



## wwPDB/EMDatabank EM Map/Model Validation Summary Report ⓘ

Oct 9, 2018 – 11:52 AM EDT

PDB ID : 6GH5  
EMDB ID: : EMD-0001  
Title : Cryo-EM structure of bacterial RNA polymerase-sigma54 holoenzyme tran-  
scription open complex  
Authors : Glyde, R.; Ye, F.Z.; Zhang, X.D.  
Deposited on : 2018-05-04  
Resolution : 3.40 Å(reported)

This is a wwPDB/EMDatabank EM Map/Model Validation Summary Report  
for a publicly released PDB/EMDB 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/EMValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

MolProbity : 4.02b-467  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et. al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20031633

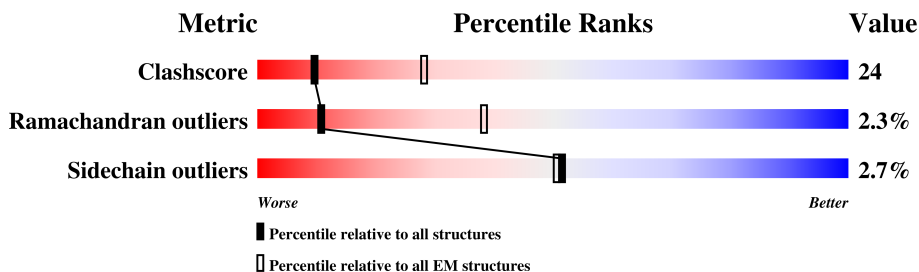
# 1 Overall quality at a glance i

The following experimental techniques were used to determine the structure:

*ELECTRON MICROSCOPY*

The reported resolution of this entry is 3.40 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)	EM structures (#Entries)
Clashscore	136327	1886
Ramachandran outliers	132723	1663
Sidechain outliers	132532	1531

The table below summarises the geometric issues observed across the polymeric chains. The red, orange, yellow and green segments on the 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	329	56% 30% 8% • 6%
1	B	329	41% 25% 5% • 29%
2	C	1342	54% 30% 13% •
3	D	1407	48% 31% 13% • •
4	E	91	47% 24% 10% • 18%
5	M	497	55% 10% • 32%
6	F	63	14% 46% 13% 27%
7	G	63	29% 44% 27%

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 28316 atoms, of which 0 are hydrogens and 0 are deuteriums.

In the tables below, 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 DNA-directed RNA polymerase subunit alpha.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
1	A	309	Total	C	N	O	S	0	0
			2302	1441	400	454	7		
1	B	235	Total	C	N	O	S	0	0
			1733	1085	301	341	6		

- Molecule 2 is a protein called DNA-directed RNA polymerase subunit beta.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
2	C	1341	Total	C	N	O	S	0	0
			10034	6289	1746	1961	38		

- Molecule 3 is a protein called DNA-directed RNA polymerase subunit beta'.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
3	D	1345	Total	C	N	O	S	0	0
			9790	6144	1746	1858	42		

- Molecule 4 is a protein called DNA-directed RNA polymerase subunit omega.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
4	E	75	Total	C	N	O	S	0	0
			565	347	110	107	1		

- Molecule 5 is a protein called RNA polymerase sigma-54 factor, RNA polymerase sigma-54 factor, RNA polymerase sigma-54 factor, RNA polymerase sigma-54 factor, RNA polymerase sigma-54 factor.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	S		
5	M	338	Total	C	N	O	S	0	0
			2002	1243	356	400	3		

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
M	-19	MET	-	initiating methionine	UNP A0A0J4U551
M	-18	GLY	-	expression tag	UNP A0A0J4U551
M	-17	SER	-	expression tag	UNP A0A0J4U551
M	-16	SER	-	expression tag	UNP A0A0J4U551
M	-15	HIS	-	expression tag	UNP A0A0J4U551
M	-14	HIS	-	expression tag	UNP A0A0J4U551
M	-13	HIS	-	expression tag	UNP A0A0J4U551
M	-12	HIS	-	expression tag	UNP A0A0J4U551
M	-11	HIS	-	expression tag	UNP A0A0J4U551
M	-10	HIS	-	expression tag	UNP A0A0J4U551
M	-9	SER	-	expression tag	UNP A0A0J4U551
M	-8	SER	-	expression tag	UNP A0A0J4U551
M	-7	GLY	-	expression tag	UNP A0A0J4U551
M	-6	LEU	-	expression tag	UNP A0A0J4U551
M	-5	VAL	-	expression tag	UNP A0A0J4U551
M	-4	PRO	-	expression tag	UNP A0A0J4U551
M	-3	ARG	-	expression tag	UNP A0A0J4U551
M	-2	GLY	-	expression tag	UNP A0A0J4U551
M	-1	SER	-	expression tag	UNP A0A0J4U551
M	0	HIS	-	expression tag	UNP A0A0J4U551
M	336	ALA	ARG	engineered mutation	UNP A0A0J4U551

