



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 31, 2021 – 09:36 PM EDT

PDB ID : 1UIJ  
Title : Crystal Structure Of Soybean beta-Conglycinin Beta Homotrimer (I122M/K124W)  
Authors : Maruyama, N.; Maruyama, Y.; Tsuruki, T.; Okuda, E.; Yoshikawa, M.; Utsumi, S.  
Deposited on : 2003-07-16  
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](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  
Xtriage (Phenix) : 1.13  
EDS : 2.23.2  
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) : 2.23.2

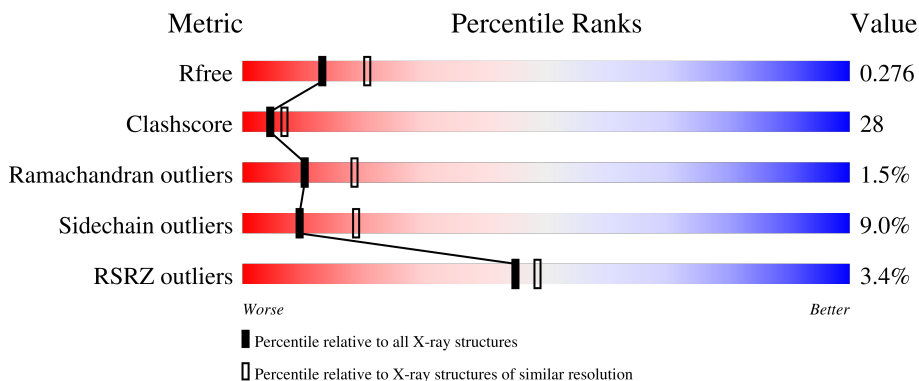
# 1 Overall quality at a glance

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.



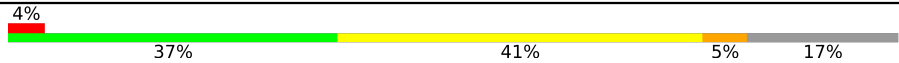
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	416	 2% 52% 33% 5% 9%
1	B	416	 2% 47% 36% 5% 11%
1	C	416	 5% 40% 38% 5% 17%
1	D	416	 2% 49% 36% 6% 9%
1	E	416	 3% 46% 39% . 11%

Continued on next page...

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
1	F	416	 <p>A horizontal bar chart representing the quality of chain. The bar is divided into five segments: a small red segment (4%), a large green segment (37%), a large yellow segment (41%), a small orange segment (5%), and a medium grey segment (17%).</p>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 18133 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 beta subunit of beta conglycinin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	377	3087	1947	550	589	1	0	0	0
1	B	370	3022	1911	538	572	1	0	0	0
1	C	346	2824	1797	494	532	1	0	0	0
1	D	377	3087	1947	550	589	1	0	0	0
1	E	370	3022	1911	538	572	1	0	0	0
1	F	346	2824	1797	494	532	1	0	0	0

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	122	MET	ILE	engineered mutation	UNP P25974
A	124	TRP	LYS	engineered mutation	UNP P25974
B	122	MET	ILE	engineered mutation	UNP P25974
B	124	TRP	LYS	engineered mutation	UNP P25974
C	122	MET	ILE	engineered mutation	UNP P25974
C	124	TRP	LYS	engineered mutation	UNP P25974
D	122	MET	ILE	engineered mutation	UNP P25974
D	124	TRP	LYS	engineered mutation	UNP P25974
E	122	MET	ILE	engineered mutation	UNP P25974
E	124	TRP	LYS	engineered mutation	UNP P25974
F	122	MET	ILE	engineered mutation	UNP P25974
F	124	TRP	LYS	engineered mutation	UNP P25974

