

# Full wwPDB X-ray Structure Validation Report (i)

May 25, 2020 – 07:11 am BST

PDB ID 4M62

> Title Ontogeny of recognition specificity and functionality for the anti-HIV neutral-

> > izing antibody 4E10

: Finton, K.A.K. Authors Deposited on 2013-08-08

Resolution 1.80 Å(reported)

This is a Full wwPDB X-ray Structure Validation 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

> 1.8.5 (274361), CSD as541be (2020) Mogul

Xtriage (Phenix) 1.13 EDS 2.11

Percentile statistics 20191225.v01 (using entries in the PDB archive December 25th 2019)

> Refmac 5.8.0158

7.0.044 (Gargrove) CCP4 Engh & Huber (2001)

Ideal geometry (proteins) Ideal geometry (DNA, RNA) Parkinson et al. (1996)

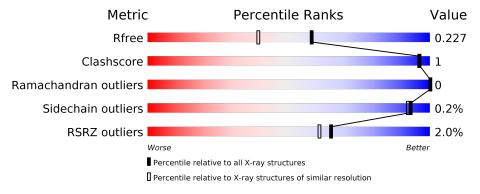
Validation Pipeline (wwPDB-VP) 2.11

### 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 1.80 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	$\begin{array}{c} \textbf{Whole archive} \\ (\#\text{Entries}) \end{array}$	$\begin{array}{c} {\rm Similar \; resolution} \\ (\#{\rm Entries, \; resolution \; range(\AA)}) \end{array}$		
$R_{free}$	130704	5950 (1.80-1.80)		
Clashscore	141614	6793 (1.80-1.80)		
Ramachandran outliers	138981	6697 (1.80-1.80)		
Sidechain outliers	138945	6696 (1.80-1.80)		
RSRZ outliers	127900	5850 (1.80-1.80)		

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% 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	L	113	100%					
1	M	113	100%					
2	Н	127	96%	•				
2	I	127	95%					
3	S	155	95%	5%				
3	Т	155	99%					



## 2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 6793 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 GEP2 FV light chain.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
1	L	113	Total 839	C 530	1,	O 170	S 2	0	1	0
1	М	113	Total 856	C 537		O 171	S 2	0	1	0

There are 10 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
L	96	SER	TRP	CONFLICT	UNP P18136
L	109	LEU	-	EXPRESSION TAG	UNP P18136
L	110	VAL	-	EXPRESSION TAG	UNP P18136
L	111	PRO	_	EXPRESSION TAG	UNP P18136
L	112	ARG	_	EXPRESSION TAG	UNP P18136
M	96	SER	TRP	CONFLICT	UNP P18136
M	109	LEU	-	EXPRESSION TAG	UNP P18136
M	110	VAL	-	EXPRESSION TAG	UNP P18136
M	111	PRO	_	EXPRESSION TAG	UNP P18136
M	112	ARG	_	EXPRESSION TAG	UNP P18136

• Molecule 2 is a protein called GEP2 FV heavy chain.

Mol	Chain	Residues	${f Atoms}$			ZeroOcc	AltConf	Trace		
2	Н	127	Total 950		N 157	O 183	S 5	0	4	0
2	I	126	Total 928	C 591	N 153	O 180	S 4	0	2	0

• Molecule 3 is a protein called T117.

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
2	C	155	Total	С	N	О	S	0	2	0
3	ا ا	199	1245	797	208	234	6	0	) 	U

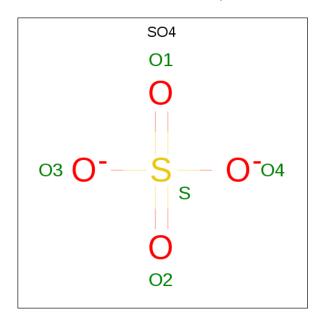
Continued on next page...



