Full wwPDB X-ray Structure Validation Report ⓘ

Sep 10, 2023 – 10:48 AM EDT

PDB ID	:	4K2P
Title	:	The Structure of a Quintuple Mutant of the Tiam1 PH-CC-Ex Domain
Authors	:	Joshi, M.; Gakhar, L.; Fuentes, E.J.
Deposited on	:	2013-04-09
Resolution	:	1.98 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

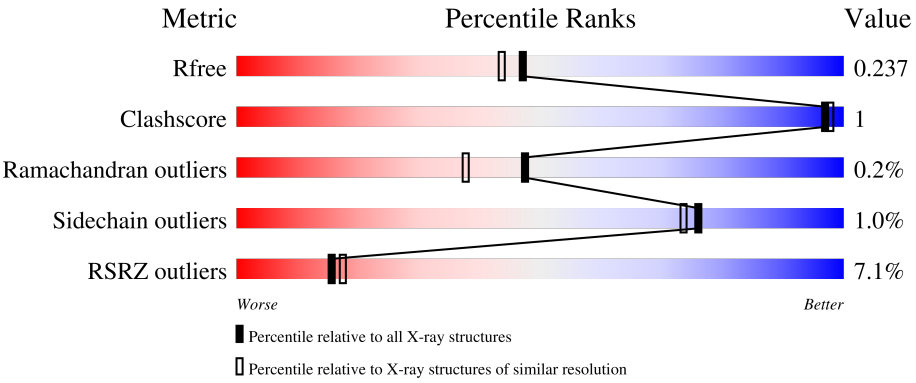
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 1.98 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	11647 (2.00-1.96)
Clashscore	141614	1014 (1.98-1.98)
Ramachandran outliers	138981	1006 (1.98-1.98)
Sidechain outliers	138945	1006 (1.98-1.98)
RSRZ outliers	127900	11410 (2.00-1.96)

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	276	5% 85% · 14%
1	B	276	7% 80% 6% 14%
1	C	276	6% 81% · 16%
1	D	276	6% 83% · 16%

2 Entry composition (i)

There are 3 unique types of molecules in this entry. The entry contains 15771 atoms, of which 7540 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 T-lymphoma invasion and metastasis-inducing protein 1.

Mol	Chain	Residues	Atoms						ZeroOcc	AltConf	Trace
1	A	238	Total	C	H	N	O	S	0	2	0
			3815	1213	1919	336	339	8			
1	B	238	Total	C	H	N	O	S	0	0	0
			3802	1210	1911	335	338	8			
1	C	233	Total	C	H	N	O	S	0	0	0
			3676	1179	1837	323	329	8			
1	D	232	Total	C	H	N	O	S	0	2	0
			3724	1188	1873	325	330	8			

There are 28 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	427	GLY	-	expression tag	UNP Q13009
A	428	SER	-	expression tag	UNP Q13009
A	580	LEU	MET	engineered mutation	UNP Q13009
A	586	LEU	MET	engineered mutation	UNP Q13009
A	596	ALA	LYS	engineered mutation	UNP Q13009
A	597	ALA	LYS	engineered mutation	UNP Q13009
A	598	ALA	LYS	engineered mutation	UNP Q13009
B	427	GLY	-	expression tag	UNP Q13009
B	428	SER	-	expression tag	UNP Q13009
B	580	LEU	MET	engineered mutation	UNP Q13009
B	586	LEU	MET	engineered mutation	UNP Q13009
B	596	ALA	LYS	engineered mutation	UNP Q13009
B	597	ALA	LYS	engineered mutation	UNP Q13009
B	598	ALA	LYS	engineered mutation	UNP Q13009
C	427	GLY	-	expression tag	UNP Q13009
C	428	SER	-	expression tag	UNP Q13009
C	580	LEU	MET	engineered mutation	UNP Q13009
C	586	LEU	MET	engineered mutation	UNP Q13009
C	596	ALA	LYS	engineered mutation	UNP Q13009
C	597	ALA	LYS	engineered mutation	UNP Q13009
C	598	ALA	LYS	engineered mutation	UNP Q13009

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
D	427	GLY	-	expression tag	UNP Q13009
D	428	SER	-	expression tag	UNP Q13009
D	580	LEU	MET	engineered mutation	UNP Q13009
D	586	LEU	MET	engineered mutation	UNP Q13009
D	596	ALA	LYS	engineered mutation	UNP Q13009
D	597	ALA	LYS	engineered mutation	UNP Q13009
D	598	ALA	LYS	engineered mutation	UNP Q13009

- Molecule 2 is CALCIUM ION (three-letter code: CA) (formula: Ca).

