W O R L D W I D E
wwPDB
PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Jan 15, 2024 – 02:00 pm GMT

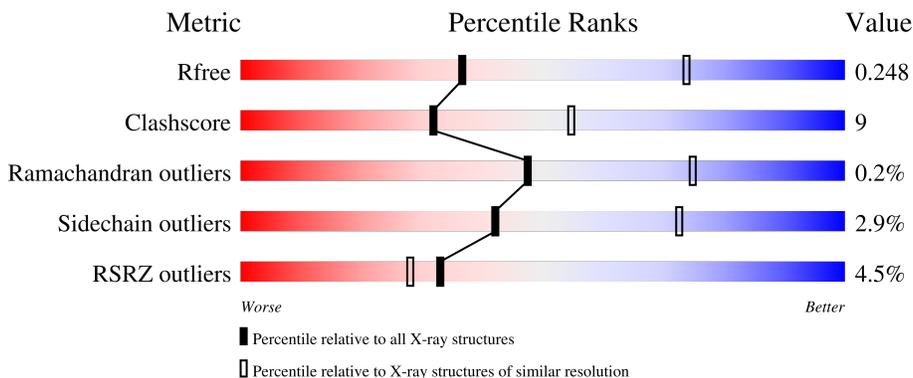|		
PDB ID	:	6YWC
Title	:	De novo designed protein 4E1H_95 in complex with 101F antibody
Authors	:	Yang, C.; Sesterhenn, F.; Pojer, F.; Correia, B.E.
Deposited on	:	2020-04-29
Resolution	:	2.85 Å(reported)

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 2.85 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		0.248
Clashscore		9
Ramachandran outliers		0.2%
Sidechain outliers		2.9%
RSRZ outliers		4.5%

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	3168 (2.90-2.82)
Clashscore	141614	3438 (2.90-2.82)
Ramachandran outliers	138981	3348 (2.90-2.82)
Sidechain outliers	138945	3351 (2.90-2.82)
RSRZ outliers	127900	3103 (2.90-2.82)

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	245	77% 13% 10%
1	D	245	5% 79% 11% 10%
2	B	237	80% 11% · 8%
2	E	237	8% 65% 25% · 8%
3	C	72	7% 47% 28% · · 19%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	72	

11%

60% 18% ·· 19%

2 Entry composition ⓘ

There are 4 unique types of molecules in this entry. The entry contains 7650 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 Antibody 101F, Heavy Chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	221	Total 1668	C 1061	N 274	O 328	S 5	0	0	0
1	D	221	Total 1668	C 1061	N 274	O 328	S 5	0	0	0

- Molecule 2 is a protein called Antibody 101F, light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	218	Total 1686	C 1060	N 284	O 337	S 5	3	0	0
2	E	218	Total 1681	C 1057	N 281	O 337	S 6	4	0	0

- Molecule 3 is a protein called De novo design 4E1H_95.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	58	Total 466	C 296	N 79	O 89	S 2	0	0	0
3	F	58	Total 466	C 296	N 79	O 89	S 2	0	0	0

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	6	Total 6	O 6	0	0
4	B	3	Total 3	O 3	0	0
4	C	1	Total 1	O 1	0	0
4	D	3	Total 3	O 3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	E	2	Total 2	O 2	0	0

3 Residue-property plots ⓘ

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: Antibody 101F, Heavy Chain

Chain A:

- Molecule 1: Antibody 101F, Heavy Chain

Chain D:

- Molecule 2: Antibody 101F, light chain

Chain B:

- Molecule 2: Antibody 101F, light chain

Chain E:

Q104 I105 I106 E107 G114 E119 I120 K121 R122 T123 P127 S128 V129 L150 F153 Y154 P155 R156 K159 V164 Q169 N172 S173 Q174 Y187 T194 L195 A198 D199 Y200 E201 K202 H203 K204 V205 V210 T211 H212 L215 S216 S217 P218 S222 F223

N224 R225 G226 E227 C228 GLY THR LYS HIS HIS HIS HIS HIS HIS HIS

• Molecule 3: De novo design 4E1H_95

Chain C:

7%
47%   28%   19%

MET K2 Y3 S9 R12 G19 I20 Q21 L22 P23 A26 L27 R32 E33 Y34 V35 L38 R39 E40 G41 S42 T45 T49 R53 V54 F55 K56 F57 K58 D59 LYS VAL GLY SER TRP GLY SER HIS HIS HIS HIS HIS

• Molecule 3: De novo design 4E1H_95

Chain F:

11%
60%   18%   19%

MET K2 Y3 R12 G13 I14 I20 Q21 L22 E29 R32 E33 Y34 V35 E36 K37 L38 R39 S42 A43 I44 D52 F57 K58 D59 LYS VAL GLY SER TRP GLY SER HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	181.15Å 65.42Å 128.53Å 90.00° 119.35° 90.00°	Depositor
Resolution (Å)	49.45 – 2.85 49.45 – 2.85	Depositor EDS
% Data completeness (in resolution range)	97.7 (49.45-2.85) 90.6 (49.45-2.85)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.17 (at 2.86Å)	Xtriage
Refinement program	PHENIX 1.13_2998, PHENIX 1.13_2998	Depositor
R, R_{free}	0.212 , 0.247 0.213 , 0.248	Depositor DCC
R_{free} test set	1954 reflections (6.41%)	wwPDB-VP
Wilson B-factor (Å^2)	44.5	Xtriage
Anisotropy	0.403	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.37 , 46.5	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	0.91	EDS
Total number of atoms	7650	wwPDB-VP
Average B, all atoms (Å^2)	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.11% 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 ⓘ

