



# wwPDB X-ray Structure Validation Summary Report ⓘ

Nov 26, 2024 – 12:07 pm GMT

PDB ID : 8R7Z  
Title : transcription factor BARHL2 homodimer with spacing four bp  
Authors : Morgunova, E.; Popov, A.; Yin, Y.; Taipale, J.  
Deposited on : 2023-11-27  
Resolution : 3.26 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

<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 : 4.02b-467  
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) : 2.40

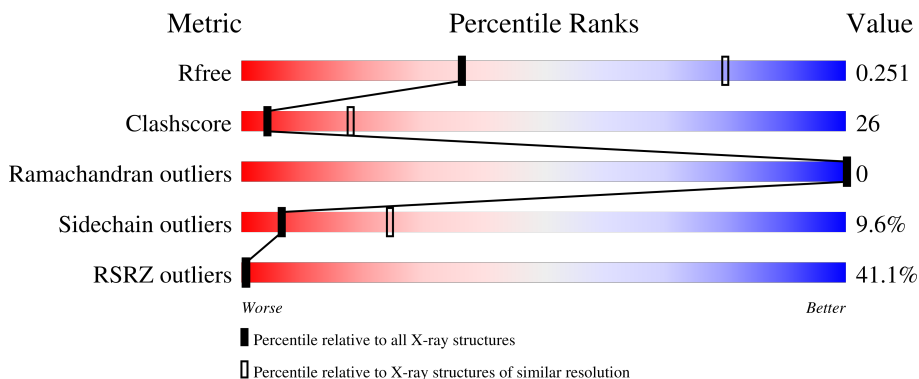
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 3.26 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.





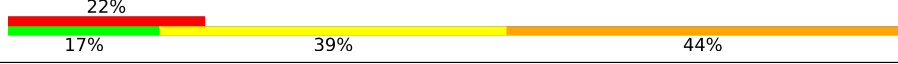
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
$R_{free}$	164625	1482 (3.30-3.22)
Clashscore	180529	1546 (3.30-3.22)
Ramachandran outliers	177936	1536 (3.30-3.22)
Sidechain outliers	177891	1535 (3.30-3.22)
RSRZ outliers	164620	1483 (3.30-3.22)

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 of the lower bar indicate the fraction of residues that contain outliers for  $\geq 3$ , 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	60	
1	E	60	
1	G	60	
1	K	60	
2	D	9	

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Length	Quality of chain
3	B	9	 11% 33% 56%
4	H	18	 17% 33% 22% 44%
5	I	18	 22% 17% 39% 44%

## 2 Entry composition [i](#)

There are 7 unique types of molecules in this entry. The entry contains 3073 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 BarH-like 2 homeobox protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	E	60	Total	C	N	O	S	0	0	0
			515	319	103	92	1			
1	A	60	Total	C	N	O	S	0	0	0
			514	317	103	93	1			
1	G	60	Total	C	N	O	S	0	0	0
			516	319	103	93	1			
1	K	48	Total	C	N	O	S	0	0	0
			414	256	79	78	1			

- Molecule 2 is a DNA chain called DNA (5'-D(P\*CP\*TP\*AP\*AP\*AP\*CP\*GP\*GP\*T)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
2	D	9	Total	C	N	O	P	0	0	0
			185	88	35	53	9			

- Molecule 3 is a DNA chain called DNA (5'-D(P\*AP\*CP\*CP\*GP\*TP\*TP\*TP\*AP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
3	B	9	Total	C	N	O	P	0	0	0
			184	88	32	55	9			

- Molecule 4 is a DNA chain called DNA (5'-D(P\*CP\*TP\*AP\*AP\*TP\*TP\*GP\*CP\*TP\*AP\*CP\*CP\*GP\*TP\*TP\*TP\*AP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
4	H	18	Total	C	N	O	P	0	0	0
			366	176	61	111	18			

- Molecule 5 is a DNA chain called DNA (5'-D(P\*CP\*TP\*AP\*AP\*AP\*CP\*GP\*GP\*TP\*AP\*GP\*CP\*AP\*AP\*TP\*TP\*AP\*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	P			
5	I	18	372	177	72	105	18	0	0	0

- Molecule 6 is SODIUM ION (three-letter code: NA) (formula: Na) (labeled as "Ligand of Interest" by depositor).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
6	E	1	Total	Na	0	0
			1	1		