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	A	2	Total Ca 2 2	0	0
2	B	1	Total Ca 1 1	0	0
2	C	1	Total Ca 1 1	0	0
2	D	1	Total Ca 1 1	0	0

- Molecule 3 is water.

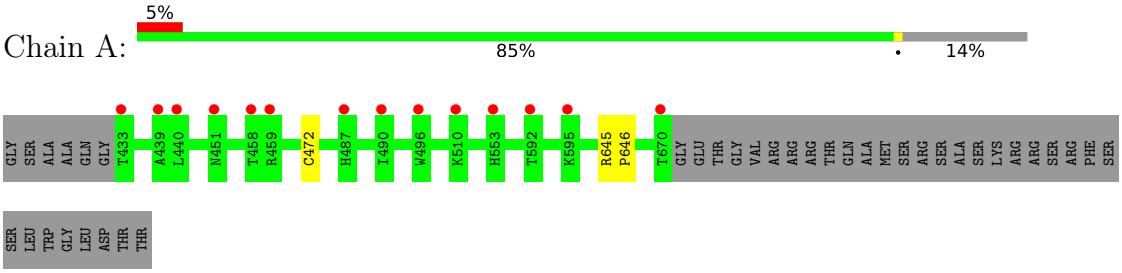
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	192	Total O 192 192	0	0
3	B	195	Total O 195 195	0	0
3	C	201	Total O 201 201	0	0
3	D	161	Total O 161 161	0	0

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.

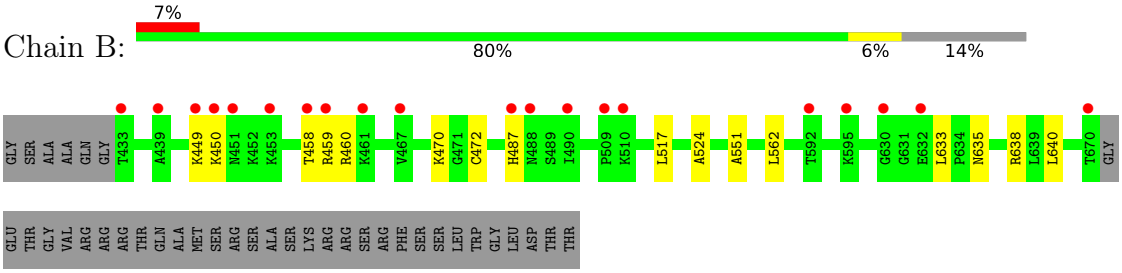
• Molecule 1: T-lymphoma invasion and metastasis-inducing protein 1

Chain A:



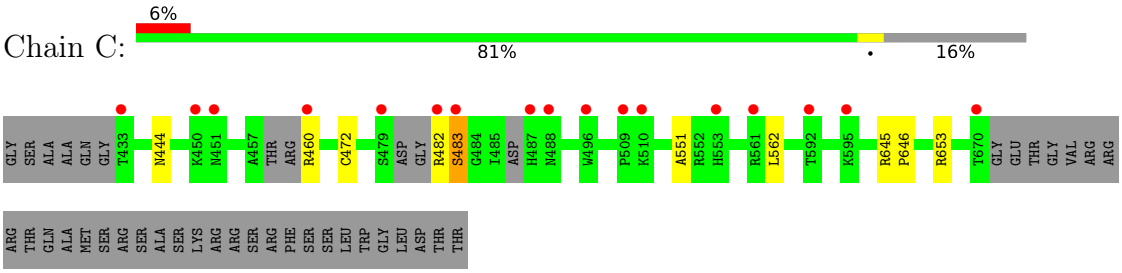
• Molecule 1: T-lymphoma invasion and metastasis-inducing protein 1

Chain B:



