Full wwPDB X-ray Structure Validation Report (i)

Apr 22, 2025 – 01:36 pm BST

PDB ID	:	9HS7 / pdb_00009hs7
Title	:	Anti-HIV-1 chimeric miniprotein mimicking the N-terminal half of gp41 NHR with an extended region targeting the MPER
Authors	:	Camara-Artigas, A.; Gavira, J.A.; Conejero-Lara, F.; Polo-Megias, D.; Salinas-Garcia, M.C.
Deposited on	:	2024-12-18
Resolution	:	1.70 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

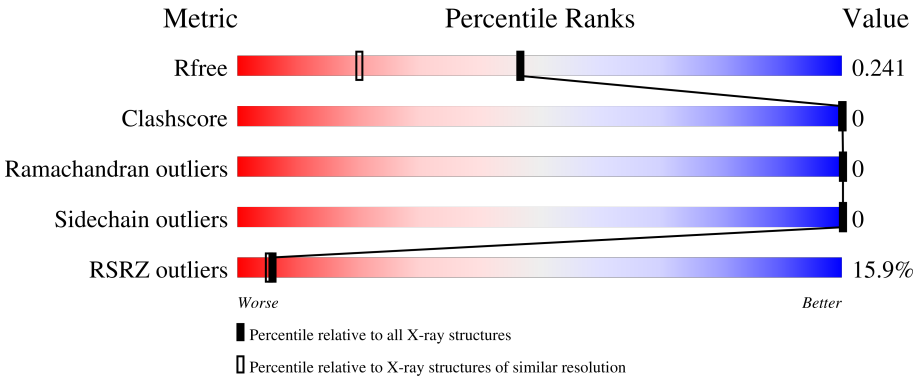
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.003 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.42

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 1.70 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}	164625	5161 (1.70-1.70)
Clashscore	180529	5671 (1.70-1.70)
Ramachandran outliers	177936	5594 (1.70-1.70)
Sidechain outliers	177891	5594 (1.70-1.70)
RSRZ outliers	164620	5159 (1.70-1.70)

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	140	14%, 89%, 10%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 2062 atoms, of which 1036 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 Transmembrane protein gp41.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
			Total	C	H	N	O	S			
1	A	126	2010	600	1026	190	186	8	0	2	0

There are 71 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1	MET	-	initiating methionine	UNP P19551
A	3	CYS	MET	engineered mutation	UNP P19551
A	10	LYS	LEU	engineered mutation	UNP P19551
A	17	LYS	LEU	engineered mutation	UNP P19551
A	24	LYS	GLN	engineered mutation	UNP P19551
A	31	LYS	ALA	engineered mutation	UNP P19551
A	35	ILE	GLN	engineered mutation	UNP P19551
A	41	ARG	-	linker	UNP P19551
A	42	GLY	-	linker	UNP P19551
A	43	LEU	-	linker	UNP P19551
A	44	ILE	-	linker	UNP P19551
A	45	CYS	-	linker	UNP P19551
A	46	GLY	-	linker	UNP P19551
A	47	PRO	-	linker	UNP P19551
A	48	GLN	-	linker	UNP P19551
A	49	LEU	-	linker	UNP P19551
A	50	LEU	-	linker	UNP P19551
A	51	HIS	-	linker	UNP P19551
A	52	GLN	-	linker	UNP P19551
A	53	ILE	-	linker	UNP P19551
A	54	ALA	-	linker	UNP P19551
A	55	GLU	-	linker	UNP P19551
A	56	ILE	-	linker	UNP P19551
A	57	GLU	-	linker	UNP P19551
A	58	ARG	-	linker	UNP P19551
A	59	GLU	-	linker	UNP P19551
A	60	LEU	-	linker	UNP P19551

Continued on next page...

Continued from previous page...

