wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 05:58 PM GMT

PDB ID	:	4K0J
Title	:	X-ray crystal structure of a heavy metal efflux pump, crystal form I
Authors	:	Pak, J.E.; Stroud, R.M.; Ngonlong Ekende, E.; Vandenbussche, G.; Center for Structures of Membrane Proteins (CSMP)
Deposited on	:	2013-04-04
Resolution	:	3.00 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
http://wwpdb.org/validation/2016/XrayValidationReportHelp
with specific help available everywhere you see the (i) symbol.

The following versions of software and data (see references (1)) were used in the production of this report:

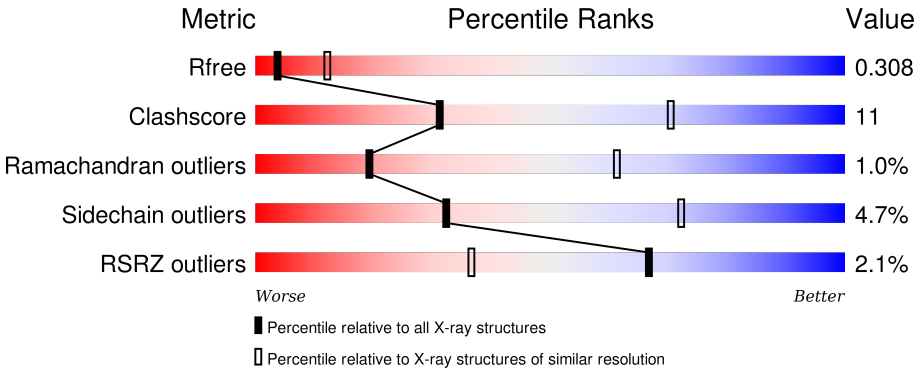
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	trunk26865

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 3.00 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.



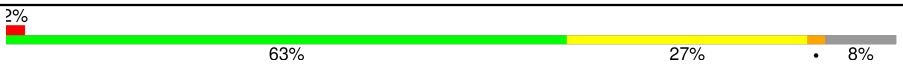
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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. 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	1045	2% 64% 26% • 8%
1	B	1045	% 65% 24% • 9%
1	C	1045	2% 63% 26% • 8%
1	D	1045	3% 65% 25% • 8%
1	E	1045	% 64% 25% • 9%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	F	1045	

2 Entry composition (i)

There are 2 unique types of molecules in this entry. The entry contains 44018 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 Heavy metal cation tricomponent efflux pump ZneA(CzcA-like).

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	958	Total	C	N	O	S	0	0	0
			7336	4709	1276	1319	32			
1	B	954	Total	C	N	O	S	0	0	0
			7314	4696	1274	1312	32			
1	C	959	Total	C	N	O	S	0	0	0
			7357	4719	1282	1324	32			
1	D	958	Total	C	N	O	S	0	0	0
			7336	4709	1276	1319	32			
1	E	954	Total	C	N	O	S	0	0	0
			7314	4696	1274	1312	32			
1	F	959	Total	C	N	O	S	0	0	0
			7357	4719	1282	1324	32			

There are 36 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	1040	HIS	-	EXPRESSION TAG	UNP Q1LCD8
A	1041	HIS	-	EXPRESSION TAG	UNP Q1LCD8
A	1042	HIS	-	EXPRESSION TAG	UNP Q1LCD8
A	1043	HIS	-	EXPRESSION TAG	UNP Q1LCD8
A	1044	HIS	-	EXPRESSION TAG	UNP Q1LCD8
A	1045	HIS	-	EXPRESSION TAG	UNP Q1LCD8
B	1040	HIS	-	EXPRESSION TAG	UNP Q1LCD8
B	1041	HIS	-	EXPRESSION TAG	UNP Q1LCD8
B	1042	HIS	-	EXPRESSION TAG	UNP Q1LCD8
B	1043	HIS	-	EXPRESSION TAG	UNP Q1LCD8
B	1044	HIS	-	EXPRESSION TAG	UNP Q1LCD8
B	1045	HIS	-	EXPRESSION TAG	UNP Q1LCD8
C	1040	HIS	-	EXPRESSION TAG	UNP Q1LCD8
C	1041	HIS	-	EXPRESSION TAG	UNP Q1LCD8
C	1042	HIS	-	EXPRESSION TAG	UNP Q1LCD8
C	1043	HIS	-	EXPRESSION TAG	UNP Q1LCD8
C	1044	HIS	-	EXPRESSION TAG	UNP Q1LCD8

