



# wwPDB X-ray Structure Validation Summary Report ⓘ

May 16, 2020 – 06:46 pm BST

PDB ID : 5KSU  
Title : Crystal structure of HLA-DQ2.5-CLIP1 at 2.73 resolution  
Authors : Nguyen, T.-B.; Jayaraman, P.; Bergseng, E.; Madhusudhan, M.S.; Kim, C.-Y.;  
Sollid, L.M.  
Deposited on : 2016-07-10  
Resolution : 2.73 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 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	:	2.11
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)	:	2.11

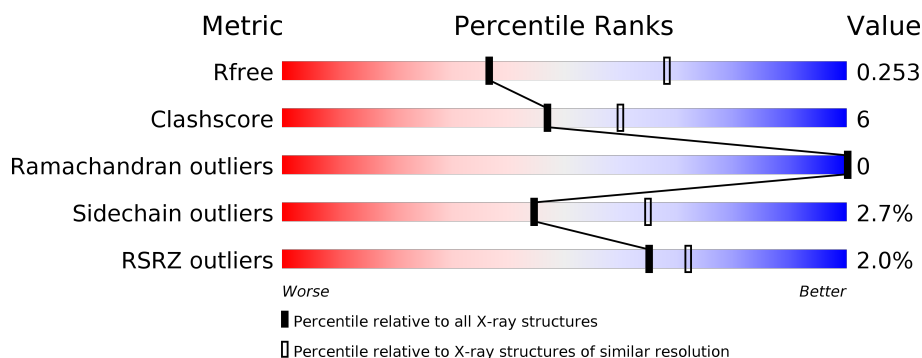
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.73 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	1271 (2.76-2.72)
Clashscore	141614	1322 (2.76-2.72)
Ramachandran outliers	138981	1297 (2.76-2.72)
Sidechain outliers	138945	1298 (2.76-2.72)
RSRZ outliers	127900	1243 (2.76-2.72)

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 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	199	<div> <div>2%</div> <div> <div></div> <div>78%</div> <div>14%</div> <div>9%</div> </div> </div>
1	D	199	<div> <div>3%</div> <div> <div></div> <div>73%</div> <div>18%</div> <div>9%</div> </div> </div>
2	B	204	<div> <div>0%</div> <div> <div></div> <div>75%</div> <div>14%</div> <div>12%</div> </div> </div>
2	E	204	<div> <div>0%</div> <div> <div></div> <div>74%</div> <div>13%</div> <div>12%</div> </div> </div>
3	C	15	<div> <div>7%</div> <div> <div></div> <div>67%</div> <div>27%</div> <div>7%</div> </div> </div>
3	F	15	<div> <div>7%</div> <div> <div></div> <div>60%</div> <div>27%</div> <div>7%</div> </div> </div>

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 6139 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 II histocompatibility antigen, DQ alpha 1 chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	182	Total	C	N	O	S	0	0	0
			1435	924	232	276	3			
1	D	182	Total	C	N	O	S	0	0	0
			1441	927	235	276	3			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	192	VAL	-	expression tag	UNP P01909
A	193	ASP	-	expression tag	UNP P01909
A	194	ILE	-	expression tag	UNP P01909
A	195	GLU	-	expression tag	UNP P01909
A	196	GLY	-	expression tag	UNP P01909
A	197	ARG	-	expression tag	UNP P01909
D	192	VAL	-	expression tag	UNP P01909
D	193	ASP	-	expression tag	UNP P01909
D	194	ILE	-	expression tag	UNP P01909
D	195	GLU	-	expression tag	UNP P01909
D	196	GLY	-	expression tag	UNP P01909
D	197	ARG	-	expression tag	UNP P01909

- Molecule 2 is a protein called MHC class II HLA-DQ-beta-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	180	Total	C	N	O	S	0	0	0
			1465	927	260	271	7			
2	E	179	Total	C	N	O	S	0	0	0
			1466	927	263	269	7			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	199	VAL	-	expression tag	UNP O19712
B	200	ASP	-	expression tag	UNP O19712
B	201	ILE	-	expression tag	UNP O19712
B	202	GLU	-	expression tag	UNP O19712
B	203	GLY	-	expression tag	UNP O19712
B	204	ARG	-	expression tag	UNP O19712
E	199	VAL	-	expression tag	UNP O19712
E	200	ASP	-	expression tag	UNP O19712
E	201	ILE	-	expression tag	UNP O19712
E	202	GLU	-	expression tag	UNP O19712
E	203	GLY	-	expression tag	UNP O19712
E	204	ARG	-	expression tag	UNP O19712

