



Full wwPDB X-ray Structure Validation Report ⓘ

Aug 10, 2026 – 05:32 pm BST

PDB ID : 9SAH / pdb_00009sah
Title : Structure of a fusion of biosynthetic proteins E4 and NB2
Authors : Li de la Sierra-Gallay, I.; Mesneau, A.
Deposited on : 2025-08-07
Resolution : 2.20 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 ⓘ 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 ⓘ](#)) were used in the production of this report:

MolProbity	:	4-5-2 with Phenix2.0
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	2.0
EDS	:	3.0
Percentile statistics	:	20250101.v01 (using entries in the PDB archive January 1st 2025)
CCP4	:	9.0.015 (Gargrove)
Density-Fitness	:	1.0.12
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.50

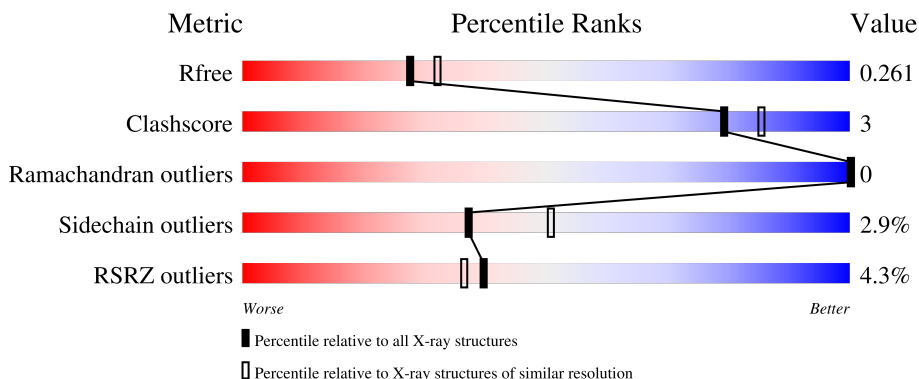
1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

X-RAY DIFFRACTION

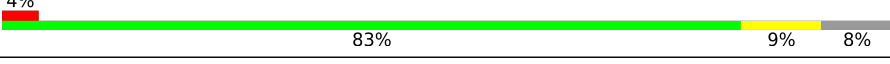
The reported resolution of this entry is 2.20 Å.

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}	180053	6164 (2.20-2.20)
Clashscore	190562	6851 (2.20-2.20)
Ramachandran outliers	187476	6768 (2.20-2.20)
Sidechain outliers	187428	6769 (2.20-2.20)
RSRZ outliers	180081	6166 (2.20-2.20)

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 $\leq 5\%$. 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	499	
1	B	499	

2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 7361 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 E4-NB2 synthetic fusion protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	461	Total	C	N	O	S	0	0	0
			3519	2175	647	694	3			
1	B	460	Total	C	N	O	S	0	0	0
			3510	2169	645	693	3			

- Molecule 2 is DI(HYDROXYETHYL)ETHER (CCD ID: PEG) (formula: C₄H₁₀O₃).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			7	4	3		

- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	176	Total	O	0	0
			176	176		

Continued on next page...

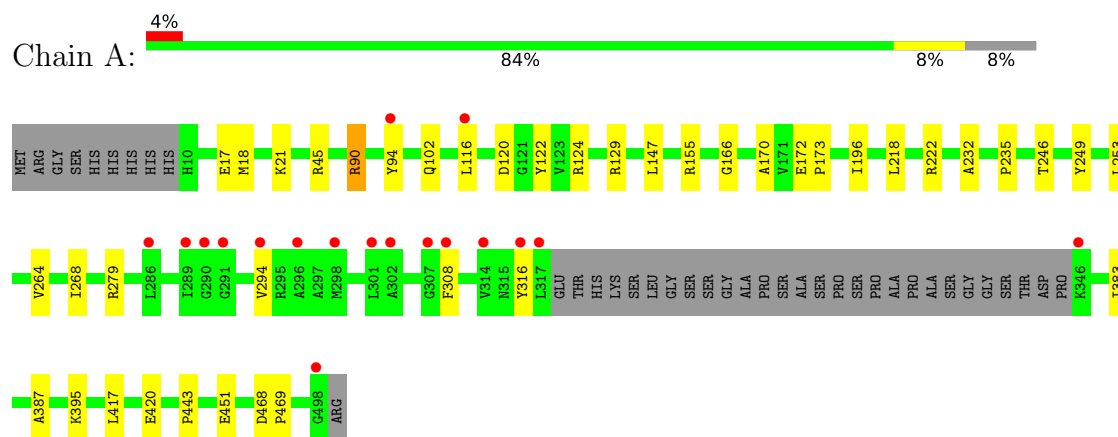
Continued from previous page...