 $Continued\ from\ previous\ page...$ 

Mol	Chain	Residues	Atoms			ZeroOcc	AltConf	Trace		
3	Т	155	Total 1234	C 793	N 208	O 227	S	0	3	0
			1254	195	200	221	O			

• Molecule 4 is SULFATE ION (three-letter code: SO4) (formula: O<sub>4</sub>S).



Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	L	1	Total O S 5 4 1	0	0
4	L	1	Total O S 5 4 1	0	0
4	M	1	Total O S 5 4 1	0	0
4	S	1	Total O S 5 4 1	0	0
4	S	1	Total O S 5 4 1	0	0
4	Т	1	Total O S 5 4 1	0	0
4	Т	1	Total O S 5 4 1	0	0
4	Т	1	Total O S 5 4 1	0	0
4	Т	1	Total O S 5 4 1	0	0

• Molecule 5 is water.



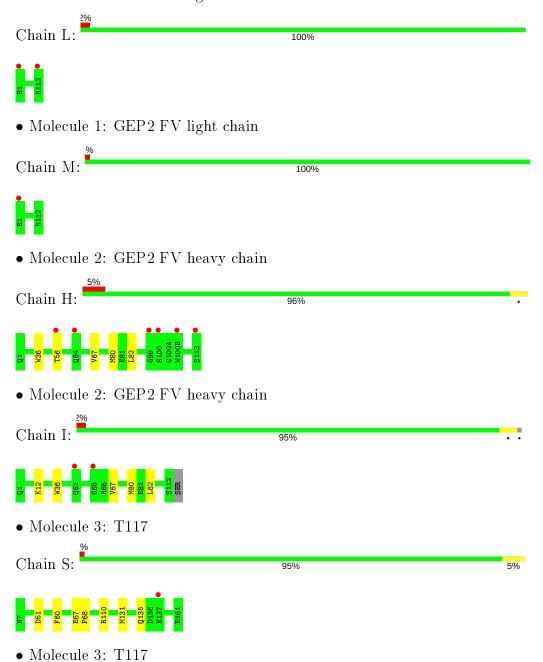
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
5	L	97	Total O 97 97	0	0
5	Н	110	Total O 110 110	0	0
5	M	105	Total O 105 105	0	0
5	I	109	Total O 109 109	0	0
5	S	147	Total O 147 147	0	0
5	Т	128	Total O 128 128	0	0



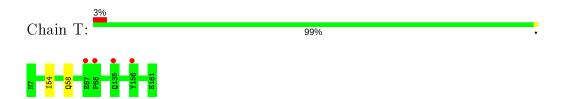
## 3 Residue-property plots (i)

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 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: GEP2 FV light chain









# 4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants	69.88Å 103.48Å 109.30Å	Depositor
a, b, c, $\alpha$ , $\beta$ , $\gamma$	90.00° 90.00° 90.00°	Depositor
Resolution (Å)	68.88 - 1.80	Depositor
resolution (A)	69.88 - 1.80	EDS
% Data completeness	98.0 (68.88-1.80)	Depositor
(in resolution range)	98.0 (69.88-1.80)	EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	(Not available)	Depositor
$< I/\sigma(I) > 1$	4.56 (at 1.80Å)	Xtriage
Refinement program	REFMAC 5.7.0029	Depositor
D D.	0.189 , 0.216	Depositor
$R, R_{free}$	0.198 , $0.227$	DCC
$R_{free}$ test set	3668  reflections  (5.05%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	19.2	Xtriage
Anisotropy	0.040	Xtriage
Bulk solvent $k_{sol}(e/Å^3)$ , $B_{sol}(Å^2)$	0.35, 42.8	EDS
L-test for twinning <sup>2</sup>	$ < L >=0.53, < L^2>=0.37$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	6793	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: The largest off-origin peak in the Patterson function is 4.24% of the height of the origin peak. No significant pseudotranslation is detected.

<sup>&</sup>lt;sup>2</sup>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.



<sup>&</sup>lt;sup>1</sup>Intensities estimated from amplitudes.

