Full wwPDB X-ray Structure Validation Report ⓘ

May 12, 2020 – 11:31 pm BST

PDB ID	:	1GO3
Title	:	Structure of an archeal homolog of the eukaryotic RNA polymerase II RPB4/RPB7 complex
Authors	:	Todone, F.; Brick, P.; Werner, F.; Weinzierl, R.O.J.; Onesti, S.
Deposited on	:	2001-10-17
Resolution	:	1.75 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:

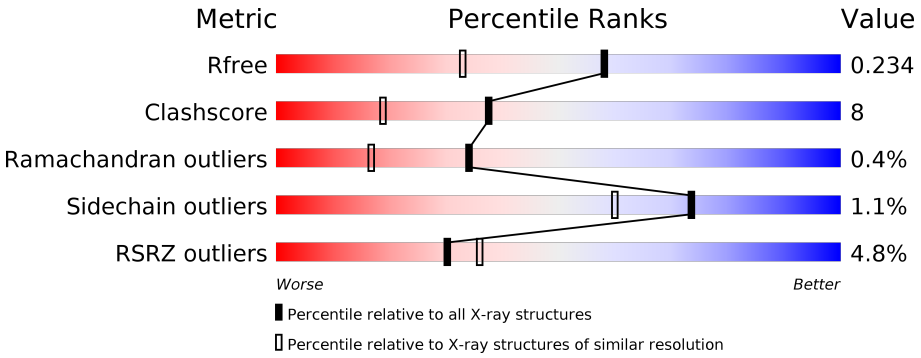
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.11

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 1.75 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	2340 (1.76-1.76)
Clashscore	141614	2466 (1.76-1.76)
Ramachandran outliers	138981	2437 (1.76-1.76)
Sidechain outliers	138945	2437 (1.76-1.76)
RSRZ outliers	127900	2298 (1.76-1.76)

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	E	187	6% 83% 11% • 5%
1	M	187	4% 81% 12% 7%
2	F	107	7% 79% 20% •
2	N	107	% 92% 7% ••

2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4695 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 DNA-DIRECTED RNA POLYMERASE SUBUNIT E.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	E	177	Total	C	N	O	S	0	0	0
			1356	888	209	254	5			
1	M	174	Total	C	N	O	S	0	0	0
			1344	877	208	254	5			

- Molecule 2 is a protein called DNA-DIRECTED RNA POLYMERASE SUBUNIT F.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	F	107	Total	C	N	O	S	0	0	0
			836	539	132	162	3			
2	N	106	Total	C	N	O	S	0	0	0
			831	538	135	156	2			

- Molecule 3 is water.

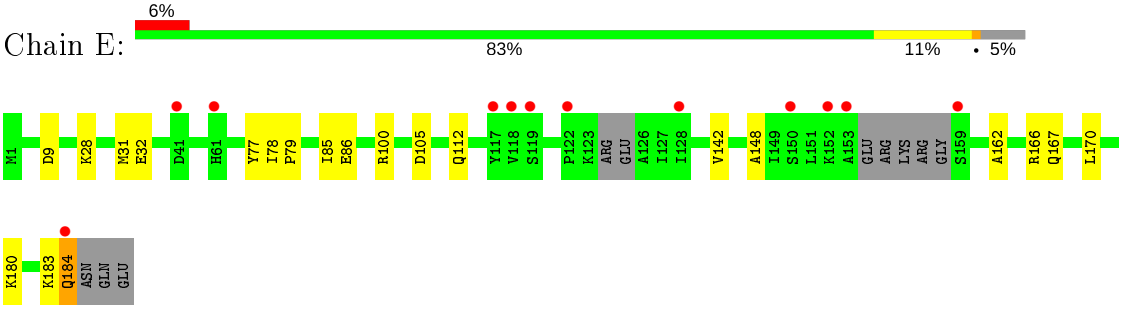
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	E	78	Total	O	0	0
			78	78		
3	F	66	Total	O	0	0
			66	66		
3	M	103	Total	O	0	0
			103	103		
3	N	81	Total	O	0	0
			81	81		

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.

