



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 22, 2017 – 05:03 PM EDT

PDB ID : 4N1E  
Title : Structural evidence for antigen receptor evolution  
Authors : Langley, D.B.; Rouet, R.; Stock, D.; Christ, D.  
Deposited on : 2013-10-04  
Resolution : 2.23 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](mailto:validation@mail.wwpdb.org)

A user guide is available at

<http://wwpdb.org/validation/2016/XrayValidationReportHelp>

with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20029077
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
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-20029077

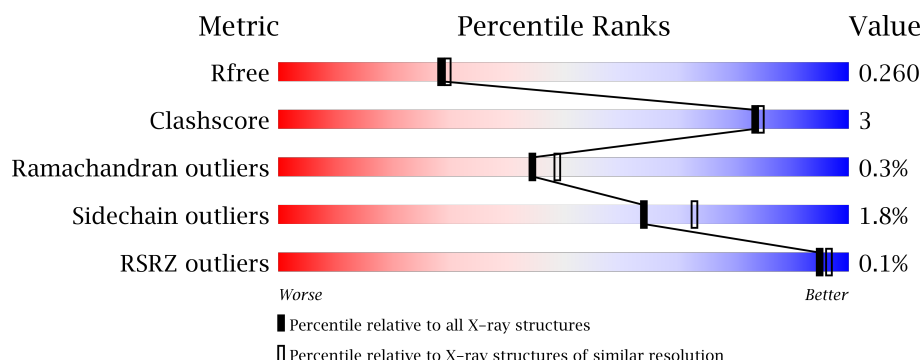
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.23 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))
$R_{free}$	100719	1804 (2.26-2.22)
Clashscore	112137	1957 (2.26-2.22)
Ramachandran outliers	110173	1916 (2.26-2.22)
Sidechain outliers	110143	1917 (2.26-2.22)
RSRZ outliers	101464	1809 (2.26-2.22)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	109	<div> <div> <div></div> <div>77%</div> <div>9%</div> <div>14%</div> </div> </div>
1	B	109	<div> <div>89%</div> <div>9%</div> <div>.</div> </div>
1	C	109	<div> <div>88%</div> <div>8%</div> <div>..</div> </div>
1	D	109	<div> <div>91%</div> <div>7%</div> <div>..</div> </div>
1	E	109	<div> <div>91%</div> <div>6%</div> <div>.</div> </div>

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	109	<div><div></div><div>86%</div><div>10%</div><div>.</div></div>
1	G	109	<div><div></div><div>93%</div><div>7%</div><div></div></div>
1	H	109	<div><div></div><div>94%</div><div></div><div>..</div></div>
2	I	129	<div><div></div><div>95%</div><div></div><div>...</div></div>
2	J	129	<div><div></div><div>96%</div><div></div><div>..</div></div>
2	K	129	<div><div></div><div>%</div><div>88%</div><div>10%</div><div>.</div></div>
2	L	129	<div><div></div><div>88%</div><div>9%</div><div>..</div></div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 9967 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 immunoglobulin variable light chain domain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	94	Total	C	N	O	S	0	0	0
			662	419	105	135	3			
1	B	107	Total	C	N	O	S	0	0	0
			770	485	122	160	3			
1	C	106	Total	C	N	O	S	0	0	0
			750	472	123	152	3			
1	D	109	Total	C	N	O	S	0	0	0
			810	511	133	163	3			
1	E	106	Total	C	N	O	S	0	0	0
			762	480	125	154	3			
1	F	105	Total	C	N	O	S	0	0	0
			752	477	119	153	3			
1	G	109	Total	C	N	O	S	0	0	0
			798	505	131	159	3			
1	H	107	Total	C	N	O	S	0	0	0
			760	480	122	155	3			

- Molecule 2 is a protein called Lysozyme C.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	I	127	Total	C	N	O	S	0	0	0
			937	580	175	172	10			
2	J	128	Total	C	N	O	S	0	0	0
			967	592	182	183	10			
2	K	127	Total	C	N	O	S	0	0	0
			953	587	179	177	10			
2	L	127	Total	C	N	O	S	0	0	0
			916	563	169	174	10			

