



# Full wwPDB X-ray Structure Validation Report ⓘ

Aug 2, 2016 – 12:02 PM EDT

PDB ID : 5CY5  
Title : Crystal structure of the T33-51H designed self-assembling protein tetrahedron  
Authors : Cannon, K.A.; Cascio, D.; Park, R.; Boyken, S.; King, N.; Yeates, T.O.  
Deposited on : 2015-07-30  
Resolution : 3.40 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20027939
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
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)	:	rb-20027939

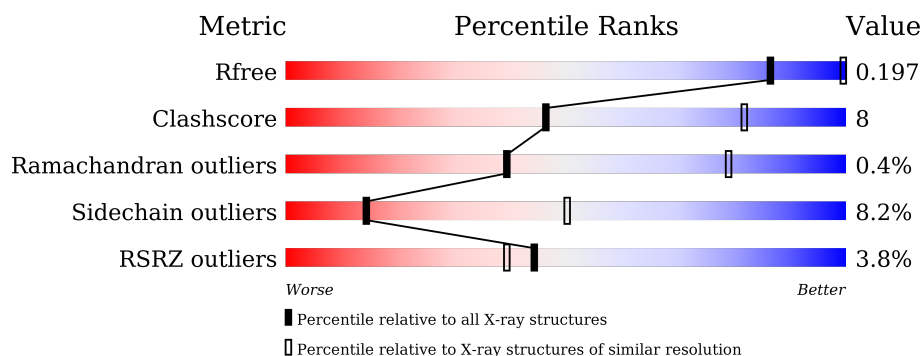
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.40 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)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	91344	1476 (3.50-3.30)
Clashscore	102246	1611 (3.50-3.30)
Ramachandran outliers	100387	1571 (3.50-3.30)
Sidechain outliers	100360	1571 (3.50-3.30)
RSRZ outliers	91569	1485 (3.50-3.30)

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	B	185	<div> <div>4%</div> <div> <div></div> <div>66%</div> <div>14%</div> <div>•</div> <div>19%</div> </div> </div>
2	A	172	<div> <div>2%</div> <div> <div></div> <div>55%</div> <div>20%</div> <div>5%</div> <div>20%</div> </div> </div>

## 2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 2253 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 T33-51H-B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	B	149	Total	C	N	O	S	0	0	0
			1176	749	206	215	6			

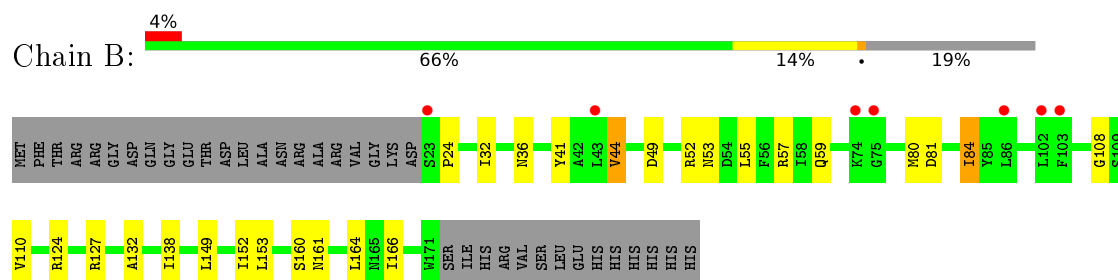
- Molecule 2 is a protein called T33-51H-A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	A	137	Total	C	N	O	S	0	0	0
			1077	687	181	205	4			

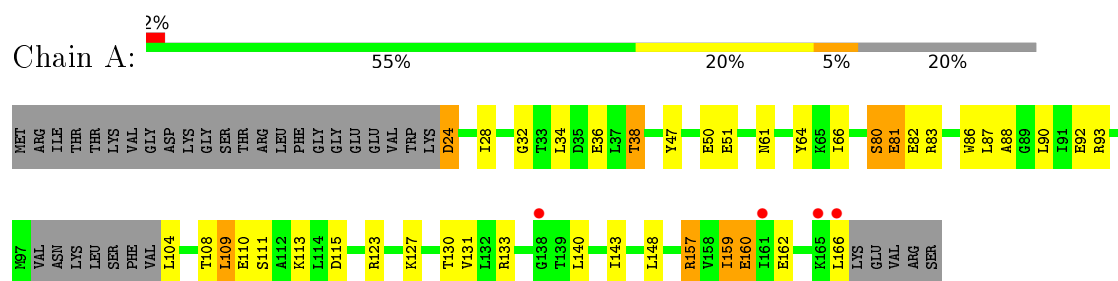
### 3 Residue-property plots

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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: T33-51H-B



#### • Molecule 2: T33-51H-A



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	106.70Å 106.70Å 106.70Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	75.45 – 3.40 75.45 – 3.40	Depositor EDS
% Data completeness (in resolution range)	100.0 (75.45-3.40) 100.0 (75.45-3.40)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.62 (at 3.41Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.9_1692)	Depositor
R, $R_{free}$	0.156 , 0.197 0.156 , 0.197	Depositor DCC
$R_{free}$ test set	578 reflections (9.99%)	DCC
Wilson B-factor (Å <sup>2</sup> )	88.6	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.36 , 50.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.41$ , $\langle L^2 \rangle = 0.23$	Xtriage
Estimated twinning fraction	0.294 for l,-k,h	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	2253	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.04% 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	B	0.40	0/1192	0.53	0/1608
2	A	0.39	0/1087	0.55	0/1457
All	All	0.39	0/2279	0.54	0/3065

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no planarity outliers.