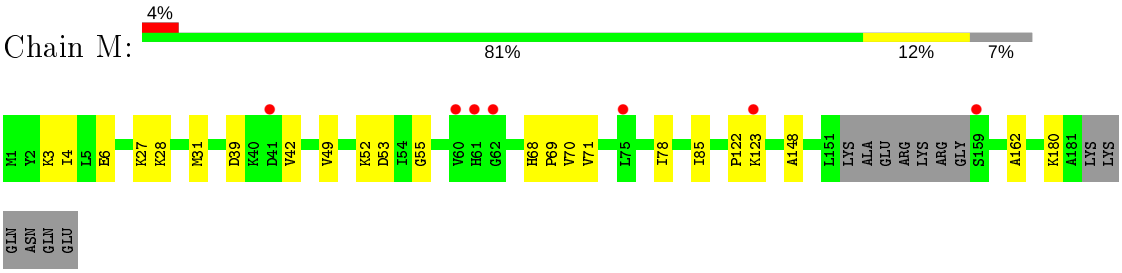
- Molecule 1: DNA-DIRECTED RNA POLYMERASE SUBUNIT E

Chain E:



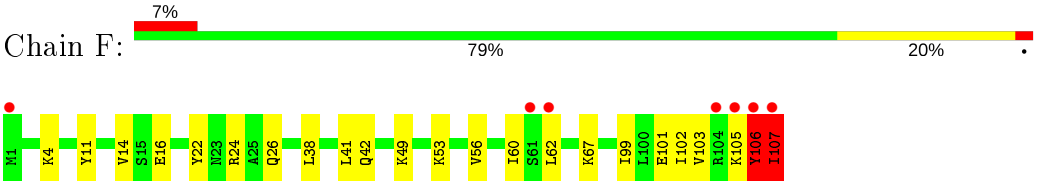
- Molecule 1: DNA-DIRECTED RNA POLYMERASE SUBUNIT E

Chain M:



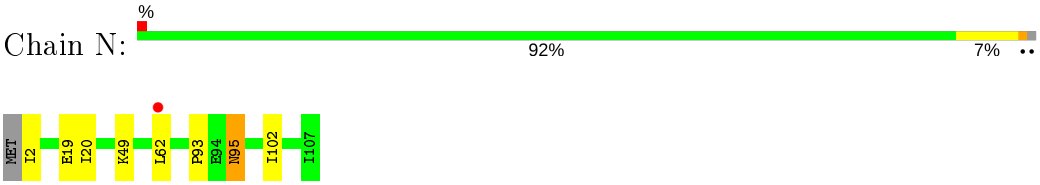
- Molecule 2: DNA-DIRECTED RNA POLYMERASE SUBUNIT F

Chain F:



- Molecule 2: DNA-DIRECTED RNA POLYMERASE SUBUNIT F

Chain N:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43	Depositor
Cell constants a, b, c, α, β, γ	92.39Å 92.39Å 91.11Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	27.80 – 1.75 27.82 – 1.75	Depositor EDS
% Data completeness (in resolution range)	99.6 (27.80-1.75) 99.7 (27.82-1.75)	Depositor EDS
R_{merge}	0.09	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.23 (at 1.75Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, R_{free}	0.217 , 0.238 0.213 , 0.234	Depositor DCC
R_{free} test set	3899 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	16.9	Xtriage
Anisotropy	0.434	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.39 , 49.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.48$, $< L^2 > = 0.30$	Xtriage
Estimated twinning fraction	0.022 for -h,-l,-k 0.012 for -h,l,k 0.012 for l,-k,h 0.023 for -l,-k,-h 0.057 for h,-k,-l	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4695	wwPDB-VP
Average B, all atoms (Å^2)	20.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.93% 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

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	E	0.27	0/1379	0.62	0/1866
1	M	0.28	0/1368	0.66	0/1853
2	F	0.39	0/845	0.72	4/1137 (0.4%)
2	N	0.26	0/840	0.50	0/1129
All	All	0.30	0/4432	0.63	4/5985 (0.1%)

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
2	F	0	1

There are no bond length outliers.

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	F	107	ILE	CG1-CB-CG2	-7.85	94.12	111.40
2	F	106	TYR	CA-CB-CG	-7.73	98.72	113.40
2	F	107	ILE	N-CA-C	-7.20	91.56	111.00
2	F	106	TYR	CB-CA-C	5.32	121.03	110.40

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	F	106	TYR	Sidechain

5.2 Too-close contacts (i)

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	E	1356	0	1345	16	0
1	M	1344	0	1340	18	0
2	F	836	0	846	27	0
2	N	831	0	849	10	0
3	E	78	0	0	0	0
3	F	66	0	0	0	0
3	M	103	0	0	1	0
3	N	81	0	0	0	0
All	All	4695	0	4380	67	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 8.

