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 (i)

May 14, 2020 – 07:38 pm BST

PDB ID	:	2XNY
Title	:	A fragment of streptococcal M1 protein in complex with human fibrinogen
Authors	:	Macheboeuf, P.; Y Fu, C.; Zinkernagel, A.S.; Johnson, J.E.; Nizet, V.; Ghosh, P.
Deposited on	:	2010-08-06
Resolution	:	7.50 Å(reported)

This is a wwPDB X-ray Structure Validation Summary 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 (i) symbol.

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

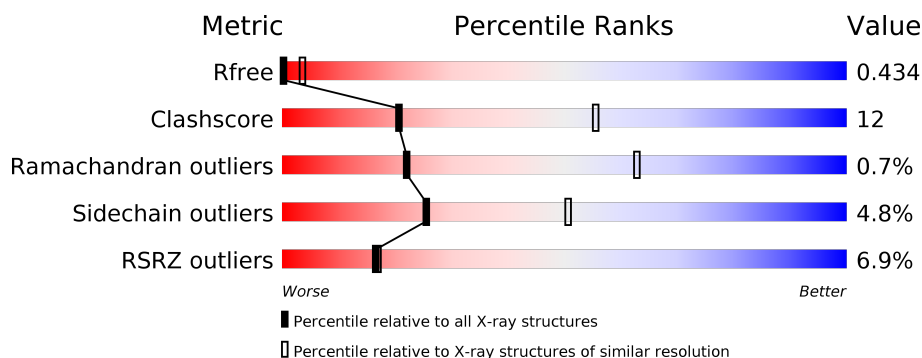|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.11
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 7.50 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	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	130704	1004 (10.00-3.90)
Clashscore	141614	1069 (10.00-3.90)
Ramachandran outliers	138981	1002 (10.00-3.90)
Sidechain outliers	138945	1002 (10.00-3.86)
RSRZ outliers	127900	1004 (9.50-3.80)

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	A	87	5% · 55% · 9% · 34%
1	D	87	6% · 55% · 9% · 34%
2	B	328	7% · 77% · 11% · 9%
2	E	328	4% · 78% · 11% · 9%
3	C	319	7% · 80% · 9% · 10%
3	F	319	5% · 80% · 8% · 10%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
4	M	102	6% ... 8% 21% 7% · 64%
4	N	102	5% ... 15% 15% 6% 65%

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 10956 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 FIBRINOGEN ALPHA CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	57	Total 463	C 283	N 88	O 89	S 3	0	0	0
1	D	57	Total 463	C 283	N 88	O 89	S 3	0	0	0

- Molecule 2 is a protein called FIBRINOGEN BETA CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	298	Total 2392	C 1494	N 422	O 454	S 22	0	0	0
2	E	298	Total 2392	C 1494	N 422	O 454	S 22	0	0	0

- Molecule 3 is a protein called FIBRINOGEN GAMMA CHAIN.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	288	Total 2311	C 1467	N 390	O 443	S 11	0	0	0
3	F	288	Total 2311	C 1467	N 390	O 443	S 11	0	0	0

- Molecule 4 is a protein called M PROTEIN.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
4	M	37	Total 315	C 192	N 58	O 65	0	0	0
4	N	36	Total 309	C 189	N 55	O 65	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
M	40	MET	-	expression tag	UNP Q48WD8
M	41	VAL	-	expression tag	UNP Q48WD8
M	136	HIS	-	expression tag	UNP Q48WD8
M	137	HIS	-	expression tag	UNP Q48WD8
M	138	HIS	-	expression tag	UNP Q48WD8
M	139	HIS	-	expression tag	UNP Q48WD8
M	140	HIS	-	expression tag	UNP Q48WD8
M	141	HIS	-	expression tag	UNP Q48WD8
N	40	MET	-	expression tag	UNP Q48WD8
N	41	VAL	-	expression tag	UNP Q48WD8
N	136	HIS	-	expression tag	UNP Q48WD8
N	137	HIS	-	expression tag	UNP Q48WD8
N	138	HIS	-	expression tag	UNP Q48WD8
N	139	HIS	-	expression tag	UNP Q48WD8
N	140	HIS	-	expression tag	UNP Q48WD8
N	141	HIS	-	expression tag	UNP Q48WD8