- Molecule 6 is a DNA chain called nifH promoter template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
6	F	46	944	445	182	271	46	0	0

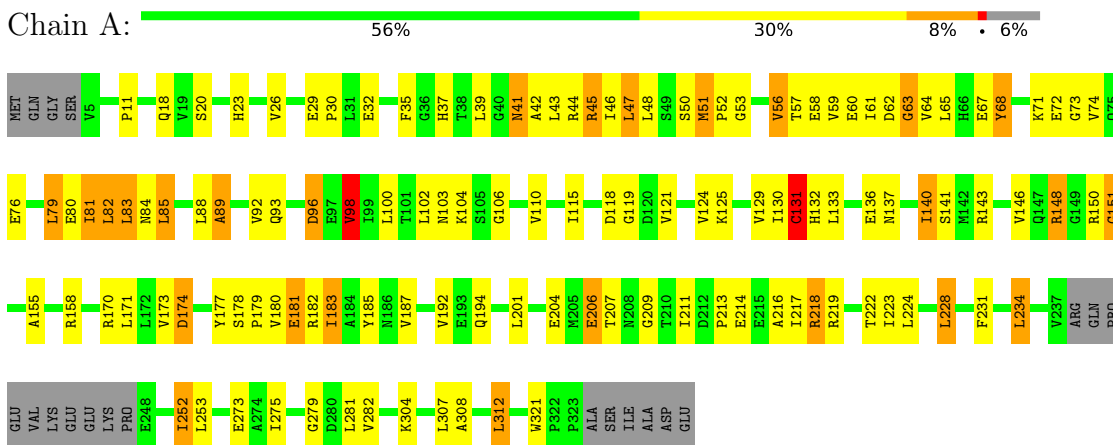
- Molecule 7 is a DNA chain called nifH promoter non-template DNA.

Mol	Chain	Residues	Atoms					AltConf	Trace
			Total	C	N	O	P		
7	G	46	946	448	173	279	46	0	0

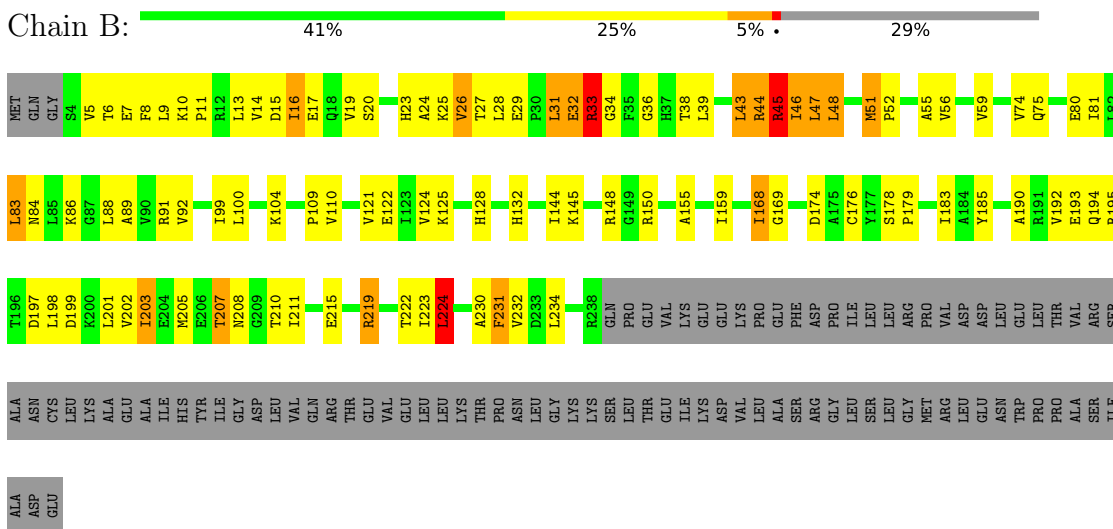
### 3 Residue-property plots [\(i\)](#)

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.

