



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 15, 2017 – 03:20 am GMT

PDB ID : 2J04  
Title : THE TAU60-TAU91 SUBCOMPLEX OF YEAST TRANSCRIPTION FACTOR IIIC  
Authors : Mylona, A.; Fernandez-Tornero, C.; Legrand, P.; Muller, C.W.  
Deposited on : 2006-07-31  
Resolution : 3.20 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

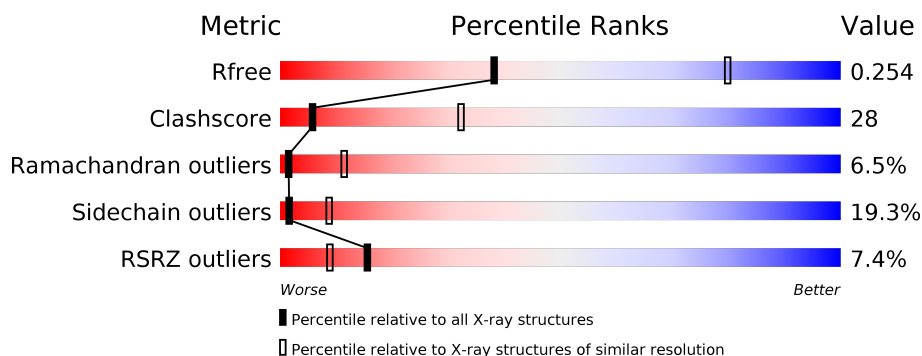
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.20 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	1015 (3.22-3.18)
Clashscore	112137	1009 (3.20-3.20)
Ramachandran outliers	110173	1118 (3.22-3.18)
Sidechain outliers	110143	1117 (3.22-3.18)
RSRZ outliers	101464	1020 (3.22-3.18)

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	588	<div> <div>7%</div> <div> <div></div> <div>52%</div> <div>34%</div> <div>11%</div> <div>.</div> </div> </div>
1	C	588	<div> <div>7%</div> <div> <div></div> <div>48%</div> <div>36%</div> <div>13%</div> <div>.</div> </div> </div>
2	B	524	<div> <div>7%</div> <div> <div></div> <div>43%</div> <div>32%</div> <div>13%</div> <div>.</div> <div>10%</div> </div> </div>
2	D	524	<div> <div>6%</div> <div> <div></div> <div>43%</div> <div>33%</div> <div>12%</div> <div>.</div> <div>10%</div> </div> </div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 16935 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 HYPOTHETICAL PROTEIN YPL007C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	587	Total	C	N	O	S	0	0	0
			4760	3058	777	902	23			
1	C	586	Total	C	N	O	S	0	0	0
			4751	3052	775	901	23			

- Molecule 2 is a protein called YDR362CP.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	472	Total	C	N	O	S	0	0	1
			3715	2384	613	703	15			
2	D	469	Total	C	N	O	S	0	0	1
			3704	2379	609	701	15			

