W O R L D W I D E

wwPDB

P R O T E I N   D A T A   B A N K

wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 27, 2023 – 07:08 AM EDT

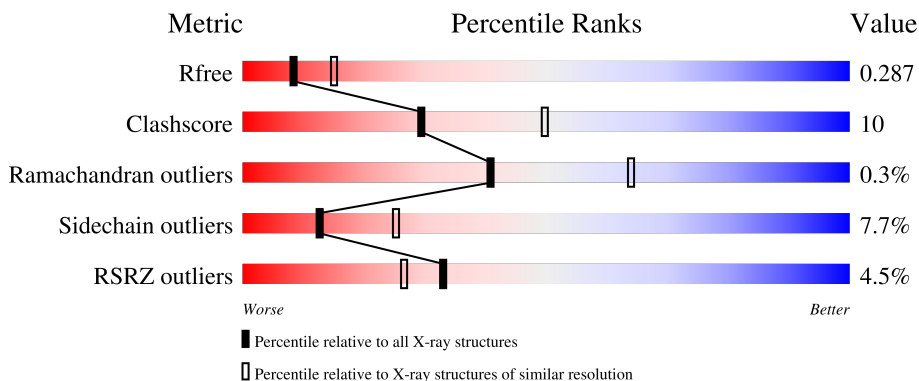|		
PDB ID	:	3RZL
Title	:	Duplex Interrogation by a Direct DNA Repair Protein in the Search of Damage
Authors	:	Yi, C.; Chen, B.; Qi, B.; Zhang, W.; Jia, G.; Zhang, L.; Li, C.; Dinner, A.; Yang, C.; He, C.
Deposited on	:	2011-05-11
Resolution	:	2.60 Å(reported)

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 2.60 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.
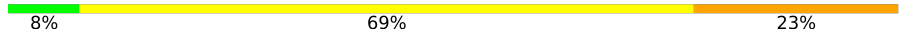
Metric	Percentile Ranks	Value
Rfree		0.287
Clashscore		10
Ramachandran outliers		0.3%
Sidechain outliers		7.7%
RSRZ outliers		4.5%

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}	130704	3163 (2.60-2.60)
Clashscore	141614	3518 (2.60-2.60)
Ramachandran outliers	138981	3455 (2.60-2.60)
Sidechain outliers	138945	3455 (2.60-2.60)
RSRZ outliers	127900	3104 (2.60-2.60)

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 <=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	208	3%, 76%, 14%, · 7%
1	D	208	6%, 62%, 31%, · ·
2	B	13	46%, 15%, 38%
2	E	13	8%, 38%, 38%, 23%
3	C	13	54%, 38%, 8%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	13	8% 69% 23%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 4288 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 Alpha-ketoglutarate-dependent dioxygenase alkB homolog 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	194	Total	C	N	O	S	0	3	0
			1585	1016	289	277	3			
1	D	199	Total	C	N	O	S	0	4	0
			1625	1040	297	285	3			

There are 12 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	54	SER	-	expression tag	UNP Q6NS38
A	55	HIS	-	expression tag	UNP Q6NS38
A	67	SER	CYS	engineered mutation	UNP Q6NS38
A	165	SER	CYS	engineered mutation	UNP Q6NS38
A	169	CYS	GLY	engineered mutation	UNP Q6NS38
A	192	SER	CYS	engineered mutation	UNP Q6NS38
D	54	SER	-	expression tag	UNP Q6NS38
D	55	HIS	-	expression tag	UNP Q6NS38
D	67	SER	CYS	engineered mutation	UNP Q6NS38
D	165	SER	CYS	engineered mutation	UNP Q6NS38
D	169	CYS	GLY	engineered mutation	UNP Q6NS38
D	192	SER	CYS	engineered mutation	UNP Q6NS38

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

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	13	Total	C	N	O	P	0	0	0
			263	127	47	77	12			
2	E	13	Total	C	N	O	P	0	0	0
			263	127	47	77	12			