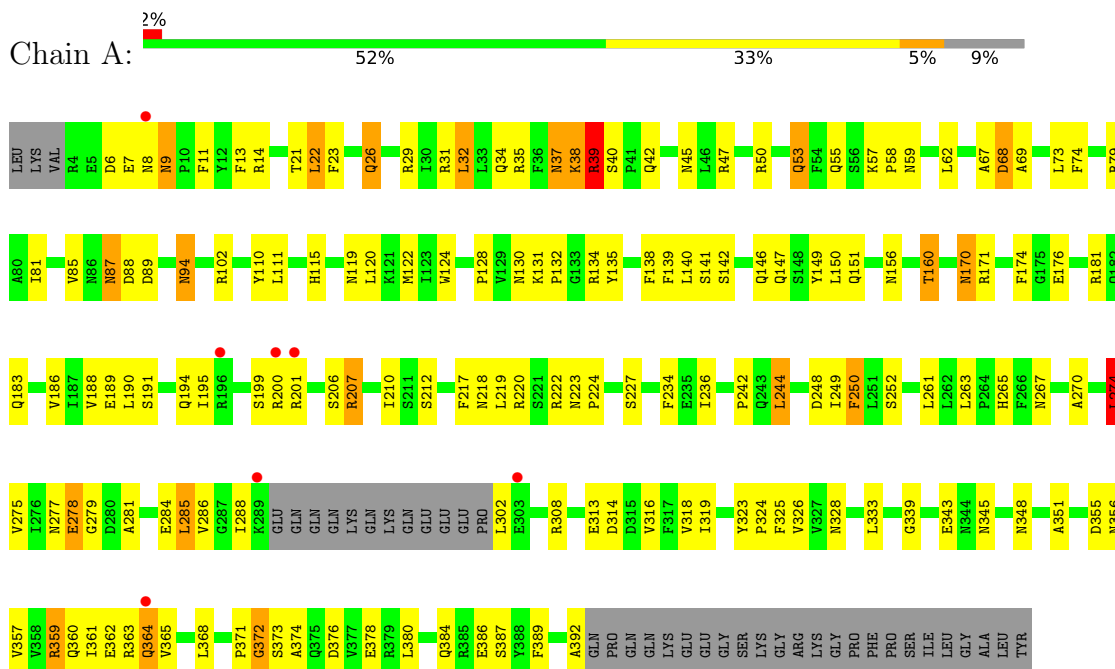
- Molecule 2 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	62	Total O 62 62	0	0
2	B	53	Total O 53 53	0	0
2	C	19	Total O 19 19	0	0
2	D	65	Total O 65 65	0	0
2	E	47	Total O 47 47	0	0
2	F	21	Total O 21 21	0	0

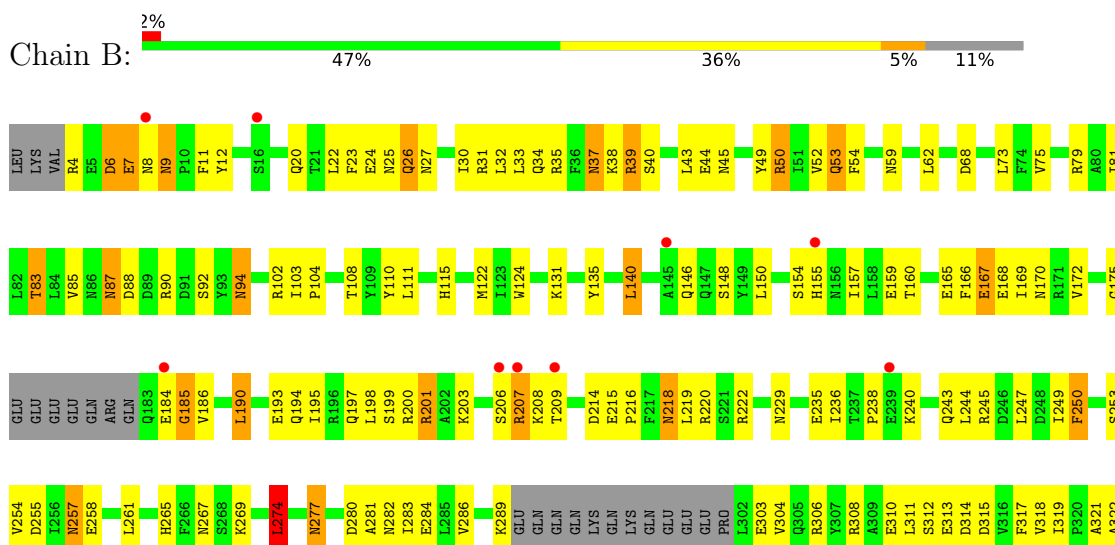
### 3 Residue-property plots [i](#)

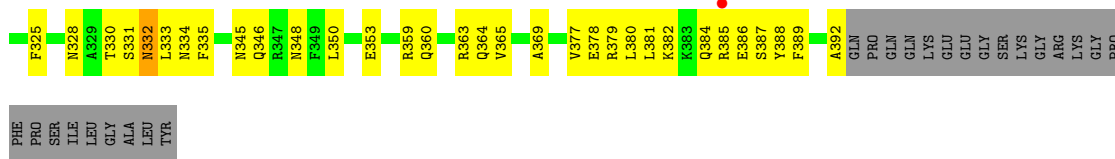
These plots are drawn for all protein, RNA, DNA and oligosaccharide chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. 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: beta subunit of beta conglycinin

