WORLDWIDE wwPDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 27, 2023 – 02:01 PM EDT

PDB ID	:	3HKI
Title	:	Crystal structure of murine thrombin mutant W215A/E217A in complex with the extracellular fragment of human PAR1
Authors	:	Gandhi, P.S.; Page, M.J.; Chen, Z.; Bush-Pelc, L.; Di Cera, E.
Deposited on	:	2009-05-23
Resolution	:	2.20 Å(reported)

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

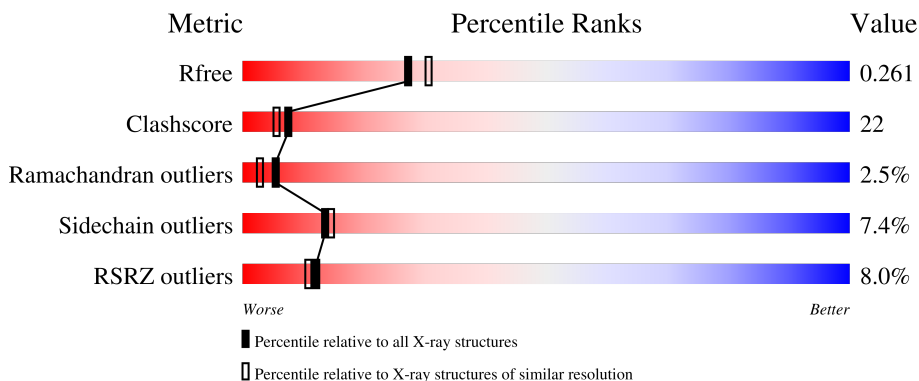|		
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35

1   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 2.20 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		22
Ramachandran outliers		2.5%
Sidechain outliers		7.4%
RSRZ outliers		8.0%

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	4898 (2.20-2.20)
Clashscore	141614	5594 (2.20-2.20)
Ramachandran outliers	138981	5503 (2.20-2.20)
Sidechain outliers	138945	5504 (2.20-2.20)
RSRZ outliers	127900	4800 (2.20-2.20)

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	44	9% / 45% 41% 5% 9%
1	D	44	11% / 70% 27% •
2	B	258	7% / 56% 37% 5% •
2	E	258	7% / 65% 29% • •
3	C	21	10% / 43% 19% 10% 29%

<div align="center">

Continued on next page...

</div>

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	21	

14%

14% 14% 5% 10% 57%

2   Entry composition ⓘ

There are 5 unique types of molecules in this entry. The entry contains 5254 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 light chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	40	Total	C	N	O	S	0	0	0
			328	212	51	64	1			
1	D	43	Total	C	N	O	S	0	0	0
			348	224	54	69	1			

- Molecule 2 is a protein called Thrombin heavy chain.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	255	Total	C	N	O	S	0	0	0
			2074	1328	373	361	12			
2	E	252	Total	C	N	O	S	0	0	0
			2032	1301	365	354	12			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
B	215	ALA	TRP	engineered mutation	UNP P19221
B	217	ALA	GLU	engineered mutation	UNP P19221
E	215	ALA	TRP	engineered mutation	UNP P19221
E	217	ALA	GLU	engineered mutation	UNP P19221

- Molecule 3 is a protein called Proteinase-activated receptor 1.

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	Trace
3	C	15	Total	C	N	O	0	0	0
			139	87	22	30			
3	F	9	Total	C	N	O	0	0	0
			88	59	11	18			

- Molecule 4 is 2-acetamido-2-deoxy-beta-D-glucopyranose (three-letter code: NAG) (formula: $C_8H_{15}NO_6$).

NAG

O6
OH

O1
HO

O5
O

C6

C1(R) C5(R)

C2(R) C4(S)

N2 C3(R)

