wwPDB X-ray Structure Validation Summary Report (i)

Feb 19, 2016 – 08:28 PM GMT

PDB ID	:	4WUU
Title	:	Structure of ESK1 in complex with HLA-A*0201/WT1
Authors	:	Ataie, N.J.; Ng, H.L.
Deposited on	:	2014-11-03
Resolution	:	3.05 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

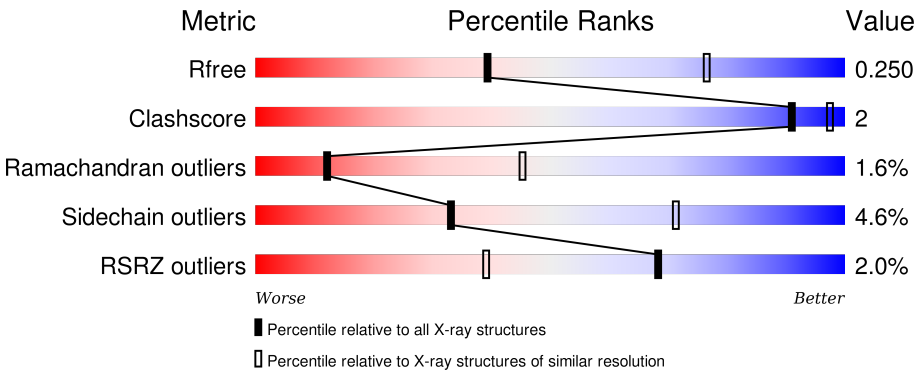
MolProbity	:	4.02b-467
Mogul	:	unknown
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026982
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20026982

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 3.05 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	1995 (3.08-3.00)
Clashscore	102246	2351 (3.08-3.00)
Ramachandran outliers	100387	2272 (3.08-3.00)
Sidechain outliers	100360	2275 (3.08-3.00)
RSRZ outliers	91569	2013 (3.08-3.00)

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 >=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 <=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	296	2% 85% 7% 7%
2	B	100	92% 5% • •
3	C	9	11% 89% 11%
4	D	216	% 90% 7% • •
5	E	223	3% 87% 9% • •

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 12529 atoms, of which 6109 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, A-2 alpha chain.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	275	4339	1403	2093	409	425	9	0	0	0

There are 20 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	0	MET	-	initiating methionine	UNP P01892
A	277	GLY	-	expression tag	UNP P01892
A	278	SER	-	expression tag	UNP P01892
A	279	GLY	-	expression tag	UNP P01892
A	280	GLY	-	expression tag	UNP P01892
A	281	GLY	-	expression tag	UNP P01892
A	282	LEU	-	expression tag	UNP P01892
A	283	ASN	-	expression tag	UNP P01892
A	284	ASP	-	expression tag	UNP P01892
A	285	ILE	-	expression tag	UNP P01892
A	286	PHE	-	expression tag	UNP P01892
A	287	GLU	-	expression tag	UNP P01892
A	288	ALA	-	expression tag	UNP P01892
A	289	GLN	-	expression tag	UNP P01892
A	290	LYS	-	expression tag	UNP P01892
A	291	ILE	-	expression tag	UNP P01892
A	292	GLY	-	expression tag	UNP P01892
A	293	TRP	-	expression tag	UNP P01892
A	294	HIS	-	expression tag	UNP P01892
A	295	GLU	-	expression tag	UNP P01892

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

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
2	B	99	1620	528	791	140	158	3	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

- Molecule 3 is a protein called ARG-MET-PHE-PRO-ASN-ALA-PRO-TYR-LEU.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
3	C	9	156	52	78	13	12	1	0	0	0

- Molecule 4 is a protein called ESK1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
4	D	212	3082	978	1514	266	320	4	0	0	0

- Molecule 5 is a protein called IMMUNOGLOBULIN HEAVY CHAIN.