All (67) 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)
2:F:107:ILE:CG1	2:F:107:ILE:CD1	1.78	1.62
2:F:105:LYS:O	2:F:107:ILE:N	1.95	0.99
2:F:105:LYS:C	2:F:107:ILE:HG12	1.89	0.92
2:F:106:TYR:O	2:F:107:ILE:HG23	1.79	0.81
2:F:105:LYS:O	2:F:105:LYS:HG2	1.83	0.79
2:F:106:TYR:O	2:F:107:ILE:HD13	1.90	0.72
1:E:148:ALA:HB3	1:E:162:ALA:HB3	1.73	0.71
2:F:107:ILE:CG2	2:F:107:ILE:CD1	2.71	0.67
1:M:148:ALA:HB3	1:M:162:ALA:HB3	1.77	0.66
1:M:122:PRO:HG2	1:M:123:LYS:HD2	1.77	0.65
2:F:102:ILE:O	2:F:105:LYS:HB3	1.98	0.63
1:E:28:LYS:O	1:E:32:GLU:HG3	1.99	0.63
1:M:27:LYS:O	1:M:31:MET:HG3	1.99	0.61
2:F:107:ILE:CD1	2:F:107:ILE:HG23	2.32	0.59
1:E:85:ILE:C	1:E:85:ILE:HD12	2.23	0.59
1:E:167:GLN:HB2	1:E:170:LEU:HD12	1.85	0.59
2:N:62:LEU:HD11	2:N:102:ILE:CD1	2.34	0.58
2:N:62:LEU:HD11	2:N:102:ILE:HD12	1.86	0.57
1:E:77:TYR:CE2	1:E:79:PRO:HG3	2.40	0.56

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:183:LYS:C	1:E:184:GLN:HG2	2.25	0.56
2:F:105:LYS:C	2:F:107:ILE:N	2.59	0.55
1:M:85:ILE:C	1:M:85:ILE:HD12	2.28	0.54
1:M:70:VAL:HG12	1:M:71:VAL:N	2.22	0.54
2:N:19:GLU:CD	2:N:49:LYS:HD3	2.28	0.53
2:N:95:ASN:H	2:N:95:ASN:HD22	1.55	0.53
2:F:56:VAL:O	2:F:60:ILE:HG13	2.08	0.53
2:N:95:ASN:ND2	2:N:95:ASN:H	2.07	0.52
2:F:38:LEU:O	2:F:42:GLN:HG3	2.10	0.52
1:M:28:LYS:HA	1:M:31:MET:CE	2.40	0.51
2:F:22:TYR:O	2:F:26:GLN:HG2	2.09	0.51
2:F:106:TYR:C	2:F:107:ILE:CG1	2.78	0.51
2:F:103:VAL:C	2:F:105:LYS:H	2.15	0.50
2:F:106:TYR:N	2:F:107:ILE:HG12	2.26	0.50
1:M:28:LYS:HA	1:M:31:MET:HE3	1.93	0.49
2:F:11:TYR:CZ	2:F:67:LYS:HE2	2.48	0.48
2:N:95:ASN:N	2:N:95:ASN:HD22	2.09	0.48
2:F:105:LYS:CA	2:F:107:ILE:HG12	2.44	0.48
2:F:49:LYS:O	2:F:53:LYS:HG3	2.14	0.47
1:E:180:LYS:O	1:E:184:GLN:N	2.45	0.47
2:F:62:LEU:HD11	2:F:102:ILE:CD1	2.45	0.47
2:N:93:PRO:HB2	2:N:95:ASN:ND2	2.30	0.47
2:N:19:GLU:OE2	2:N:49:LYS:HD3	2.15	0.46
1:M:39:ASP:HB3	1:M:42:VAL:HB	1.98	0.46
1:M:70:VAL:CG1	1:M:71:VAL:N	2.78	0.45
1:M:123:LYS:N	1:M:123:LYS:HD2	2.31	0.45
2:F:62:LEU:HD11	2:F:102:ILE:HD12	1.99	0.45
1:M:6:GLU:OE1	1:M:52:LYS:HE2	2.17	0.44
1:E:86:GLU:HA	1:E:142:VAL:O	2.18	0.44
1:M:42:VAL:HG21	2:N:2:ILE:CG2	2.48	0.44
2:F:16:GLU:HA	2:F:49:LYS:HG3	1.99	0.43
2:F:99:ILE:O	2:F:103:VAL:HG23	2.18	0.43
1:M:68:HIS:HE1	3:M:2039:HOH:O	2.02	0.42
1:M:55:GLY:HA3	1:M:69:PRO:HG2	2.01	0.42
1:E:184:GLN:HE21	1:E:184:GLN:HB3	1.66	0.42
1:M:78:ILE:O	1:M:78:ILE:HG23	2.18	0.42
1:M:4:ILE:HD11	2:N:20:ILE:HD13	2.02	0.42
2:F:101:GLU:O	2:F:105:LYS:HB2	2.19	0.42
1:E:78:ILE:O	1:E:78:ILE:HG23	2.20	0.42
1:E:9:ASP:HA	2:F:4:LYS:HG2	2.02	0.42
1:E:148:ALA:HB3	1:E:162:ALA:CB	2.48	0.41