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.
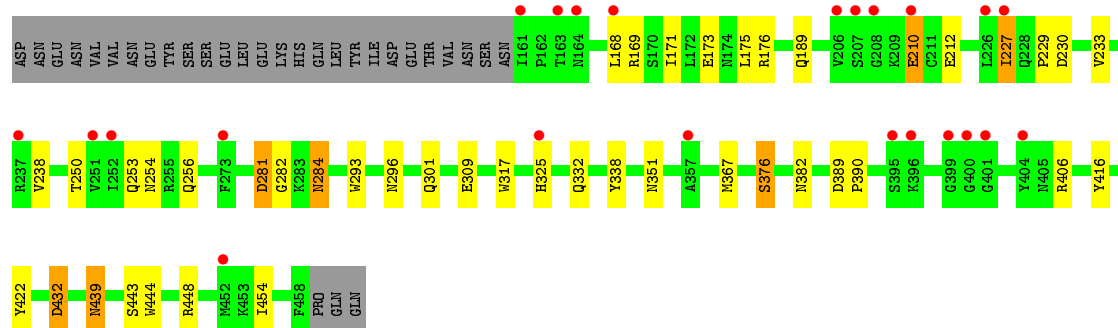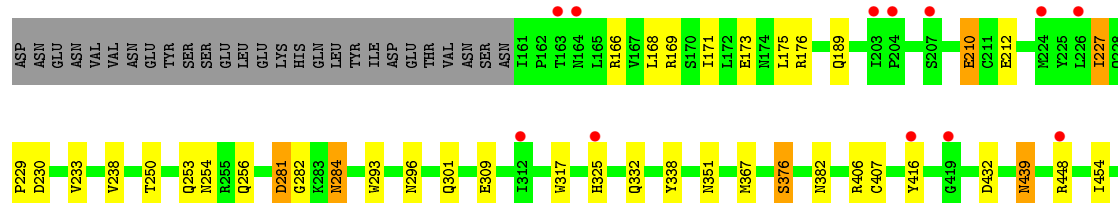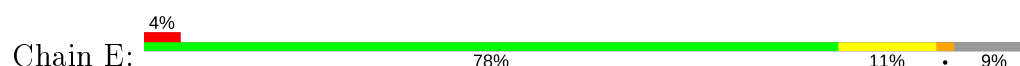
- Molecule 1: FIBRINOGEN ALPHA CHAIN

Chain A:

- Molecule 1: FIBRINOGEN ALPHA CHAIN

Chain D:

- Molecule 2: FIBRINOGEN BETA CHAIN

Chain B:

- Molecule 2: FIBRINOGEN BETA CHAIN

Chain E:

F458 PRO GLN GLN

- **Molecule 3: FIBRINOGEN GAMMA CHAIN**

Chain C:

7%
80%
9%
10%

LYS MET LEU GLU GLU ILE MET LYS TYR GLU ALA SER ILE LEU THR HIS ASP S105 R108 Y109 L110 Q111 I121 I148 K151 F168 I169 Q176 Q177 F178 L179 E183 N189 G190 W191 F194 L198 G229 I233 Q239 Y244 E249 M264

G292 F293 D297 T305 S306 H307 N308 N317 E323 G324 N325 E328 W335 M336 N337 K338 C339 H340 A341 G342 G351 G352 T353 P360 N361 N365 K381 T382 T383 F389 N390 R391 L392 THR ILE GLY GLU GLY GLN GLN HIS HIS LEU GLY GLY ALA LYS

- **Molecule 3: FIBRINOGEN GAMMA CHAIN**

Chain F:

5%
80%
8%
10%

LYS MET LEU GLU GLU ILE MET LYS TYR GLU ALA SER ILE LEU THR HIS ASP S105 R108 Y109 L110 Q111 E112 I113 Q136 I148 K151 I169 F178 L179 V180 N189 F194 L198 G229 I233 Q239 Y244 E249 M264 F265 K266 V267 L276

G284 D297 D298 H307 N317 E323 G324 N325 W335 M336 N337 A341 G342 G362 N365 W369 M379 K380 K381 T382 T383 F389 N390 R391 L392 THR ILE GLY GLU GLY GLN GLN HIS HIS LEU GLY GLY ALA LYS

- **Molecule 4: M PROTEIN**

Chain M:

6%
8%
21%
7%
64%

MET VAL ASN GLY ASP GLY ASN PRO ARG GLU VAL ILE GLU ASP LEU ALA ALA ASN ASN PRO ALA ILE GLN ASN ILE ARG LEU ARG HIS GLU ASN LYS ASP LEU LYS ALA ARG LEU GLU ASN ALA MET GLU VAL ALA GLY ARG ASP PHE LYS ARG ALA GLU GLU LEU GLU K96 A97 K98 Q99

A100 L101 E102 R105 K106 D107 L108 E109 T110 K111 L112 K113 E114 L115 Q116 Q117 D118 Y119 D120 L121 A122 K123 E124 S125 T126 W128 D129 R130 Q131 R132 LEU GLU LYS HIS HIS HIS HIS HIS HIS

- **Molecule 4: M PROTEIN**

Chain N:

5%
15%
15%
6%
65%

MET VAL ASN GLY ASP GLY ASN PRO ARG GLU VAL ILE GLU ASP LEU ALA ALA ASN ASN PRO ALA ILE GLN ASN ILE ARG LEU ARG HIS GLU ASN LYS ASP LEU LYS ALA ARG LEU GLU ASN ALA MET GLU VAL ALA GLY ARG ASP PHE LYS ARG ALA GLU GLU LEU GLU LYS ALA LYS GLN

A100 Q104 R105 K106 D107 L108 E109 T110 K111 L112 K113 E114 L115 D118 Y119 D120 L121 E124 S125 T126 D129 R130 Q131 R132 L133 E134 K135 HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32 2 1	Depositor
Cell constants a, b, c, α, β, γ	165.69Å 165.69Å 289.65Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	142.86 – 7.50 47.83 – 6.00	Depositor EDS
% Data completeness (in resolution range)	92.1 (142.86-7.50) 43.9 (47.83-6.00)	Depositor EDS
R_{merge}	0.19	Depositor
R_{sym}	(Not available)	Depositor
$<I/\sigma(I)>$ [1]	3.15 (at 6.15Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, R_{free}	0.424 , 0.408 0.423 , 0.434	Depositor DCC
R_{free} test set	247 reflections (4.69%)	wwPDB-VP
Wilson B-factor (Å^2)	98.8	Xtriage
Anisotropy	0.564	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.30 , 49.4	EDS
L-test for twinning[2]	$<|L|> = 0.45$, $<L^2> = 0.27$	Xtriage
Estimated twinning fraction	0.086 for -h,-k,l	Xtriage
F_o,F_c correlation	0.65	EDS
Total number of atoms	10956	wwPDB-VP
Average B, all atoms (Å^2)	34.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 10.05% 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 (i)

5.1 Standard geometry (i)

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	0/463	0.52	0/617
1	D	0.32	0/463	0.52	0/617
2	B	0.42	0/2453	0.55	1/3312 (0.0%)
2	E	0.42	0/2453	0.55	1/3312 (0.0%)
3	C	0.50	0/2375	0.58	0/3211
3	F	0.50	0/2375	0.58	0/3211
4	M	1.86	8/317 (2.5%)	1.35	6/421 (1.4%)
4	N	1.63	3/311 (1.0%)	1.31	5/414 (1.2%)
All	All	0.60	11/11210 (0.1%)	0.63	13/15115 (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
4	N	0	1

The worst 5 of 11 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	M	124	GLU	CD-OE2	12.07	1.39	1.25
4	M	118	ASP	CG-OD1	11.15	1.51	1.25
4	M	132	ARG	CZ-NH1	9.32	1.45	1.33
4	M	107	ASP	CG-OD2	8.91	1.45	1.25
4	M	127	SER	CB-OG	8.71	1.53	1.42

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
4	M	132	ARG	NE-CZ-NH2	-9.71	115.44	120.30
4	N	120	ASP	CB-CG-OD1	8.35	125.82	118.30
4	M	118	ASP	CB-CG-OD2	-7.96	111.14	118.30

Continued on next page...

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	M	107	ASP	CB-CG-OD1	-6.89	112.09	118.30
4	M	107	ASP	CB-CG-OD2	-6.86	112.13	118.30

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
4	N	124	GLU	Peptide

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	463	0	482	9	0
1	D	463	0	482	9	0
2	B	2392	0	2260	39	3
2	E	2392	0	2257	79	3
3	C	2311	0	2163	46	5
3	F	2311	0	2161	58	0
4	M	315	0	310	112	0
4	N	309	0	301	43	0
All	All	10956	0	10416	263	8

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 12.