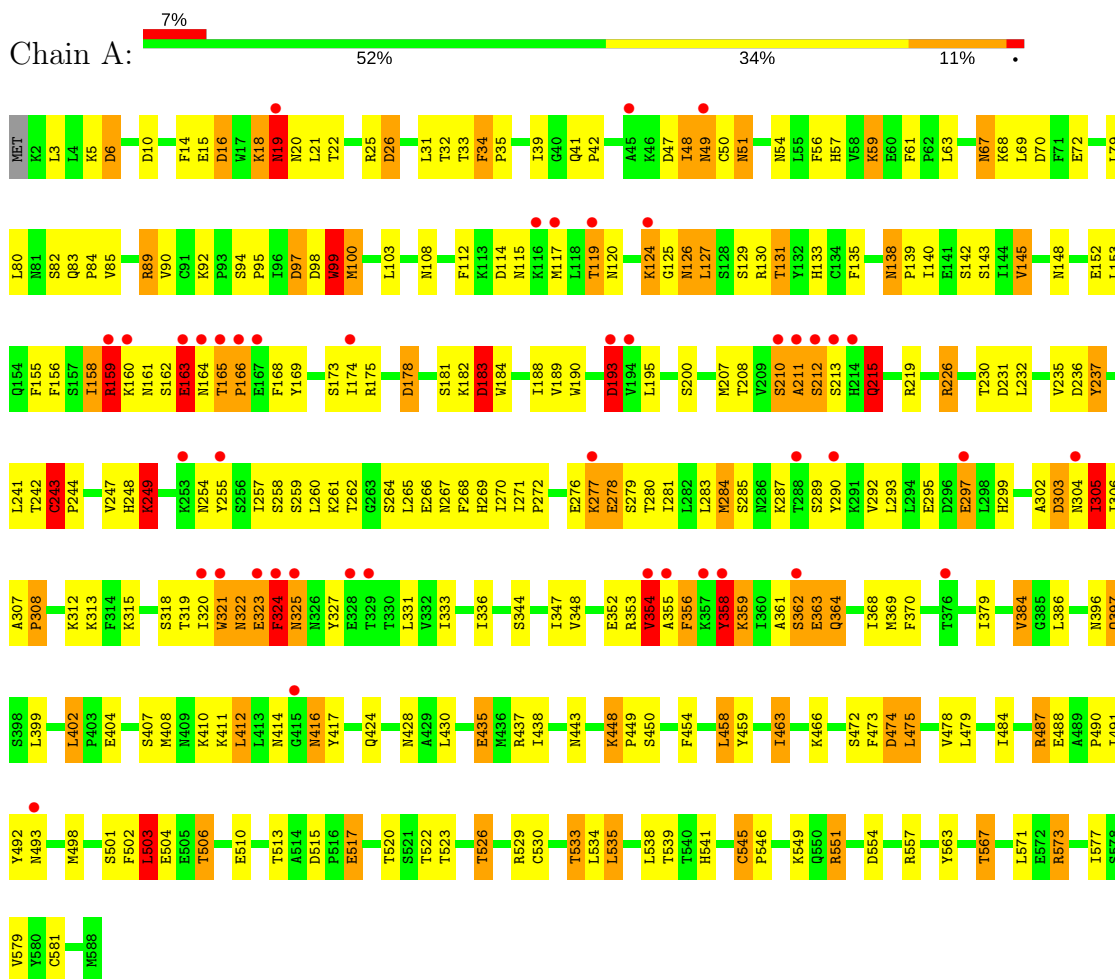
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total	O	0	0
			2	2		
3	B	1	Total	O	0	0
			1	1		
3	C	1	Total	O	0	0
			1	1		
3	D	1	Total	O	0	0
			1	1		