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

W O R L D W I D E
wwPDB
PROTEIN DATA BANK

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	13	Total	C	N	O	P	0	0	0
			263	127	49	75	12			
3	F	13	Total	C	N	O	P	0	0	0
			263	127	49	75	12			

- Molecule 4 is propane-1-thiol (three-letter code: XL3) (formula: C_3H_8S).
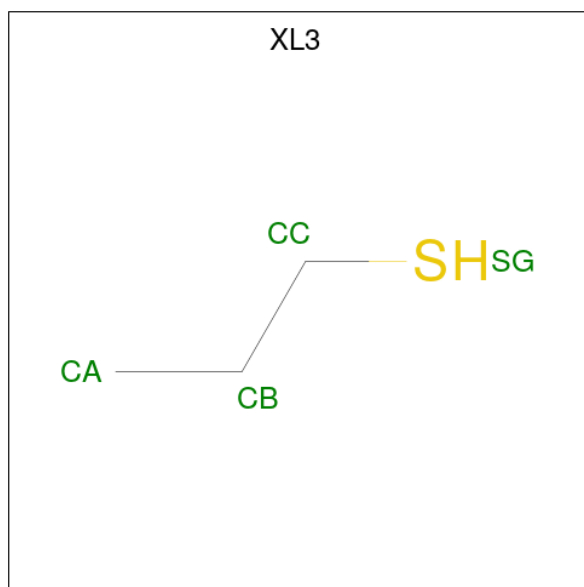
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
4	B	1	Total	C	S	0	0
			4	3	1		
4	E	1	Total	C	S	0	0
			4	3	1		

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	13	Total	O	0	0
			13	13		
5	C	1	Total	O	0	0
			1	1		
5	D	3	Total	O	0	0
			3	3		
5	F	1	Total	O	0	0
			1	1		

3 Residue-property plots ⓘ

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: Alpha-ketoglutarate-dependent dioxygenase alkB homolog 2
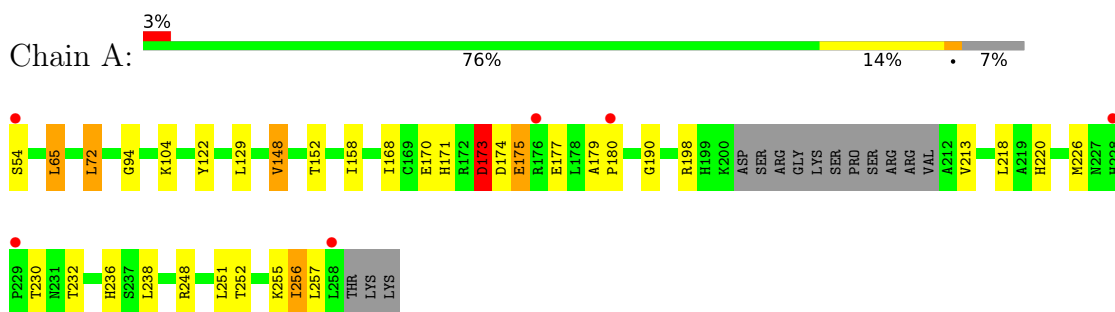
• Molecule 1: Alpha-ketoglutarate-dependent dioxygenase alkB homolog 2
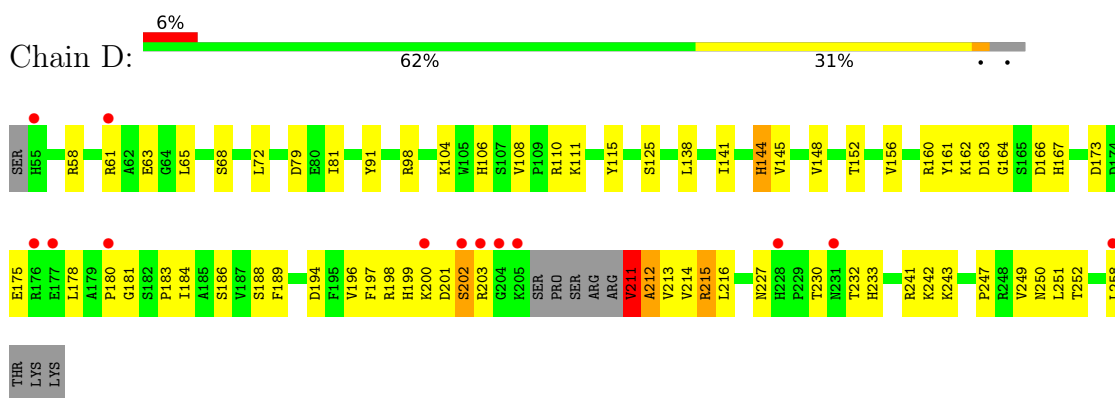
• Molecule 2: 5'-D(*AP*TP*GP*TP*AP*TP*CP*AP*CP*TP*GP*CP*G)-3'
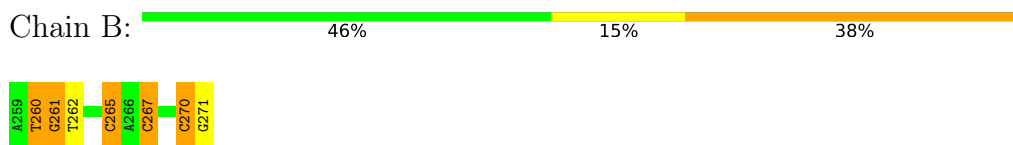
• Molecule 2: 5'-D(*AP*TP*GP*TP*AP*TP*CP*AP*CP*TP*GP*CP*G)-3'