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	B	149	Total	O	0	0
			149	149		

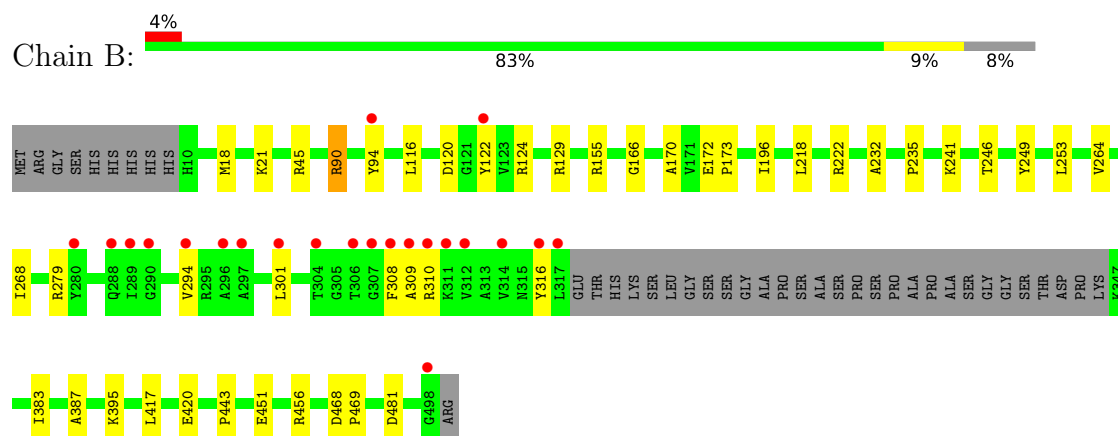
3 Residue-property plots [i](#)

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: E4-NB2 synthetic fusion protein



- Molecule 1: E4-NB2 synthetic fusion protein



4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α , β , γ	50.84Å 117.81Å 154.24Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.28 – 2.20 48.28 – 2.20	Depositor EDS
% Data completeness (in resolution range)	100.0 (48.28-2.20) 100.0 (48.28-2.20)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ ¹	1.64 (at 2.20Å)	Xtriage
Refinement program	BUSTER 2.10.4 (10-JUL-2024)	Depositor
R, R_{free}	0.232 , 0.260 0.226 , 0.261	Depositor DCC
R_{free} test set	2400 reflections (5.00%)	wwPDB-VP
Wilson B-factor (Å ²)	35.9	Xtriage
Anisotropy	0.220	Xtriage
Bulk solvent k_{sol} (e/Å ³), B_{sol} (Å ²)	0.35 , 42.1	EDS
L-test for twinning ²	$\langle L \rangle = 0.50$, $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o, F_c correlation	0.95	EDS
Total number of atoms	7361	wwPDB-VP
Average B, all atoms (Å ²)	46.0	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 50.18 % 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 6.7073e-05. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

¹Intensities estimated from amplitudes.

²Theoretical 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

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	0.72	0/3548	1.07	0/4777
1	B	0.73	0/3539	1.08	0/4766
All	All	0.73	0/7087	1.07	0/9543

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

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	3519	0	3594	19	0
1	B	3510	0	3581	19	0
2	A	7	0	10	0	0
3	A	176	0	0	1	0
3	B	149	0	0	0	0
All	All	7361	0	7185	38	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.

All (38) 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:B:279:ARG:HB3	1:B:308:PHE:CE1	2.39	0.58
1:A:129:ARG:HD2	1:A:420:GLU:HB2	1.89	0.55
1:A:279:ARG:HB3	1:A:308:PHE:CE2	2.43	0.54
1:B:456:ARG:NH1	1:B:481:ASP:OD1	2.41	0.54
1:A:90:ARG:NH2	1:A:122:TYR:HE2	2.06	0.52
1:B:90:ARG:NH2	1:B:122:TYR:HE2	2.07	0.52
1:A:166:GLY:HA2	1:A:196:ILE:HG12	1.92	0.52
1:A:222:ARG:NH1	1:A:387:ALA:O	2.43	0.51
1:B:90:ARG:NH2	1:B:122:TYR:CE2	2.79	0.50
1:B:222:ARG:NH1	1:B:387:ALA:O	2.44	0.50
1:A:147:LEU:O	1:A:155:ARG:HD2	2.12	0.50
1:B:166:GLY:HA2	1:B:196:ILE:HG12	1.93	0.50
1:A:90:ARG:NH2	1:A:122:TYR:CE2	2.79	0.49
1:A:249:TYR:CZ	1:A:253:LEU:HD11	2.50	0.47
1:B:264:VAL:O	1:B:268:ILE:HG12	2.18	0.44
1:A:264:VAL:O	1:A:268:ILE:HG12	2.17	0.44
1:B:170:ALA:C	1:B:173:PRO:HD2	2.42	0.44
1:A:116:LEU:O	1:A:124:ARG:HD2	2.17	0.44
1:A:122:TYR:CD1	1:A:451:GLU:HB3	2.53	0.44
1:A:170:ALA:C	1:A:173:PRO:HD2	2.43	0.43
1:A:18:MET:HA	1:A:21:LYS:HE3	2.00	0.43
1:A:120:ASP:OD2	1:A:122:TYR:HB2	2.19	0.43
1:B:116:LEU:O	1:B:124:ARG:HD2	2.18	0.43
1:B:232:ALA:C	1:B:235:PRO:HD2	2.44	0.43
1:A:232:ALA:C	1:A:235:PRO:HD2	2.44	0.43
1:B:122:TYR:CD1	1:B:451:GLU:HB3	2.54	0.43
1:B:249:TYR:CZ	1:B:253:LEU:HD11	2.54	0.42
1:B:18:MET:HA	1:B:21:LYS:HE3	2.01	0.42
1:A:246:THR:HG21	1:A:383:ILE:HG22	2.02	0.41
1:A:417:LEU:HD23	1:A:443:PRO:HB2	2.01	0.41
1:B:120:ASP:OD2	1:B:122:TYR:HB2	2.19	0.41
1:A:468:ASP:HA	1:A:469:PRO:HD3	1.94	0.41
1:B:129:ARG:HD2	1:B:420:GLU:HB2	2.02	0.41
1:B:246:THR:HG21	1:B:383:ILE:HG22	2.01	0.41
1:A:102:GLN:NE2	3:A:610:HOH:O	2.53	0.41
1:B:301:LEU:HD21	1:B:309:ALA:HB1	2.02	0.41
1:B:417:LEU:HD23	1:B:443:PRO:HB2	2.02	0.41
1:B:468:ASP:HA	1:B:469:PRO:HD3	1.95	0.40

