



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 15, 2023 – 06:00 PM EDT

PDB ID : 1TH1  
Title : Beta-catenin in complex with a phosphorylated APC 20aa repeat fragment  
Authors : Xing, Y.; Clements, W.K.; Le Trong, I.; Hinds, T.R.; Stenkamp, R.; Kimelman, D.; Xu, W.  
Deposited on : 2004-05-31  
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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

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 : 2.35  
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) : 2.35

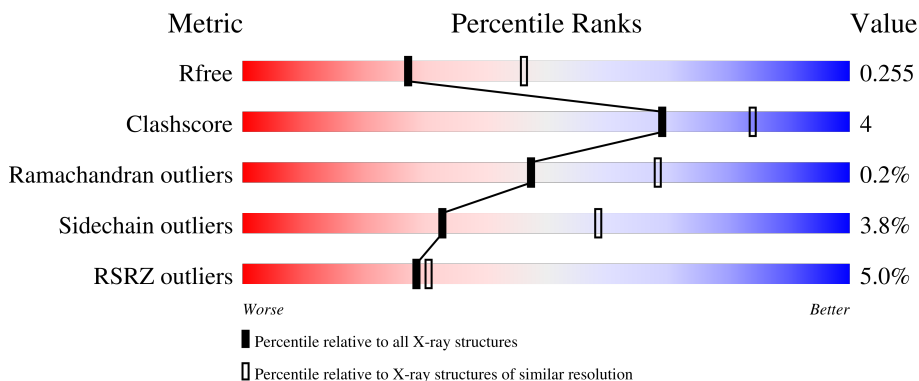
# 1 Overall quality at a glance

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	532	
1	B	532	
2	C	179	
2	D	179	

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 8667 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-catenin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	514	3912	2456	710	720	26	0	0	0
1	B	513	3904	2452	709	717	26	0	0	0

- Molecule 2 is a protein called Adenomatous polyposis coli protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace	
			Total	C	N	O	P				S
2	C	56	416	249	65	96	5	1	0	0	0
2	D	54	389	232	61	90	5	1	0	0	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	1487	TPO	THR	modified residue	UNP P25054
C	1504	SEP	SER	modified residue	UNP P25054
C	1505	SEP	SER	modified residue	UNP P25054
C	1507	SEP	SER	modified residue	UNP P25054
C	1510	SEP	SER	modified residue	UNP P25054
D	1487	TPO	THR	modified residue	UNP P25054
D	1504	SEP	SER	modified residue	UNP P25054
D	1505	SEP	SER	modified residue	UNP P25054
D	1507	SEP	SER	modified residue	UNP P25054
D	1510	SEP	SER	modified residue	UNP P25054

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	22	Total	O	0	0
			22	22		

*Continued on next page...*

*Continued from previous page...*

<b>Mol</b>	<b>Chain</b>	<b>Residues</b>	<b>Atoms</b>		<b>ZeroOcc</b>	<b>AltConf</b>
3	B	20	Total 20	O 20	0	0
3	C	2	Total 2	O 2	0	0
3	D	2	Total 2	O 2	0	0



● Molecule 2: Adenomatous polyposis coli protein



SER LYS SER GLY ALA GLN THR PRO LYS SER PRO PRO GLU HIS TYR VAL GLN THR PRO LEU MET PHE SER ARG CYS THR SER VAL SER LEU ASP SER PHE GLU SER ARG SER ILE ALA SER VAL SER GLN GLU PRO CYS SER GLY MET VAL SER GLY ILE ILE SER PRO SER

ASP LEU PRO ASP SER PRO GLY GLN THR MET THR SER ARG SER LYS THR PRO PRO PRO PRO GLN THR ALA GLN THR LYS SER ARG GLU GLY LYS Q1469 A1470 A1471 Q1477 R1478 V1479 L1482 P1483 D1484 A1485 D1486