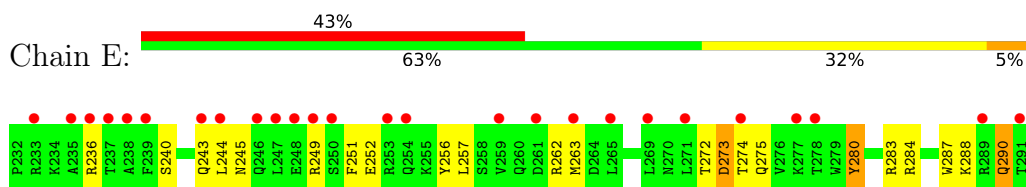
- Molecule 7 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
7	D	1	Total	O	0	0
			1	1		
7	B	3	Total	O	0	0
			3	3		
7	G	1	Total	O	0	0
			1	1		
7	H	1	Total	O	0	0
			1	1		

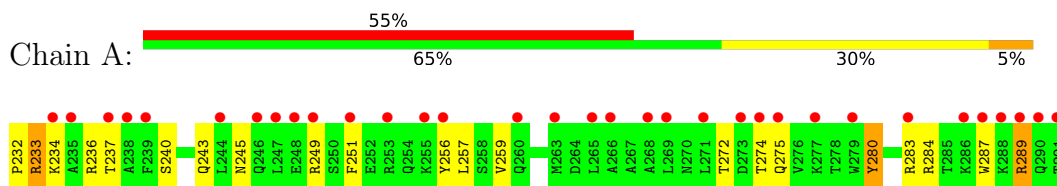
### 3 Residue-property plots [i](#)

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.

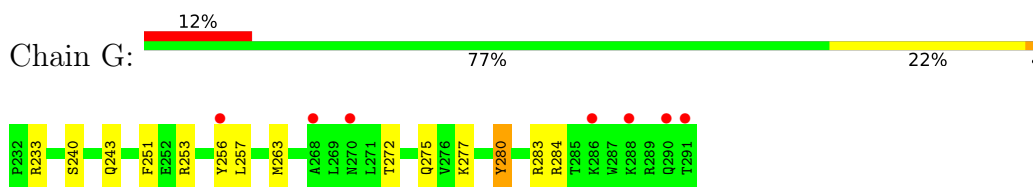
- Molecule 1: BarH-like 2 homeobox protein



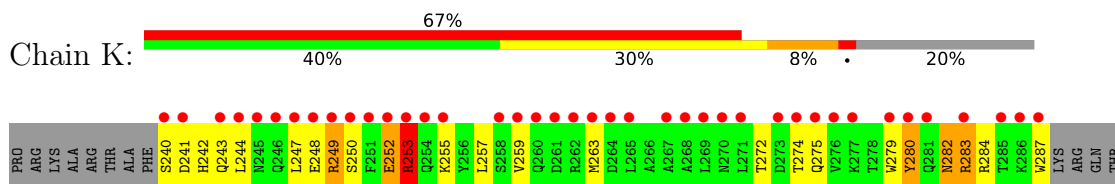
- Molecule 1: BarH-like 2 homeobox protein



- Molecule 1: BarH-like 2 homeobox protein



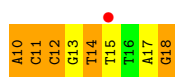
- Molecule 1: BarH-like 2 homeobox protein



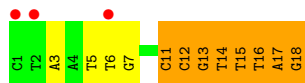
- Molecule 2: DNA (5'-D(P\*CP\*TP\*AP\*AP\*AP\*CP\*GP\*GP\*T)-3')



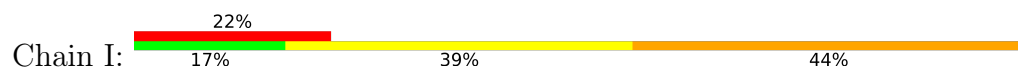
- Molecule 3: DNA (5'-D(P\*AP\*CP\*CP\*GP\*TP\*TP\*TP\*AP\*G)-3')



- Molecule 4: DNA (5'-D(P\*CP\*TP\*AP\*AP\*TP\*TP\*GP\*CP\*TP\*AP\*CP\*CP\*GP\*TP\*TP\*TP\*AP\*G)-3')



- Molecule 5: DNA (5'-D(P\*CP\*TP\*AP\*AP\*AP\*CP\*GP\*GP\*TP\*AP\*GP\*CP\*AP\*AP\*TP\*TP\*AP\*G)-3')



## 4 Data and refinement statistics i

Property	Value	Source
Space group	I 21 3	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	159.96Å 159.96Å 159.96Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	43.00 – 3.26 43.00 – 3.26	Depositor EDS
% Data completeness (in resolution range)	99.3 (43.00-3.26) 99.4 (43.00-3.26)	Depositor EDS
$R_{merge}$	0.18	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.01 (at 3.25Å)	Xtrriage
Refinement program	REFMAC 5.8.0419	Depositor
R, $R_{free}$	0.198 , 0.252 0.203 , 0.251	Depositor DCC
$R_{free}$ test set	529 reflections (4.87%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	141.9	Xtrriage
Anisotropy	0.000	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.21 , 999.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.46$ , $\langle L^2 \rangle = 0.29$	Xtrriage
Estimated twinning fraction	0.043 for -l,-k,-h	Xtrriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	3073	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	177.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.62% 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 [i](#)

