wwPDB X-ray Structure Validation Summary Report (i)

Dec 17, 2023 – 10:51 pm GMT

PDB ID	:	2YPL
Title	:	Structural features underlying T-cell receptor sensitivity to concealed MHC class I micropolymorphisms
Authors	:	Stewart-Jones, G.B.; Simpson, P.; van der Merwe, P.A.; Easterbrook, P.; McMichael, A.J.; Rowland-Jones, S.L.; Jones, E.Y.; Gillespie, G.M.
Deposited on	:	2012-10-30
Resolution	:	2.40 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*
A user guide is available at
https://www.wwpdb.org/validation/2017/XrayValidationReportHelp
with specific help available everywhere you see the (i) 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 (1)) were used in the production of this report:

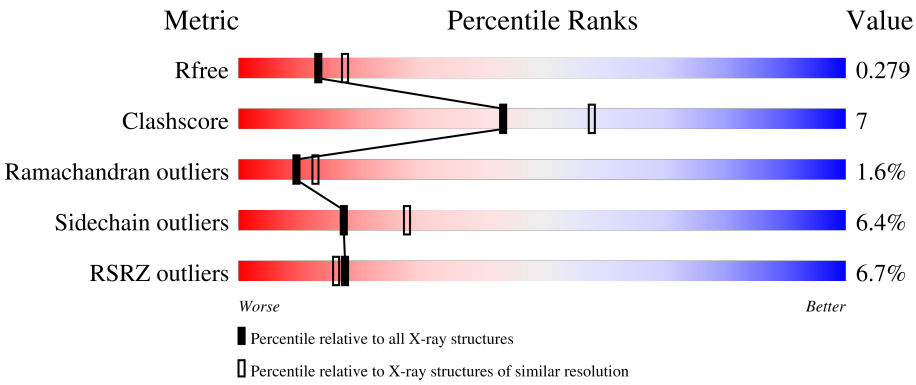
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.36
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

1 Overall quality at a glance (i)

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 2.40 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}	130704	3907 (2.40-2.40)
Clashscore	141614	4398 (2.40-2.40)
Ramachandran outliers	138981	4318 (2.40-2.40)
Sidechain outliers	138945	4319 (2.40-2.40)
RSRZ outliers	127900	3811 (2.40-2.40)

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 >=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 <=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	%, 86%, 13%
2	B	99	5%, 81%, 18%, •
3	C	11	18%, 64%, 36%
4	D	199	13%, 77%, 17%, 6%, •
5	E	238	8%, 83%, 14%, •

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 6796 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-57 ALPHA CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	274	2228	1394	404	421	9	0	0	0

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	116	TYR	SER	conflict	UNP P18465
A	143	ILE	THR	conflict	UNP P18465

- Molecule 2 is a protein called BETA-2-MICROGLOBULIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
2	B	99	829	528	140	158	3	0	0	0

- Molecule 3 is a protein called KF11 P24 GAG PEPTIDE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
3	C	11	89	61	12	15	1	0	0	0

- Molecule 4 is a protein called AGA T-CELL RECEPTOR ALPHA CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
4	D	199	1565	981	254	321	9	0	0	0

- Molecule 5 is a protein called AGA T-CELL RECEPTOR BETA CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
5	E	238	1890	1191	331	361	7	0	0	0

- Molecule 6 is water.

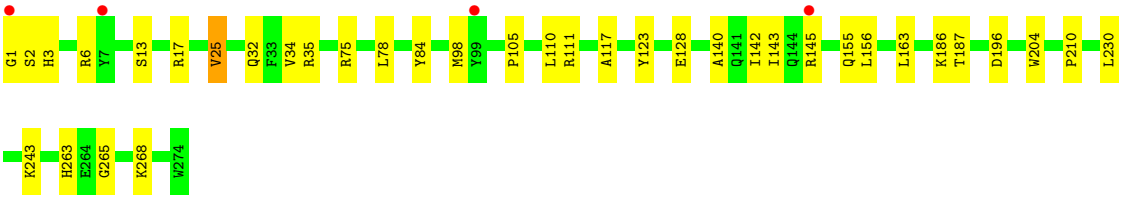
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
6	A	77	77	77	0	0
6	B	16	16	16	0	0
6	C	2	2	2	0	0
6	D	51	51	51	0	0
6	E	49	49	49	0	0

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.

