Full wwPDB X-ray Structure Validation Report ⓘ

Aug 22, 2020 – 07:11 AM BST

PDB ID	:	6KI6
Title	:	Crystal structure of BCL11A in complex with gamma-globin -115 HPFH region
Authors	:	Li, F.D.; Yang, Y.; Shi, Y.Y.
Deposited on	:	2019-07-17
Resolution	:	2.50 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 following versions of software and data (see references ⓘ) were used in the production of this report:

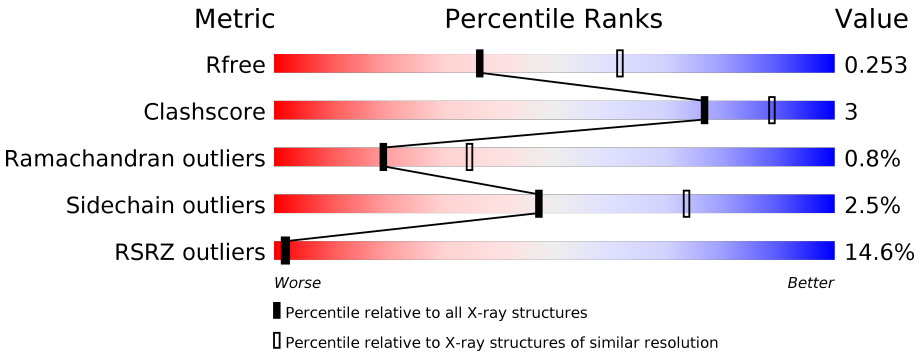
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.13.1
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.13.1

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 2.50 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}	130704	4661 (2.50-2.50)
Clashscore	141614	5346 (2.50-2.50)
Ramachandran outliers	138981	5231 (2.50-2.50)
Sidechain outliers	138945	5233 (2.50-2.50)
RSRZ outliers	127900	4559 (2.50-2.50)

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	A	109	3% 72% 26%
1	B	109	19% 38% 10% 52%
2	C	13	85% 15%
2	F	13	8% 100%
3	D	13	8% 85% 15%
3	E	13	8% 92% 8%

2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 2179 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 B-cell lymphoma/leukemia 11A.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	81	Total	C	N	O	S	0	0	0
			655	406	120	118	11			
1	B	52	Total	C	N	O	S	0	0	0
			396	240	74	75	7			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	727	GLY	-	expression tag	UNP Q9H165
A	728	SER	-	expression tag	UNP Q9H165
A	729	HIS	-	expression tag	UNP Q9H165
A	730	MET	-	expression tag	UNP Q9H165
B	727	GLY	-	expression tag	UNP Q9H165
B	728	SER	-	expression tag	UNP Q9H165
B	729	HIS	-	expression tag	UNP Q9H165
B	730	MET	-	expression tag	UNP Q9H165

- Molecule 2 is a DNA chain called DNA (5'-D(*AP*TP*AP*TP*TP*GP*GP*TP*CP*AP*AP*GP*G)-3').

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	13	Total	C	N	O	P	0	0	0
			268	129	51	76	12			
2	F	13	Total	C	N	O	P	0	0	0
			268	129	51	76	12			

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	D	13	Total	C	N	O	P	0	0	0
			259	126	45	76	12			

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	E	13	Total	C	N	O	P	0	0	0
			259	126	45	76	12			

- Molecule 4 is ZINC ION (three-letter code: ZN) (formula: Zn) (labeled as "Ligand of Interest" by author).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	2	Total	Zn	0	0
			2	2		
4	A	3	Total	Zn	0	0
			3	3		

- Molecule 5 is water.

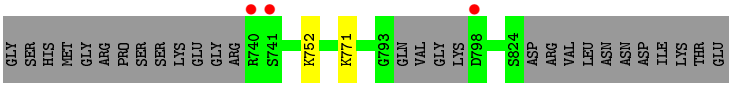
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	37	Total	O	0	0
			37	37		
5	B	1	Total	O	0	0
			1	1		
5	C	15	Total	O	0	0
			15	15		
5	D	11	Total	O	0	0
			11	11		
5	E	1	Total	O	0	0
			1	1		
5	F	4	Total	O	0	0
			4	4		

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.