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	2	Total 2	O 2	0	0
3	B	21	Total 21	O 21	0	0
3	C	2	Total 2	O 2	0	0
3	D	29	Total 29	O 29	0	0
3	E	13	Total 13	O 13	0	0
3	F	12	Total 12	O 12	0	0
3	G	9	Total 9	O 9	0	0
3	H	9	Total 9	O 9	0	0
3	I	5	Total 5	O 5	0	0
3	J	16	Total 16	O 16	0	0
3	K	10	Total 10	O 10	0	0
3	L	2	Total 2	O 2	0	0

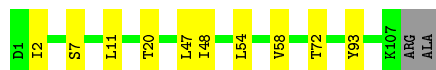
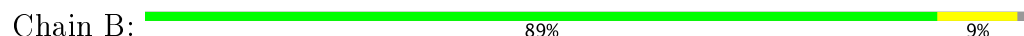
### 3 Residue-property plots [i](#)

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 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: immunoglobulin variable light chain domain



- Molecule 1: immunoglobulin variable light chain domain



- Molecule 1: immunoglobulin variable light chain domain



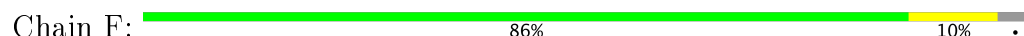
- Molecule 1: immunoglobulin variable light chain domain



- Molecule 1: immunoglobulin variable light chain domain



- Molecule 1: immunoglobulin variable light chain domain





- Molecule 1: immunoglobulin variable light chain domain

Chain G: 93% 7%



- Molecule 1: immunoglobulin variable light chain domain

Chain H: 94% ..



- Molecule 2: Lysozyme C

Chain I: 95% ...



- Molecule 2: Lysozyme C

Chain J: 96% ..



- Molecule 2: Lysozyme C

Chain K: 88% 10% .



- Molecule 2: Lysozyme C

Chain L: 88% 9% ..



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	114.92Å 39.41Å 178.30Å 90.00° 106.38° 90.00°	Depositor
Resolution (Å)	45.66 – 2.23 45.67 – 2.23	Depositor EDS
% Data completeness (in resolution range)	95.5 (45.66-2.23) 95.6 (45.67-2.23)	Depositor EDS
$R_{merge}$	(Not available)	Depositor
$R_{sym}$	0.08	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.01 (at 2.22Å)	Xtriage
Refinement program	REFMAC 5.6.0111	Depositor
R, $R_{free}$	0.223 , 0.265 0.219 , 0.260	Depositor DCC
$R_{free}$ test set	3655 reflections (5.29%)	DCC
Wilson B-factor (Å <sup>2</sup> )	32.0	Xtriage
Anisotropy	0.828	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.30 , 33.2	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Estimated twinning fraction	0.018 for h,-k,-h-l	Xtriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	9967	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	35.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.46% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  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	0.63	0/677	0.70	0/927
1	B	0.83	0/787	0.74	0/1076
1	C	0.63	0/768	0.71	0/1052
1	D	0.84	0/828	0.90	2/1126 (0.2%)
1	E	0.70	1/780 (0.1%)	0.73	0/1067
1	F	0.64	0/770	0.71	0/1056
1	G	0.62	0/816	0.69	0/1112
1	H	0.55	0/777	0.66	0/1066
2	I	0.70	0/957	0.68	0/1301
2	J	0.81	0/987	0.76	0/1339
2	K	0.74	0/973	0.71	0/1322
2	L	0.71	0/936	0.68	0/1275
All	All	0.71	1/10056 (0.0%)	0.72	2/13719 (0.0%)

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	D	0	1

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	35	TRP	CD2-CE2	5.42	1.47	1.41

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	D	108	ARG	NE-CZ-NH2	-7.74	116.43	120.30
1	D	108	ARG	NE-CZ-NH1	7.69	124.14	120.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	108	ARG	Peptide

## 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	662	0	562	5	0
1	B	770	0	709	8	0
1	C	750	0	649	4	0
1	D	810	0	776	5	0
1	E	762	0	678	3	0
1	F	752	0	676	6	0
1	G	798	0	745	5	0
1	H	760	0	675	2	0
2	I	937	0	855	2	0
2	J	967	0	895	2	0
2	K	953	0	879	7	0
2	L	916	0	798	9	0
3	A	2	0	0	0	0
3	B	21	0	0	0	0
3	C	2	0	0	0	0
3	D	29	0	0	0	0
3	E	13	0	0	0	0
3	F	12	0	0	0	0
3	G	9	0	0	0	0
3	H	9	0	0	0	0
3	I	5	0	0	0	0
3	J	16	0	0	0	0
3	K	10	0	0	1	0
3	L	2	0	0	0	0
All	All	9967	0	8897	53	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 3.