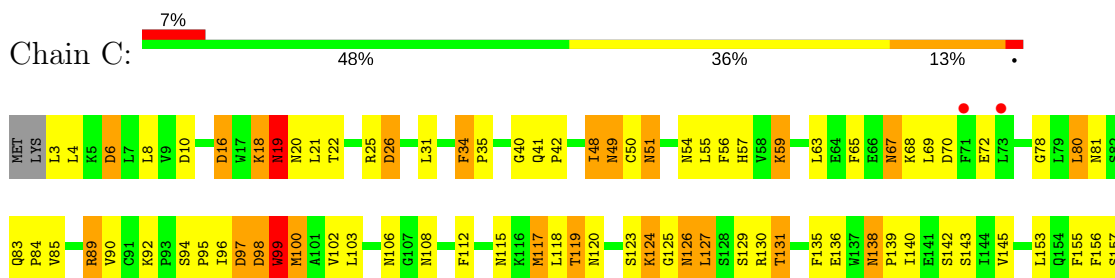
### 3 Residue-property plots

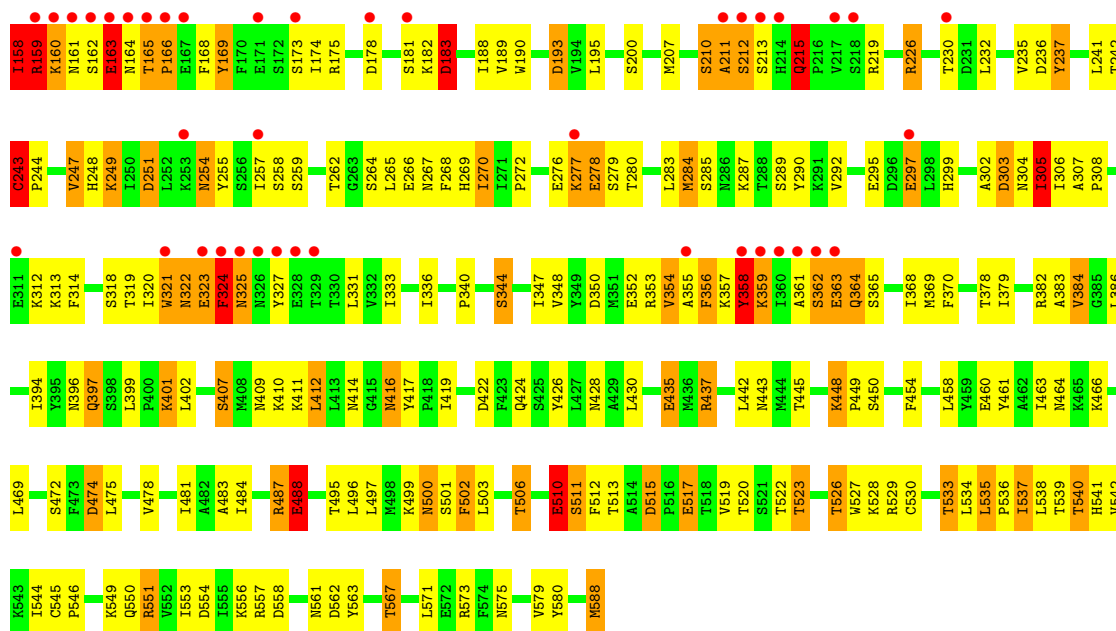
These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry and electron density. Residues are 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: HYPOTHETICAL PROTEIN YPL007C

