



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 9, 2023 – 06:24 PM EDT

PDB ID : 7TUE  
Title : Crystal structure of Tapasin in complex with HLA-B\*44:05 (T73C)  
Authors : Jiang, J.; Natarajan, K.; Kim, E.; Boyd, L.F.; Margulies, D.H.  
Deposited on : 2022-02-02  
Resolution : 3.10 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  
Xtriage (Phenix) : 1.13  
EDS : 2.35.1  
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

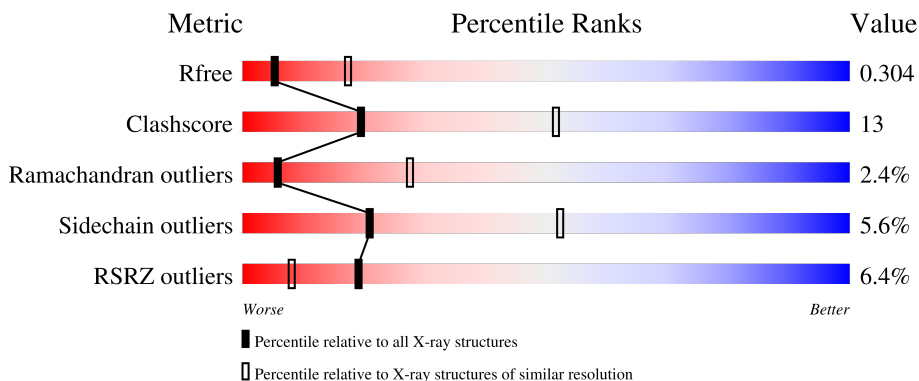
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.10 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	1094 (3.10-3.10)
Clashscore	141614	1184 (3.10-3.10)
Ramachandran outliers	138981	1141 (3.10-3.10)
Sidechain outliers	138945	1141 (3.10-3.10)
RSRZ outliers	127900	1067 (3.10-3.10)

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	274	 5% 58% 40%
2	D	416	 6% 61% 23% 13%
3	B	100	 8% 58% 29% 12%

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 5630 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 HLA class I histocompatibility antigen, B alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	274	2182	1358	393	423	8	0	0	0

- Molecule 2 is a protein called Tapasin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	D	361	2716	1746	473	486	11	0	0	0

There are 35 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
D	382	GLY	-	expression tag	UNP O15533
D	383	GLY	-	expression tag	UNP O15533
D	384	LEU	-	expression tag	UNP O15533
D	385	GLU	-	expression tag	UNP O15533
D	386	VAL	-	expression tag	UNP O15533
D	387	LEU	-	expression tag	UNP O15533
D	388	PHE	-	expression tag	UNP O15533
D	389	GLN	-	expression tag	UNP O15533
D	390	GLY	-	expression tag	UNP O15533
D	391	PRO	-	expression tag	UNP O15533
D	392	GLY	-	expression tag	UNP O15533
D	393	GLY	-	expression tag	UNP O15533
D	394	GLY	-	expression tag	UNP O15533
D	395	LEU	-	expression tag	UNP O15533
D	396	ASN	-	expression tag	UNP O15533
D	397	ASP	-	expression tag	UNP O15533
D	398	ILE	-	expression tag	UNP O15533
D	399	PHE	-	expression tag	UNP O15533
D	400	GLU	-	expression tag	UNP O15533
D	401	ALA	-	expression tag	UNP O15533

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	402	GLN	-	expression tag	UNP O15533
D	403	LYS	-	expression tag	UNP O15533
D	404	ILE	-	expression tag	UNP O15533
D	405	GLU	-	expression tag	UNP O15533
D	406	TRP	-	expression tag	UNP O15533
D	407	HIS	-	expression tag	UNP O15533
D	408	GLU	-	expression tag	UNP O15533
D	409	GLY	-	expression tag	UNP O15533
D	410	GLY	-	expression tag	UNP O15533
D	411	HIS	-	expression tag	UNP O15533
D	412	HIS	-	expression tag	UNP O15533
D	413	HIS	-	expression tag	UNP O15533
D	414	HIS	-	expression tag	UNP O15533
D	415	HIS	-	expression tag	UNP O15533
D	416	HIS	-	expression tag	UNP O15533

- Molecule 3 is a protein called Beta-2-microglobulin.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	B	88	732	464	126	140	2	0	0	0

