wwPDB X-ray Structure Validation Summary Report (i)

May 22, 2020 – 01:37 am BST

PDB ID	:	3WS3
Title	:	Crystal Structure of H-2D in complex with an insulin derived peptide
Authors	:	Kumar, P.R.; Mukherjee, G.; Samanta, D.; DiLorenzo, T.P.; Almo, S.C.; Immune Function Network; New York Structural Genomics Research Consortium (NYSGRC)
Deposited on	:	2014-02-28
Resolution	:	2.33 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 following versions of software and data (see references (1)) were used in the production of this report:

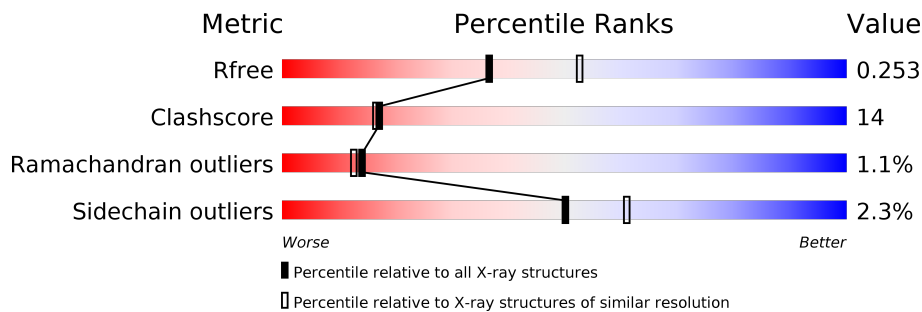|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 2.33 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	Percentile Ranks	Value
Rfree		0.253
Clashscore		14
Ramachandran outliers		1.1%
Sidechain outliers		2.3%

Worse　　　　　　　　　　　　　　　　　　　　　*Better*

▮ Percentile relative to all X-ray structures

▯ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	2096 (2.36-2.32)
Clashscore	141614	2193 (2.36-2.32)
Ramachandran outliers	138981	2159 (2.36-2.32)
Sidechain outliers	138945	2160 (2.36-2.32)

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 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%

Mol	Chain	Length	Quality of chain
1	A	273	63%　　31%　　• • •
1	C	273	70%　　27%　　• •
2	B	100	73%　　27%
2	D	100	70%　　29%　　•
3	E	9	78%　　22%
3	F	9	56%　　44%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 6355 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 H-2 class I histocompatibility antigen, D-B alpha chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	266	Total 2205	C 1394	N 388	O 414	S 9	0	1	0
1	C	271	Total 2233	C 1410	N 395	O 419	S 9	0	0	0

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	100	Total 829	C 529	N 139	O 153	S 8	0	0	0
2	D	100	Total 829	C 529	N 139	O 153	S 8	0	0	0

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	20	MET	-	EXPRESSION TAG	UNP P01887
B	105	ASP	ALA	SEE REMARK 999	UNP P01887
D	20	MET	-	EXPRESSION TAG	UNP P01887
D	105	ASP	ALA	SEE REMARK 999	UNP P01887

- Molecule 3 is a protein called Insulin derived 9-mer peptide.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	9	Total 77	C 49	N 11	O 16	S 1	0	0	0
3	F	9	Total 77	C 49	N 11	O 16	S 1	0	0	0

- Molecule 4 is water.

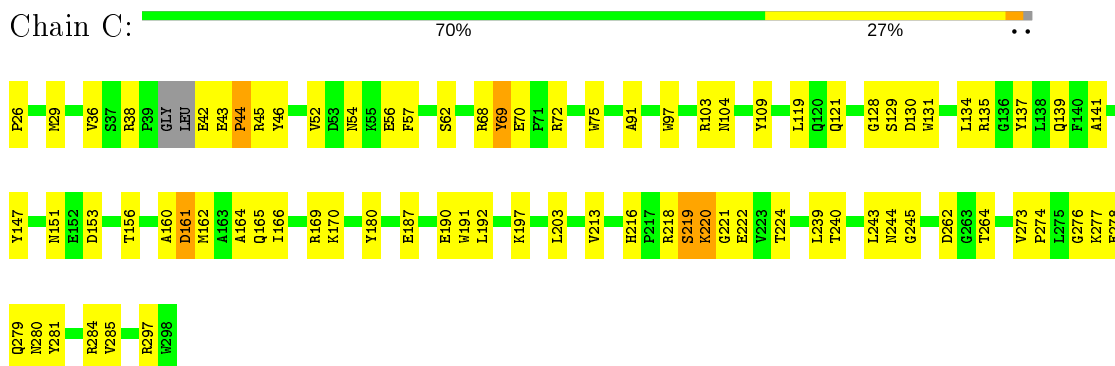| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	36	Total 36	O 36	0	0
4	B	11	Total 11	O 11	0	0
4	C	33	Total 33	O 33	0	0
4	D	23	Total 23	O 23	0	0
4	E	2	Total 2	O 2	0	0