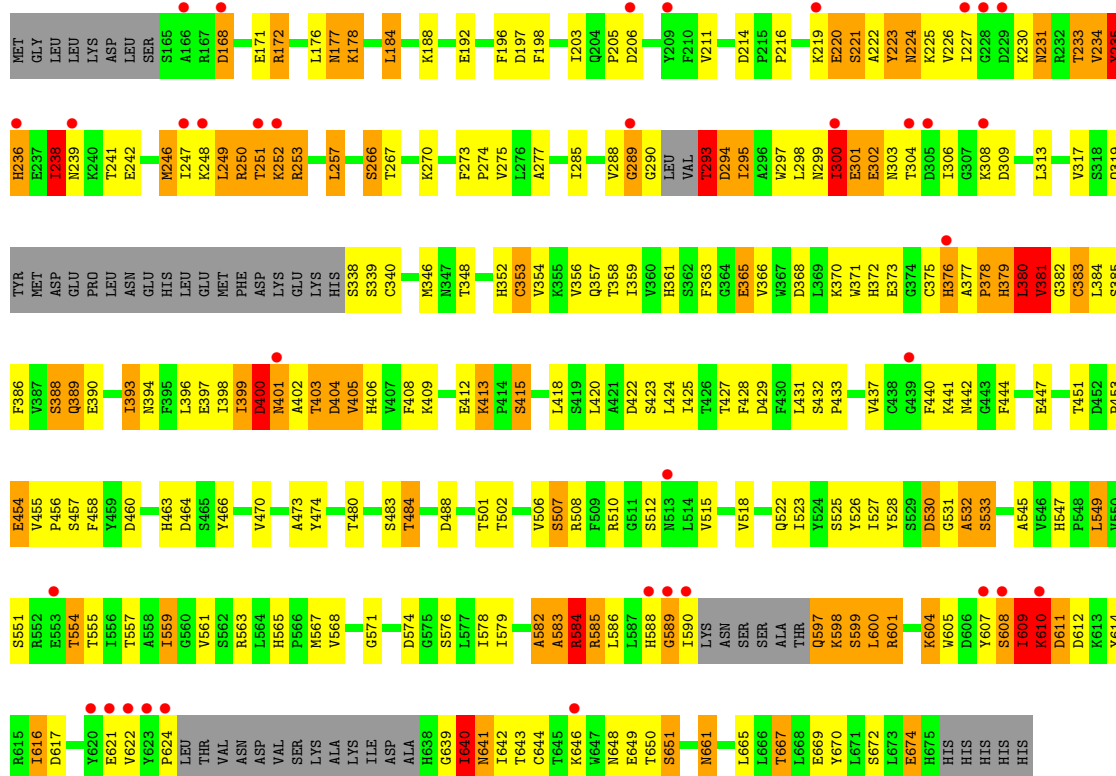


#### • Molecule 1: HYPOTHETICAL PROTEIN YPL007C





### • Molecule 2: YDR362CP



### • Molecule 2: YDR362CP



P624	T554	P453	L380	L313	I238	MET
THR	T555	E454	G381	V317	GLY	
THR	I556	V455	C382	K240	N239	LEU
VAL	T557	P456	C383	K241	T241	LEU
ASN	A558	S457	L384	Q318	E242	LYS
ASP	I559	P458	S385	Y320	MET	ASP
VAL	R563	H463	F386	ASP		N245
SER		D464	V387	ASP	M246	SER
LYS	L564	D464	S388	GLU	I247	SER
ALA	H565	Q389	Q389	PRO	K248	ALA
LYS	P566	S469	E390	ASN	L249	R167
ILE	M567	V470	I393	GLU	R250	
ASP	W568	A473		N394	GLU	T251
A637	L569	V474	F395	HIS	K252	K169
H638	D574	S475	L396	LEU	R253	I170
G639		D476	E397	GLU	K253	E171
T640	I578	D479	I398	MET	L257	R172
N641		I579	I399	PHE	I258	LYS
I642	T580	T480	ASP	LYS	L258	L176
T643	T580	T480	D400	LYS	D262	N177
C644	N581	A401	N401	GLU	K178	ASP
T645	A582	S483	A402	LYS	S266	L184
K646	A583	T484	T403	HIS	T267	K188
W647	R584	D488	D404	S338	K270	
N648	R585		V405	S339	ASP	I271
E649	L586	H406	C340	LYS	E272	E192
T650	L587	T501	V407	K345	F273	F196
S651	H588	V506	F408	M346	P274	D197
N661	G589	S507	K409	K347	V275	F198
	ILE	R508	E412	T348	L276	T203
S662	LYS	F509	K413	H352	A277	
L665	ASN	R510	P414	C353	I285	P205
T667	SER	G511	S415	V354	V288	D206
L668	ALA	S512	L418	K355	GLY	V211
S672	THR	N513	S419	V356	LEU	
	GLN	L514	L420	Q357	VAL	K219
L673	K598	V515	S423	I359	D294	E220
E674	S599	C520	L424	V360	T295	S221
H675	L600	P521	I425	H361	A296	A222
HIS	R601	Q522	T426	D426	W297	Y223
HIS	K604	I523	Y524	F363	L298	N224
HIS	D606	H525	F428	G364	N299	K225
HIS	Y607	S526	D429	E365	V266	V226
HIS	S608	I527	F430	V366	I300	I227
	I609	L431	L431	W367	E301	G228
HIS	K610	D530	S432	D368	N303	K230
HIS	D611	G531	P433	I369	E302	N231
HIS	D612	A532	V437	K370	N304	R232
HIS	K613	S533	Y437	W371	T304	T233
HIS	Y614	F440		H372	D305	V234
	K615	A545	F440	H372	I306	H236
HIS	I616	V546	K441	E373	G307	E237
	V620	H547	E447	G374	K308	F233
HIS	E621	P548	L549	H376	D309	V234
HIS	V622	V549	T451	A377	I310	Y235
HIS	V623		D451	H378	Q311	H236

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	61.42Å 125.80Å 210.47Å 90.00° 94.49° 90.00°	Depositor
Resolution (Å)	208.51 – 3.20 34.91 – 3.20	Depositor EDS
% Data completeness (in resolution range)	98.7 (208.51-3.20) 98.8 (34.91-3.20)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.06 (at 3.18Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.212 , 0.257 0.209 , 0.254	Depositor DCC
$R_{free}$ test set	2619 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	77.5	Xtriage
Anisotropy	0.356	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 130.4	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.93	EDS
Total number of atoms	16935	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	70.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.63% 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