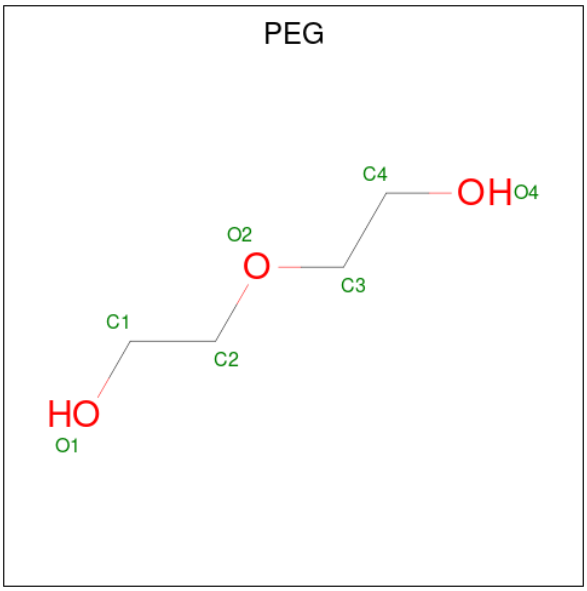
Chain	Residue	Modelled	Actual	Comment	Reference
A	61	ASN	-	linker	UNP P19551
A	62	ASN	-	linker	UNP P19551
A	63	GLN	-	linker	UNP P19551
A	64	GLU	-	linker	UNP P19551
A	65	GLN	-	linker	UNP P19551
A	66	GLU	-	linker	UNP P19551
A	67	ILE	-	linker	UNP P19551
A	68	GLY	-	linker	UNP P19551
A	69	SER	-	linker	UNP P19551
A	70	LEU	-	linker	UNP P19551
A	71	LYS	-	linker	UNP P19551
A	72	GLN	-	linker	UNP P19551
A	73	ARG	-	linker	UNP P19551
A	74	ALA	-	linker	UNP P19551
A	75	GLN	-	linker	UNP P19551
A	76	VAL	-	linker	UNP P19551
A	77	GLU	-	linker	UNP P19551
A	78	LYS	-	linker	UNP P19551
A	79	THR	-	linker	UNP P19551
A	80	MET	-	linker	UNP P19551
A	81	SER	-	linker	UNP P19551
A	82	ALA	-	linker	UNP P19551
A	83	ALA	-	linker	UNP P19551
A	84	SER	-	linker	UNP P19551
A	85	GLY	-	linker	UNP P19551
A	86	CYS	-	linker	UNP P19551
A	87	GLY	-	linker	UNP P19551
A	90	PRO	ALA	engineered mutation	UNP P19551
A	96	ASP	THR	engineered mutation	UNP P19551
A	107	ASP	VAL	engineered mutation	UNP P19551
A	114	LYS	LEU	engineered mutation	UNP P19551
A	120	ILE	GLN	engineered mutation	UNP P19551
A	121	LYS	GLN	engineered mutation	UNP P19551
A	128	CYS	VAL	engineered mutation	UNP P19551
A	131	GLY	-	expression tag	UNP P19551
A	132	GLY	-	expression tag	UNP P19551
A	133	GLY	-	expression tag	UNP P19551
A	134	SER	-	expression tag	UNP P19551
A	135	HIS	-	expression tag	UNP P19551
A	136	HIS	-	expression tag	UNP P19551
A	137	HIS	-	expression tag	UNP P19551
A	138	HIS	-	expression tag	UNP P19551

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
A	139	HIS	-	expression tag	UNP P19551
A	140	HIS	-	expression tag	UNP P19551

- Molecule 2 is DI(HYDROXYETHYL)ETHER (CCD ID: PEG) (formula: $C_4H_{10}O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
2	A	1	Total	C	H	O	0	0
			17	4	10	3		