T1487 T1496 PRO ASP PHE SER CYS SER S1504 S1505 I1506 S1507 S1510 E1513 S1529 GLU ASN ASP ASN GLY ASN THR GLU SER GLU

## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	85.16Å 133.80Å 82.17Å 90.00° 110.85° 90.00°	Depositor
Resolution (Å)	47.98 – 2.50 47.98 – 2.22	Depositor EDS
% Data completeness (in resolution range)	85.1 (47.98-2.50) 87.3 (47.98-2.22)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	0.10	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.12 (at 2.22Å)	Xtrriage
Refinement program	CNS 1.1	Depositor
R, $R_{free}$	0.225 , 0.257 0.225 , 0.255	Depositor DCC
$R_{free}$ test set	3029 reflections (3.57%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	38.0	Xtrriage
Anisotropy	0.534	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 30.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.33$	Xtrriage
Estimated twinning fraction	0.019 for l,-k,h	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	8667	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	49.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.35% 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](#)

Bond lengths and bond angles in the following residue types are not validated in this section: TPO, SEP

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	0.55	0/3966	0.72	0/5381
1	B	0.55	0/3958	0.72	1/5370 (0.0%)
2	C	0.64	0/366	0.78	0/495
2	D	0.60	0/340	0.85	0/462
All	All	0.55	0/8630	0.73	1/11708 (0.0%)

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	B	0	1

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	B	624	ASP	CB-CG-OD1	7.40	124.96	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	B	624	ASP	Mainchain



## 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	3912	0	4050	29	0
1	B	3904	0	4046	35	0
2	C	416	0	358	2	0
2	D	389	0	318	7	0
3	A	22	0	0	0	0
3	B	20	0	0	1	0
3	C	2	0	0	0	0
3	D	2	0	0	0	0
All	All	8667	0	8772	65	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 4.

The worst 5 of 65 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:A:483:ASN:HD22	1:A:486:ARG:HH21	1.22	0.87
1:B:364:GLN:HE21	1:B:364:GLN:H	1.26	0.84
1:B:205:THR:HG22	1:B:207:ASP:H	1.45	0.81
1:B:280:GLN:HE21	1:B:280:GLN:H	1.27	0.78
1:A:483:ASN:HD22	1:A:486:ARG:NH2	1.83	0.76

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	510/532 (96%)	502 (98%)	8 (2%)	0	100	100
1	B	509/532 (96%)	503 (99%)	4 (1%)	2 (0%)	34	54
2	C	47/179 (26%)	45 (96%)	2 (4%)	0	100	100
2	D	46/179 (26%)	41 (89%)	5 (11%)	0	100	100
All	All	1112/1422 (78%)	1091 (98%)	19 (2%)	2 (0%)	47	68

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	181	LYS
1	B	182	GLU

### 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	424/441 (96%)	410 (97%)	14 (3%)	38	64
1	B	423/441 (96%)	405 (96%)	18 (4%)	29	53
2	C	35/155 (23%)	32 (91%)	3 (9%)	10	20
2	D	29/155 (19%)	29 (100%)	0	100	100
All	All	911/1192 (76%)	876 (96%)	35 (4%)	33	58

5 of 35 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	559	GLN
1	B	578	HIS
2	C	1489	LEU
1	A	580	LEU
1	A	559	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 24 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	287	ASN
1	B	326	ASN
1	B	322	GLN
1	B	364	GLN
1	A	395	GLN

### 5.3.3 RNA [i](#)

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains [i](#)

10 non-standard protein/DNA/RNA residues 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	SEP	C	1504	2	8,9,10	1.15	0	8,12,14	3.75	4 (50%)
2	SEP	C	1510	2	8,9,10	1.25	1 (12%)	8,12,14	3.77	4 (50%)
2	TPO	D	1487	2	8,10,11	2.03	2 (25%)	10,14,16	0.78	0
2	SEP	D	1504	2	8,9,10	1.27	1 (12%)	8,12,14	4.40	4 (50%)
2	SEP	D	1507	2	8,9,10	1.04	0	8,12,14	3.81	4 (50%)
2	TPO	C	1487	2	8,10,11	1.60	1 (12%)	10,14,16	0.94	0
2	SEP	D	1510	2	8,9,10	1.17	0	8,12,14	4.12	4 (50%)
2	SEP	D	1505	2	8,9,10	1.20	1 (12%)	8,12,14	4.03	4 (50%)
2	SEP	C	1505	2	8,9,10	1.30	2 (25%)	8,12,14	4.01	4 (50%)
2	SEP	C	1507	2	8,9,10	1.09	1 (12%)	8,12,14	4.68	4 (50%)