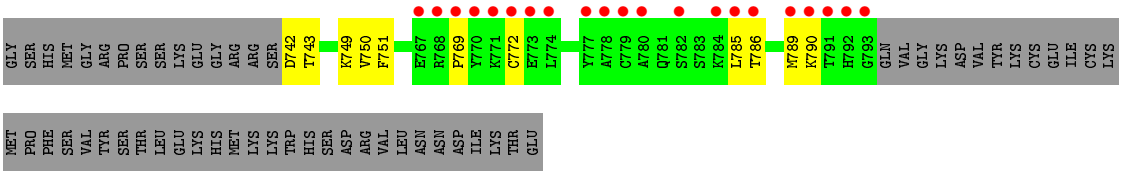
- Molecule 1: B-cell lymphoma/leukemia 11A

Chain A:



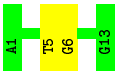
- Molecule 1: B-cell lymphoma/leukemia 11A

Chain B:



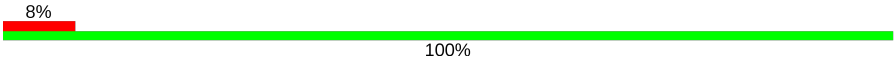
- Molecule 2: DNA (5'-D(*AP*TP*AP*TP*TP*GP*GP*TP*CP*AP*AP*GP*G)-3')

Chain C:



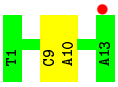
- Molecule 2: DNA (5'-D(*AP*TP*AP*TP*TP*GP*GP*TP*CP*AP*AP*GP*G)-3')

Chain F:



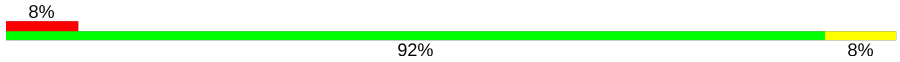
- Molecule 3: DNA (5'-D(*TP*CP*CP*TP*TP*GP*AP*CP*CP*AP*AP*TP*A)-3')

Chain D:



• Molecule 3: DNA (5'-D(*TP*CP*CP*TP*TP*GP*AP*CP*CP*AP*AP*TP*A)-3')

Chain E:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	59.55Å 59.55Å 213.37Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	37.08 – 2.50 37.08 – 2.50	Depositor EDS
% Data completeness (in resolution range)	99.1 (37.08-2.50) 99.1 (37.08-2.50)	Depositor EDS
R_{merge}	0.08	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.33 (at 2.51Å)	Xtriage
Refinement program	PHENIX 1.14_3260	Depositor
R, R_{free}	0.207 , 0.251 0.207 , 0.253	Depositor DCC
R_{free} test set	745 reflections (4.70%)	wwPDB-VP
Wilson B-factor (Å^2)	38.1	Xtriage
Anisotropy	0.007	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 60.8	EDS
L-test for twinning[2]	$< \|L\| > = 0.50$, $< L^2 > = 0.33$	Xtriage
Estimated twinning fraction	0.031 for -h,-k,l	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	2179	wwPDB-VP
Average B, all atoms (Å^2)	60.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.14% 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: ZN

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	0.25	0/670	0.40	0/895
1	B	0.24	0/403	0.43	0/541
2	C	0.53	0/301	0.94	0/464
2	F	0.53	0/301	0.91	0/464
3	D	0.51	0/289	0.97	0/443
3	E	0.51	0/289	0.95	0/443
All	All	0.41	0/2253	0.76	0/3250