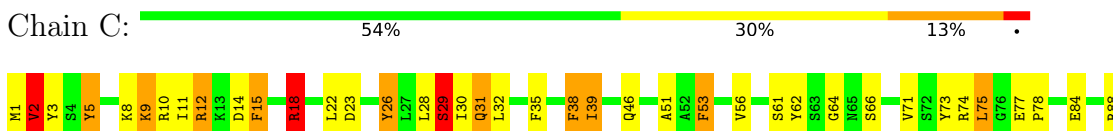
- Molecule 1: DNA-directed RNA polymerase subunit alpha



- Molecule 1: DNA-directed RNA polymerase subunit alpha



- Molecule 2: DNA-directed RNA polymerase subunit beta





ALA	I1320	H1218	Q1108	T931	L863	A784	V706	G613	Y537	A476	Q340	Y144	G81
SER	S1321	D1219	L1109	M932	L864	A787	I707	L614	R538	Q477	M341	V145	G82
ALA	A1322	I1220	G1118	V966	H865	R788	M708	K615	S539	L478	L342	V146	V83
GLU	S1324	L1221	T1119	V967	H866	K789	D710	P616	G540	E479	L343	L147	L84
LEU	F1325	R1222	T1120	N968	Q867	T790	G711	V618	A541	R481	K344	Q157	C85
LEU	Q1326	L1223	L1121	N969	C868	K791	D712	V619	A542	R482	G345	Q158	G88
ASN	T1327	R1224	A1122	I975	C869	N792	Q716	F620	S543	A483	R346	K179	G89
ALA	T1328	V1229	T976	T976	L872	S793	V717	A621	H545	M484	D348	M180	V90
GLY	V1329	Q1126	Q980	T980	D878	G794	S718	D622	H546	M485	Y349	L279	E91
LEU	R1330	GLU	I985	T985	A879	V795	F719	Q623	A547	S486	S350	A184	V92
GLY	V1331	SER	I986	I986	V880	T797	I722	I624	V548	T487	G351	L185	T93
GLY	L1332	GLY	D886	K881	V881	R798	Y723	Y626	H550	M489	R352	L186	Q94
SER	T1333	THR	F987	R882	K882	R799	Y724	T627	R551	V490	S353	L188	T95
ASP	V1337	LYS	F988	R883	V883	L800	M725	D628	R552	L491	I355	L189	K96
ASN	V1338	LYS	F989	R884	V884	V801	R726	F629	S492	F431	T356	L205	V97
GLU	V1339	LYS	F990	R885	V885	D802	R731	A630	E554	F432	V357	L206	V98
ALA	T1340	I1134	V1011	V886	V886	M803	G732	Y631	E555	A433	G358	T208	F100
ALA	R1341	G1013	A1012	S887	S887	V803	G733	A632	Y557	I434	R359	M209	R101
D1342	D1342	V1031	I1046	D806	D806	D806	Q736	A633	T567	Q435	Y360	M209	M102
L1344	L1344	V1032	I1047	L807	L807	L807	I737	R634	S568	A436	L361	K215	G103
L1345	L1345	D1042	D1042	V808	V808	V808	R738	R635	R569	F437	L362	K216	G104
L1346	L1346	D1043	D1043	V809	V809	V809	Q739	S635	I500	F438	L363	L217	I105
L1347	L1347	I1044	I1044	T810	T810	T810	L740	S636	V501	E439	L364	L218	I106
L1348	L1348	I1045	I1045	E811	E811	E811	A741	V639	P502	V440	Q300	K219	F105
L1349	L1349	I1046	I1046	D812	D812	D812	G742	G640	S503	L442	C306	K220	L107
M1350	M1350	Q1049	Q1049	R813	R813	R813	M743	F641	G504	T442	A305	R221	A108
L1351	L1351	THR	THR	C814	C814	C814	G744	D642	D505	L443	L366	I222	S109