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	SEP	C	1504	2	-	1/5/8/10	-
2	SEP	C	1510	2	-	1/5/8/10	-
2	TPO	D	1487	2	-	3/9/11/13	-
2	SEP	D	1504	2	-	1/5/8/10	-
2	SEP	D	1507	2	-	0/5/8/10	-
2	TPO	C	1487	2	-	1/9/11/13	-
2	SEP	D	1510	2	-	1/5/8/10	-
2	SEP	D	1505	2	-	1/5/8/10	-
2	SEP	C	1505	2	-	1/5/8/10	-
2	SEP	C	1507	2	-	1/5/8/10	-

The worst 5 of 9 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	D	1487	TPO	P-OG1	3.80	1.66	1.59
2	C	1487	TPO	P-O1P	2.73	1.59	1.50
2	D	1487	TPO	P-O1P	2.57	1.58	1.50
2	C	1505	SEP	P-O1P	2.15	1.57	1.50
2	D	1504	SEP	P-O1P	2.07	1.57	1.50

The worst 5 of 32 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	C	1507	SEP	OG-CB-CA	7.62	115.56	108.14
2	C	1507	SEP	P-OG-CB	-7.42	97.86	118.30
2	D	1504	SEP	OG-CB-CA	7.10	115.06	108.14
2	C	1505	SEP	P-OG-CB	-6.89	99.32	118.30
2	D	1510	SEP	P-OG-CB	-6.79	99.61	118.30

There are no chirality outliers.

5 of 11 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	1487	TPO	O-C-CA-CB
2	C	1504	SEP	N-CA-CB-OG
2	C	1505	SEP	N-CA-CB-OG
2	C	1507	SEP	N-CA-CB-OG
2	C	1510	SEP	N-CA-CB-OG

There are no ring outliers.

No monomer is involved in short contacts.

## 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	514/532 (96%)	-0.11	11 (2%) 63 66	19, 36, 80, 117	0
1	B	513/532 (96%)	0.11	29 (5%) 23 25	21, 43, 98, 120	0
2	C	51/179 (28%)	1.10	12 (23%) 0 0	32, 77, 96, 97	0
2	D	49/179 (27%)	0.74	4 (8%) 11 11	34, 82, 101, 103	0
All	All	1127/1422 (79%)	0.08	56 (4%) 28 30	19, 42, 97, 120	0

The worst 5 of 56 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	552	SER	6.0
1	A	146	ALA	5.9
1	B	551	THR	5.6
1	B	560	PHE	5.3
1	B	148	LEU	5.2

### 6.2 Non-standard residues in protein, DNA, RNA chains [i](#)

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. 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	RSCC	RSR	B-factors(Å <sup>2</sup> )	Q<0.9
2	SEP	D	1504	10/11	0.69	0.21	77,82,88,89	0
2	TPO	D	1487	11/12	0.75	0.27	60,72,84,84	0
2	SEP	C	1504	10/11	0.78	0.23	87,91,98,98	0
2	SEP	D	1505	10/11	0.81	0.14	67,75,85,86	0
2	SEP	D	1510	10/11	0.86	0.14	50,58,70,71	0
2	SEP	C	1510	10/11	0.87	0.16	64,70,79,80	0
2	SEP	C	1505	10/11	0.90	0.19	78,85,92,92	0

*Continued on next page...*

*Continued from previous page...*

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors( $\text{\AA}^2$ )	Q<0.9
2	TPO	C	1487	11/12	0.93	0.12	43,49,61,61	0
2	SEP	C	1507	10/11	0.95	0.14	55,58,61,62	0
2	SEP	D	1507	10/11	0.97	0.18	39,43,45,48	0

### 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.