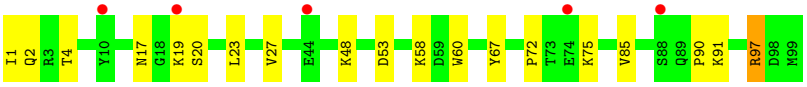
• Molecule 1: HLA CLASS I HISTOCOMPATIBILITY ANTIGEN, B-57 ALPHA CHAIN

Chain A:



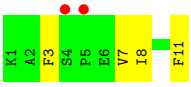
• Molecule 2: BETA-2-MICROGLOBULIN

Chain B:



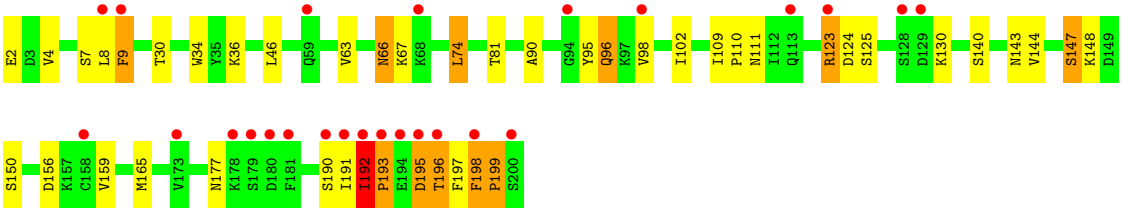
• Molecule 3: KF11 P24 GAG PEPTIDE

Chain C:



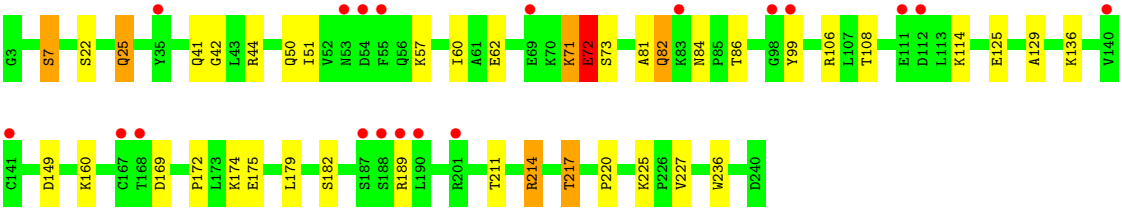
• Molecule 4: AGA T-CELL RECEPTOR ALPHA CHAIN

Chain D:



• Molecule 5: AGA T-CELL RECEPTOR BETA CHAIN

Chain E:


G3
S7
S22
Q25
Y35
Q41
G42
L43
R44
Q50
I51
V52
N53
D54
F55
Q56
K57
I60
A61
E62
E69
K70
K71
E72
S73
A81
Q82
K83
N84
P85
T86
G98
Y99
R106
L107
T108
E111
D112
L113
K114
E125
A129
K136
V140
C141
D149
K160
C167
T168
D169
P172
L173
K174
E175
L179
S182
S187
S188
R189
L190
R201
T211
R214
T217
P220
K225
P226
V227
W236
D240

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	56.35Å 75.61Å 241.53Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.77 – 2.40 29.77 – 2.40	Depositor EDS
% Data completeness (in resolution range)	98.7 (29.77-2.40) 98.7 (29.77-2.40)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	4.35 (at 2.39Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
R, R_{free}	0.221 , 0.277 0.222 , 0.279	Depositor DCC
R_{free} test set	2054 reflections (5.02%)	wwPDB-VP
Wilson B-factor (Å^2)	53.1	Xtriage
Anisotropy	0.574	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 39.6	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.95	EDS
Total number of atoms	6796	wwPDB-VP
Average B, all atoms (Å^2)	66.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.85% of the height of the origin peak. No significant pseudotranslation is detected.*

[1]Intensities estimated from amplitudes.

