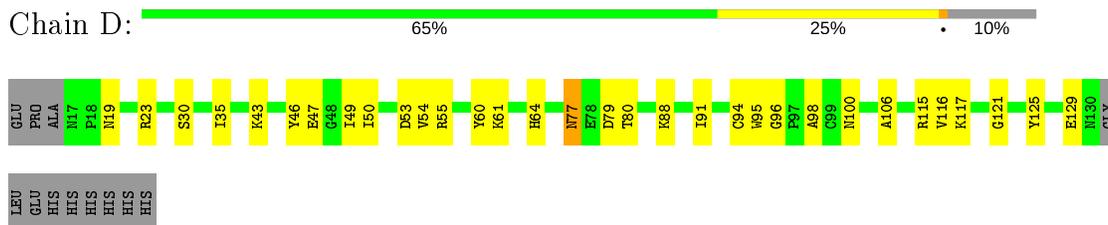
- Molecule 1: BH2092 protein



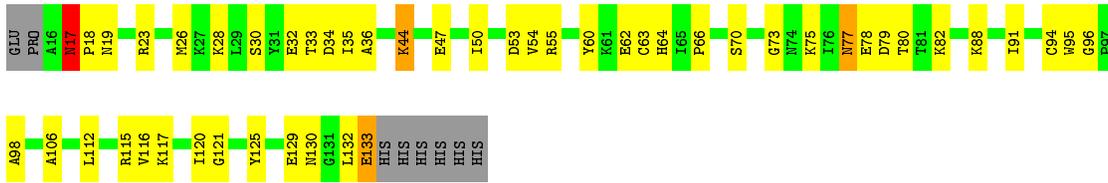
- Molecule 1: BH2092 protein



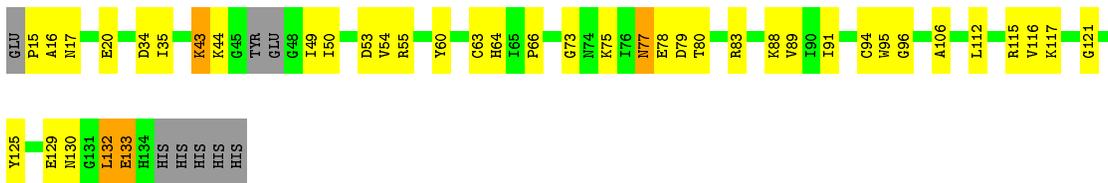
- Molecule 1: BH2092 protein



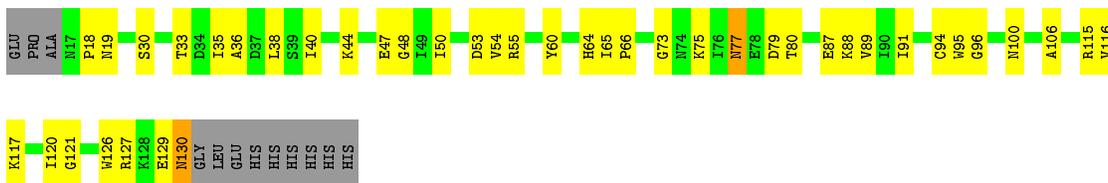
• Molecule 1: BH2092 protein



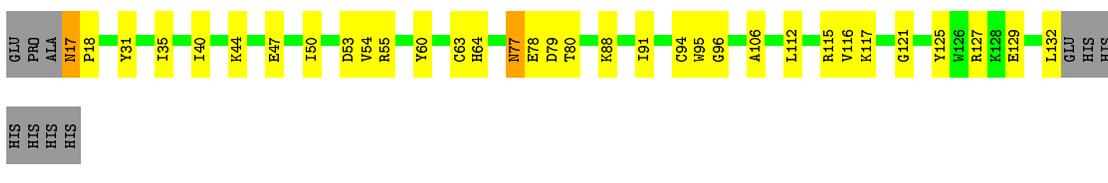
• Molecule 1: BH2092 protein



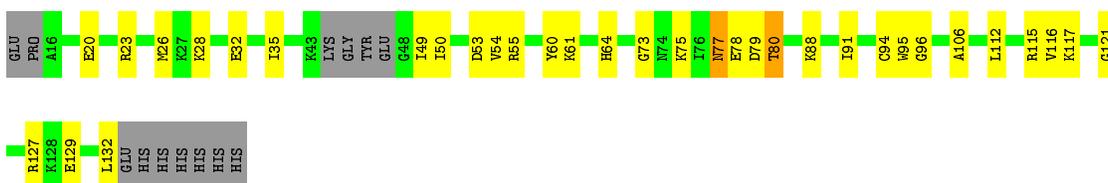
• Molecule 1: BH2092 protein



• Molecule 1: BH2092 protein



• Molecule 1: BH2092 protein



## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 31	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	114.24Å 114.24Å 85.45Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.79 – 2.91 28.56 – 2.91	Depositor EDS
% Data completeness (in resolution range)	93.9 (19.79-2.91) 99.5 (28.56-2.91)	Depositor EDS
$R_{merge}$	0.17	Depositor
$R_{sym}$	0.17	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.74 (at 2.90Å)	Xtriage
Refinement program	CNS 1.2, REFMAC	Depositor
R, $R_{free}$	0.250 , 0.292 0.261 , 0.299	Depositor DCC
$R_{free}$ test set	5410 reflections (9.88%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	33.1	Xtriage
Anisotropy	0.009	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 2.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.39$ , $\langle L^2 \rangle = 0.22$	Xtriage
Estimated twinning fraction	0.098 for -h,-k,l 0.099 for h,-h-k,-l 0.286 for -k,-h,-l	Xtriage
$F_o, F_c$ correlation	0.87	EDS
Total number of atoms	8353	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	31.0	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.*