### 5.1 Standard geometry [i](#)

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

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	0.46	0/522	1.00	2/700 (0.3%)
1	E	0.50	0/524	1.06	3/703 (0.4%)
1	G	0.34	0/525	0.83	0/704
1	K	0.34	0/421	0.96	0/567
2	D	0.93	0/207	2.84	17/317 (5.4%)
3	B	0.85	0/205	2.23	14/314 (4.5%)
4	H	1.30	5/408 (1.2%)	1.99	18/627 (2.9%)
5	I	1.07	2/418 (0.5%)	1.83	20/643 (3.1%)
All	All	0.76	7/3230 (0.2%)	1.57	74/4575 (1.6%)

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
1	G	0	1
1	K	0	1
All	All	0	3

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	H	17	DA	N3-C4	7.78	1.39	1.34
4	H	17	DA	C6-N1	7.54	1.40	1.35
4	H	17	DA	N7-C5	6.76	1.43	1.39
4	H	17	DA	N9-C8	6.73	1.43	1.37
5	I	1	DC	N1-C6	-6.63	1.33	1.37

The worst 5 of 74 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	D	3	DA	O5'-P-OP2	-22.26	83.99	110.70
2	D	3	DA	O5'-P-OP1	17.93	132.22	110.70
2	D	5	DA	O5'-P-OP1	-12.48	94.47	105.70
4	H	17	DA	O5'-P-OP1	-11.58	95.28	105.70
5	I	7	DG	OP2-P-O3'	10.22	127.69	105.20

There are no chirality outliers.

All (3) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	289	ARG	Sidechain
1	G	253	ARG	Sidechain
1	K	253	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	514	0	510	12	0
1	E	515	0	516	13	0
1	G	516	0	516	10	0
1	K	414	0	402	49	0
2	D	185	0	102	9	2
3	B	184	0	103	22	2
4	H	366	0	206	23	0
5	I	372	0	203	34	0
6	E	1	0	0	0	0
7	B	3	0	0	0	0
7	D	1	0	0	1	0
7	G	1	0	0	0	0
7	H	1	0	0	0	0
All	All	3073	0	2558	147	2

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 26.

The worst 5 of 147 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:K:255:LYS:NZ	1:K:287:TRP:CZ2	2.15	1.13
1:K:283:ARG:HH21	1:K:283:ARG:HA	1.10	1.12
4:H:17:DA:H4'	4:H:18:DG:OP1	1.51	1.07
1:K:249:ARG:O	1:K:252:GLU:HG3	1.62	0.96
5:I:8:DG:O3'	1:K:242:HIS:ND1	2.03	0.91

All (2) 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:D:9:DT:O3'	3:B:10:DA:OP1[15_455]	1.94	0.26
2:D:9:DT:O3'	3:B:10:DA:P[15_455]	1.96	0.24

## 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	58/60 (97%)	55 (95%)	3 (5%)	0	100	100
1	E	58/60 (97%)	55 (95%)	3 (5%)	0	100	100
1	G	58/60 (97%)	54 (93%)	4 (7%)	0	100	100
1	K	46/60 (77%)	42 (91%)	4 (9%)	0	100	100
All	All	220/240 (92%)	206 (94%)	14 (6%)	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	54/55 (98%)	49 (91%)	5 (9%)	7	25
1	E	55/55 (100%)	50 (91%)	5 (9%)	7	26
1	G	55/55 (100%)	52 (94%)	3 (6%)	18	44
1	K	45/55 (82%)	38 (84%)	7 (16%)	2	9
All	All	209/220 (95%)	189 (90%)	20 (10%)	7	24

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	K	252	GLU
1	K	280	TYR
1	K	283	ARG
1	K	282	ASN
1	A	249	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 6 such sidechains are listed below:

Mol	Chain	Res	Type
1	K	243	GLN
1	K	245	ASN
1	K	282	ASN
1	A	290	GLN
1	E	245	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 oligosaccharides in this entry.

## 5.6 Ligand geometry [i](#)

Of 1 ligands modelled in this entry, 1 is monoatomic - leaving 0 for Mogul analysis.