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 the various outlier classes displayed in the second graphic. The second graphic shows the sequence view annotated by issues in geometry. 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. 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.
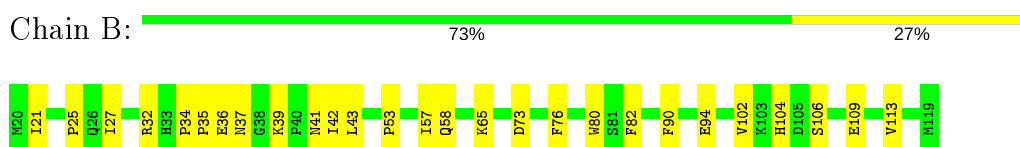
- Molecule 1: H-2 class I histocompatibility antigen, D-B alpha chain
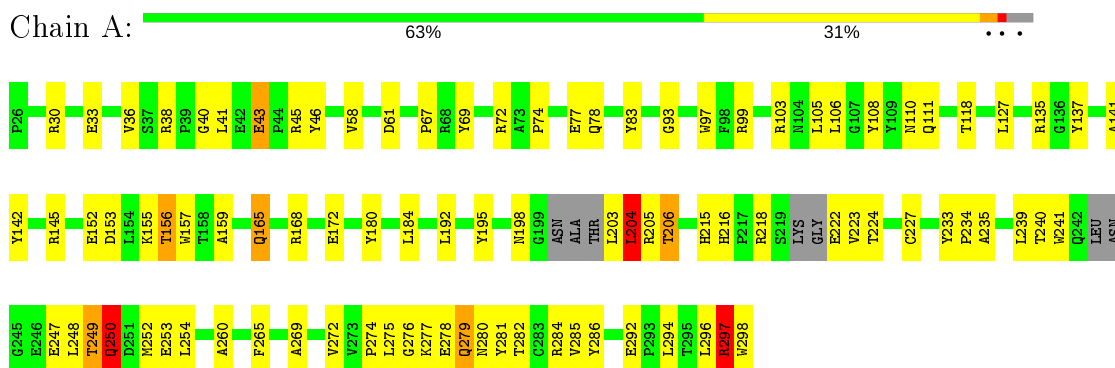
Chain A:

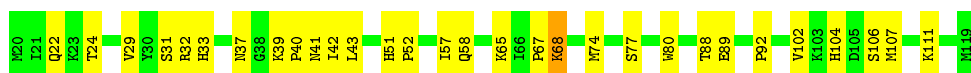- Molecule 1: H-2 class I histocompatibility antigen, D-B alpha chain

Chain C:

- Molecule 2: Beta-2-microglobulin

Chain B:

- Molecule 2: Beta-2-microglobulin

Chain D:

M20 I21 Q22 K23 T24 V29 Y30 S31 R32 H33 N37 G38 K39 P40 N41 I42 L43 H51 P52 I57 Q58 K65 I66 P67 K68 M74 S77 W80 T88 E89 P92 V102 K103 H104 D105 S106 M107 K111 M119

• Molecule 3: Insulin derived 9-mer peptide

Chain E:

78%　　　22%

Y1 Q2 G8 L9

• Molecule 3: Insulin derived 9-mer peptide

Chain F:

56%　　　44%

Y1 N5 G8 L9
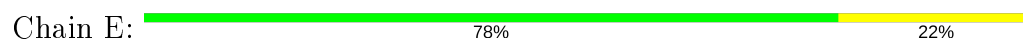
4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	87.22Å 101.13Å 117.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	46.45 – 2.33 46.45 – 1.93	Depositor EDS
% Data completeness (in resolution range)	74.6 (46.45-2.33) 88.0 (46.45-1.93)	Depositor EDS
R_{merge}	0.15	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.19 (at 1.94Å)	Xtriage
Refinement program	PHENIX 1.8.4_1496	Depositor
R, R_{free}	0.217 , 0.294 0.247 , 0.253	Depositor DCC
R_{free} test set	3381 reflections (4.89%)	wwPDB-VP
Wilson B-factor (Å^2)	25.1	Xtriage
Anisotropy	0.519	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.32 , 48.6	EDS
L-test for twinning[2]	$< |L| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	6355	wwPDB-VP
Average B, all atoms (Å^2)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 32.65 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 9.0861e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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	0.55	1/2268 (0.0%)	0.77	4/3073 (0.1%)
1	C	0.49	0/2298	0.66	1/3117 (0.0%)
2	B	0.44	0/855	0.60	0/1158
2	D	0.49	0/855	0.66	0/1158
3	E	0.44	0/78	0.46	0/103
3	F	0.40	0/78	0.58	0/103
All	All	0.50	1/6432 (0.0%)	0.69	5/8712 (0.1%)

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
1	A	0	3

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	297	ARG	CG-CD	-5.42	1.38	1.51

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	204	LEU	CB-CG-CD1	8.15	124.86	111.00
1	A	204	LEU	CA-CB-CG	7.60	132.78	115.30
1	A	297	ARG	NE-CZ-NH1	-7.28	116.66	120.30
1	C	130	ASP	N-CA-C	6.04	127.29	111.00
1	A	206	THR	N-CA-C	5.14	124.88	111.00

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	203	LEU	Peptide
1	A	204	LEU	Peptide
1	A	250	GLN	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	2205	0	2069	85	0
1	C	2233	0	2103	54	0
2	B	829	0	802	16	1
2	D	829	0	802	19	0
3	E	77	0	70	2	0
3	F	77	0	70	5	0
4	A	36	0	0	5	0
4	B	11	0	0	1	0
4	C	33	0	0	4	1
4	D	23	0	0	2	0
4	E	2	0	0	0	0
All	All	6355	0	5916	169	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 14.

The worst 5 of 169 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:145:ARG:HD2	2:B:21:ILE:HD13	1.55	0.87
1:C:129:SER:HA	1:C:131:TRP:H	1.42	0.84
1:C:103:ARG:NH2	4:C:317:HOH:O	2.11	0.80
1:C:68:ARG:NH1	1:C:70:GLU:OE2	2.15	0.79
1:A:279:GLN:H	1:A:297:ARG:NH2	1.83	0.76

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:94:GLU:OE1	4:C:317:HOH:O[2_555]	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	259/273 (95%)	239 (92%)	16 (6%)	4 (2%)	10	7
1	C	267/273 (98%)	250 (94%)	13 (5%)	4 (2%)	10	7
2	B	98/100 (98%)	95 (97%)	3 (3%)	0	100	100
2	D	98/100 (98%)	95 (97%)	3 (3%)	0	100	100
3	E	7/9 (78%)	6 (86%)	1 (14%)	0	100	100
3	F	7/9 (78%)	6 (86%)	1 (14%)	0	100	100
All	All	736/764 (96%)	691 (94%)	37 (5%)	8 (1%)	14	13

5 of 8 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	206	THR
1	A	249	THR
1	C	220	LYS
1	A	279	GLN
1	C	44	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	228/232 (98%)	220 (96%)	8 (4%)	36	45
1	C	231/232 (100%)	227 (98%)	4 (2%)	60	72
2	B	95/95 (100%)	94 (99%)	1 (1%)	73	83
2	D	95/95 (100%)	93 (98%)	2 (2%)	53	65
3	E	8/8 (100%)	8 (100%)	0	100	100
3	F	8/8 (100%)	8 (100%)	0	100	100
All	All	665/670 (99%)	650 (98%)	15 (2%)	50	61

5 of 15 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	254	LEU
1	A	297	ARG
1	C	190	GLU
1	A	250	GLN
1	C	139	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	117	HIS
2	D	22	GLN
1	C	165	GLN
1	A	280	ASN
1	C	121	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 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)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands (i)

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers (i)

Unable to reproduce the depositors R factor - this section is therefore empty.