[2]Theoretical values of $< |L| >$, $< L^2 >$ 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	0.41	0/2290	0.61	0/3112
2	B	0.36	0/852	0.53	0/1152
3	C	0.50	0/92	0.56	0/122
4	D	0.48	1/1595 (0.1%)	0.68	0/2159
5	E	0.37	0/1941	0.58	0/2637
All	All	0.41	1/6770 (0.0%)	0.61	0/9182

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
4	D	0	2

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	D	190	SER	CB-OG	-8.70	1.30	1.42

There are no bond angle outliers.

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	D	192	ILE	Peptide
4	D	8	LEU	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	2228	0	2102	26	0
2	B	829	0	794	7	0
3	C	89	0	92	5	0
4	D	1565	0	1497	39	0
5	E	1890	0	1813	24	0
6	A	77	0	0	0	0
6	B	16	0	0	1	0
6	C	2	0	0	0	0
6	D	51	0	0	7	0
6	E	49	0	0	3	0
All	All	6796	0	6298	87	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 7.

The worst 5 of 87 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)
4:D:192:ILE:HG23	4:D:193:PRO:HD2	1.26	1.16
1:A:155:GLN:HE22	3:C:7:VAL:H	1.10	0.98
4:D:123:ARG:HD3	4:D:123:ARG:H	1.37	0.86
4:D:7:SER:HB2	4:D:9:PHE:HB2	1.56	0.86
4:D:63:VAL:HG11	6:D:2017:HOH:O	1.75	0.84

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%)	263 (97%)	8 (3%)	1 (0%)	34	48
2	B	97/99 (98%)	95 (98%)	2 (2%)	0	100	100
3	C	9/11 (82%)	9 (100%)	0	0	100	100
4	D	197/199 (99%)	172 (87%)	17 (9%)	8 (4%)	3	2
5	E	236/238 (99%)	220 (93%)	12 (5%)	4 (2%)	9	11
All	All	811/821 (99%)	759 (94%)	39 (5%)	13 (2%)	9	13

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	9	PHE
4	D	193	PRO
4	D	199	PRO
4	D	192	ILE
5	E	72	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	228/228 (100%)	219 (96%)	9 (4%)	32	50
2	B	94/94 (100%)	85 (90%)	9 (10%)	8	12
3	C	10/10 (100%)	10 (100%)	0	100	100
4	D	181/181 (100%)	167 (92%)	14 (8%)	13	20
5	E	207/207 (100%)	193 (93%)	14 (7%)	16	25
All	All	720/720 (100%)	674 (94%)	46 (6%)	17	28

5 of 46 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	148	LYS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
5	E	50	GLN
4	D	156	ASP
5	E	7	SER
5	E	57	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
4	D	111	ASN
5	E	17	GLN
5	E	171	GLN
5	E	50	GLN
5	E	56	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)

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 (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, 95th 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^2)	Q<0.9
1	A	274/274 (100%)	-0.07	4 (1%)	73	72	37, 53, 82, 105	0
2	B	99/99 (100%)	0.28	5 (5%)	28	26	48, 72, 105, 110	0
3	C	11/11 (100%)	1.19	2 (18%)	1	1	46, 50, 61, 71	0
4	D	199/199 (100%)	0.72	25 (12%)	3	3	33, 63, 119, 173	0
5	E	238/238 (100%)	0.43	19 (7%)	12	11	42, 68, 112, 118	0
All	All	821/821 (100%)	0.33	55 (6%)	17	16	33, 61, 105, 173	0

The worst 5 of 55 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
4	D	193	PRO	13.1
4	D	194	GLU	9.3
4	D	129	ASP	5.9
4	D	192	ILE	5.7
1	A	1	GLY	4.9

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.