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<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	60/60 (100%)	2.56	33 (55%) 0 0	107, 149, 204, 235	0
1	E	60/60 (100%)	2.22	26 (43%) 1 1	93, 120, 187, 245	0
1	G	60/60 (100%)	0.71	7 (11%) 10 10	193, 241, 298, 318	0
1	K	48/60 (80%)	4.86	40 (83%) 0 0	62, 82, 99, 121	48 (100%)
2	D	9/9 (100%)	1.15	2 (22%) 3 2	109, 130, 162, 181	0
3	B	9/9 (100%)	0.33	1 (11%) 12 11	111, 139, 153, 153	0
4	H	18/18 (100%)	1.36	3 (16%) 5 5	147, 194, 264, 265	0
5	I	18/18 (100%)	2.08	4 (22%) 3 2	126, 197, 267, 278	0
All	All	282/294 (95%)	2.26	116 (41%) 1 1	62, 147, 267, 318	48 (17%)

The worst 5 of 116 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	K	249	ARG	17.7
1	K	268	ALA	15.6
1	K	255	LYS	10.8
5	I	14	DA	9.7
1	K	244	LEU	9.6

### 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](#)

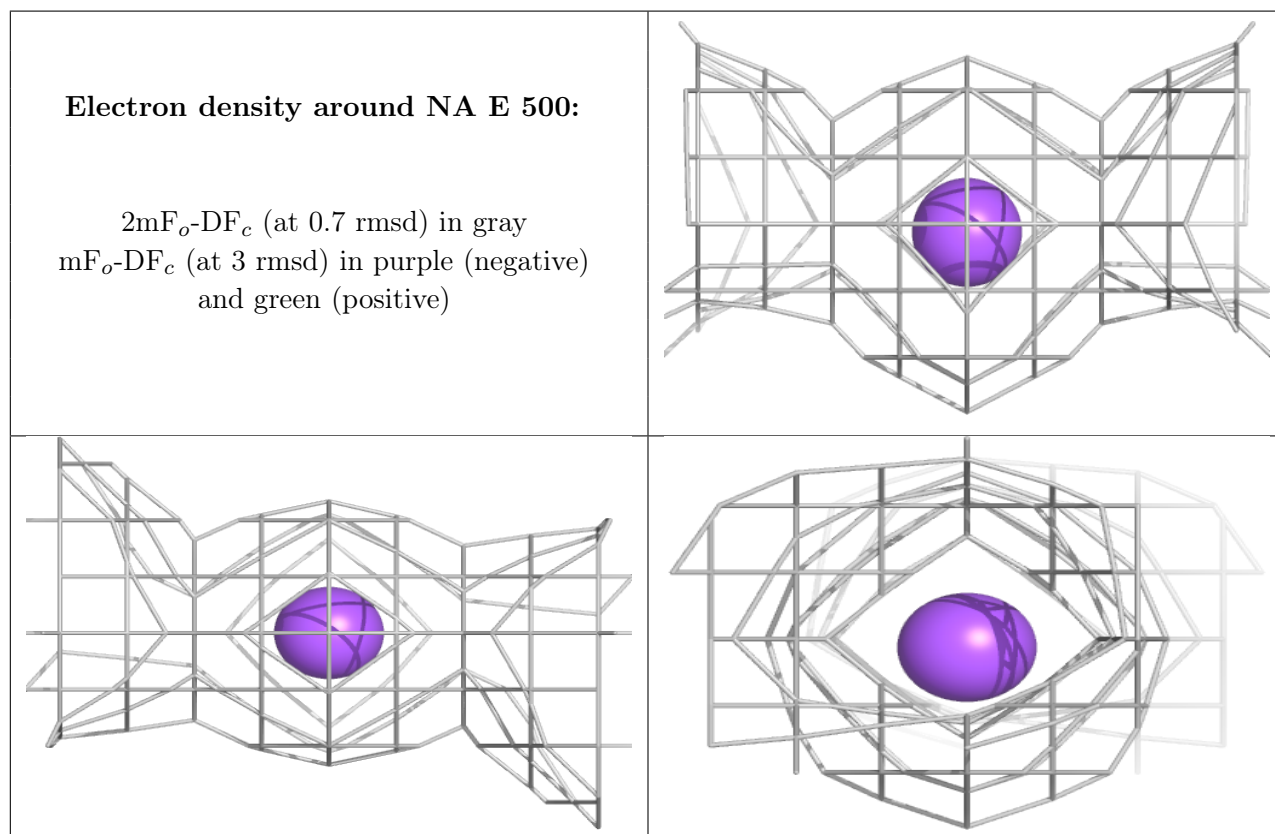
There are no monosaccharides in this entry.

## 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<sup>th</sup> 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
6	NA	E	500	1/1	0.98	0.05	101,101,101,101	1

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.



## 6.5 Other polymers [i](#)

There are no such residues in this entry.