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	655	0	628	2	0
1	B	396	0	365	5	0
2	C	268	0	149	1	0
2	F	268	0	149	0	0
3	D	259	0	149	1	0
3	E	259	0	149	1	0
4	A	3	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
4	B	2	0	0	0	0
5	A	37	0	0	0	0
5	B	1	0	0	0	0
5	C	15	0	0	0	0
5	D	11	0	0	0	0
5	E	1	0	0	0	0
5	F	4	0	0	0	0
All	All	2179	0	1589	10	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 (10) 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:772:CYS:HB2	1:B:785:LEU:HD11	1.64	0.79
1:B:743:THR:HG22	1:B:750:VAL:HG12	1.77	0.65
1:A:752:LYS:N	1:A:752:LYS:HD2	2.22	0.55
2:C:5:DT:H2"	2:C:6:DG:C8	2.47	0.49
1:B:772:CYS:SG	1:B:789:MET:CE	3.00	0.49
3:E:13:DA:H2'	3:E:13:DA:N3	2.31	0.46
1:B:749:LYS:HD3	1:B:751:PHE:CZ	2.53	0.43
3:D:9:DC:H2"	3:D:10:DA:C8	2.55	0.42
1:B:786:THR:O	1:B:789:MET:HB2	2.20	0.42
1:A:752:LYS:HD2	1:A:752:LYS:H	1.85	0.41

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	77/109 (71%)	77 (100%)	0	0	100 100
1	B	50/109 (46%)	46 (92%)	3 (6%)	1 (2%)	7 12
All	All	127/218 (58%)	123 (97%)	3 (2%)	1 (1%)	19 35

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	769	PRO

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	75/100 (75%)	74 (99%)	1 (1%)	69 87
1	B	43/100 (43%)	41 (95%)	2 (5%)	26 49
All	All	118/200 (59%)	115 (98%)	3 (2%)	47 73

All (3) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	771	LYS
1	B	742	ASP
1	B	790	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

There are no monosaccharides in this entry.

5.6 Ligand geometry

Of 5 ligands modelled in this entry, 5 are 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

There are no such residues in this entry.

5.8 Polymer linkage issues

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	81/109 (74%)	0.14	3 (3%) 41 45	20, 32, 63, 97	0
1	B	52/109 (47%)	1.85	21 (40%) 0 0	40, 75, 120, 161	0
2	C	13/13 (100%)	-0.15	0 100 100	30, 34, 56, 61	0
2	F	13/13 (100%)	0.46	1 (7%) 13 13	63, 87, 104, 122	0
3	D	13/13 (100%)	0.17	1 (7%) 13 13	31, 52, 62, 131	0
3	E	13/13 (100%)	0.71	1 (7%) 13 13	77, 96, 129, 168	0
All	All	185/270 (68%)	0.66	27 (14%) 2 2	20, 49, 114, 168	0

All (27) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	772	CYS	8.1
1	B	773	GLU	6.1
1	B	770	TYR	5.8
1	B	793	GLY	5.7
1	B	785	LEU	5.6
1	B	774	LEU	4.5
1	B	792	HIS	4.2
3	E	1	DT	3.9
1	B	782	SER	3.7
3	D	13	DA	3.5
1	B	777	TYR	3.4
1	B	769	PRO	3.3
1	B	778	ALA	3.3
2	F	13	DG	3.3
1	B	786	THR	3.0
1	B	789	MET	2.8
1	B	768	ARG	2.8
1	B	771	LYS	2.8
1	B	790	LYS	2.8

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	B	791	THR	2.7
1	A	740	ARG	2.7
1	B	784	LYS	2.6
1	A	741	SER	2.5
1	B	780	ALA	2.3
1	B	767	GLU	2.3
1	B	779	CYS	2.1
1	A	798	ASP	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 monosaccharides in this entry.

6.4 Ligands (i)

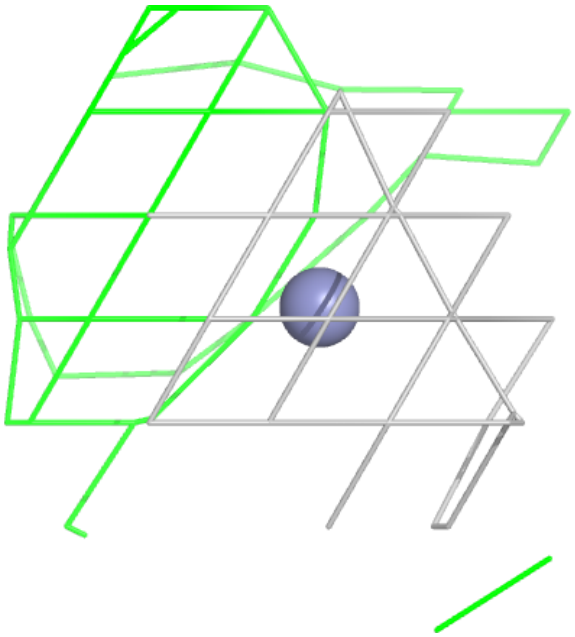
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95^{th} percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9'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