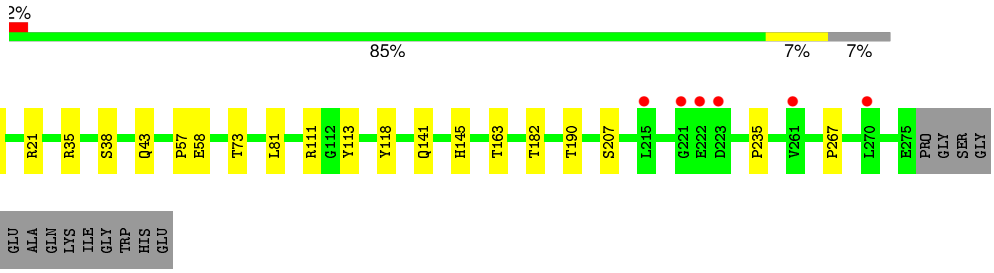
Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
5	E	223	3332	1078	1633	279	334	8	0	0	0

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 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.

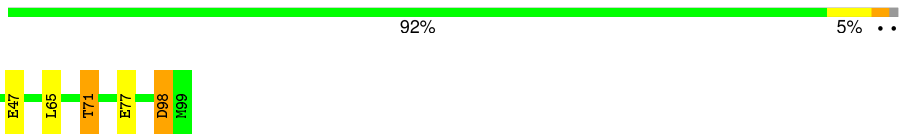
- Molecule 1: HLA class I histocompatibility antigen, A-2 alpha chain

Chain A:



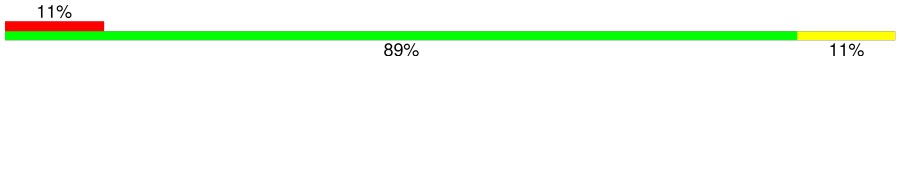
- Molecule 2: Beta-2-microglobulin

Chain B:



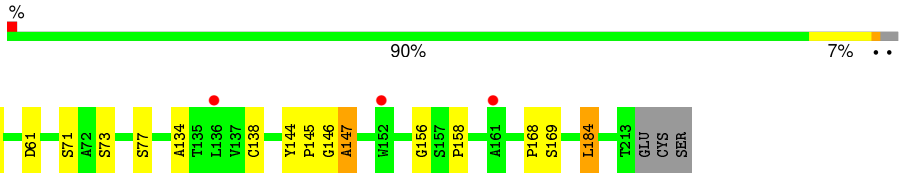
- Molecule 3: ARG-MET-PHE-PRO-ASN-ALA-PRO-TYR-LEU

Chain C:



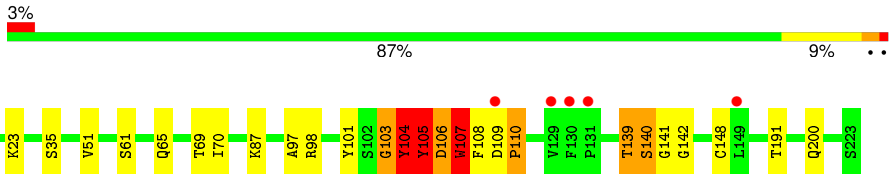
- Molecule 4: ESK1

Chain D:



- Molecule 5: IMMUNOGLOBULIN HEAVY CHAIN

Chain E:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	70.08Å 118.26Å 126.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.44 – 3.05 37.63 – 3.05	Depositor EDS
% Data completeness (in resolution range)	99.2 (33.44-3.05) 97.3 (37.63-3.05)	Depositor EDS
R_{merge}	0.17	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.19 (at 3.06Å)	Xtriage
Refinement program	PHENIX (1.10.1_2155: ???)	Depositor
R, R_{free}	0.198 , 0.254 0.198 , 0.250	Depositor DCC
R_{free} test set	1944 reflections (9.74%)	DCC
Wilson B-factor (Å^2)	77.8	Xtriage
Anisotropy	0.086	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 14.4	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.47$, $< L^2 > = 0.30$	Xtriage
Outliers	0 of 19974 reflections	Xtriage
F_o,F_c correlation	0.91	EDS
Total number of atoms	12529	wwPDB-VP
Average B, all atoms (Å^2)	82.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.64% 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, 0.375 respectively for untwinned datasets, and 0.333, 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	0.27	0/2311	0.46	0/3137
2	B	0.27	0/852	0.46	0/1152
3	C	0.33	0/81	0.58	0/108
4	D	0.28	0/1607	0.48	0/2197
5	E	0.28	0/1749	0.54	0/2384
All	All	0.28	0/6600	0.49	0/8978

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
5	E	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
5	E	35	SER	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	2246	2093	2096	9	0
2	B	829	791	794	3	1
3	C	78	78	78	1	0
4	D	1568	1514	1514	4	1
5	E	1699	1633	1633	13	0
All	All	6420	6109	6115	28	1

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 2.

The worst 5 of 28 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)
5:E:141:GLY:N	5:E:142:GLY:HA2	2.14	0.62
5:E:105:TYR:HB3	5:E:107:TRP:CZ3	2.39	0.58
5:E:103:GLY:HA3	5:E:104:TYR:HB2	1.88	0.55
1:A:182:THR:O	1:A:182:THR:HG23	2.07	0.55
5:E:109:ASP:HB3	5:E:110:PRO:HD3	1.89	0.54

All (1) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
2:B:77:GLU:OE2	4:D:77:SER:OG[2_454]	2.06	0.14

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	273/296 (92%)	254 (93%)	17 (6%)	2 (1%)	26 67
2	B	97/100 (97%)	94 (97%)	3 (3%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	C	7/9 (78%)	7 (100%)	0	0	100	100
4	D	210/216 (97%)	184 (88%)	21 (10%)	5 (2%)	7	33
5	E	221/223 (99%)	202 (91%)	13 (6%)	6 (3%)	6	29
All	All	808/844 (96%)	741 (92%)	54 (7%)	13 (2%)	12	45

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	D	147	ALA
5	E	140	SER
1	A	207	SER
5	E	105	TYR
4	D	169	SER

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	231/246 (94%)	225 (97%)	6 (3%)	54	84
2	B	94/95 (99%)	89 (95%)	5 (5%)	28	65
3	C	8/8 (100%)	8 (100%)	0	100	100
4	D	176/180 (98%)	169 (96%)	7 (4%)	38	76
5	E	191/191 (100%)	177 (93%)	14 (7%)	17	51
All	All	700/720 (97%)	668 (95%)	32 (5%)	33	71

5 of 32 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
4	D	71	SER
4	D	184	LEU
5	E	148	CYS
4	D	138	CYS
5	E	19	ARG

Some 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 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, 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	275/296 (92%)	-0.00	6 (2%) 65 35	40, 71, 131, 162	0
2	B	99/100 (99%)	-0.23	0 100 100	38, 56, 85, 126	0
3	C	9/9 (100%)	0.63	1 (11%) 7 2	47, 61, 79, 88	0
4	D	212/216 (98%)	-0.06	3 (1%) 78 51	35, 76, 136, 167	0
5	E	223/223 (100%)	0.06	6 (2%) 58 28	32, 68, 119, 174	0
All	All	818/844 (96%)	-0.02	16 (1%) 68 39	32, 68, 131, 174	0

The worst 5 of 16 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
5	E	130	PHE	4.0
1	A	223	ASP	3.7
4	D	152	TRP	3.4
5	E	129	VAL	3.0
5	E	1	GLN	2.8

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.