<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	0.48	0/946	0.61	0/1275
1	B	0.45	0/945	0.58	0/1272
1	C	0.49	0/950	1.26	3/1279 (0.2%)
1	D	0.45	0/924	0.57	0/1244
1	E	0.47	0/950	0.60	0/1279
1	F	0.45	0/946	0.59	0/1272
1	G	0.46	0/924	0.55	0/1244
1	H	0.42	0/936	0.55	0/1260
1	I	0.46	0/905	0.56	0/1218
All	All	0.46	0/8426	0.69	3/11343 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	C	23	ARG	NE-CZ-NH1	-27.41	106.60	120.30
1	C	23	ARG	NE-CZ-NH2	25.63	133.12	120.30
1	C	23	ARG	CD-NE-CZ	12.99	141.78	123.60

There are no chirality outliers.

There are no planarity outliers.

### 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	928	0	927	43	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	928	0	930	35	0
1	C	933	0	935	56	1
1	D	907	0	910	41	0
1	E	933	0	935	59	0
1	F	929	0	934	47	1
1	G	907	0	910	35	0
1	H	919	0	924	28	0
1	I	890	0	897	41	0
2	A	14	0	0	1	0
2	B	9	0	0	2	0
2	C	6	0	0	1	0
2	D	10	0	0	1	0
2	E	11	0	0	4	0
2	F	7	0	0	2	0
2	G	10	0	0	4	0
2	H	2	0	0	0	0
2	I	10	0	0	6	0
All	All	8353	0	8302	324	1

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 20.

The worst 5 of 324 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:C:78:GLU:OE1	1:E:26:MSE:HE2	1.37	1.24
1:I:49:ILE:HD12	1:I:91:ILE:HD12	1.37	1.07
1:E:115:ARG:HH12	1:F:15:PRO:HB2	1.17	1.05
1:D:23:ARG:NH2	1:I:20:GLU:HB2	1.73	1.03
1:G:35:ILE:HG21	1:G:129:GLU:HG3	1.42	0.98

All (1) 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:C:22:TYR:OH	1:F:83:ARG:NH1[2_455]	2.11	0.09

## 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	116/126 (92%)	108 (93%)	7 (6%)	1 (1%)	17	46
1	B	115/126 (91%)	109 (95%)	6 (5%)	0	100	100
1	C	116/126 (92%)	108 (93%)	7 (6%)	1 (1%)	17	46
1	D	112/126 (89%)	108 (96%)	4 (4%)	0	100	100
1	E	116/126 (92%)	109 (94%)	6 (5%)	1 (1%)	17	46
1	F	114/126 (90%)	106 (93%)	8 (7%)	0	100	100
1	G	112/126 (89%)	103 (92%)	9 (8%)	0	100	100
1	H	114/126 (90%)	105 (92%)	9 (8%)	0	100	100
1	I	109/126 (86%)	102 (94%)	7 (6%)	0	100	100
All	All	1024/1134 (90%)	958 (94%)	63 (6%)	3 (0%)	41	70

All (3) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	17	ASN
1	A	15	PRO
1	C	132	LEU

### 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	96/104 (92%)	92 (96%)	4 (4%)	30	62

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	97/104 (93%)	94 (97%)	3 (3%)	40	72
1	C	97/104 (93%)	93 (96%)	4 (4%)	30	63
1	D	95/104 (91%)	93 (98%)	2 (2%)	53	80
1	E	97/104 (93%)	91 (94%)	6 (6%)	18	45
1	F	97/104 (93%)	91 (94%)	6 (6%)	18	45
1	G	95/104 (91%)	92 (97%)	3 (3%)	39	71
1	H	96/104 (92%)	93 (97%)	3 (3%)	40	72
1	I	93/104 (89%)	91 (98%)	2 (2%)	52	80
All	All	863/936 (92%)	830 (96%)	33 (4%)	33	65

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	44	LYS
1	E	133	GLU
1	H	80	THR
1	E	77	ASN
1	E	80	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 36 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	130	ASN
1	E	111	GLN
1	I	77	ASN
1	E	77	ASN
1	F	77	ASN

### 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 [i](#)

There are no chain breaks in this entry.

## 6 Fit of model and data

### 6.1 Protein, DNA and RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.2 Non-standard residues in protein, DNA, RNA chains

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.3 Carbohydrates

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.4 Ligands

Unable to reproduce the depositors R factor - this section is therefore empty.

### 6.5 Other polymers

Unable to reproduce the depositors R factor - this section is therefore empty.