The worst 5 of 263 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)
3:F:110:LEU:CD2	4:M:115:LEU:CD2	1.93	1.45
2:E:166:ARG:HB3	4:M:123:LYS:CE	1.53	1.36
2:E:166:ARG:CB	4:M:123:LYS:HZ1	1.39	1.36
2:E:166:ARG:CB	4:M:123:LYS:NZ	1.86	1.35
2:E:173:GLU:OE2	4:M:116:GLN:HA	1.19	1.35

The worst 5 of 8 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)
3:C:328:GLU:OE2	3:C:328:GLU:OE2[4_556]	1.59	0.61
3:C:328:GLU:OE1	3:C:328:GLU:OE2[4_556]	1.63	0.57
2:B:444:TRP:NE1	2:E:458:PHE:CA[5_445]	1.66	0.54
2:B:444:TRP:CZ2	2:E:458:PHE:O[5_445]	1.68	0.52
3:C:328:GLU:CD	3:C:328:GLU:OE2[4_556]	1.78	0.42

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	55/87 (63%)	49 (89%)	6 (11%)	0	100	100
1	D	55/87 (63%)	49 (89%)	6 (11%)	0	100	100
2	B	296/328 (90%)	281 (95%)	14 (5%)	1 (0%)	41	77
2	E	296/328 (90%)	281 (95%)	14 (5%)	1 (0%)	41	77
3	C	286/319 (90%)	272 (95%)	12 (4%)	2 (1%)	22	63
3	F	286/319 (90%)	272 (95%)	12 (4%)	2 (1%)	22	63
4	M	35/102 (34%)	20 (57%)	12 (34%)	3 (9%)	1	11
4	N	34/102 (33%)	24 (71%)	9 (26%)	1 (3%)	4	29
All	All	1343/1672 (80%)	1248 (93%)	85 (6%)	10 (1%)	22	63

5 of 10 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
4	M	127	SER
4	M	99	GLN
4	M	107	ASP
3	C	110	LEU
3	C	111	GLN

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	52/82 (63%)	50 (96%)	2 (4%)	33	57
1	D	52/82 (63%)	50 (96%)	2 (4%)	33	57
2	B	256/286 (90%)	247 (96%)	9 (4%)	36	59
2	E	256/286 (90%)	247 (96%)	9 (4%)	36	59
3	C	242/267 (91%)	235 (97%)	7 (3%)	42	64
3	F	242/267 (91%)	235 (97%)	7 (3%)	42	64
4	M	34/89 (38%)	24 (71%)	10 (29%)	0	2
4	N	34/89 (38%)	24 (71%)	10 (29%)	0	2
All	All	1168/1448 (81%)	1112 (95%)	56 (5%)	25	50

5 of 56 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	351	ASN
3	F	317	ASN
4	N	121	LEU
2	E	376	SER
2	E	454	ILE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 54 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	365	ASN
2	E	253	GLN
3	F	319	ASN
3	C	390	ASN
2	E	164	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 (i)

There are no carbohydrates in this entry.

5.6 Ligand geometry (i)

There are no ligands in this entry.

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	57/87 (65%)	0.58	4 (7%)	16	17	34, 65, 96, 97	0
1	D	57/87 (65%)	0.66	5 (8%)	10	12	34, 65, 96, 97	0
2	B	298/328 (90%)	0.36	23 (7%)	13	15	17, 32, 76, 102	0
2	E	298/328 (90%)	0.08	12 (4%)	38	34	17, 32, 76, 102	0
3	C	288/319 (90%)	0.31	23 (7%)	12	14	13, 20, 64, 100	0
3	F	288/319 (90%)	0.38	16 (5%)	24	24	13, 20, 64, 100	0
4	M	37/102 (36%)	0.80	6 (16%)	1	4	24, 37, 82, 116	0
4	N	36/102 (35%)	0.88	5 (13%)	2	6	21, 32, 70, 86	0
All	All	1359/1672 (81%)	0.34	94 (6%)	16	17	13, 28, 84, 116	0

The worst 5 of 94 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	207	SER	7.6
2	B	400	GLY	5.4
4	N	106	LYS	4.9
2	B	206	VAL	4.8
2	E	203	ILE	4.1

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.