- Molecule 3 is a protein called HLA class II histocompatibility antigen gamma chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	14	Total	C	N	O	S	0	0	0
			108	69	19	17	3			
3	F	14	Total	C	N	O	S	0	0	0
			108	69	19	17	3			

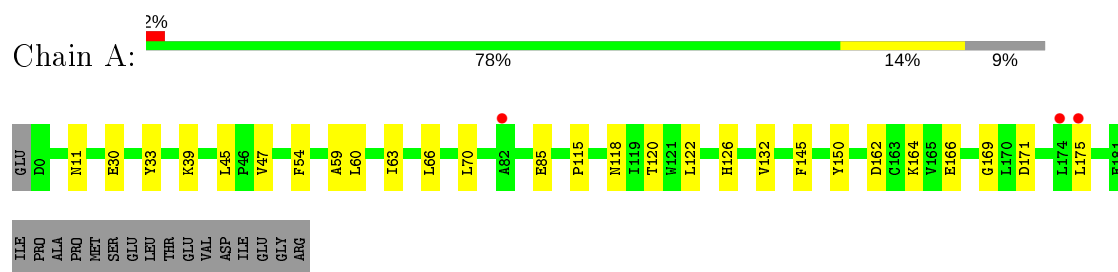
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	32	Total	O	0	0
			32	32		
4	B	33	Total	O	0	0
			33	33		
4	C	2	Total	O	0	0
			2	2		
4	D	21	Total	O	0	0
			21	21		
4	E	23	Total	O	0	0
			23	23		
4	F	5	Total	O	0	0
			5	5		

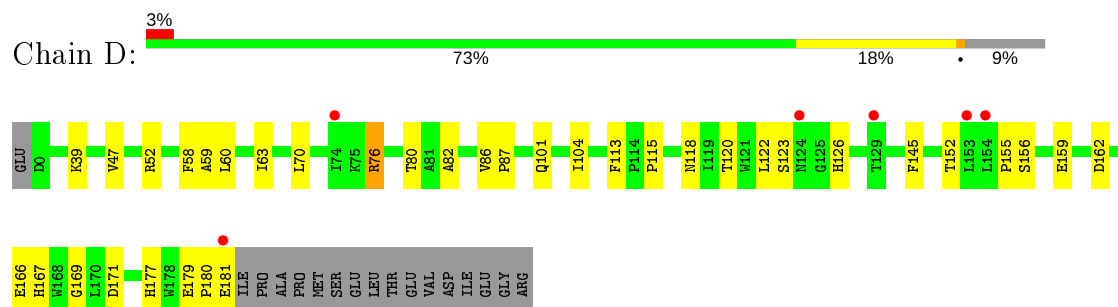
### 3 Residue-property plots [i](#)

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 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.

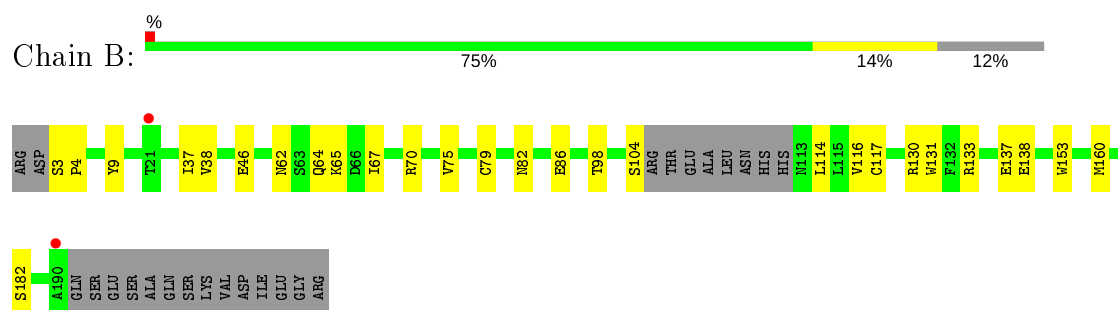
- Molecule 1: HLA class II histocompatibility antigen, DQ alpha 1 chain



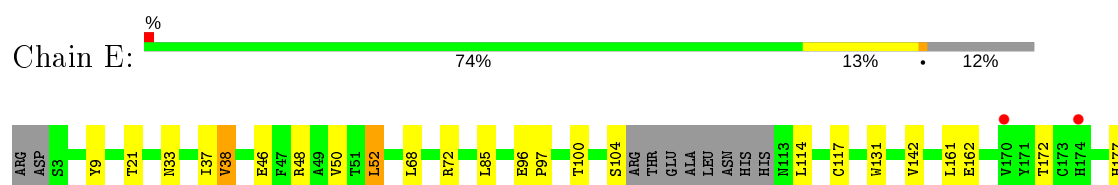
- Molecule 1: HLA class II histocompatibility antigen, DQ alpha 1 chain