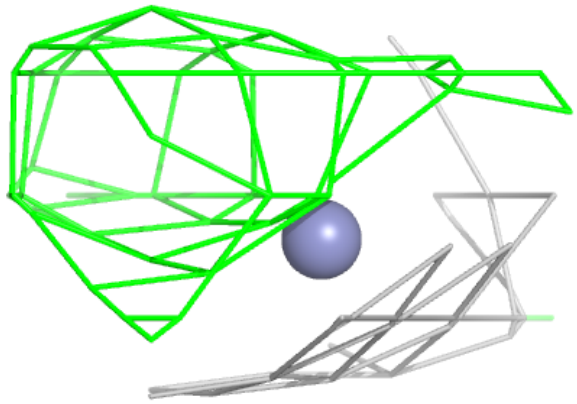
Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	ZN	B	902	1/1	0.69	0.16	151,151,151,151	0
4	ZN	B	901	1/1	0.99	0.15	39,39,39,39	0
4	ZN	A	901	1/1	0.99	0.18	31,31,31,31	0
4	ZN	A	903	1/1	0.99	0.15	29,29,29,29	0
4	ZN	A	902	1/1	1.00	0.16	24,24,24,24	0

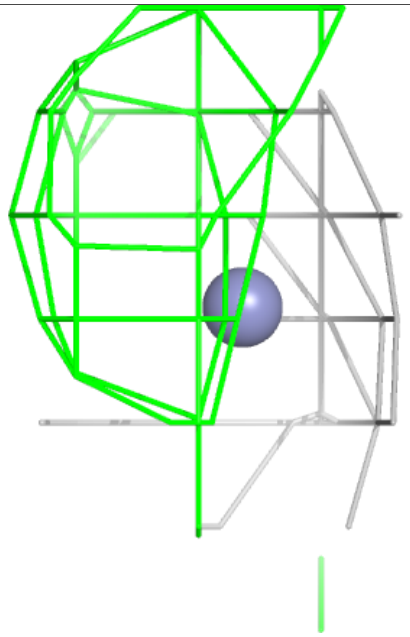
The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation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

Electron density around ZN B 902:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)