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

Chain C:

54%   38%   8%

T272 C273 G274 A280 T281 A282 C283 A284
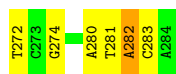
• Molecule 3: 5'-D(*TP*CP*GP*CP*AP*GP*TP*IP*AP*TP*AP*CP*A)-3'

Chain F:
8%   69%   23%

T272 C273 G274 C275 A276 G277 T278 I279 A280 T281 A282 C283 A284

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, α, β, γ	55.93Å 65.01Å 167.90Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 2.60 19.97 – 2.60	Depositor EDS
% Data completeness (in resolution range)	94.5 (20.00-2.60) 94.5 (19.97-2.60)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.39 (at 2.59Å)	Xtriage
Refinement program	REFMAC 5.5.0109	Depositor
R, R_{free}	0.221 , 0.289 0.217 , 0.287	Depositor DCC
R_{free} test set	944 reflections (5.12%)	wwPDB-VP
Wilson B-factor (Å^2)	61.8	Xtriage
Anisotropy	0.074	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 34.2	EDS
L-test for twinning[2]	$< |L| > = 0.49, < L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4288	wwPDB-VP
Average B, all atoms (Å^2)	62.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.25% 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: XL3

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	0.66	0/1635	0.78	1/2213 (0.0%)
1	D	0.61	0/1677	0.84	7/2268 (0.3%)
2	B	1.15	0/294	1.73	8/452 (1.8%)
2	E	1.07	0/294	1.79	9/452 (2.0%)
3	C	1.20	1/270 (0.4%)	2.05	12/412 (2.9%)
3	F	3.70	16/270 (5.9%)	2.53	24/412 (5.8%)
All	All	1.18	17/4440 (0.4%)	1.28	61/6209 (1.0%)

The worst 5 of 17 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	F	272	DT	N1-C2	36.64	1.67	1.38
3	F	272	DT	C4-C5	32.30	1.74	1.45
3	F	272	DT	C5-C7	12.95	1.57	1.50
3	F	272	DT	N1-C6	11.98	1.46	1.38
3	F	272	DT	C5-C6	10.70	1.41	1.34

The worst 5 of 61 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	F	272	DT	C5-C6-N1	-16.19	113.98	123.70
3	F	272	DT	C6-C5-C7	-16.02	113.29	122.90
3	C	282	DA	C1'-O4'-C4'	-11.71	98.39	110.10
3	C	282	DA	O4'-C4'-C3'	-11.20	99.28	106.00
1	D	181	GLY	N-CA-C	10.93	140.41	113.10

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	1585	0	1576	24	0
1	D	1625	0	1621	38	0
2	B	263	0	148	9	0
2	E	263	0	148	4	0
3	C	263	0	147	3	0
3	F	263	0	147	5	0
4	B	4	0	6	3	0
4	E	4	0	6	0	0
5	A	13	0	0	0	0
5	C	1	0	0	0	0
5	D	3	0	0	1	0
5	F	1	0	0	0	0
All	All	4288	0	3799	77	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 10.