There are no symmetry-related clashes.

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	457/499 (92%)	454 (99%)	3 (1%)	0	100	100
1	B	456/499 (91%)	454 (100%)	2 (0%)	0	100	100
All	All	913/998 (92%)	908 (100%)	5 (0%)	0	100	100

There are no Ramachandran outliers to report.

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	343/372 (92%)	334 (97%)	9 (3%)	40	55
1	B	342/372 (92%)	331 (97%)	11 (3%)	34	47
All	All	685/744 (92%)	665 (97%)	20 (3%)	37	51

All (20) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	17	GLU
1	A	45	ARG
1	A	90	ARG
1	A	94	TYR
1	A	172	GLU
1	A	218	LEU
1	A	294	VAL
1	A	316	TYR

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	395	LYS
1	B	45	ARG
1	B	90	ARG
1	B	94	TYR
1	B	155	ARG
1	B	172	GLU
1	B	218	LEU
1	B	241	LYS
1	B	294	VAL
1	B	310	ARG
1	B	316	TYR
1	B	395	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	91	GLN
1	A	102	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 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	501	-	6,6,6	0.24	0	5,5,5	0.14	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	501	-	-	1/4/4/4	-

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	A	501	PEG	O2-C3-C4-O4

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, 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(Å ²)	Q<0.9
1	A	461/499 (92%)	0.31	18 (3%) 43 40	23, 42, 98, 114	0
1	B	460/499 (92%)	0.29	22 (4%) 35 32	23, 41, 98, 111	0
All	All	921/998 (92%)	0.30	40 (4%) 40 36	23, 41, 98, 114	0

All (40) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	317	LEU	6.0
1	A	317	LEU	5.7
1	B	316	TYR	4.3
1	B	308	PHE	4.2
1	A	308	PHE	4.0
1	A	314	VAL	3.7
1	A	316	TYR	3.4
1	B	312	VAL	3.4
1	A	294	VAL	3.3
1	B	290	GLY	3.3
1	B	289	ILE	3.2
1	B	122	TYR	3.1
1	B	314	VAL	3.1
1	A	296	ALA	2.9
1	B	280	TYR	2.8
1	B	306	THR	2.8
1	B	309	ALA	2.7
1	B	301	LEU	2.7
1	A	298	MET	2.6
1	B	304	THR	2.6
1	A	301	LEU	2.6
1	A	289	ILE	2.6
1	A	290	GLY	2.5
1	B	498	GLY	2.4

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	A	291	GLY	2.4
1	B	296	ALA	2.4
1	B	294	VAL	2.4
1	A	94	TYR	2.3
1	B	311	LYS	2.2
1	A	302	ALA	2.2
1	A	498	GLY	2.2
1	B	288	GLN	2.2
1	B	307	GLY	2.2
1	B	297	ALA	2.1
1	A	286	LEU	2.1
1	B	94	TYR	2.1
1	A	307	GLY	2.1
1	A	116	LEU	2.0
1	B	310	ARG	2.0
1	A	346	LYS	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 oligosaccharides 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, 95th 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($\text{\AA}^2$)	Q<0.9
2	PEG	A	501	7/7	0.74	0.22	64,64,64,64	0

6.5 Other polymers [i](#)

There are no such residues in this entry.