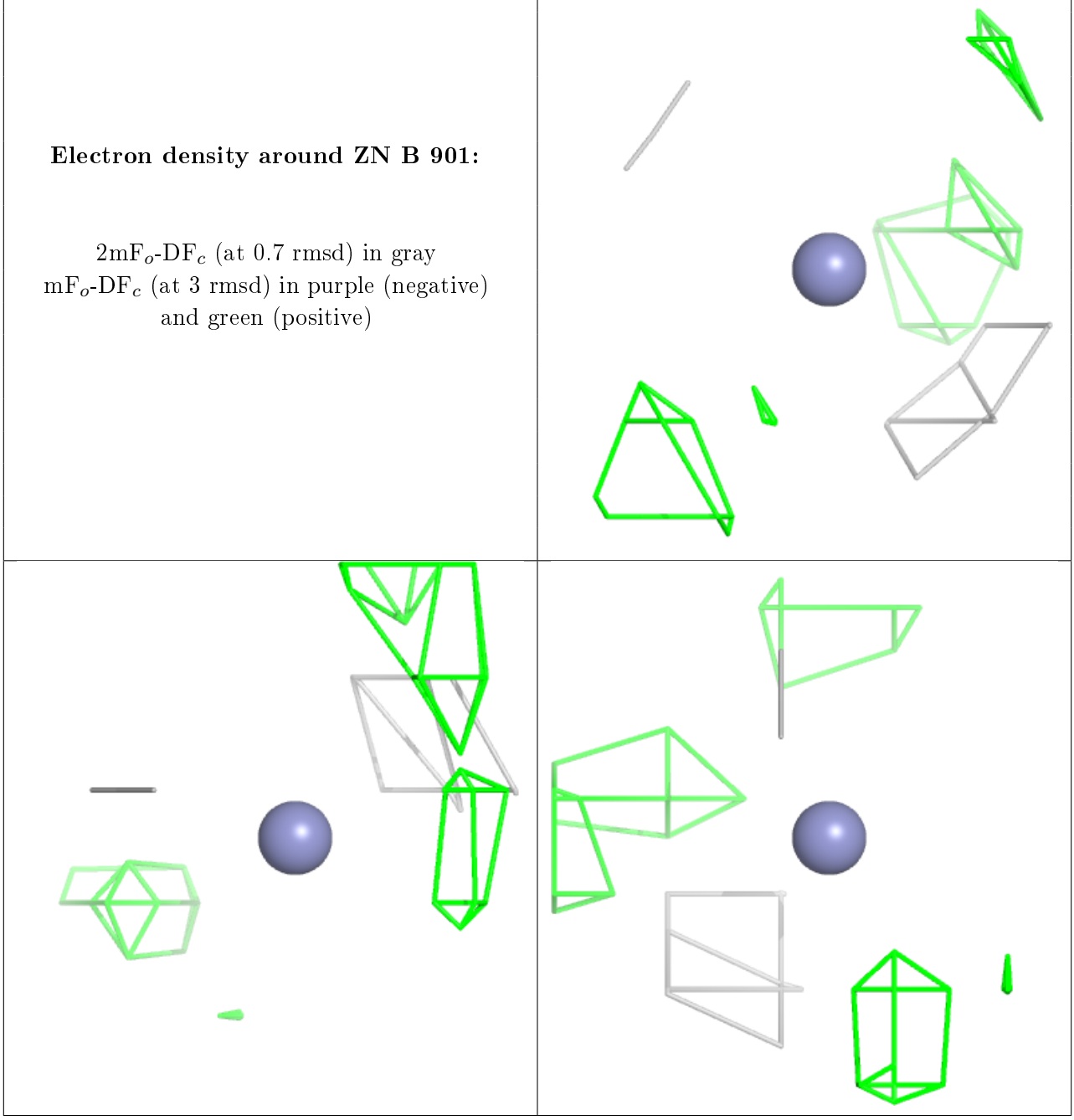






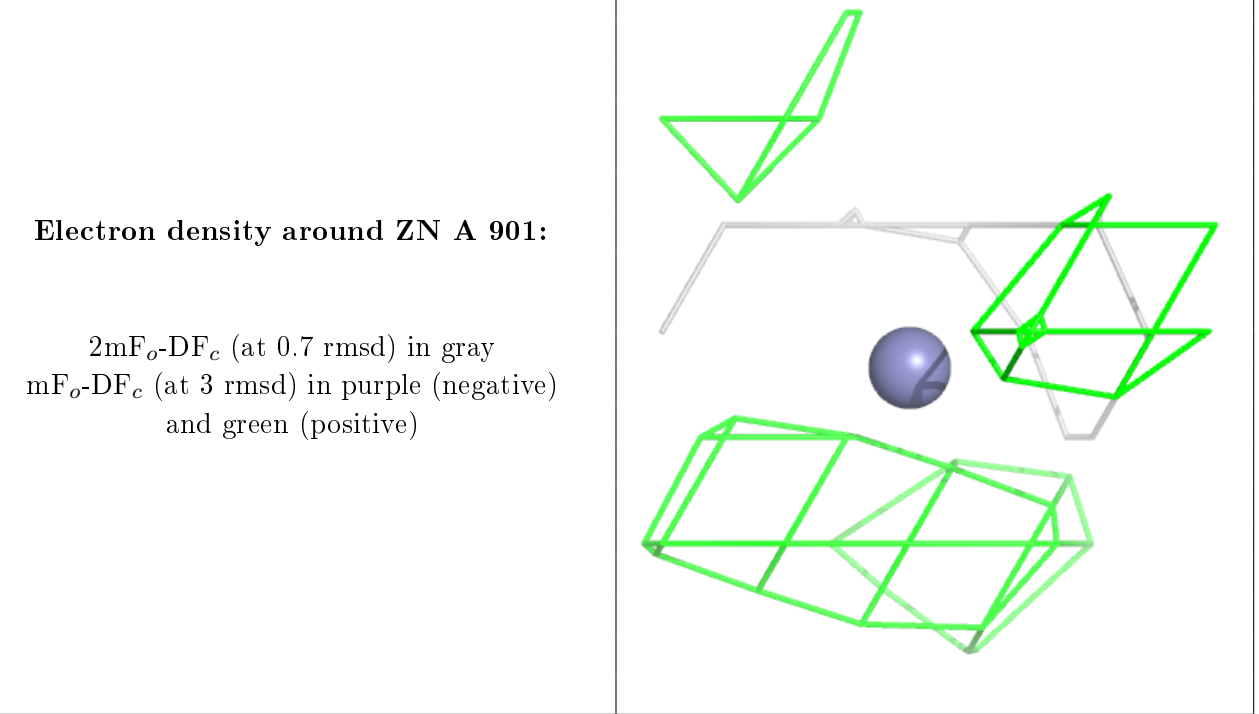
Electron density around ZN B 901:

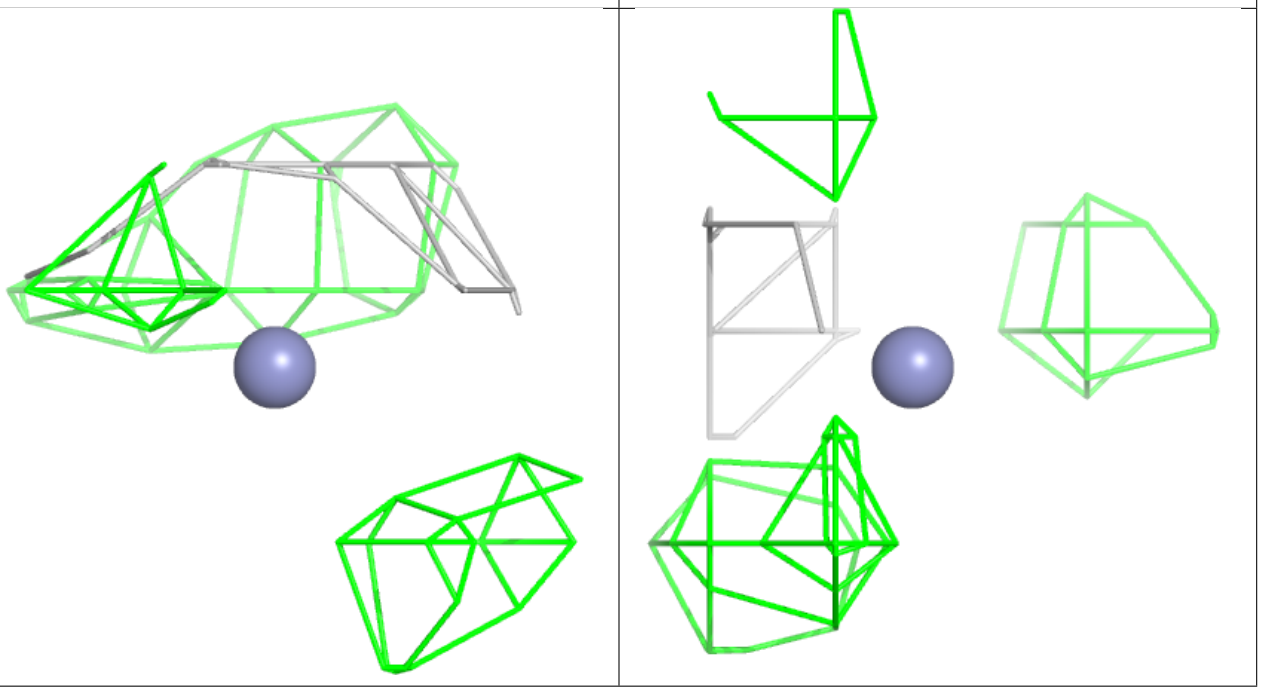
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