V1352	V1352	ASP	ASP	H814	H814	H814	R745	D643	R506	E444	L367	K222	P110
V1353	V1353	GLU	GLU	M822	M822	M822	L746	D644	V506	Q444	L368	L223	T111
V1354	V1354	LEU	LEU	V825	V825	V825	K749	I646	V507	A445	L369	L224	A112
R1355	R1355	THR	THR	R826	R826	R826	P750	V649	G509	T447	M372	E225	H113
L1356	L1356	GLY	GLY	I827	I827	I827	P751	K649	M510	Q448	A373	A226	L114
L1357	L1357	LEU	LEU	D901	D901	D901	P752	V653	Y511	L449	L376	F227	W115
A1359	A1359	S1057	S1057	L903	L903	L903	P753	I653	Y512	H450	L377	S230	F116
K1263	K1263	S1058	S1058	A904	A904	A904	P754	V654	M513	H451	L378	S231	L117
S1271	S1271	R1067	R1067	R905	R905	R905	T755	E680	T514	L452	F377	K233	K118
S1272	S1272	THR	THR	I908	I908	I908	A761	E681	R515	V453	P379	P294	L120
S1273	S1273	ALA	ALA	O909	O909	O909	M762	T684	G586	C454	F380	E235	P121
G1277	G1277	GLY	GLY	N910	N910	N910	F763	V673	C517	A455	I391	E236	P122
V1280	V1280	GLY	GLY	K911	K911	K911	R764	V674	V518	A456	Y392	W236	S122
R1304	R1304	LYS	LYS	E912	E912	E912	E765	T674	V519	A457	K321	M237	S123
D1305	D1305	ASP	ASP	E913	E913	E913	G766	A675	A520	M458	G383	L238	I124
ALA	ALA	LEU	LEU	A914	A914	A914	L767	V678	K821	V459	L385	T240	G125
ALA	ALA	LEU	LEU	I915	I915	I915	M768	R678	G522	F461	E386	P243	L126
ALA	ALA	PRO	PRO	I916	I916	I916	V769	V693	E523	D462	L324	V244	L127
ALA	ALA	VAL	VAL	G917	G917	G917	L770	V694	E524	D463	L325	V245	M130
ALA	ALA	THR	THR	V917	V917	V917	Q771	K681	G525	G464	R388	L244	P131
ALA	ALA	ASP	ASP	I918	I918	I918	T772	V685	M526	D465	G389	L245	L132
ALA	ALA	LEU	LEU	I919	I919	I919	F773	V686	V526	M466	L390	P246	L133
ALA	ALA	ILE	ILE	Q921	Q921	Q921	I774	V687	K599	M467	A391	L249	R133
GLN	GLN	PRO	PRO	G924	G924	G924	S775	V693	A600	V468	I394	L250	D134
VAL	VAL	THR	THR	E925	E925	E925	T776	V694	I601	H469	R395	R250	L135
THR	THR	MET	MET	P926	P926	P926	H777	M697	L605	K531	G396	P251	I136
ALA	ALA	PRO	PRO	G927	G927	G927	R780	M698	Y609	V471	A396	L252	R137
GLU	GLU	ASP	ASP	T928	T928	T928	K781	V701	R610	L472	K337	V253	V138
ALA	ALA	ASP	ASP	Q929	Q929	Q929	L782	Q702	E554	L473	R337	F260	L139
SER	SER	SER	SER	L930	L930	L930	L783	Q703	E555	L474	F338	A261	Y140
												T262	S143