- Molecule 2: MHC class II HLA-DQ-beta-1

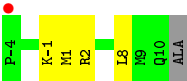


- Molecule 2: MHC class II HLA-DQ-beta-1





● Molecule 3: HLA class II histocompatibility antigen gamma chain



● Molecule 3: HLA class II histocompatibility antigen gamma chain



## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	128.86Å 69.21Å 146.69Å 90.00° 110.29° 90.00°	Depositor
Resolution (Å)	39.16 – 2.73 39.16 – 2.73	Depositor EDS
% Data completeness (in resolution range)	93.8 (39.16-2.73) 93.8 (39.16-2.73)	Depositor EDS
$R_{merge}$	0.10	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.45 (at 2.73Å)	Xtriage
Refinement program	PHENIX dev_1137	Depositor
R, $R_{free}$	0.190 , 0.249 0.194 , 0.253	Depositor DCC
$R_{free}$ test set	1525 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	57.1	Xtriage
Anisotropy	0.825	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.32 , 51.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	6139	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	44.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.97% 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	0.46	0/1477	0.64	0/2022
1	D	0.48	0/1483	0.64	0/2029
2	B	0.52	0/1498	0.64	0/2036
2	E	0.47	0/1499	0.65	0/2036
3	C	0.60	0/109	0.96	0/144
3	F	0.53	0/109	0.83	0/144
All	All	0.49	0/6175	0.65	0/8411

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	1435	0	1365	18	0
1	D	1441	0	1376	28	0
2	B	1465	0	1427	19	0
2	E	1466	0	1435	13	0
3	C	108	0	125	2	0
3	F	108	0	125	4	0
4	A	32	0	0	0	0
4	B	33	0	0	0	0
4	C	2	0	0	0	0
4	D	21	0	0	1	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	E	23	0	0	0	0
4	F	5	0	0	0	0
All	All	6139	0	5853	74	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 6.

The worst 5 of 74 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:D:180:PRO:O	1:D:181:GLU:HB2	1.61	1.00
1:D:180:PRO:O	1:D:181:GLU:CB	2.20	0.86
1:D:118:ASN:HB2	1:D:166:GLU:HB2	1.75	0.69
2:E:85:LEU:HD12	3:F:1:MET:SD	2.31	0.69
2:B:70:ARG:HH11	2:B:70:ARG:HG3	1.58	0.68

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	180/199 (90%)	175 (97%)	5 (3%)	0	100	100
1	D	180/199 (90%)	174 (97%)	6 (3%)	0	100	100
2	B	176/204 (86%)	168 (96%)	8 (4%)	0	100	100
2	E	175/204 (86%)	165 (94%)	10 (6%)	0	100	100
3	C	12/15 (80%)	12 (100%)	0	0	100	100
3	F	12/15 (80%)	12 (100%)	0	0	100	100
All	All	735/836 (88%)	706 (96%)	29 (4%)	0	100	100

There are no Ramachandran outliers to report.

### 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	A	162/181 (90%)	161 (99%)	1 (1%)	86	91
1	D	163/181 (90%)	161 (99%)	2 (1%)	71	83
2	B	162/186 (87%)	159 (98%)	3 (2%)	57	74
2	E	163/186 (88%)	155 (95%)	8 (5%)	25	43
3	C	13/13 (100%)	12 (92%)	1 (8%)	13	23
3	F	13/13 (100%)	10 (77%)	3 (23%)	1	1
All	All	676/760 (89%)	658 (97%)	18 (3%)	44	65

5 of 18 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	38	VAL
2	E	50	VAL
2	E	182	SER
1	D	120	THR
2	E	21	THR

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

### 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 [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	A	182/199 (91%)	0.03	3 (1%) 72 78	16, 35, 69, 128	0
1	D	182/199 (91%)	0.12	6 (3%) 46 52	20, 48, 92, 138	0
2	B	180/204 (88%)	0.00	2 (1%) 80 85	19, 39, 72, 97	0
2	E	179/204 (87%)	-0.02	2 (1%) 80 85	22, 40, 72, 103	0
3	C	14/15 (93%)	0.09	1 (7%) 16 17	29, 38, 72, 106	0
3	F	14/15 (93%)	0.23	1 (7%) 16 17	27, 38, 59, 77	0
All	All	751/836 (89%)	0.04	15 (1%) 65 72	16, 40, 79, 138	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	174	LEU	3.5
1	D	181	GLU	2.7
3	F	-4	PRO	2.7
2	B	190	ALA	2.6
1	A	175	LEU	2.3

### 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 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.