The worst 5 of 53 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:C:100:GLN:HA	1:D:43:ALA:HB2	1.75	0.68
1:B:2:ILE:HD11	1:B:93:TYR:CD2	2.34	0.62
1:B:20:THR:HG23	1:B:72:THR:HG23	1.82	0.60
1:A:8:PRO:O	1:A:102:THR:HG23	2.01	0.60
1:B:2:ILE:CD1	1:B:93:TYR:CD2	2.89	0.56

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	88/109 (81%)	84 (96%)	4 (4%)	0	100	100
1	B	105/109 (96%)	99 (94%)	6 (6%)	0	100	100
1	C	104/109 (95%)	97 (93%)	3 (3%)	4 (4%)	4	1
1	D	107/109 (98%)	102 (95%)	5 (5%)	0	100	100
1	E	104/109 (95%)	101 (97%)	3 (3%)	0	100	100
1	F	103/109 (94%)	98 (95%)	5 (5%)	0	100	100
1	G	107/109 (98%)	103 (96%)	4 (4%)	0	100	100
1	H	105/109 (96%)	101 (96%)	4 (4%)	0	100	100
2	I	125/129 (97%)	122 (98%)	3 (2%)	0	100	100
2	J	126/129 (98%)	123 (98%)	3 (2%)	0	100	100
2	K	125/129 (97%)	122 (98%)	3 (2%)	0	100	100
2	L	125/129 (97%)	120 (96%)	5 (4%)	0	100	100
All	All	1324/1388 (95%)	1272 (96%)	48 (4%)	4 (0%)	44	49

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	16	GLY
1	C	17	ASP
1	C	30	ALA
1	C	15	VAL

### 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	60/90 (67%)	59 (98%)	1 (2%)	66	74
1	B	80/90 (89%)	79 (99%)	1 (1%)	73	82
1	C	69/90 (77%)	68 (99%)	1 (1%)	71	80
1	D	86/90 (96%)	82 (95%)	4 (5%)	30	32
1	E	73/90 (81%)	73 (100%)	0	100	100
1	F	74/90 (82%)	72 (97%)	2 (3%)	50	59
1	G	80/90 (89%)	79 (99%)	1 (1%)	73	82
1	H	73/90 (81%)	71 (97%)	2 (3%)	50	59
2	I	90/105 (86%)	88 (98%)	2 (2%)	57	66
2	J	99/105 (94%)	98 (99%)	1 (1%)	80	86
2	K	95/105 (90%)	94 (99%)	1 (1%)	78	85
2	L	85/105 (81%)	84 (99%)	1 (1%)	75	83
All	All	964/1140 (85%)	947 (98%)	17 (2%)	64	73

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	14	SER
1	F	81	GLU
2	I	99	VAL
1	D	70	ASP
2	J	1	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such

sidechains are listed below:

Mol	Chain	Res	Type
1	C	27	GLN
1	H	55	GLN
2	I	113	ASN
2	L	106	ASN

### 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 ⓘ

### 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, 95<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	94/109 (86%)	0.05	1 (1%) 80 81	20, 42, 68, 84	0
1	B	107/109 (98%)	-0.30	0 100 100	16, 26, 38, 45	0
1	C	106/109 (97%)	-0.23	0 100 100	22, 41, 56, 66	0
1	D	109/109 (100%)	-0.27	0 100 100	14, 21, 36, 45	0
1	E	106/109 (97%)	-0.26	0 100 100	21, 31, 46, 51	0
1	F	105/109 (96%)	-0.34	0 100 100	21, 31, 47, 55	0
1	G	109/109 (100%)	-0.26	0 100 100	21, 35, 52, 67	0
1	H	107/109 (98%)	-0.29	0 100 100	23, 34, 49, 56	0
2	I	127/129 (98%)	-0.23	0 100 100	28, 42, 61, 78	0
2	J	128/129 (99%)	-0.16	0 100 100	20, 30, 43, 51	0
2	K	127/129 (98%)	-0.12	1 (0%) 86 86	24, 33, 47, 60	0
2	L	127/129 (98%)	-0.15	0 100 100	30, 47, 58, 71	0
All	All	1352/1388 (97%)	-0.21	2 (0%) 95 96	14, 34, 55, 84	0

All (2) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	K	47	THR	3.0
1	A	20	THR	2.1

### 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 [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.