Electron density around ZN A 901:

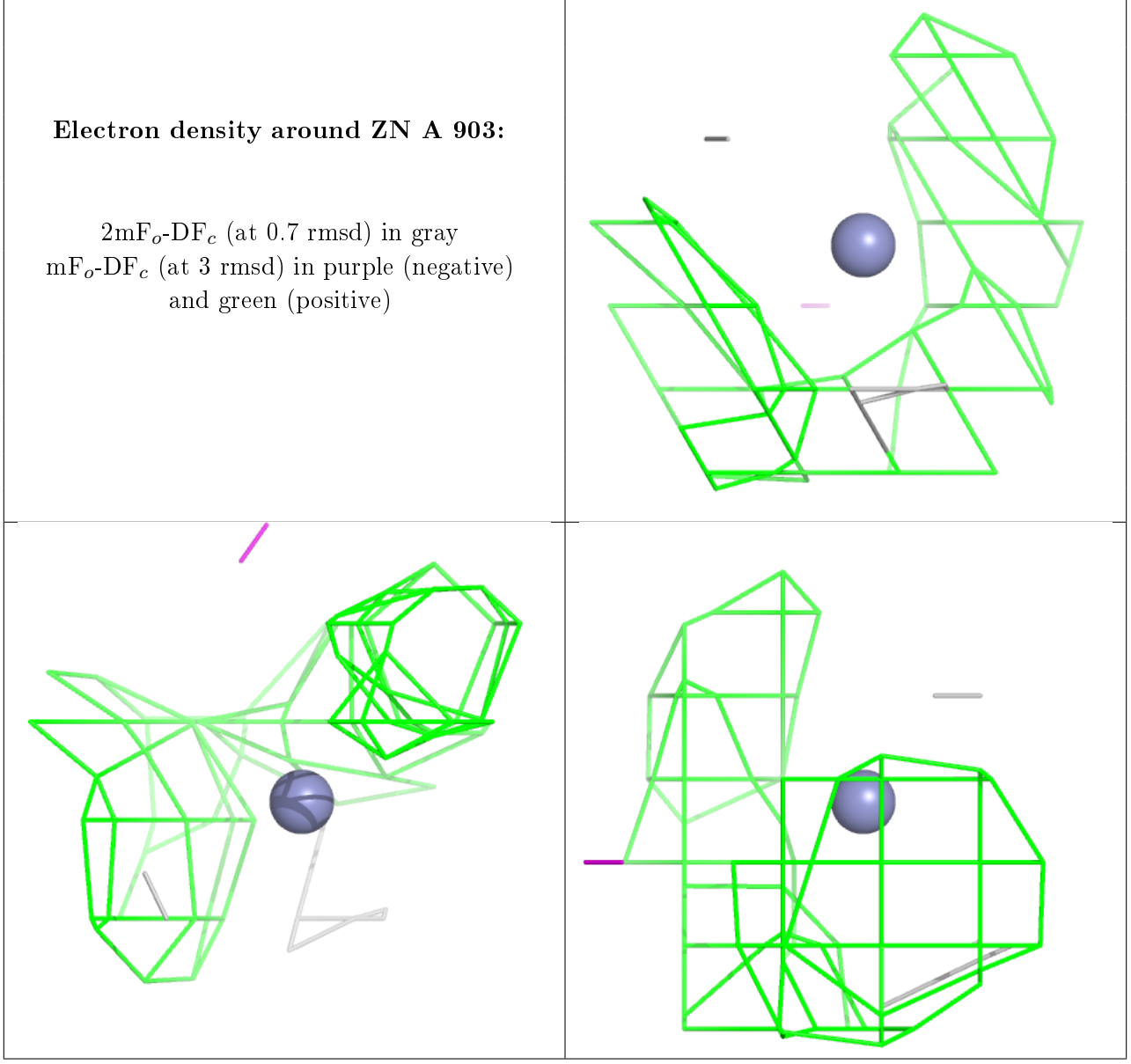
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)