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	1045	HIS	-	EXPRESSION TAG	UNP Q1LCD8
D	1040	HIS	-	EXPRESSION TAG	UNP Q1LCD8
D	1041	HIS	-	EXPRESSION TAG	UNP Q1LCD8
D	1042	HIS	-	EXPRESSION TAG	UNP Q1LCD8
D	1043	HIS	-	EXPRESSION TAG	UNP Q1LCD8
D	1044	HIS	-	EXPRESSION TAG	UNP Q1LCD8
D	1045	HIS	-	EXPRESSION TAG	UNP Q1LCD8
E	1040	HIS	-	EXPRESSION TAG	UNP Q1LCD8
E	1041	HIS	-	EXPRESSION TAG	UNP Q1LCD8
E	1042	HIS	-	EXPRESSION TAG	UNP Q1LCD8
E	1043	HIS	-	EXPRESSION TAG	UNP Q1LCD8
E	1044	HIS	-	EXPRESSION TAG	UNP Q1LCD8
E	1045	HIS	-	EXPRESSION TAG	UNP Q1LCD8
F	1040	HIS	-	EXPRESSION TAG	UNP Q1LCD8
F	1041	HIS	-	EXPRESSION TAG	UNP Q1LCD8
F	1042	HIS	-	EXPRESSION TAG	UNP Q1LCD8
F	1043	HIS	-	EXPRESSION TAG	UNP Q1LCD8
F	1044	HIS	-	EXPRESSION TAG	UNP Q1LCD8
F	1045	HIS	-	EXPRESSION TAG	UNP Q1LCD8

- Molecule 2 is ZINC ION (three-letter code: ZN) (formula: Zn).

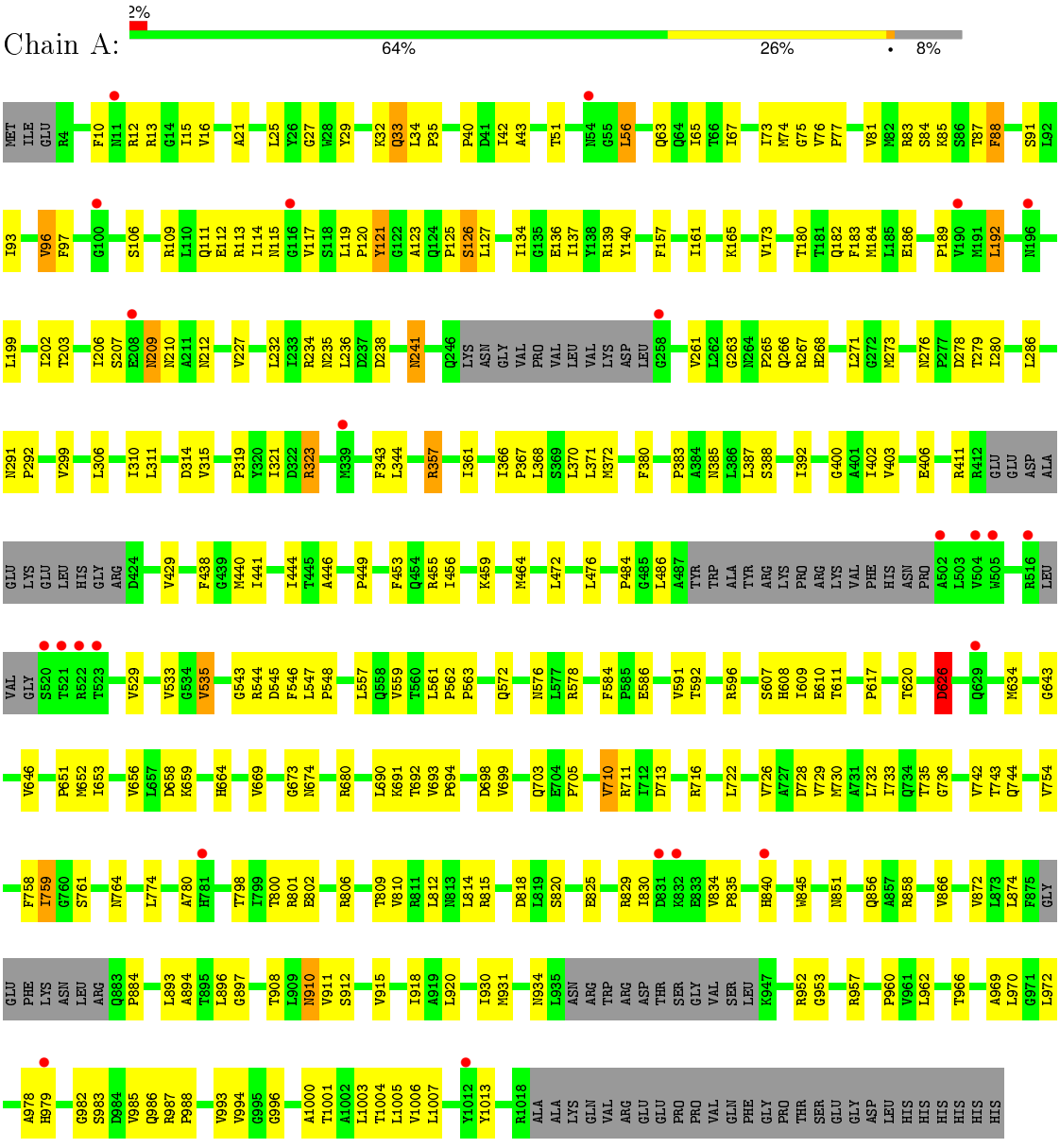
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
2	B	1	Total Zn 1 1	0	0
2	A	1	Total Zn 1 1	0	0
2	D	1	Total Zn 1 1	0	0
2	E	1	Total Zn 1 1	0	0