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:E:100:ARG:HA	1:E:105:ASP:OD1	2.20	0.41
2:F:14:VAL:HG13	2:F:41:LEU:HD23	2.02	0.41
1:E:112:GLN:C	1:E:166:ARG:HG3	2.41	0.41
1:M:3:LYS:HE3	1:M:78:ILE:HG21	2.03	0.41
1:M:53:ASP:O	1:M:70:VAL:HG13	2.21	0.41
1:E:183:LYS:O	1:E:184:GLN:HG2	2.21	0.41
1:E:31:MET:HE3	2:F:24:ARG:NH2	2.37	0.40

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	E	171/187 (91%)	168 (98%)	3 (2%)	0	100 100
1	M	170/187 (91%)	168 (99%)	1 (1%)	1 (1%)	25 10
2	F	105/107 (98%)	102 (97%)	2 (2%)	1 (1%)	15 4
2	N	104/107 (97%)	104 (100%)	0	0	100 100
All	All	550/588 (94%)	542 (98%)	6 (1%)	2 (0%)	34 17

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	M	49	VAL
2	F	106	TYR

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	E	139/163 (85%)	138 (99%)	1 (1%)	84	75
1	M	141/163 (86%)	140 (99%)	1 (1%)	84	75
2	F	84/92 (91%)	82 (98%)	2 (2%)	49	26
2	N	82/92 (89%)	81 (99%)	1 (1%)	71	56
All	All	446/510 (88%)	441 (99%)	5 (1%)	73	60

All (5) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	E	184	GLN
2	F	106	TYR
2	F	107	ILE
1	M	180	LYS
2	N	95	ASN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (7) such sidechains are listed below:

Mol	Chain	Res	Type
1	E	68	HIS
1	E	184	GLN
1	M	61	HIS
1	M	68	HIS
1	M	109	HIS
1	M	112	GLN
2	N	95	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

There are no carbohydrates in this entry.

5.6 Ligand geometry

There are no ligands in this entry.

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	E	177/187 (94%)	0.27	12 (6%) 17 22	11, 18, 37, 43	0
1	M	174/187 (93%)	0.08	7 (4%) 38 45	9, 15, 32, 48	0
2	F	107/107 (100%)	0.76	7 (6%) 18 24	11, 21, 41, 51	0
2	N	106/107 (99%)	0.21	1 (0%) 84 89	9, 18, 35, 40	0
All	All	564/588 (95%)	0.29	27 (4%) 30 36	9, 18, 37, 51	0

All (27) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	F	107	ILE	22.8
2	F	106	TYR	9.3
1	E	117	TYR	4.7
2	F	104	ARG	4.5
1	E	184	GLN	4.2
1	M	62	GLY	4.1
1	M	60	VAL	4.0
1	M	159	SER	3.9
2	F	1	MET	3.9
1	E	153	ALA	3.8
1	M	123	LYS	3.0
2	F	105	LYS	2.8
1	E	122	PRO	2.6
2	N	62	LEU	2.5
1	E	128	ILE	2.4
2	F	62	LEU	2.4
1	E	119	SER	2.4
1	M	41	ASP	2.4
1	E	159	SER	2.3
1	M	75	LEU	2.3
1	E	61	HIS	2.2

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	E	152	LYS	2.2
1	M	61	HIS	2.1
1	E	41	ASP	2.1
2	F	61	SER	2.1
1	E	118	VAL	2.0
1	E	150	SER	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 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.