### 5 Model quality (i)

#### 5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: SO4

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		
MIOI		RMSZ	# Z >5	RMSZ	# Z  > 5	
1	L	0.26	0/861	0.45	0/1173	
1	M	0.25	0/878	0.44	0/1193	
2	Н	0.25	0/983	0.43	0/1333	
2	I	0.25	0/955	0.43	0/1296	
3	S	0.24	0/1282	0.40	0/1751	
3	Т	0.24	0/1274	0.40	0/1741	
All	All	0.25	0/6233	0.42	0/8487	

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	L	839	0	808	0	0
1	M	856	0	840	0	0
2	Н	950	0	937	5	0
2	I	928	0	906	2	0
3	S	1245	0	1204	7	0
3	Т	1234	0	1198	1	0
4	L	10	0	0	0	0

Continued on next page...



$\alpha \cdots$	· ·	•	
Continued	trom	mromanne	maaa
-	110116	DICUIUU	$Du_iu_{C}$

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	M	5	0	0	0	0
4	S	10	0	0	0	0
4	Т	20	0	0	0	0
5	Н	110	0	0	0	0
5	I	109	0	0	0	0
5	L	97	0	0	0	0
5	M	105	0	0	0	0
5	S	147	0	0	2	0
5	Т	128	0	0	0	0
All	All	6793	0	5893	12	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 1.

All (12) close contacts within the same asymmetric unit are listed below, sorted by their clash magnitude.

Atom-1	Atom-2	$egin{array}{ll}  ext{Interatomic} \  ext{distance} \ ( ext{\AA}) \end{array}$	Clash overlap (Å)
2:H:56:THR:CG2	3:S:60:PHE:CD2	2.86	0.59
2:H:56:THR:HG22	3:S:60:PHE:CE2	2.41	0.56
2:H:56:THR:HG22	3:S:60:PHE:CD2	2.42	0.55
2:I:36:TRP:CE2	2:I:80:MET:HB2	2.52	0.45
3:T:54:ILE:O	3:T:58:GLN:HG3	2.17	0.45
3:S:67:GLU:HB2	3:S:68:PRO:HA	2.01	0.43
2:H:36:TRP:CE2	2:H:80[A]:MET:HB2	2.54	0.42
2:I:67:VAL:HG22	2:I:82:LEU:HD13	2.02	0.41
3:S:110:ARG:NH2	5:S:412:HOH:O	2.53	0.41
2:H:67:VAL:HG22	2:H:82[A]:LEU:HD13	2.01	0.41
3:S:51:ASP:CG	5:S:360:HOH:O	2.59	0.40
3:S:131:MET:O	3:S:135:GLN:HG3	2.22	0.40

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 r	number	of	residues	for	which	the	backbone	conformation	was
analysed, and the total numb	er of	residues								

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Perce	$\mathbf{ntiles}$
1	L	112/113 (99%)	108 (96%)	4 (4%)	0	100	100
1	М	112/113 (99%)	109 (97%)	3 (3%)	0	100	100
2	Н	129/127~(102%)	127~(98%)	2 (2%)	0	100	100
2	I	126/127~(99%)	125 (99%)	1 (1%)	0	100	100
3	S	156/155~(101%)	153~(98%)	3 (2%)	0	100	100
3	${ m T}$	$156/155 \; (101\%)$	156 (100%)	0	0	100	100
All	All	791/790 (100%)	778 (98%)	13 (2%)	0	100	100

There are no Ramachandran outliers to report.

#### 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	Percentil	les
1	L	91/95~(96%)	91 (100%)	0	100 10	00
1	M	95/95~(100%)	95 (100%)	0	100 10	00
2	Н	99/98 (101%)	99 (100%)	0	100 10	00
2	I	95/98 (97%)	94 (99%)	1 (1%)	73 68	}
3	S	138/138 (100%)	138 (100%)	0	100 10	00
3	Т	135/138 (98%)	135 (100%)	0	100 10	00
All	All	653/662 (99%)	652 (100%)	1 (0%)	93 92	?

All (1) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	I	12	LYS

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)

9 ligands are modelled in this entry.

