Full wwPDB X-ray Structure Validation Report (i)

Nov 16, 2017 – 12:11 AM EST

PDB ID	:	5DWS
Title	:	Crystal Structure of ITCH WW3 domain in complex with TXNIP peptide
Authors	:	Liu, Y.; Tempel, W.; Dong, A.; Bountra, C.; Arrowsmith, C.H.; Edwards, A.M.; Min, J.; Structural Genomics Consortium (SGC)
Deposited on	:	unknown
Resolution	:	1.65 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

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

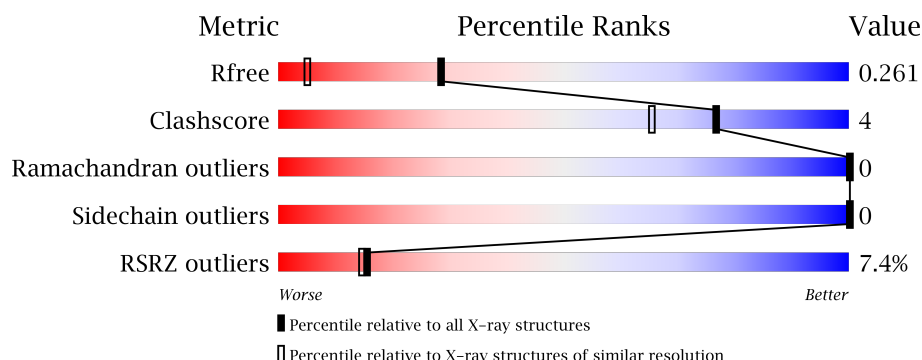|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20030345
Percentile statistics	:	20161228.v01 (using entries in the PDB archive December 28th 2016)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	rb-20030345

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 1.65 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	Percentile Ranks	Value
Rfree		0.261
Clashscore		4
Ramachandran outliers		0
Sidechain outliers		0
RSRZ outliers		7.4%

Worse *Better*

▮ Percentile relative to all X-ray structures

⬚ Percentile relative to X-ray structures of similar resolution

Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R$_{free}$	100719	1368 (1.66-1.66)
Clashscore	112137	1468 (1.66-1.66)
Ramachandran outliers	110173	1438 (1.66-1.66)
Sidechain outliers	110143	1438 (1.66-1.66)
RSRZ outliers	101464	1371 (1.66-1.66)

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. 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	47	13% / 70% 6% · 21%
1	C	47	6% / 74% 11% 15%
1	E	47	9% / 74% 26%
1	G	47	72% 6% 21%
2	B	14	71% 7% 21%

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	D	14	71% 7% 21%
2	F	14	71% 7% 21%
2	H	14	7% 79% 7% 14%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
3	UNX	A	502	-	-	-	X
3	UNX	A	503	-	-	-	X
3	UNX	E	501	-	-	-	X
3	UNX	G	502	-	-	-	X
3	UNX	G	506[A]	-	-	-	X
3	UNX	G	509	-	-	-	X
3	UNX	G	510	-	-	-	X
3	UNX	G	511	-	-	-	X
3	UNX	G	513	-	-	-	X

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 1689 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 E3 ubiquitin-protein ligase Itchy homolog.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
1	A	37	Total 316	C 198	N 62	O 56	0	5	1
1	C	40	Total 343	C 213	N 67	O 63	0	4	0
1	E	35	Total 306	C 192	N 59	O 55	0	4	0
1	G	37	Total 314	C 198	N 62	O 54	0	4	1

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	428	GLY	-	expression tag	UNP Q96J02
A	429	ARG	-	expression tag	UNP Q96J02
A	430	ARG	-	expression tag	UNP Q96J02
A	431	ALA	-	expression tag	UNP Q96J02
A	432	SER	-	expression tag	UNP Q96J02
A	433	VAL	-	expression tag	UNP Q96J02
A	434	GLU	-	expression tag	UNP Q96J02
A	435	LEU	-	expression tag	UNP Q96J02
C	428	GLY	-	expression tag	UNP Q96J02
C	429	ARG	-	expression tag	UNP Q96J02
C	430	ARG	-	expression tag	UNP Q96J02
C	431	ALA	-	expression tag	UNP Q96J02
C	432	SER	-	expression tag	UNP Q96J02
C	433	VAL	-	expression tag	UNP Q96J02
C	434	GLU	-	expression tag	UNP Q96J02
C	435	LEU	-	expression tag	UNP Q96J02
E	428	GLY	-	expression tag	UNP Q96J02
E	429	ARG	-	expression tag	UNP Q96J02
E	430	ARG	-	expression tag	UNP Q96J02
E	431	ALA	-	expression tag	UNP Q96J02
E	432	SER	-	expression tag	UNP Q96J02