5.1   Standard geometry ⓘ

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	0.26	0/1711	0.51	0/2333
1	D	0.29	0/1711	0.56	0/2333
2	B	0.27	0/1726	0.49	0/2347
2	E	0.34	0/1721	0.68	5/2341 (0.2%)
3	C	0.38	0/473	0.71	1/633 (0.2%)
3	F	0.32	0/473	0.61	1/633 (0.2%)
All	All	0.30	0/7815	0.58	7/10620 (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
3	C	0	2

There are no bond length outliers.

The worst 5 of 7 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	C	55	PHE	CB-CG-CD2	-8.13	115.11	120.80
2	E	59	LYS	N-CA-CB	7.13	123.43	110.60
2	E	59	LYS	CD-CE-NZ	-6.44	96.88	111.70
2	E	59	LYS	CB-CG-CD	-6.04	95.91	111.60
3	F	38	LEU	CA-CB-CG	5.64	128.26	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
3	C	54	VAL	Peptide
3	C	55	PHE	Sidechain

5.2 Too-close contacts ⓘ

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	1668	0	1645	26	0
1	D	1668	0	1645	19	0
2	B	1686	0	1625	17	1
2	E	1681	0	1617	48	1
3	C	466	0	466	20	0
3	F	466	0	466	12	0
4	A	6	0	0	0	0
4	B	3	0	0	0	0
4	C	1	0	0	0	0
4	D	3	0	0	0	0
4	E	2	0	0	0	0
All	All	7650	0	7464	134	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 9.

The worst 5 of 134 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)
3:C:53:ARG:HG2	3:C:55:PHE:HE2	1.18	1.03
2:E:201:GLU:HA	2:E:225:ARG:NH1	1.74	1.01
3:C:53:ARG:HG2	3:C:55:PHE:CE2	1.98	0.96
2:E:201:GLU:CA	2:E:225:ARG:NH1	2.30	0.94
3:C:22:LEU:HD12	3:C:23:PRO:O	1.69	0.93

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:79:SER:OG	2:E:215:LEU:O[3_455]	2.08	0.12

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	219/245 (89%)	214 (98%)	5 (2%)	0	100	100
1	D	219/245 (89%)	212 (97%)	7 (3%)	0	100	100
2	B	216/237 (91%)	208 (96%)	8 (4%)	0	100	100
2	E	216/237 (91%)	207 (96%)	7 (3%)	2 (1%)	17	43
3	C	56/72 (78%)	54 (96%)	2 (4%)	0	100	100
3	F	56/72 (78%)	54 (96%)	2 (4%)	0	100	100
All	All	982/1108 (89%)	949 (97%)	31 (3%)	2 (0%)	47	75

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	E	216	SER
2	E	217	SER

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	191/211 (90%)	190 (100%)	1 (0%)	88	96
1	D	191/211 (90%)	188 (98%)	3 (2%)	62	84
2	B	188/203 (93%)	183 (97%)	5 (3%)	44	74
2	E	188/203 (93%)	180 (96%)	8 (4%)	29	59
3	C	50/62 (81%)	46 (92%)	4 (8%)	12	31

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	F	50/62 (81%)	46 (92%)	4 (8%)	12	31
All	All	858/952 (90%)	833 (97%)	25 (3%)	42	72

5 of 25 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	36	SER
2	E	156	ARG
3	F	57	PHE
2	E	79	SER
2	E	174	GLN

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	190	GLN
1	D	13	GLN
1	D	218	ASN
2	E	56	GLN

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 ⓘ

There are no monosaccharides in this entry.

5.6 Ligand geometry ⓘ

There are no ligands in this entry.

5.7 Other polymers ⓘ

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 ⓘ

6.1 Protein, DNA and RNA chains ⓘ

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(Å²)	Q<0.9
1	A	221/245 (90%)	-0.02	0	100	100	25, 38, 62, 76	0
1	D	221/245 (90%)	0.31	13 (5%)	22	17	31, 49, 87, 101	0
2	B	218/237 (91%)	-0.05	0	100	100	27, 42, 58, 77	2 (0%)
2	E	218/237 (91%)	0.65	19 (8%)	10	7	43, 69, 90, 111	2 (0%)
3	C	58/72 (80%)	0.79	5 (8%)	10	7	42, 75, 94, 97	0
3	F	58/72 (80%)	1.01	8 (13%)	2	2	55, 81, 108, 111	0
All	All	994/1108 (89%)	0.30	45 (4%)	33	28	25, 51, 89, 111	4 (0%)

The worst 5 of 45 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
3	C	42	SER	6.4
2	E	228	CYS	4.8
2	E	227	GLU	4.7
2	E	36	SER	4.4
1	D	147	SER	4.3

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

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.