HN

OH
O4

C7

OH
O3

C8

O
O7
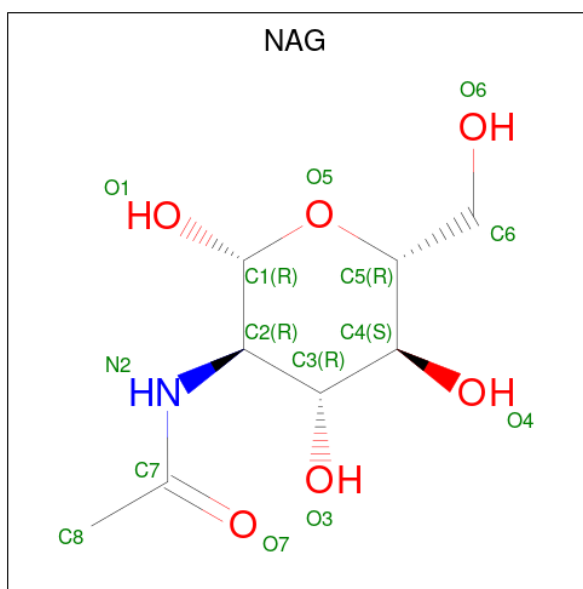
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
4	B	1	Total 14	C 8	N 1	O 5	0	0
4	E	1	Total 14	C 8	N 1	O 5	0	0

- Molecule 5 is water.

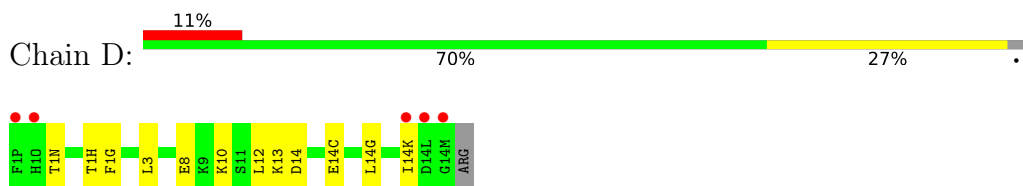| Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	9	Total 9	O 9	0	0
5	B	86	Total 86	O 86	0	0
5	C	9	Total 9	O 9	0	0
5	D	12	Total 12	O 12	0	0
5	E	96	Total 96	O 96	0	0
5	F	5	Total 5	O 5	0	0

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

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.
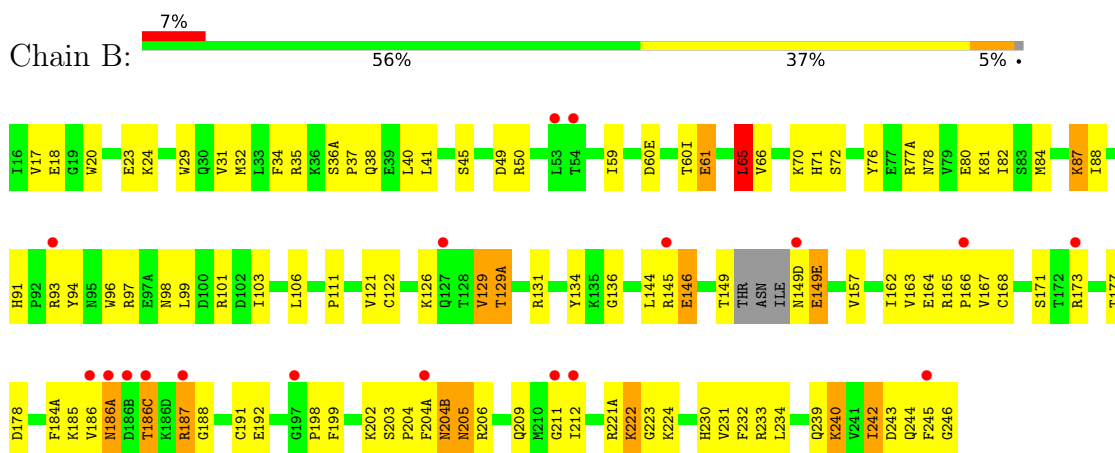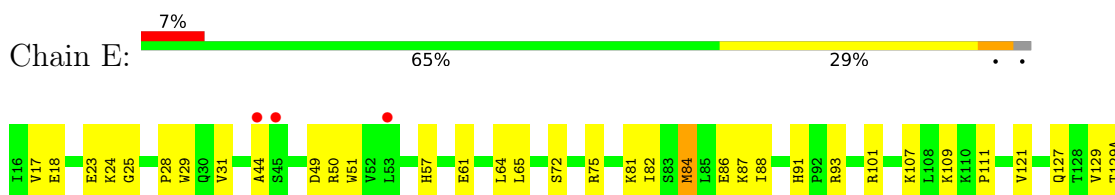
- Molecule 1: Thrombin light chain
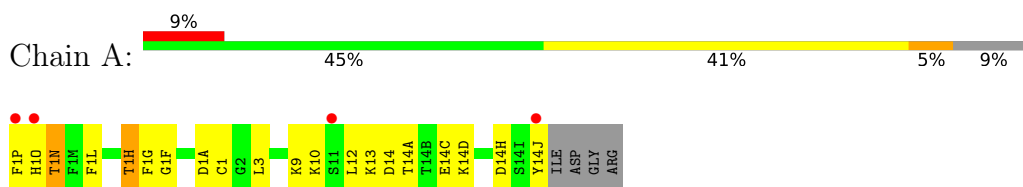
- Molecule 1: Thrombin light chain