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
E	433	VAL	-	expression tag	UNP Q96J02
E	434	GLU	-	expression tag	UNP Q96J02
E	435	LEU	-	expression tag	UNP Q96J02
G	428	GLY	-	expression tag	UNP Q96J02
G	429	ARG	-	expression tag	UNP Q96J02
G	430	ARG	-	expression tag	UNP Q96J02
G	431	ALA	-	expression tag	UNP Q96J02
G	432	SER	-	expression tag	UNP Q96J02
G	433	VAL	-	expression tag	UNP Q96J02
G	434	GLU	-	expression tag	UNP Q96J02
G	435	LEU	-	expression tag	UNP Q96J02

- Molecule 2 is a protein called txnip.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	11	Total	C	N	O	S	0	1	1
			79	51	12	14	2			
2	D	11	Total	C	N	O	S	0	2	1
			81	52	12	14	3			
2	F	11	Total	C	N	O	S	0	0	1
			74	48	11	13	2			
2	H	12	Total	C	N	O	S	0	0	1
			77	50	11	14	2			

- Molecule 3 is UNKNOWN ATOM OR ION (three-letter code: UNX) (formula: X).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	G	13	Total	X	0	1
			13	13		
3	E	3	Total	X	0	0
			3	3		
3	H	1	Total	X	0	0
			1	1		
3	C	4	Total	X	0	0
			4	4		
3	A	4	Total	X	0	0
			4	4		
3	F	1	Total	X	0	0
			1	1		

- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	11	Total 11	O 11	0	1
4	B	3	Total 3	O 3	0	0
4	C	17	Total 17	O 17	0	0
4	D	3	Total 3	O 3	0	0
4	E	12	Total 12	O 12	0	0
4	F	4	Total 4	O 4	0	1
4	G	23	Total 23	O 23	0	0

# 3 Residue-property plots ⓘ

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.
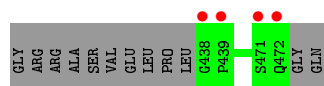
- Molecule 1: E3 ubiquitin-protein ligase Itchy homolog
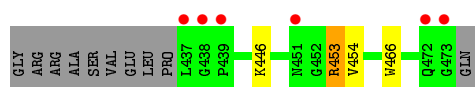
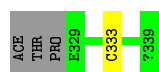- Molecule 1: E3 ubiquitin-protein ligase Itchy homolog
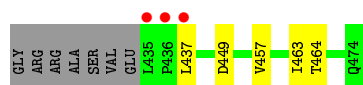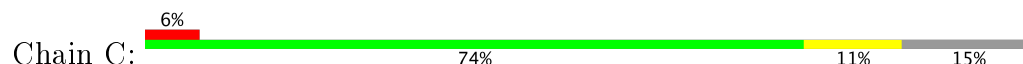
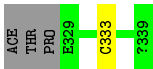- Molecule 1: E3 ubiquitin-protein ligase Itchy homolog

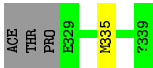- Molecule 1: E3 ubiquitin-protein ligase Itchy homolog
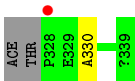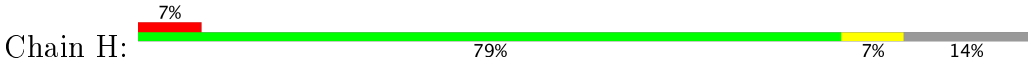

- Molecule 2: txnip

- Molecule 2: txnip

• Molecule 2: txnip

Chain F:

71%   7%   21%

• Molecule 2: txnip

Chain H:

7%   79%   7%   14%

4   Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	59.73Å 30.14Å 60.76Å 90.00° 109.20° 90.00°	Depositor
Resolution (Å)	30.00 – 1.65 28.69 – 1.65	Depositor EDS
% Data completeness (in resolution range)	97.8 (30.00-1.65) 97.8 (28.69-1.65)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	2.47 (at 1.65Å)	Xtriage
Refinement program	REFMAC	Depositor
R, R_{free}	0.222 , 0.257 0.233 , 0.261	Depositor DCC
R_{free} test set	1252 reflections (5.38%)	DCC
Wilson B-factor (Å^2)	21.8	Xtriage
Anisotropy	0.245	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.38 , 48.9	EDS
L-test for twinning [2]	$<|L|> = 0.48, <L^2> = 0.31$	Xtriage
Estimated twinning fraction	0.001 for l,-k,h	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	1689	wwPDB-VP
Average B, all atoms (Å^2)	26.0	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 66.08 % 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 6.4015e-06. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

---

[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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: UNX, NH2

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	1.00	2/357 (0.6%)	0.89	0/483
1	C	0.86	0/370	1.02	1/502 (0.2%)
1	E	0.81	0/340	0.83	0/461
1	G	0.91	0/349	1.05	2/471 (0.4%)
2	B	0.88	0/80	0.75	0/110
2	D	0.71	0/85	0.81	0/116
2	F	0.72	0/75	0.80	0/103
2	H	0.72	0/78	0.76	0/107
All	All	0.87	2/1734 (0.1%)	0.92	3/2353 (0.1%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	453[A]	ARG	CB-CG	5.63	1.67	1.52
1	A	453[B]	ARG	CB-CG	5.63	1.67	1.52

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	C	449	ASP	CB-CG-OD1	5.45	123.21	118.30
1	G	462[A]	ARG	NE-CZ-NH1	5.16	122.88	120.30
1	G	462[B]	ARG	NE-CZ-NH1	5.16	122.88	120.30

There are no chirality outliers.

There are no planarity outliers.

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	A	316	0	291	3	0
1	C	343	0	320	3	0
1	E	306	0	287	0	0
1	G	314	0	303	1	0
2	B	79	0	66	3	0
2	D	81	0	67	3	0
2	F	74	0	66	1	0
2	H	77	0	67	1	0
3	A	4	0	0	0	0
3	C	4	0	0	0	0
3	E	3	0	0	0	0
3	F	1	0	0	0	0
3	G	13	0	0	0	0
3	H	1	0	0	0	0
4	A	11	0	0	0	0
4	B	3	0	0	0	0
4	C	17	0	0	0	0
4	D	3	0	0	0	0
4	E	12	0	0	0	0
4	F	4	0	0	0	0
4	G	23	0	0	0	0
All	All	1689	0	1467	11	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 4.

All (11) 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:B:333:CYS:HB3	2:D:333[B]:CYS:SG	2.10	0.92
1:A:446[B]:LYS:HE2	1:A:454:VAL:HG11	1.78	0.66
2:B:333:CYS:CB	2:D:333[B]:CYS:SG	2.85	0.64
2:B:333:CYS:SG	2:D:333[B]:CYS:SG	3.00	0.58
2:F:335:MET:HG3	2:H:330:ALA:HB3	1.84	0.58
1:A:446[B]:LYS:HE2	1:A:454:VAL:CG1	2.45	0.45

Continued on next page...

Continued from previous page...

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:C:457:VAL:HG22	1:C:464:THR:HG22	1.99	0.45
1:G:444:TRP:HA	1:G:457:VAL:O	2.17	0.45
1:C:463:ILE:C	1:C:463:ILE:HD12	2.39	0.42
1:A:453[B]:ARG:HD3	1:A:466:TRP:CH2	2.55	0.41
1:C:437[B]:LEU:HD12	1:C:437[B]:LEU:HA	1.79	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	A	40/47 (85%)	40 (100%)	0	0	100	100
1	C	42/47 (89%)	42 (100%)	0	0	100	100
1	E	37/47 (79%)	37 (100%)	0	0	100	100
1	G	39/47 (83%)	39 (100%)	0	0	100	100
2	B	9/14 (64%)	9 (100%)	0	0	100	100
2	D	10/14 (71%)	10 (100%)	0	0	100	100
2	F	9/14 (64%)	9 (100%)	0	0	100	100
2	H	10/14 (71%)	10 (100%)	0	0	100	100
All	All	196/244 (80%)	196 (100%)	0	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	36/41 (88%)	36 (100%)	0	100	100
1	C	38/41 (93%)	38 (100%)	0	100	100
1	E	35/41 (85%)	35 (100%)	0	100	100
1	G	36/41 (88%)	36 (100%)	0	100	100
2	B	8/11 (73%)	8 (100%)	0	100	100
2	D	8/11 (73%)	8 (100%)	0	100	100
2	F	8/11 (73%)	8 (100%)	0	100	100
2	H	8/11 (73%)	8 (100%)	0	100	100
All	All	177/208 (85%)	177 (100%)	0	100	100

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

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 (i)

There are no carbohydrates in this entry.