### 5.1 Standard geometry

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	0.83	5/4872 (0.1%)	0.94	15/6610 (0.2%)
1	C	0.80	5/4863 (0.1%)	0.95	16/6599 (0.2%)
2	B	0.76	4/3798 (0.1%)	0.98	21/5153 (0.4%)
2	D	0.77	4/3788 (0.1%)	0.95	15/5140 (0.3%)
All	All	0.79	18/17321 (0.1%)	0.95	67/23502 (0.3%)

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	2
1	C	0	3
2	B	0	8
2	D	0	7
All	All	0	20

The worst 5 of 18 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	413	LYS	CD-CE	18.55	1.97	1.51
2	B	413	LYS	CD-CE	14.62	1.87	1.51
1	A	249	LYS	CD-CE	13.65	1.85	1.51
1	A	297	GLU	CD-OE1	11.57	1.38	1.25
1	C	249	LYS	CD-CE	11.35	1.79	1.51

The worst 5 of 67 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	413	LYS	CD-CE-NZ	-8.44	92.29	111.70
2	B	413	LYS	CD-CE-NZ	-8.38	92.42	111.70
2	D	413	LYS	CG-CD-CE	-7.55	89.24	111.90

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	6	ASP	CB-CG-OD2	7.20	124.78	118.30
1	A	231	ASP	CB-CG-OD2	7.00	124.60	118.30

There are no chirality outliers.

5 of 20 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	243	CYS	Peptide
1	A	98	ASP	Peptide
2	B	277	ALA	Peptide
2	B	293	THR	Peptide
2	B	379	HIS	Peptide

## 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	4760	0	4722	220	0
1	C	4751	0	4709	244	0
2	B	3715	0	3674	258	0
2	D	3704	0	3660	238	0
3	A	2	0	0	0	0
3	B	1	0	0	0	0
3	C	1	0	0	0	0
3	D	1	0	0	0	0
All	All	16935	0	16765	942	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 942 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:249:LYS:CD	1:C:249:LYS:CE	1.79	1.60
2:B:413:LYS:CD	2:B:413:LYS:CE	1.87	1.52

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:249:LYS:CE	1:A:249:LYS:CD	1.85	1.50
2:B:413:LYS:NZ	2:B:413:LYS:CE	1.74	1.47
2:D:413:LYS:CE	2:D:413:LYS:NZ	1.76	1.45

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	585/588 (100%)	475 (81%)	74 (13%)	36 (6%)	2	13
1	C	584/588 (99%)	475 (81%)	69 (12%)	40 (7%)	1	10
2	B	462/524 (88%)	362 (78%)	68 (15%)	32 (7%)	1	9
2	D	459/524 (88%)	356 (78%)	76 (17%)	27 (6%)	2	15
All	All	2090/2224 (94%)	1668 (80%)	287 (14%)	135 (6%)	1	11

5 of 135 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	80	LEU
1	A	124	LYS
1	A	159	ARG
1	A	164	ASN
1	A	166	PRO

### 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	543/544 (100%)	434 (80%)	109 (20%)	1	7
1	C	542/544 (100%)	434 (80%)	108 (20%)	1	7
2	B	415/465 (89%)	340 (82%)	75 (18%)	2	10
2	D	414/465 (89%)	336 (81%)	78 (19%)	2	9
All	All	1914/2018 (95%)	1544 (81%)	370 (19%)	1	9

5 of 370 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	B	549	LEU
1	C	159	ARG
2	D	455	VAL
2	B	586	LEU
1	C	22	THR

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 60 such sidechains are listed below:

Mol	Chain	Res	Type
2	B	641	ASN
1	C	106	ASN
2	D	565	HIS
1	C	41	GLN
1	C	138	ASN

### 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 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 [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	587/588 (99%)	0.43	44 (7%) 15 8	61, 71, 71, 81	0
1	C	586/588 (99%)	0.49	42 (7%) 16 10	60, 71, 71, 87	0
2	B	472/524 (90%)	0.63	36 (7%) 15 8	70, 71, 71, 72	0
2	D	469/524 (89%)	0.60	34 (7%) 16 10	71, 71, 71, 75	0
All	All	2114/2224 (95%)	0.53	156 (7%) 15 9	60, 71, 71, 87	0

The worst 5 of 156 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	211	ALA	6.5
2	D	589	GLY	6.1
1	C	325	ASN	5.8
1	A	328	GLU	5.7
1	C	324	PHE	5.7

### 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.