• Molecule 4: DNA-directed RNA polymerase subunit omega





## 4 Experimental information

Property	Value	Source
Reconstruction method	SINGLE PARTICLE	Depositor
Imposed symmetry	POINT, C1	Depositor
Number of particles used	79678	Depositor
Resolution determination method	FSC 0.143 CUT-OFF	Depositor
CTF correction method	PHASE FLIPPING AND AMPLITUDE CORRECTION	Depositor
Microscope	FEI TITAN KRIOS	Depositor
Voltage (kV)	300	Depositor
Electron dose ( $e^-/\text{\AA}^2$ )	45	Depositor
Minimum defocus (nm)	Not provided	Depositor
Maximum defocus (nm)	Not provided	Depositor
Magnification	Not provided	Depositor
Image detector	GATAN K2 SUMMIT (4k x 4k)	Depositor

## 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  >2	RMSZ	# Z  >2
1	A	1.46	24/2331 (1.0%)	1.65	48/3177 (1.5%)
1	B	1.15	8/1752 (0.5%)	1.41	29/2385 (1.2%)
2	C	1.84	276/10187 (2.7%)	1.96	384/13822 (2.8%)
3	D	1.84	272/9923 (2.7%)	1.98	394/13482 (2.9%)
4	E	1.21	3/567 (0.5%)	1.72	15/767 (2.0%)
5	M	0.93	11/1764 (0.6%)	1.02	8/2445 (0.3%)
6	F	0.70	1/1060 (0.1%)	1.31	17/1633 (1.0%)
7	G	0.54	0/1060	0.84	1/1635 (0.1%)
All	All	1.66	595/28644 (2.1%)	1.81	896/39346 (2.3%)

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	6
2	C	0	28
3	D	0	28
4	E	0	1
5	M	0	1
All	All	0	64

The worst 5 of 595 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	D	462	ASP	CB-CG	18.37	1.90	1.51
3	D	99	ARG	CZ-NH1	17.98	1.56	1.33
2	C	1216	ARG	CZ-NH1	17.86	1.56	1.33
2	C	555	TYR	CE2-CZ	16.61	1.60	1.38
3	D	333	GLY	N-CA	16.11	1.70	1.46

The worst 5 of 896 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	1216	ARG	NE-CZ-NH2	-24.92	107.84	120.30
2	C	1301	ARG	NE-CZ-NH1	23.30	131.95	120.30
2	C	1301	ARG	NE-CZ-NH2	-23.28	108.66	120.30
2	C	149	LEU	CB-CG-CD1	-23.16	71.63	111.00
3	D	422	LEU	CB-CG-CD2	-22.53	72.70	111.00

There are no chirality outliers.

5 of 64 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	131	CYS	Mainchain
1	A	151	GLY	Mainchain
1	A	47	LEU	Mainchain
1	A	63	GLY	Peptide
1	A	93	GLN	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	2302	0	2266	94	0
1	B	1733	0	1720	80	0
2	C	10034	0	9668	444	0
3	D	9790	0	9539	583	0
4	E	565	0	563	27	0
5	M	2002	0	1412	62	0
6	F	944	0	513	99	0
7	G	946	0	518	42	0
All	All	28316	0	26199	1305	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 24.

The worst 5 of 1305 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)
7:G:14:DT:H2''	7:G:15:DT:C7	1.31	1.55

*Continued on next page...*

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:C:1066:MET:CE	2:C:1232:MET:HE2	1.31	1.53
2:C:1260:GLY:CA	3:D:346:ARG:HH12	1.20	1.52
3:D:333:GLY:CA	3:D:333:GLY:N	1.70	1.51
3:D:316:ILE:CD1	3:D:317:THR:H	1.23	1.50

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 PDB entries followed by that with respect to all EM entries.

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	305/329 (93%)	269 (88%)	32 (10%)	4 (1%)	13	50
1	B	233/329 (71%)	209 (90%)	22 (9%)	2 (1%)	19	58
2	C	1339/1342 (100%)	1183 (88%)	128 (10%)	28 (2%)	8	40
3	D	1335/1407 (95%)	1127 (84%)	169 (13%)	39 (3%)	5	33
4	E	73/91 (80%)	68 (93%)	3 (4%)	2 (3%)	5	34
5	M	283/497 (57%)	236 (83%)	40 (14%)	7 (2%)	6	36
All	All	3568/3995 (89%)	3092 (87%)	394 (11%)	82 (2%)	11	38

5 of 82 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	252	ILE
2	C	29	SER
2	C	347	ILE
2	C	541	GLU
2	C	1155	VAL

### 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 PDB entries followed by that with respect to all EM entries.

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	244/286 (85%)	241 (99%)	3 (1%)	74 88
1	B	183/286 (64%)	178 (97%)	5 (3%)	48 77
2	C	1029/1157 (89%)	1003 (98%)	26 (2%)	50 79
3	D	961/1168 (82%)	925 (96%)	36 (4%)	37 70
4	E	56/75 (75%)	55 (98%)	1 (2%)	62 83
5	M	113/393 (29%)	113 (100%)	0	100 100
All	All	2586/3365 (77%)	2515 (97%)	71 (3%)	51 77

5 of 71 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	1243	MET
3	D	93	THR
3	D	987	GLU
2	C	1319	MET
3	D	28	ASP

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 27 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	688	GLN
2	C	1136	GLN
3	D	907	HIS
2	C	799	ASN
1	B	23	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 [i](#)

There are no chain breaks in this entry.