W O R L D W I D E

wwPDB

PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report ⓘ

Jan 12, 2025 – 08:08 PM EST

PDB ID	:	9C50
Title	:	Replacement of a single residue changes the primary specificity of thrombin
Authors	:	Dei Rossi, A.; Deavila, S.; Mohammed, B.M.; Korolev, S.; Di Cera, E.
Deposited on	:	2024-06-05
Resolution	:	2.50 Å(reported)

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

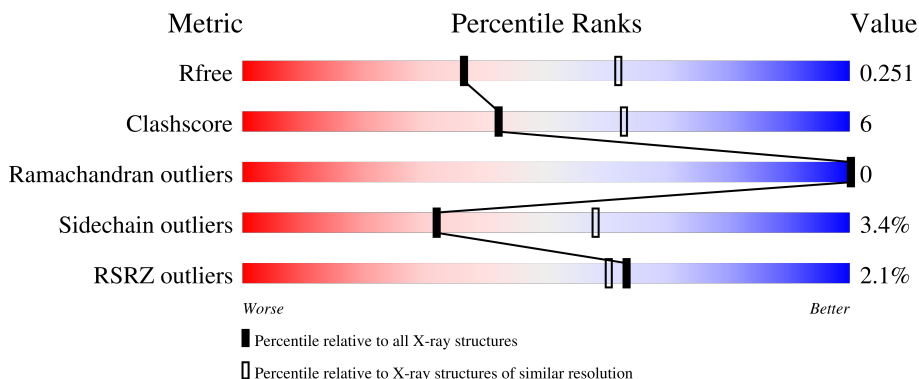|		
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.21
EDS	:	3.0
Percentile statistics	:	20231227.v01 (using entries in the PDB archive December 27th 2023)
CCP4	:	9.0.004 (Gargrove)
Density-Fitness	:	1.0.11
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (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 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	Percentile Ranks	Value
Rfree		0.251
Clashscore		6
Ramachandran outliers		0
Sidechain outliers		3.4%
RSRZ outliers		2.1%

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$_{free}$	164625	5504 (2.50-2.50)
Clashscore	180529	6282 (2.50-2.50)
Ramachandran outliers	177936	6191 (2.50-2.50)
Sidechain outliers	177891	6193 (2.50-2.50)
RSRZ outliers	164620	5504 (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 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	36	11% / 67% 8% 25%
1	C	36	6% / 56% 19% 25%
2	B	273	% / 76% 14% • 8%
2	D	273	% / 81% 13% 5%
3	E	3	100%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	3	33% 67%

2 Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 4743 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 Thrombin A-chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	27	Total	C	N	O	S	0	0	0
			222	140	36	45	1			
1	C	27	Total	C	N	O	S	0	0	0
			222	140	36	45	1			

- Molecule 2 is a protein called Thrombin B-chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	250	Total	C	N	O	S	0	0	0
			2025	1292	358	361	14			
2	D	258	Total	C	N	O	S	0	0	0
			2091	1333	367	377	14			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	248	TYR	-	expression tag	UNP P00734
B	249	LEU	-	expression tag	UNP P00734
B	250	GLU	-	expression tag	UNP P00734
B	251	ASP	-	expression tag	UNP P00734
B	252	GLN	-	expression tag	UNP P00734
B	253	VAL	-	expression tag	UNP P00734
B	254	ASP	-	expression tag	UNP P00734
B	255	PRO	-	expression tag	UNP P00734
B	256	ARG	-	expression tag	UNP P00734
B	257	LEU	-	expression tag	UNP P00734
B	258	ILE	-	expression tag	UNP P00734
B	259	ASP	-	expression tag	UNP P00734
B	260	GLY	-	expression tag	UNP P00734
B	261	LYS	-	expression tag	UNP P00734
D	248	TYR	-	expression tag	UNP P00734
D	249	LEU	-	expression tag	UNP P00734

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	250	GLU	-	expression tag	UNP P00734
D	251	ASP	-	expression tag	UNP P00734
D	252	GLN	-	expression tag	UNP P00734
D	253	VAL	-	expression tag	UNP P00734
D	254	ASP	-	expression tag	UNP P00734
D	255	PRO	-	expression tag	UNP P00734
D	256	ARG	-	expression tag	UNP P00734
D	257	LEU	-	expression tag	UNP P00734
D	258	ILE	-	expression tag	UNP P00734
D	259	ASP	-	expression tag	UNP P00734
D	260	GLY	-	expression tag	UNP P00734
D	261	LYS	-	expression tag	UNP P00734

- Molecule 3 is a protein called FPF.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	E	3	Total 30	C 24	N 3	O 3	0	0	0
3	F	3	Total 30	C 24	N 3	O 3	0	0	0

- Molecule 4 is SODIUM ION (three-letter code: NA) (formula: Na).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	B	1	Total 1	Na 1	0	0
4	D	1	Total 1	Na 1	0	0

- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total 1	O 1	0	0
5	B	43	Total 43	O 43	0	0
5	C	6	Total 6	O 6	0	0
5	D	71	Total 71	O 71	0	0

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.
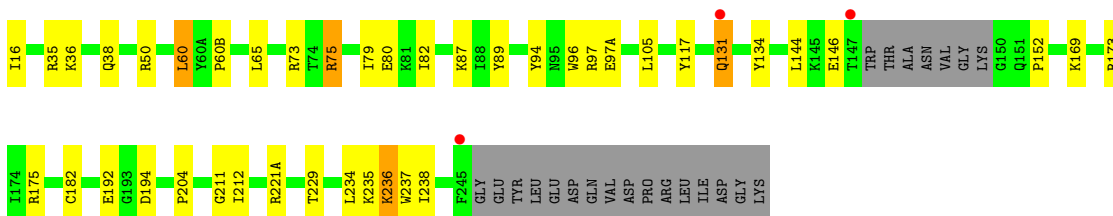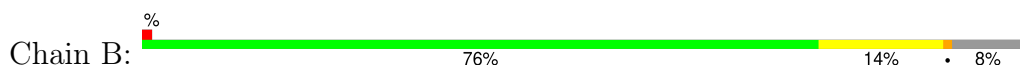
• Molecule 1: Thrombin A-chain
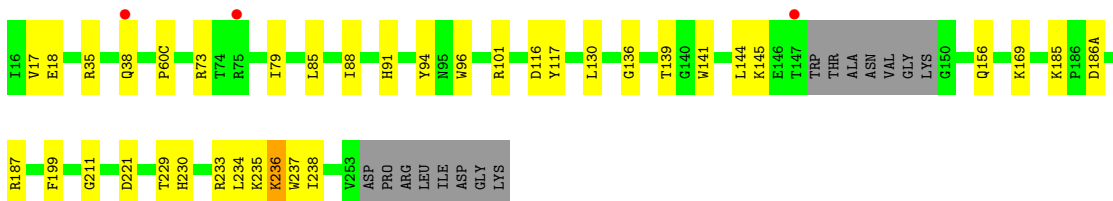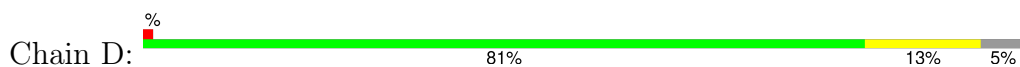
Chain A:

• Molecule 1: Thrombin A-chain

Chain C:

• Molecule 2: Thrombin B-chain

Chain B:

• Molecule 2: Thrombin B-chain

Chain D:

• Molecule 3: FPF

Chain E:

100%

F1 P2 F3

• Molecule 3: FPF

Chain F:

33% 67%

F1 P2 F3

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	131.15Å 76.73Å 89.59Å 90.00° 113.77° 90.00°	Depositor
Resolution (Å)	30.85 – 2.50 30.85 – 2.50	Depositor EDS
% Data completeness (in resolution range)	89.8 (30.85-2.50) 89.8 (30.85-2.50)	Depositor EDS
R_{merge}	0.10	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.37 (at 2.51Å)	Xtriage
Refinement program	PHENIX 1.21rc1_5127	Depositor
R, R_{free}	0.190 , 0.245 0.196 , 0.251	Depositor DCC
R_{free} test set	1448 reflections (5.08%)	wwPDB-VP
Wilson B-factor (Å^2)	39.8	Xtriage
Anisotropy	0.165	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.34 , 33.7	EDS
L-test for twinning[2]	$< |L| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	4743	wwPDB-VP
Average B, all atoms (Å^2)	41.0	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 19.91 % 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 9.7712e-03. 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 ⓘ

5.1   Standard geometry ⓘ

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

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	0.51	0/224	0.80	0/298
1	C	0.49	0/224	0.74	0/298
2	B	0.46	0/2077	0.67	0/2806
2	D	0.46	0/2144	0.65	0/2897
3	E	4.35	4/19 (21.1%)	1.92	0/25
3	F	4.35	4/19 (21.1%)	1.93	0/25
All	All	0.61	8/4707 (0.2%)	0.69	0/6349

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	B	0	1

The worst 5 of 8 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	E	2	PRO	N-CD	-9.53	1.34	1.47
3	E	2	PRO	N-CA	9.50	1.63	1.47
3	F	2	PRO	N-CD	-9.49	1.34	1.47
3	F	2	PRO	N-CA	9.47	1.63	1.47
3	F	1	PHE	C-N	9.22	1.51	1.34

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	50	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	222	0	225	0	1
1	C	222	0	225	4	0
2	B	2025	0	2000	26	1
2	D	2091	0	2056	25	0
3	E	30	0	26	4	0
3	F	30	0	26	1	0
4	B	1	0	0	0	0
4	D	1	0	0	0	0
5	A	1	0	0	0	0
5	B	43	0	0	2	0
5	C	6	0	0	0	0
5	D	71	0	0	4	0
All	All	4743	0	4558	57	1

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 6.