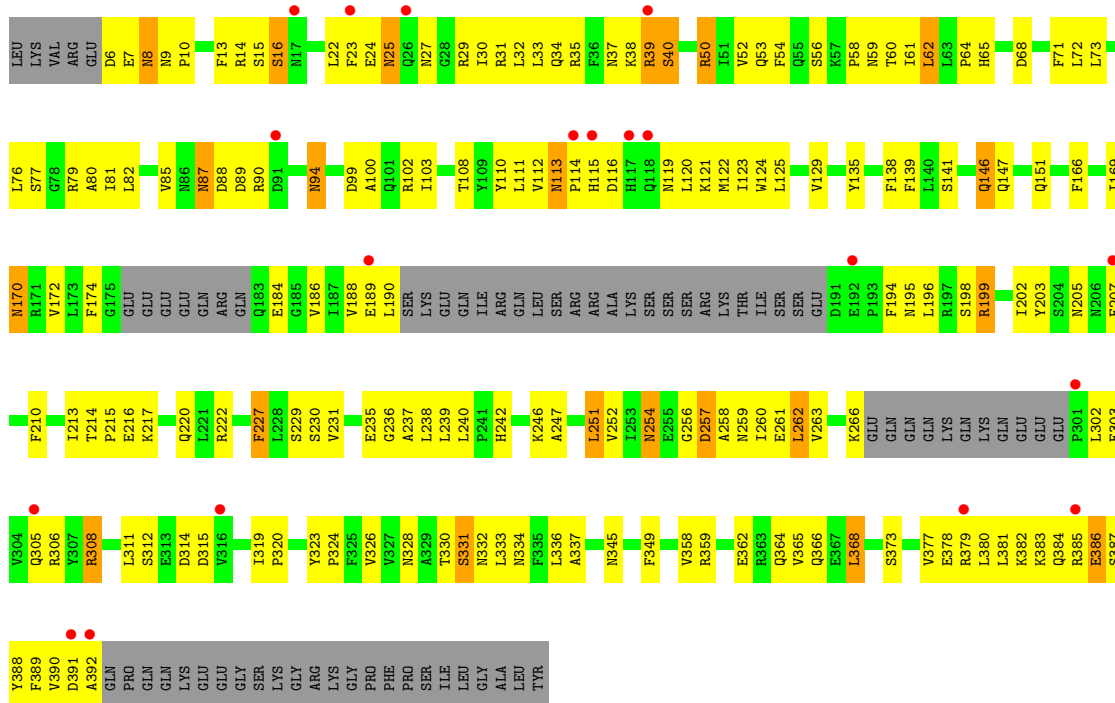


- Molecule 1: beta subunit of beta conglycinin

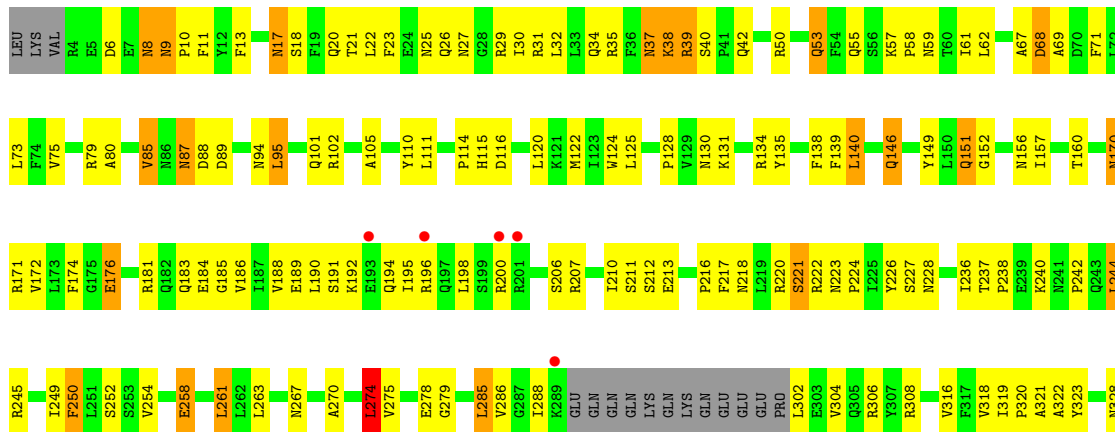




• Molecule 1: beta subunit of beta conglycinin



• Molecule 1: beta subunit of beta conglycinin







V365	Q366	E367	L368	A369	F370	P371	D376	L380	L381	K382	K385	Q384	R385	E386	S387	Y388	F389	V390	D391	A392	GLN	PRO	GLN	GLN	GLN	LYS	GLU	GLU	GLU	GLY	SER	LYS	LYS	GLY	ARG	ARG	LYS	LYS	GLY	PRO	PHE	PRO	SER	SER	ILE	ILE	LEU	GLY	ALA	LEU	TYR
------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	------	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----	-----

## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	130.26Å 62.58Å 159.01Å 90.00° 90.44° 90.00°	Depositor
Resolution (Å)	8.00 – 2.50 20.00 – 2.50	Depositor EDS
% Data completeness (in resolution range)	(Not available) (8.00-2.50) 93.8 (20.00-2.50)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.85 (at 2.50Å)	Xtrriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.221 , 0.273 0.224 , 0.276	Depositor DCC
$R_{free}$ test set	4277 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	39.4	Xtrriage
Anisotropy	0.638	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 61.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.018 for h,-k,-l	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	18133	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	43.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 43.83 % 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 1.6659e-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	0.39	0/3150	0.68	1/4260 (0.0%)
1	B	0.40	0/3084	0.67	1/4171 (0.0%)
1	C	0.34	0/2886	0.61	0/3910
1	D	0.39	0/3150	0.68	1/4260 (0.0%)
1	E	0.39	0/3084	0.65	0/4171
1	F	0.35	0/2886	0.62	1/3910 (0.0%)
All	All	0.38	0/18240	0.66	4/24682 (0.0%)

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	274	LEU	CA-CB-CG	6.46	130.16	115.30
1	F	251	LEU	CA-CB-CG	5.72	128.46	115.30
1	D	274	LEU	CA-CB-CG	5.45	127.84	115.30
1	A	274	LEU	CA-CB-CG	5.34	127.59	115.30

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	3087	0	3004	150	0
1	B	3022	0	2950	157	0
1	C	2824	0	2739	202	0

*Continued on next page...*

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	D	3087	0	3004	163	0
1	E	3022	0	2950	176	0
1	F	2824	0	2739	210	0
2	A	62	0	0	3	0
2	B	53	0	0	3	0
2	C	19	0	0	0	0
2	D	65	0	0	4	0
2	E	47	0	0	5	0
2	F	21	0	0	2	0
All	All	18133	0	17386	1002	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 28.

The worst 5 of 1002 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:59:ASN:HD22	1:B:195:ILE:HD12	1.03	1.07
1:E:90:ARG:HH11	1:F:362:GLU:HG2	1.19	1.05
1:E:59:ASN:HD22	1:E:195:ILE:HD12	1.20	1.05
1:A:26:GLN:H	1:A:26:GLN:HE21	1.04	1.03
1:A:183:GLN:HG2	1:A:188:VAL:HG11	1.42	1.01

There are no symmetry-related clashes.

## 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	373/416 (90%)	348 (93%)	21 (6%)	4 (1%)	14 26
1	B	364/416 (88%)	343 (94%)	14 (4%)	7 (2%)	8 13

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	338/416 (81%)	306 (90%)	25 (7%)	7 (2%)	7	11
1	D	373/416 (90%)	342 (92%)	27 (7%)	4 (1%)	14	26
1	E	364/416 (88%)	340 (93%)	20 (6%)	4 (1%)	14	26
1	F	338/416 (81%)	295 (87%)	37 (11%)	6 (2%)	8	14
All	All	2150/2496 (86%)	1974 (92%)	144 (7%)	32 (2%)	10	18

5 of 32 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	6	ASP
1	C	7	GLU
1	D	6	ASP
1	E	6	ASP
1	E	201	ARG

### 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	341/375 (91%)	309 (91%)	32 (9%)	8	17
1	B	334/375 (89%)	301 (90%)	33 (10%)	8	15
1	C	311/375 (83%)	287 (92%)	24 (8%)	13	25
1	D	341/375 (91%)	304 (89%)	37 (11%)	6	12
1	E	334/375 (89%)	307 (92%)	27 (8%)	11	23
1	F	311/375 (83%)	287 (92%)	24 (8%)	13	25
All	All	1972/2250 (88%)	1795 (91%)	177 (9%)	9	19

5 of 177 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	254	VAL
1	E	244	LEU

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type
1	D	275	VAL
1	E	22	LEU
1	E	328	ASN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 175 such sidechains are listed below:

Mol	Chain	Res	Type
1	D	345	ASN
1	E	334	ASN
1	D	384	GLN
1	E	94	ASN
1	F	9	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 monosaccharides 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 [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	377/416 (90%)	-0.16	7 (1%) 66 69	18, 34, 67, 82	0
1	B	370/416 (88%)	-0.02	10 (2%) 54 58	14, 36, 62, 76	0
1	C	346/416 (83%)	0.39	19 (5%) 25 26	25, 54, 74, 85	0
1	D	377/416 (90%)	-0.14	7 (1%) 66 69	17, 32, 67, 80	0
1	E	370/416 (88%)	0.04	14 (3%) 40 43	15, 39, 69, 78	0
1	F	346/416 (83%)	0.36	18 (5%) 27 29	28, 53, 75, 84	0
All	All	2186/2496 (87%)	0.07	75 (3%) 45 48	14, 40, 70, 85	0

The worst 5 of 75 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	117	HIS	5.3
1	C	26	GLN	4.5
1	F	301	PRO	4.4
1	E	207	ARG	4.3
1	D	200	ARG	4.1

### 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 monosaccharides 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.