- Molecule 2: Thrombin heavy chain

- Molecule 2: Thrombin heavy chain

S129B L129C L130 R131 G136 L144 R145 E146 T147 W148 T149 THR ASN ILE N149D E149E I150 Q151 V157 R165 K169 T172 R173 I174 R175 I176 F184A K185 V186 N186A D186B T186C K186D R187 G188 D189 E192 S195 F199 S203 P204 F204A N204B N205 R206 M210 A217

G219 C220 ASP ARG LYS G223 H230 V231 F232 R233 L234 K235 R236 W237 K240 V241 I242 D243 Q244 F245 G246

• Molecule 3: Proteinase-activated receptor 1

Chain C:

10%

43%   19%   10%   29%

SER PHE LEU LEU R46 N47 F55 W56 E57 D58 E59 E60 LYS ASN

• Molecule 3: Proteinase-activated receptor 1

Chain F:

14%

14%   14%   5%   10%   57%

SER PHE LEU LEU ARG ASN PRO ASN ASP K51 F55 W56 E57 D58 E59 GLU LYS ASN

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 21 2	Depositor
Cell constants a, b, c, α, β, γ	231.42Å 50.99Å 80.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	26.23 – 2.20 26.23 – 2.10	Depositor EDS
% Data completeness (in resolution range)	88.9 (26.23-2.20) 87.0 (26.23-2.10)	Depositor EDS
R_{merge}	0.05	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.73 (at 2.10Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.228 , 0.261 0.228 , 0.261	Depositor DCC
R_{free} test set	4121 reflections (8.11%)	wwPDB-VP
Wilson B-factor (Å^2)	32.3	Xtriage
Anisotropy	0.068	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.36 , 56.5	EDS
L-test for twinning[2]	$< |L| > = 0.46, < L^2 > = 0.29$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	5254	wwPDB-VP
Average B, all atoms (Å^2)	41.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.07% 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: NAG

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	0.58	0/335	0.79	0/447
1	D	0.52	0/355	0.76	0/474
2	B	0.52	0/2128	0.80	2/2878 (0.1%)
2	E	0.55	0/2083	0.78	0/2816
3	C	0.52	0/144	0.71	0/195
3	F	0.78	0/92	1.69	3/124 (2.4%)
All	All	0.54	0/5137	0.81	5/6934 (0.1%)

There are no bond length outliers.

All (5) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	F	58	ASP	N-CA-C	11.14	141.07	111.00
2	B	65	LEU	CA-CB-CG	6.90	131.16	115.30
3	F	58	ASP	C-N-CA	6.83	138.79	121.70
3	F	59	GLU	N-CA-C	6.48	128.51	111.00
2	B	223	GLY	N-CA-C	-6.11	97.84	113.10

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	328	0	318	21	0
1	D	348	0	336	14	0
2	B	2074	0	2071	120	0
2	E	2032	0	2027	76	0
3	C	139	0	111	8	0
3	F	88	0	69	4	0
4	B	14	0	13	0	0
4	E	14	0	13	0	0
5	A	9	0	0	2	0
5	B	86	0	0	8	0
5	C	9	0	0	0	0
5	D	12	0	0	1	0
5	E	96	0	0	6	0
5	F	5	0	0	0	0
All	All	5254	0	4958	221	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 22.

The worst 5 of 221 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:E:18:GLU:HG3	2:E:187:ARG:HB2	1.38	1.03
2:B:18:GLU:HG3	2:B:187:ARG:HG3	1.46	0.95
2:B:171:SER:HB2	2:B:224:LYS:HD2	1.55	0.88
2:E:17:VAL:HG12	2:E:144:LEU:O	1.78	0.83
2:B:204(B):ASN:ND2	2:B:206:ARG:H	1.78	0.81

There are no symmetry-related clashes.