The worst 5 of 57 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:D:234:LEU:HA	2:D:236:LYS:HE3	1.42	1.01
2:D:185:LYS:NZ	5:D:401:HOH:O	1.99	0.96
2:D:236:LYS:HD2	2:D:237:TRP:H	1.36	0.88
2:B:234:LEU:HA	2:B:236:LYS:HE3	1.62	0.82
2:D:236:LYS:HD2	2:D:237:TRP:N	2.04	0.72

All (1) 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)
1:A:14(A):LYS:NZ	2:B:97(A):GLU:O[4_555]	1.93	0.27

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	25/36 (69%)	22 (88%)	3 (12%)	0	100	100
1	C	25/36 (69%)	24 (96%)	1 (4%)	0	100	100
2	B	246/273 (90%)	238 (97%)	8 (3%)	0	100	100
2	D	254/273 (93%)	243 (96%)	11 (4%)	0	100	100
3	E	1/3 (33%)	1 (100%)	0	0	100	100
3	F	1/3 (33%)	1 (100%)	0	0	100	100
All	All	552/624 (88%)	529 (96%)	23 (4%)	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	25/31 (81%)	23 (92%)	2 (8%)	10	20
1	C	25/31 (81%)	24 (96%)	1 (4%)	27	51
2	B	219/238 (92%)	210 (96%)	9 (4%)	26	50
2	D	226/238 (95%)	221 (98%)	5 (2%)	47	73

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
3	E	2/2 (100%)	2 (100%)	0	100	100
3	F	2/2 (100%)	2 (100%)	0	100	100
All	All	499/542 (92%)	482 (97%)	17 (3%)	32	58

5 of 17 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	D	169	LYS
2	D	236	LYS
2	B	173	ARG
2	B	182	CYS
2	B	204	PRO

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (1) such sidechains are listed below:

Mol	Chain	Res	Type
2	D	244	GLN

5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

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

2 non-standard protein/DNA/RNA residues 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	$\#\|Z\| > 2$
3	PCS	E	3	3,2	11,12,13	1.59	1 (9%)	12,15,16	1.90	2 (16%)
3	PCS	F	3	3,2	11,12,13	1.59	1 (9%)	12,15,16	1.90	2 (16%)

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
3	PCS	E	3	3,2	-	4/7/8/10	0/1/1/1
3	PCS	F	3	3,2	-	4/7/8/10	0/1/1/1

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	F	3	PCS	CT-C	3.77	1.59	1.49
3	E	3	PCS	CT-C	3.76	1.59	1.49

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	E	3	PCS	CB-CA-C	5.13	116.88	110.72
3	F	3	PCS	CB-CA-C	5.12	116.86	110.72
3	F	3	PCS	CG-CB-CA	-3.44	107.10	114.13
3	E	3	PCS	CG-CB-CA	-3.43	107.11	114.13

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
3	E	3	PCS	C-CA-CB-CG
3	F	3	PCS	C-CA-CB-CG
3	E	3	PCS	CA-CB-CG-CD2
3	F	3	PCS	CA-CB-CG-CD2
3	F	3	PCS	CA-CB-CG-CD1

There are no ring outliers.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	E	3	PCS	3	0

5.5 Carbohydrates (i)

There are no oligosaccharides in this entry.

5.6 Ligand geometry ⓘ

Of 2 ligands modelled in this entry, 2 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, 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	27/36 (75%)	0.58	4 (14%)	7	7	40, 45, 58, 81	0
1	C	27/36 (75%)	-0.24	2 (7%)	22	20	31, 37, 47, 61	0
2	B	250/273 (91%)	-0.03	3 (1%)	76	73	29, 44, 61, 81	0
2	D	258/273 (94%)	-0.41	3 (1%)	76	73	24, 35, 51, 83	0
3	E	2/3 (66%)	0.34	0	100	100	35, 35, 35, 36	0
3	F	2/3 (66%)	-0.27	0	100	100	35, 35, 35, 36	0
All	All	566/624 (90%)	-0.18	12 (2%)	63	60	24, 40, 59, 83	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	14(J)	TYR	6.5
1	A	14(K)	ILE	5.4
2	D	147	THR	4.4
2	B	131	GLN	4.4
2	B	147	THR	3.6

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

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(Å²)	Q<0.9
3	PCS	E	3	12/13	0.89	0.12	32,36,40,41	0
3	PCS	F	3	12/13	0.96	0.08	32,36,40,41	0

6.3 Carbohydrates ⓘ

There are no monosaccharides in this entry.

6.4 Ligands ⓘ

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	NA	D	301	1/1	0.88	0.10	50,50,50,50	0
4	NA	B	301	1/1	0.90	0.07	41,41,41,41	0

6.5 Other polymers ⓘ

There are no such residues in this entry.

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