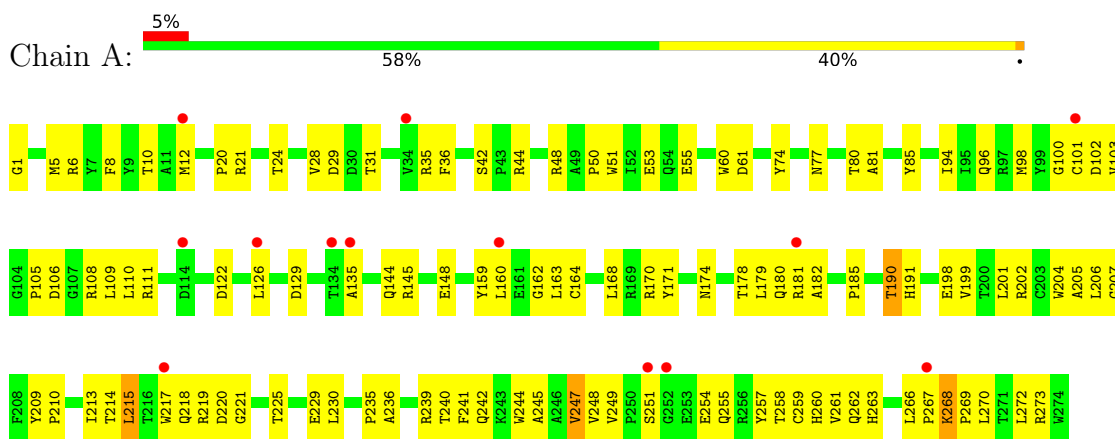
There is a discrepancy between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	0	MET	-	initiating methionine	UNP P61769

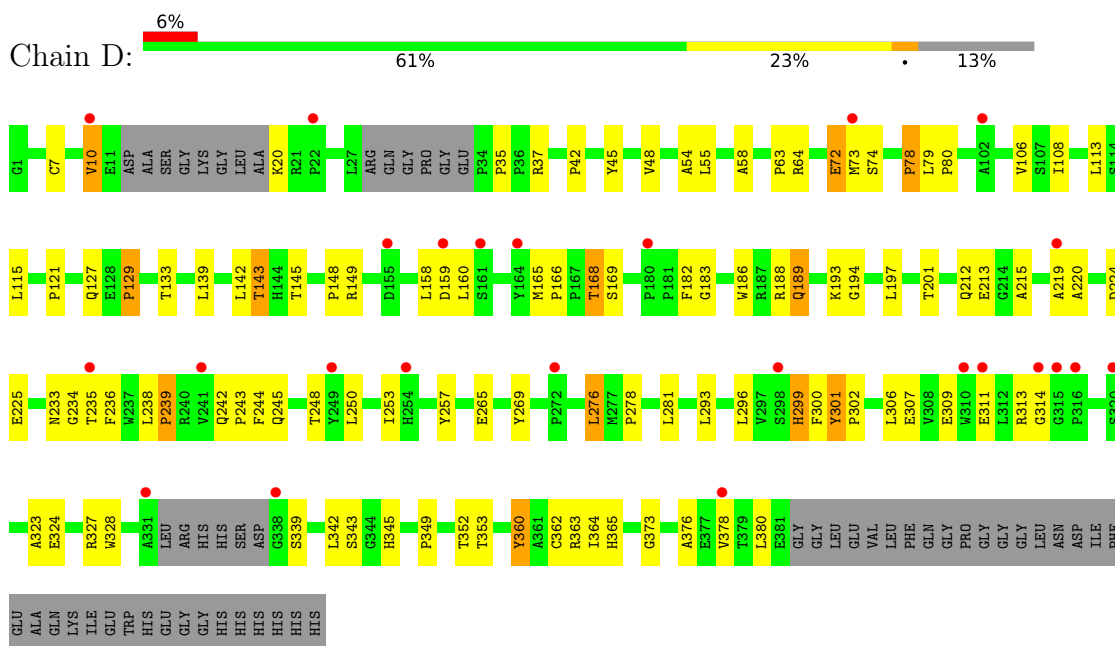
### 3 Residue-property plots [i](#)

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: HLA class I histocompatibility antigen, B alpha chain