5.6   Ligand geometry (i)

Of 26 ligands modelled in this entry, 26 are unknown - 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, 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	37/47 (78%)	0.64	6 (16%)	2	2	18, 24, 57, 68	2 (5%)
1	C	40/47 (85%)	0.42	3 (7%)	15	14	16, 22, 35, 53	1 (2%)
1	E	35/47 (74%)	0.51	4 (11%)	6	5	18, 26, 45, 51	2 (5%)
1	G	37/47 (78%)	0.08	0	100	100	15, 22, 39, 41	2 (5%)
2	B	10/14 (71%)	0.05	0	100	100	21, 30, 36, 41	0
2	D	10/14 (71%)	0.74	0	100	100	21, 32, 39, 40	0
2	F	10/14 (71%)	0.50	0	100	100	22, 31, 37, 40	0
2	H	11/14 (78%)	1.08	1 (9%)	10	8	23, 32, 45, 49	0
All	All	190/244 (77%)	0.45	14 (7%)	15	15	15, 26, 46, 68	7 (3%)

All (14) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	E	438	GLY	6.1
1	A	473	GLY	5.8
2	H	328	PRO	5.4
1	C	435	LEU	5.0
1	C	437[A]	LEU	4.6
1	A	472	GLN	4.1
1	A	439	PRO	3.4
1	E	472	GLN	3.4
1	A	437	LEU	2.4
1	C	436	PRO	2.3
1	E	471	SER	2.2
1	A	451	ASN	2.1
1	E	439	PRO	2.1
1	A	438	GLY	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)

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
3	UNX	G	506[A]	1/1	0.98	0.20	16.95	12,12,12,12	1
3	UNX	E	501	1/1	0.73	0.29	9.37	21,21,21,21	1
3	UNX	G	513	1/1	0.93	0.35	8.27	32,32,32,32	0
3	UNX	G	510	1/1	0.91	0.24	7.81	24,24,24,24	0
3	UNX	A	503	1/1	0.94	0.21	4.88	18,18,18,18	1
3	UNX	G	509	1/1	0.90	0.14	4.38	24,24,24,24	0
3	UNX	G	511	1/1	0.71	0.18	4.27	29,29,29,29	0
3	UNX	A	502	1/1	0.97	0.24	4.07	39,39,39,39	0
3	UNX	G	502	1/1	0.87	0.16	2.25	31,31,31,31	0
3	UNX	G	512	1/1	0.73	0.16	1.91	33,33,33,33	0
3	UNX	H	401	1/1	0.86	0.17	1.81	26,26,26,26	0
3	UNX	E	502	1/1	0.88	0.11	-0.11	26,26,26,26	0
3	UNX	C	502	1/1	0.57	0.18	-	37,37,37,37	0
3	UNX	G	501	1/1	0.93	0.23	-	28,28,28,28	0
3	UNX	F	401	1/1	0.91	0.12	-	24,24,24,24	0
3	UNX	G	505	1/1	0.97	0.25	-	29,29,29,29	0
3	UNX	G	507	1/1	0.97	0.11	-	21,21,21,21	0
3	UNX	C	504	1/1	0.75	0.15	-	30,30,30,30	0
3	UNX	G	508	1/1	0.95	0.17	-	21,21,21,21	0
3	UNX	A	501	1/1	0.97	0.09	-	25,25,25,25	0
3	UNX	E	503	1/1	0.76	0.17	-	34,34,34,34	0
3	UNX	G	503	1/1	0.90	0.13	-	23,23,23,23	0
3	UNX	C	501	1/1	0.86	0.11	-	34,34,34,34	0
3	UNX	C	503	1/1	0.93	0.24	-	35,35,35,35	0
3	UNX	A	504	1/1	0.98	0.23	-	19,19,19,19	0

Continued from previous page...

Mol	Type	Chain	Res	Atoms	RSCC	RSR	LLDF	B-factors(Å^2)	Q<0.9
3	UNX	G	504	1/1	0.76	0.12	-	36,36,36,36	0

6.5 Other polymers (i)

There are no such residues in this entry.