5.3 Torsion angles ⓘ

5.3.1 Protein backbone ⓘ

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	38/44 (86%)	30 (79%)	5 (13%)	3 (8%)	1	0
1	D	41/44 (93%)	38 (93%)	3 (7%)	0	100	100
2	B	251/258 (97%)	221 (88%)	25 (10%)	5 (2%)	7	4
2	E	246/258 (95%)	224 (91%)	18 (7%)	4 (2%)	9	7
3	C	13/21 (62%)	11 (85%)	0	2 (15%)	0	0
3	F	7/21 (33%)	4 (57%)	2 (29%)	1 (14%)	0	0
All	All	596/646 (92%)	528 (89%)	53 (9%)	15 (2%)	5	3

5 of 15 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	1(N)	THR
2	B	149(E)	GLU
2	B	222	LYS
3	C	47	ASN
2	E	147	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	36/39 (92%)	34 (94%)	2 (6%)	21	25
1	D	38/39 (97%)	37 (97%)	1 (3%)	46	58
2	B	224/227 (99%)	207 (92%)	17 (8%)	13	14
2	E	219/227 (96%)	203 (93%)	16 (7%)	14	15
3	C	15/21 (71%)	15 (100%)	0	100	100
3	F	9/21 (43%)	5 (56%)	4 (44%)	0	0
All	All	541/574 (94%)	501 (93%)	40 (7%)	13	14

5 of 40 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	E	173	ARG

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	E	240	LYS
2	E	174	ILE
2	E	186(B)	ASP
3	F	57	GLU

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 13 such sidechains are listed below:

Mol	Chain	Res	Type
2	E	151	GLN
2	E	186(A)	ASN
2	E	239	GLN
2	E	205	ASN
2	E	209	GLN

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	NAG	B	301	2	14,14,15	1.64	4 (28%)	17,19,21	2.04	6 (35%)
4	NAG	E	301	2	14,14,15	1.61	4 (28%)	17,19,21	1.87	5 (29%)

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	NAG	B	301	2	-	4/6/23/26	0/1/1/1
4	NAG	E	301	2	-	6/6/23/26	0/1/1/1

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

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
4	B	301	NAG	C3-C2	3.56	1.60	1.52
4	E	301	NAG	C3-C2	3.27	1.59	1.52
4	B	301	NAG	C4-C5	2.48	1.58	1.53
4	E	301	NAG	C4-C5	2.42	1.58	1.53
4	E	301	NAG	C1-C2	2.24	1.55	1.52

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
4	B	301	NAG	C6-C5-C4	3.92	122.19	113.00
4	B	301	NAG	O5-C1-C2	-3.83	105.23	111.29
4	E	301	NAG	O5-C1-C2	-3.82	105.26	111.29
4	B	301	NAG	C1-C2-N2	-3.64	104.27	110.49
4	E	301	NAG	C6-C5-C4	3.29	120.72	113.00

There are no chirality outliers.

5 of 10 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
4	B	301	NAG	C8-C7-N2-C2
4	B	301	NAG	O7-C7-N2-C2
4	E	301	NAG	C8-C7-N2-C2
4	E	301	NAG	O7-C7-N2-C2
4	B	301	NAG	C1-C2-N2-C7

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	40/44 (90%)	0.32	4 (10%)	7	6	32, 45, 78, 89	0
1	D	43/44 (97%)	0.35	5 (11%)	4	4	35, 47, 78, 93	0
2	B	255/258 (98%)	0.18	18 (7%)	16	14	17, 37, 65, 82	0
2	E	252/258 (97%)	0.09	17 (6%)	17	16	14, 33, 59, 90	0
3	C	15/21 (71%)	0.32	2 (13%)	3	3	26, 42, 83, 95	0
3	F	9/21 (42%)	0.71	3 (33%)	0	0	36, 45, 76, 77	0
All	All	614/646 (95%)	0.17	49 (7%)	12	11	14, 38, 69, 95	0

The worst 5 of 49 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	148	TRP	9.0
1	D	14(K)	ILE	5.9
2	E	149	THR	5.5
2	E	149(D)	ASN	5.5
1	A	1(P)	PHE	5.4

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	NAG	E	301	14/15	0.64	0.37	71,77,79,80	0
4	NAG	B	301	14/15	0.68	0.25	68,71,75,77	0

6.5 Other polymers (i)

There are no such residues in this entry.