In the following table, the Counts columns list the number of bonds (or angles) for which Mogul statistics could be retrieved, the number of bonds (or angles) that are observed in the model and the number of bonds (or angles) that are defined in the Chemical Component Dictionary. The Link column lists molecule types, if any, to which the group is linked. 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| > 2 is considered an outlier worth inspection. RMSZ is the root-mean-square of all Z scores of the bond lengths (or angles).

Mol	Tune	Chain	Chain Res Link Bond lengths		gths	Bond angles				
10101	Type	Chain	nes	Lilik	Counts	RMSZ	# Z  > 2	Counts	RMSZ	# Z  > 2
4	SO4	Т	202	-	4,4,4	0.32	0	6,6,6	0.11	0
4	SO4	Т	203	-	4,4,4	0.33	0	6,6,6	0.07	0
4	SO4	Т	204	-	4,4,4	0.32	0	6,6,6	0.05	0
4	SO4	S	202	-	4,4,4	0.33	0	6,6,6	0.12	0
4	SO4	M	201	-	4,4,4	0.33	0	6,6,6	0.06	0
4	SO4	L	201	-	4,4,4	0.31	0	6,6,6	0.07	0
4	SO4	S	201	-	4,4,4	0.32	0	6,6,6	0.11	0
4	SO4	L	202	-	4,4,4	0.34	0	6,6,6	0.04	0
4	SO4	Т	201	-	4,4,4	0.34	0	6,6,6	0.12	0

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.



No monomer is involved in short contacts.

### 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,  $95^{th}$  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 $>$	$\#\mathrm{RSRZ}{>}2$	$OWAB(A^2)$	Q < 0.9
1	L	113/113 (100%)	-0.07	2 (1%) 68 64	11, 17, 31, 51	0
1	M	113/113 (100%)	-0.10	1 (0%) 84 82	13, 20, 29, 46	0
2	Н	127/127 (100%)	0.07	6 (4%) 31 25	10, 17, 32, 42	0
2	I	126/127 (99%)	-0.02	2 (1%) 72 68	12, 17, 32, 43	0
3	S	155/155~(100%)	-0.08	1 (0%) 89 87	12, 17, 33, 47	0
3	Т	155/155~(100%)	-0.05	4 (2%) 56 51	11, 17, 34, 48	0
All	All	789/790 (99%)	-0.04	16 (2%) 65 61	10, 18, 32, 51	0

All (16) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	Н	99	GLY	6.3
2	Н	100	SER	4.6
2	I	65	GLY	3.6
3	Т	68	PRO	3.3
2	Н	113	SER	2.9
3	Т	135	GLN	2.5
1	L	112	ARG	2.5
2	Н	100(B)	TRP	2.5
2	I	61	GLN	2.5
3	Т	156	TYR	2.5
1	L	1	GLU	2.4
2	Н	56	THR	2.2
3	S	137	GLU	2.2
2	Н	64	GLN	2.2
3	Т	67	GLU	2.1
1	M	1	GLU	2.0



#### 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)

In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median,  $95^{th}$  percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	$\mathbf{B} ext{-}\mathbf{factors}(\mathbf{\mathring{A}}^2)$	Q < 0.9
4	SO4	L	202	5/5	0.74	0.24	46,51,56,58	0
4	SO4	Т	203	5/5	0.81	0.27	60,61,63,64	0
4	SO4	Т	204	5/5	0.82	0.19	63,65,68,71	0
4	SO4	L	201	5/5	0.93	0.22	46,51,53,54	0
4	SO4	Т	202	5/5	0.94	0.22	11,12,12,12	5
4	SO4	M	201	5/5	0.96	0.13	31,33,35,35	0
4	SO4	S	201	5/5	0.99	0.08	14,14,14,14	0
4	SO4	S	202	5/5	0.99	0.11	21,24,24,24	0
4	SO4	Т	201	5/5	0.99	0.10	18,19,20,20	0

### 6.5 Other polymers (i)

There are no such residues in this entry.