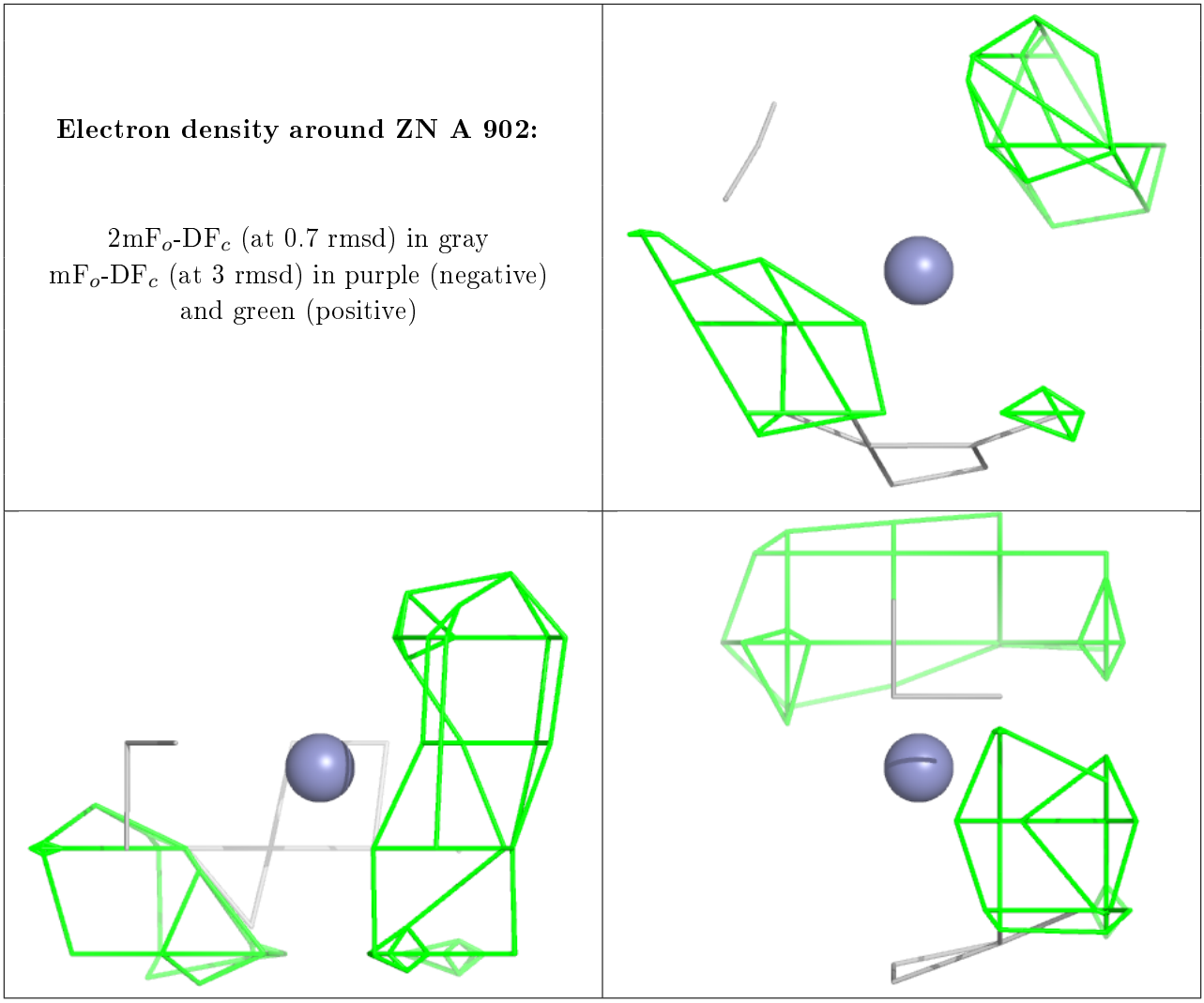
Electron density around ZN A 903:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



Electron density around ZN A 902:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative) and green (positive)



6.5 Other polymers

There are no such residues in this entry.