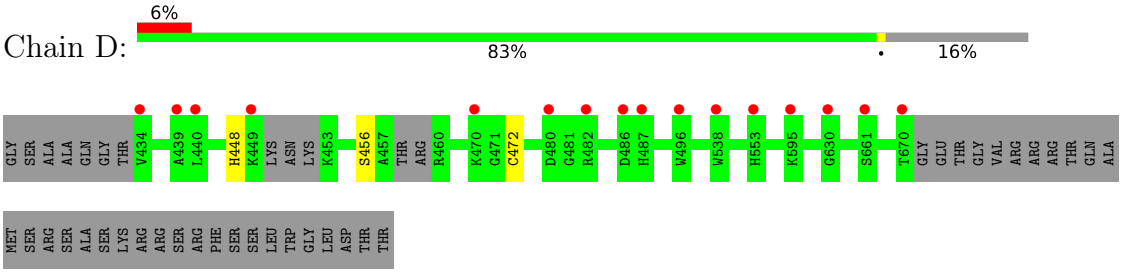
• Molecule 1: T-lymphoma invasion and metastasis-inducing protein 1

Chain C:



• Molecule 1: T-lymphoma invasion and metastasis-inducing protein 1

Chain D:



4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	80.51Å 70.11Å 87.75Å 90.00° 90.19° 90.00°	Depositor
Resolution (Å)	19.70 – 1.98 19.70 – 1.98	Depositor EDS
% Data completeness (in resolution range)	97.5 (19.70-1.98) 97.7 (19.70-1.98)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.34 (at 1.97Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, R_{free}	0.192 , 0.236 0.193 , 0.237	Depositor DCC
R_{free} test set	3306 reflections (4.97%)	wwPDB-VP
Wilson B-factor (Å^2)	23.6	Xtriage
Anisotropy	0.761	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.40 , 49.3	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	0.062 for h,-k,-l	Xtriage
F_o,F_c correlation	0.96	EDS
Total number of atoms	15771	wwPDB-VP
Average B, all atoms (Å^2)	42.0	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 44.83 % 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 1.4433e-04. 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: CA

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	0.32	0/1947	0.47	0/2625
1	B	0.32	0/1933	0.50	0/2607
1	C	0.33	0/1877	0.48	0/2531
1	D	0.31	0/1899	0.47	0/2555
All	All	0.32	0/7656	0.48	0/10318

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	1896	1919	1895	1	0
1	B	1891	1911	1906	8	0
1	C	1839	1837	1838	7	0
1	D	1851	1873	1849	1	0
2	A	2	0	0	0	0
2	B	1	0	0	0	0
2	C	1	0	0	0	0
2	D	1	0	0	0	0
3	A	192	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	195	0	0	1	0
3	C	201	0	0	0	0
3	D	161	0	0	0	0
All	All	8231	7540	7488	14	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 1.

All (14) 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:551:ALA:HA	1:B:562:LEU:CD2	2.39	0.52
1:B:449:LYS:HD2	1:C:653:ARG:HD3	1.91	0.52
1:C:482:ARG:O	1:C:483:SER:HB3	2.09	0.51
1:B:449:LYS:CD	1:C:653:ARG:HD3	2.45	0.46
1:C:645:ARG:HB3	1:C:646:PRO:HD3	1.98	0.45
1:B:635:ASN:HB3	1:B:638:ARG:HE	1.83	0.43
1:C:551:ALA:HA	1:C:562:LEU:CD2	2.48	0.43
1:B:517:LEU:O	1:B:524:ALA:HA	2.19	0.42
1:C:444:ASN:HD22	1:C:460:ARG:HH11	1.66	0.42
1:A:645:ARG:HB3	1:A:646:PRO:HD3	2.01	0.42
1:B:633:LEU:HD21	3:B:1196:HOH:O	2.19	0.42
1:B:450:LYS:HG2	1:C:653:ARG:HG3	2.02	0.42
1:B:458:THR:HB	1:B:460:ARG:HE	1.86	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	238/276 (86%)	232 (98%)	6 (2%)	0	100 100

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles
1	B	236/276 (86%)	228 (97%)	7 (3%)	1 (0%)	34 22
1	C	225/276 (82%)	220 (98%)	4 (2%)	1 (0%)	34 22
1	D	227/276 (82%)	223 (98%)	4 (2%)	0	100 100
All	All	926/1104 (84%)	903 (98%)	21 (2%)	2 (0%)	47 38