### 5.2 Too-close contacts

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	B	1176	0	1216	16	0
2	A	1077	0	1122	20	0
All	All	2253	0	2338	35	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 8.

All (35) 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)
2:A:47:TYR:HB3	2:A:110:GLU:HG3	1.82	0.60
2:A:51:GLU:OE2	2:A:93:ARG:NH2	2.34	0.60

*Continued on next page...*

*Continued from previous page...*

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:A:108:THR:H	2:A:111:SER:HB3	1.68	0.58
2:A:109:LEU:HG	2:A:110:GLU:N	2.19	0.58
1:B:49:ASP:OD1	1:B:52:ARG:NH2	2.38	0.56
2:A:109:LEU:O	2:A:113:LYS:HG2	2.05	0.56
1:B:161:ASN:HB3	1:B:166:ILE:O	2.06	0.56
2:A:34:LEU:O	2:A:38:THR:HG22	2.07	0.55
2:A:81:GLU:HG3	2:A:140:LEU:HD21	1.91	0.53
1:B:81:ASP:HA	1:B:84:ILE:HB	1.90	0.52
1:B:53:ASN:HB3	1:B:57:ARG:HH12	1.77	0.50
1:B:124:ARG:HA	1:B:127:ARG:CZ	2.42	0.49
2:A:127:LYS:O	2:A:131:VAL:HG23	2.14	0.48
2:A:61:ASN:O	2:A:64:TYR:HB3	2.13	0.48
1:B:108:GLY:HA3	1:B:161:ASN:HD21	1.79	0.48
2:A:80:SER:HB3	2:A:83:ARG:HG2	1.96	0.47
1:B:132:ALA:HB3	1:B:138:ILE:HD11	1.96	0.47
2:A:157:ARG:HA	2:A:160:GLU:OE1	2.15	0.46
2:A:162:GLU:O	2:A:166:LEU:HG	2.16	0.46
1:B:80:MET:O	1:B:84:ILE:N	2.44	0.44
1:B:124:ARG:HG3	1:B:127:ARG:HH12	1.82	0.44
2:A:32:GLY:O	2:A:36:GLU:HG3	2.18	0.44
2:A:24:ASP:O	2:A:28:ILE:HG13	2.18	0.43
1:B:41:TYR:O	1:B:44:VAL:HG12	2.19	0.43
1:B:80:MET:CE	2:A:143:ILE:HG23	2.49	0.43
1:B:55:LEU:O	1:B:59:GLN:HG3	2.18	0.42
2:A:88:ALA:O	2:A:92:GLU:HG3	2.19	0.42
1:B:149:LEU:O	1:B:152:ILE:HG22	2.20	0.42
2:A:87:LEU:HD23	2:A:87:LEU:HA	1.83	0.41
1:B:32:ILE:O	1:B:36:ASN:HB2	2.20	0.41
2:A:83:ARG:O	2:A:86:TRP:HB3	2.20	0.41
2:A:159:ILE:HG13	2:A:159:ILE:H	1.70	0.41
1:B:149:LEU:O	1:B:153:LEU:HG	2.21	0.40
1:B:160:SER:O	1:B:164:LEU:HG	2.21	0.40
2:A:90:LEU:HA	2:A:90:LEU:HD23	1.76	0.40

There are no symmetry-related clashes.

## 5.3 Torsion angles

### 5.3.1 Protein backbone

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	B	147/185 (80%)	140 (95%)	6 (4%)	1 (1%)	26	70
2	A	133/172 (77%)	128 (96%)	5 (4%)	0	100	100
All	All	280/357 (78%)	268 (96%)	11 (4%)	1 (0%)	39	79

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	24	PRO

### 5.3.2 Protein sidechains

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	B	129/161 (80%)	126 (98%)	3 (2%)	58	85
2	A	114/145 (79%)	97 (85%)	17 (15%)	4	20
All	All	243/306 (79%)	223 (92%)	20 (8%)	14	50

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	44	VAL
1	B	84	ILE
1	B	110	VAL
2	A	24	ASP
2	A	38	THR

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Res	Type
2	A	50	GLU
2	A	66	ILE
2	A	80	SER
2	A	81	GLU
2	A	82	GLU
2	A	104	LEU
2	A	109	LEU
2	A	115	ASP
2	A	123	ARG
2	A	130	THR
2	A	133	ARG
2	A	148	LEU
2	A	157	ARG
2	A	159	ILE
2	A	160	GLU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	B	29	GLN
1	B	116	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.

## 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	B	149/185 (80%)	0.59	7 (4%) 35 32	45, 74, 111, 131	0
2	A	137/172 (79%)	0.54	4 (2%) 55 50	51, 79, 118, 137	0
All	All	286/357 (80%)	0.56	11 (3%) 44 39	45, 77, 112, 137	0

All (11) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	A	165	LYS	4.8
1	B	23	SER	3.5
1	B	74	LYS	3.2
2	A	166	LEU	2.6
2	A	138	GLY	2.4
1	B	102	LEU	2.3
1	B	75	GLY	2.2
2	A	161	ILE	2.1
1	B	86	LEU	2.1
1	B	103	PHE	2.1
1	B	43	LEU	2.1

### 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 carbohydrates in this entry.

## 6.4 Ligands

There are no ligands in this entry.

## 6.5 Other polymers

There are no such residues in this entry.