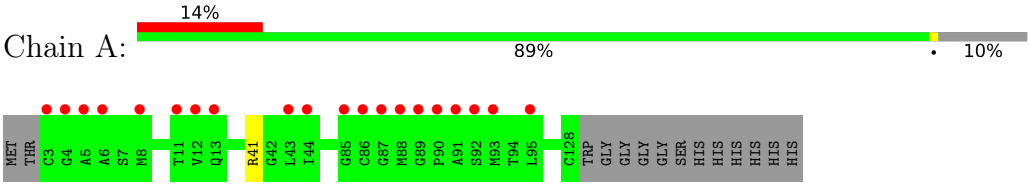
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	35	Total	O	0	0
			35	35		

3 Residue-property plots (i)

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: Transmembrane protein gp41



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 65	Depositor
Cell constants a, b, c, α, β, γ	65.44Å 65.44Å 88.77Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	19.29 – 1.70 19.29 – 1.70	Depositor EDS
% Data completeness (in resolution range)	86.9 (19.29-1.70) 48.0 (19.29-1.70)	Depositor EDS
R_{merge}	0.07	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.04 (at 1.69Å)	Xtriage
Refinement program	PHENIX 1.20.1_4487	Depositor
R, R_{free}	0.201 , 0.239 0.201 , 0.241	Depositor DCC
R_{free} test set	1179 reflections (4.82%)	wwPDB-VP
Wilson B-factor (Å^2)	25.2	Xtriage
Anisotropy	0.073	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 58.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.31$	Xtriage
Estimated twinning fraction	0.089 for h,-h-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	2062	wwPDB-VP
Average B, all atoms (Å^2)	58.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.99% 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

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

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	0.65	0/988	0.82	0/1322

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	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
1	A	41	ARG	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	984	1026	1024	0	0
2	A	7	10	10	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	A	35	0	0	0	0
All	All	1026	1036	1034	0	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 0.

There are no clashes within the asymmetric unit.

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	126/140 (90%)	121 (96%)	5 (4%)	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	Percentiles
1	A	107/115 (93%)	107 (100%)	0	100 100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	62	ASN

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 oligosaccharides in this entry.

5.6 Ligand geometry

1 ligand is 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	Type	Chain	Res	Link	Bond lengths			Bond angles		
					Counts	RMSZ	$\#\|Z\| > 2$	Counts	RMSZ	$\#\|Z\| > 2$
2	PEG	A	201	-	6,6,6	0.40	0	5,5,5	0.30	0

In the following table, the Chirals column lists the number of chiral outliers, the number of chiral centers analysed, the number of these observed in the model and the number defined in the Chemical Component Dictionary. Similar counts are reported in the Torsion and Rings columns. '-' means no outliers of that kind were identified.

Mol	Type	Chain	Res	Link	Chirals	Torsions	Rings
2	PEG	A	201	-	-	2/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	201	PEG	O2-C3-C4-O4
2	A	201	PEG	C1-C2-O2-C3

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

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^{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>	#RSRZ>2	OWAB(Å^2)	Q<0.9
1	A	126/140 (90%)	0.82	20 (15%) 6 5	19, 51, 125, 170	2 (1%)

All (20) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	3	CYS	6.2
1	A	4	GLY	6.2
1	A	90	PRO	6.0
1	A	5	ALA	5.5
1	A	43	LEU	5.3
1	A	88	MET	5.1
1	A	6	ALA	4.4
1	A	44	ILE	4.0
1	A	91	ALA	4.0
1	A	87	GLY	3.6
1	A	92	SER	3.1
1	A	89	GLY	2.8
1	A	8	MET	2.7
1	A	11	THR	2.6
1	A	85	GLY	2.5
1	A	13	GLN	2.4
1	A	93	MET	2.3
1	A	95	LEU	2.2
1	A	86	CYS	2.2
1	A	12	VAL	2.0

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 monosaccharides 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	B-factors(Å^2)	Q<0.9
2	PEG	A	201	7/7	0.85	0.15	56,69,83,83	0

6.5 Other polymers (i)

There are no such residues in this entry.