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	483	SER
1	B	459	ARG

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	204/235 (87%)	203 (100%)	1 (0%)	88 87
1	B	203/235 (86%)	199 (98%)	4 (2%)	55 48
1	C	196/235 (83%)	195 (100%)	1 (0%)	88 87
1	D	199/235 (85%)	197 (99%)	2 (1%)	76 73
All	All	802/940 (85%)	794 (99%)	8 (1%)	76 73

All (8) residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	472	CYS
1	B	470	LYS
1	B	472	CYS
1	B	487	HIS
1	B	640	LEU
1	C	472	CYS
1	D	456	SER
1	D	472	CYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (2) such sidechains are listed below:

Mol	Chain	Res	Type
1	C	444	ASN
1	C	493	HIS

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

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	238/276 (86%)	0.37	14 (5%)	22	24	20, 33, 58, 68	0
1	B	238/276 (86%)	0.48	20 (8%)	11	12	22, 34, 70, 92	0
1	C	233/276 (84%)	0.36	17 (7%)	15	16	19, 31, 64, 85	0
1	D	232/276 (84%)	0.52	16 (6%)	16	18	22, 39, 65, 83	0
All	All	941/1104 (85%)	0.43	67 (7%)	16	17	19, 34, 64, 92	0

All (67) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	451	ASN	5.9
1	D	487	HIS	4.8
1	B	487	HIS	4.5
1	A	592	THR	4.3
1	B	458	THR	4.2
1	C	483	SER	4.2
1	C	433	THR	4.2
1	C	553	HIS	4.1
1	B	449	LYS	4.0
1	C	510	LYS	3.7
1	A	487	HIS	3.6
1	D	670	THR	3.6
1	B	459	ARG	3.6
1	C	670	THR	3.5
1	B	632	GLU	3.5
1	B	670	THR	3.5
1	C	561	ARG	3.5
1	C	488	ASN	3.4
1	C	592	THR	3.4
1	A	433	THR	3.4
1	A	496	TRP	3.3

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	C	460	ARG	3.3
1	D	496	TRP	3.3
1	C	450	LYS	3.2
1	B	510	LYS	3.1
1	D	482	ARG	3.0
1	A	451	ASN	3.0
1	B	595	LYS	3.0
1	D	630	GLY	3.0
1	C	451	ASN	2.9
1	D	439	ALA	2.9
1	D	440	LEU	2.8
1	B	509	PRO	2.8
1	B	433	THR	2.8
1	B	490	ILE	2.7
1	A	510	LYS	2.7
1	A	553	HIS	2.7
1	B	450	LYS	2.7
1	A	458	THR	2.7
1	D	553	HIS	2.7
1	A	490	ILE	2.6
1	A	670	THR	2.6
1	D	480	ASP	2.6
1	C	479	SER	2.5
1	C	496	TRP	2.5
1	C	595	LYS	2.5
1	B	467	VAL	2.4
1	A	440	LEU	2.4
1	D	449[A]	LYS	2.4
1	B	453	LYS	2.3
1	B	461	LYS	2.3
1	A	459	ARG	2.3
1	D	538	TRP	2.2
1	A	439	ALA	2.2
1	D	595	LYS	2.2
1	B	630	GLY	2.1
1	D	470	LYS	2.1
1	D	661	SER	2.1
1	C	509	PRO	2.1
1	A	595	LYS	2.1
1	B	488	ASN	2.1
1	D	434	VAL	2.1
1	B	439	ALA	2.1

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type	RSRZ
1	D	486	ASP	2.1
1	C	487	HIS	2.1
1	C	482	ARG	2.0
1	B	592	THR	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
2	CA	A	1001	1/1	0.94	0.17	53,53,53,53	0
2	CA	A	1002	1/1	0.98	0.07	46,46,46,46	0
2	CA	C	1001	1/1	0.98	0.18	45,45,45,45	0
2	CA	D	1001	1/1	0.98	0.17	45,45,45,45	0
2	CA	B	1001	1/1	0.99	0.19	45,45,45,45	0

6.5 Other polymers (i)

There are no such residues in this entry.