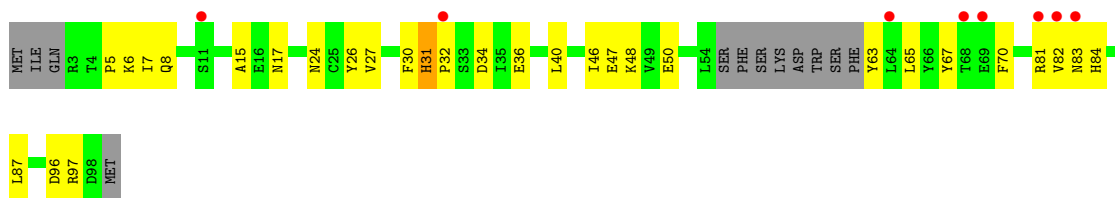


- Molecule 2: Tapasin



- Molecule 3: Beta-2-microglobulin





## 4 Data and refinement statistics i

Property	Value	Source
Space group	P 61	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	116.87Å 116.87Å 131.14Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	80.12 – 3.10 80.12 – 3.10	Depositor EDS
% Data completeness (in resolution range)	99.8 (80.12-3.10) 99.8 (80.12-3.10)	Depositor EDS
$R_{merge}$	0.34	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.23 (at 3.13Å)	Xtrriage
Refinement program	PHENIX 1.19.2_4158	Depositor
R, $R_{free}$	0.286 , 0.309 0.295 , 0.304	Depositor DCC
$R_{free}$ test set	920 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	80.3	Xtrriage
Anisotropy	0.337	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 449.5	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.31$ , $\langle L^2 \rangle = 0.14$	Xtrriage
Estimated twinning fraction	0.307 for h,-h-k,-l	Xtrriage
$F_o, F_c$ correlation	0.78	EDS
Total number of atoms	5630	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	71.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.53% 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](#)

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	0.25	0/2238	0.53	0/3047
2	D	0.25	0/2808	0.51	0/3856
3	B	0.24	0/750	0.52	0/1015
All	All	0.25	0/5796	0.52	0/7918

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 [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	2182	0	2021	72	0
2	D	2716	0	2661	57	0
3	B	732	0	702	20	0
All	All	5630	0	5384	141	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 13.

The worst 5 of 141 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:261:VAL:O	1:A:269:PRO:HA	1.73	0.89

*Continued on next page...*



Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:D:300:PHE:HZ	2:D:342:LEU:HB2	1.47	0.79
1:A:262:GLN:HG2	1:A:269:PRO:HB3	1.68	0.74
2:D:276:LEU:HB3	2:D:376:ALA:HB1	1.76	0.68
1:A:21:ARG:NH2	3:B:34:ASP:OD2	2.26	0.66

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	272/274 (99%)	238 (88%)	29 (11%)	5 (2%)	8	34
2	D	353/416 (85%)	298 (84%)	44 (12%)	11 (3%)	4	23
3	B	84/100 (84%)	75 (89%)	8 (10%)	1 (1%)	13	44
All	All	709/790 (90%)	611 (86%)	81 (11%)	17 (2%)	6	27

5 of 17 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	D	169	SER
2	D	189	GLN
2	D	302	PRO
1	A	220	ASP
2	D	121	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	222/235 (94%)	209 (94%)	13 (6%)	19	50
2	D	281/328 (86%)	265 (94%)	16 (6%)	20	52
3	B	83/95 (87%)	79 (95%)	4 (5%)	25	58
All	All	586/658 (89%)	553 (94%)	33 (6%)	21	52

5 of 33 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	360	TYR
3	B	31	HIS
3	B	70	PHE
1	A	268	LYS
1	A	263	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. There are no such sidechains identified.

### 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

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	274/274 (100%)	0.02	13 (4%) 31 15	38, 70, 93, 125	9 (3%)
2	D	361/416 (86%)	0.12	25 (6%) 16 7	42, 71, 105, 150	6 (1%)
3	B	88/100 (88%)	-0.00	8 (9%) 9 3	44, 67, 87, 97	1 (1%)
All	All	723/790 (91%)	0.07	46 (6%) 19 8	38, 70, 98, 150	16 (2%)

The worst 5 of 46 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	338	GLY	9.5
1	A	114	ASP	7.9
2	D	378	VAL	7.9
2	D	315	GLY	6.6
2	D	241	VAL	6.3

### 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.