The worst 5 of 77 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:A:174:ASP:C	1:A:175:GLU:HG2	1.74	1.03
1:A:174:ASP:O	1:A:175:GLU:HG2	1.62	0.99
1:A:54:SER:HB2	1:A:148:VAL:HG11	1.45	0.97
1:D:200:LYS:HB3	1:D:233:HIS:O	1.66	0.95
1:A:54:SER:HB3	1:A:72:LEU:HD23	1.49	0.94

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	193/208 (93%)	176 (91%)	17 (9%)	0	100	100
1	D	199/208 (96%)	185 (93%)	13 (6%)	1 (0%)	29	52
All	All	392/416 (94%)	361 (92%)	30 (8%)	1 (0%)	41	64

All (1) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	D	230	THR

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	171/181 (94%)	159 (93%)	12 (7%)	15	30
1	D	175/181 (97%)	160 (91%)	15 (9%)	10	20
All	All	346/362 (96%)	319 (92%)	27 (8%)	13	25

5 of 27 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	72	LEU
1	D	125	SER
1	D	215[B]	ARG
1	D	104	LYS
1	D	144	HIS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (4) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	228	HIS
1	A	250	ASN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	100	GLN
1	D	250	ASN

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 monosaccharides in this entry.

5.6 Ligand geometry ⓘ

2 ligands are 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	$\#|Z| > 2$
4	XL3	B	1	1,2	3,3,3	0.29	0	2,2,2	3.06	1 (50%)
4	XL3	E	1	1,2	3,3,3	0.51	0	2,2,2	4.02	1 (50%)

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
4	XL3	B	1	1,2	-	1/1/1/1	-
4	XL3	E	1	1,2	-	1/1/1/1	-

WORLDWIDE
wwPDB
PROTEIN DATA BANK

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	E	1	XL3	CB-CC-SG	-5.68	107.82	113.74
4	B	1	XL3	CB-CC-SG	-4.33	109.23	113.74

There are no chirality outliers.

All (2) torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	E	1	XL3	CA-CB-CC-SG
4	B	1	XL3	CA-CB-CC-SG

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
4	B	1	XL3	3	0

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, 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(Å²)	Q<0.9
1	A	194/208 (93%)	-0.17	6 (3%)	49	42	38, 48, 74, 91	0
1	D	199/208 (95%)	0.17	13 (6%)	18	14	41, 62, 94, 137	0
2	B	13/13 (100%)	-0.47	0	100	100	41, 66, 120, 126	0
2	E	13/13 (100%)	-0.01	1 (7%)	13	10	42, 66, 154, 161	0
3	C	12/13 (92%)	-0.58	0	100	100	53, 65, 78, 90	0
3	F	12/13 (92%)	-0.54	0	100	100	62, 78, 99, 111	0
All	All	443/468 (94%)	-0.04	20 (4%)	33	26	38, 56, 92, 161	0

The worst 5 of 20 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	D	204	GLY	11.4
1	D	203	ARG	6.9
1	D	205	LYS	4.0
1	D	202	SER	3.8
1	D	200	LYS	3.6

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

There are no non-standard protein/DNA/RNA residues in this entry.

6.3 Carbohydrates ⓘ

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, 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(Å^2)	Q<0.9
4	XL3	B	1	4/4	0.95	0.16	39,41,42,46	0
4	XL3	E	1	4/4	0.96	0.19	47,47,48,48	0

6.5 Other polymers (i)

There are no such residues in this entry.

WORLDWIDE
wwPDB
PROTEIN DATA BANK