3 Residue-property plots (i)

These plots are drawn for all protein, RNA and DNA chains in the entry. The first graphic for a chain summarises the proportions of errors 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.

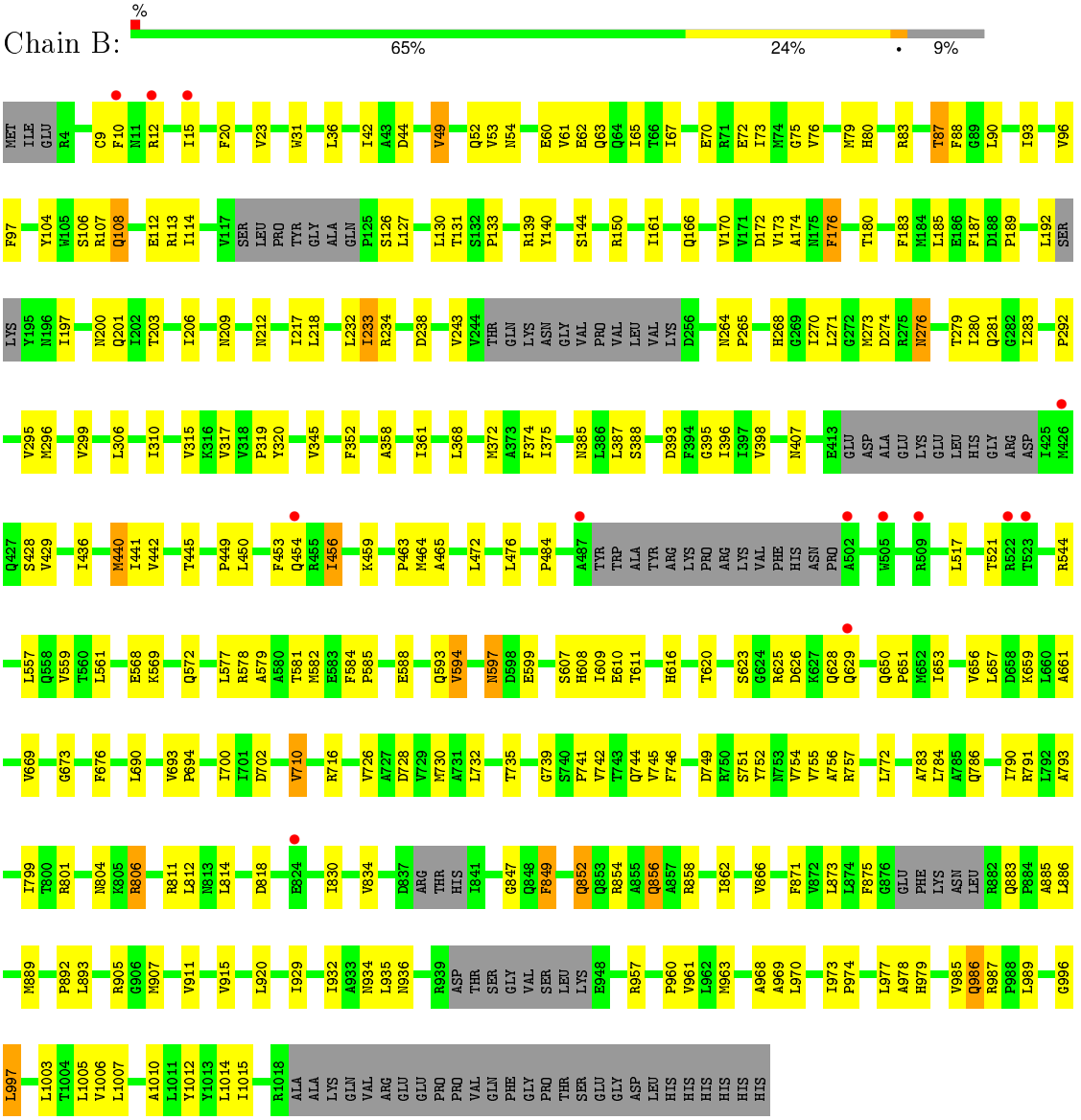
- Molecule 1: Heavy metal cation tricomponent efflux pump ZneA(CzcA-like)

Chain A:



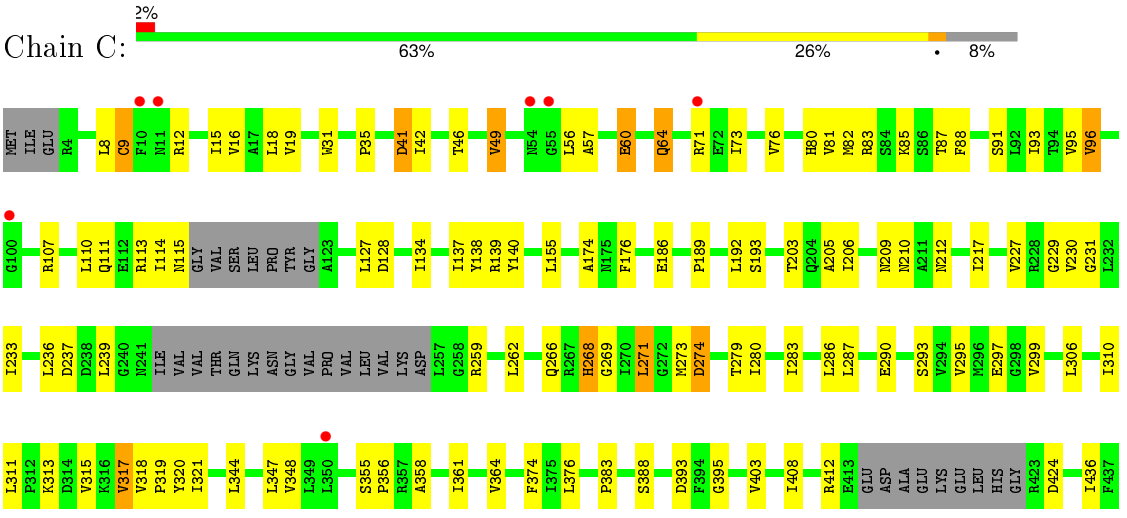
- Molecule 1: Heavy metal cation tricomponent efflux pump ZneA(CzcA-like)

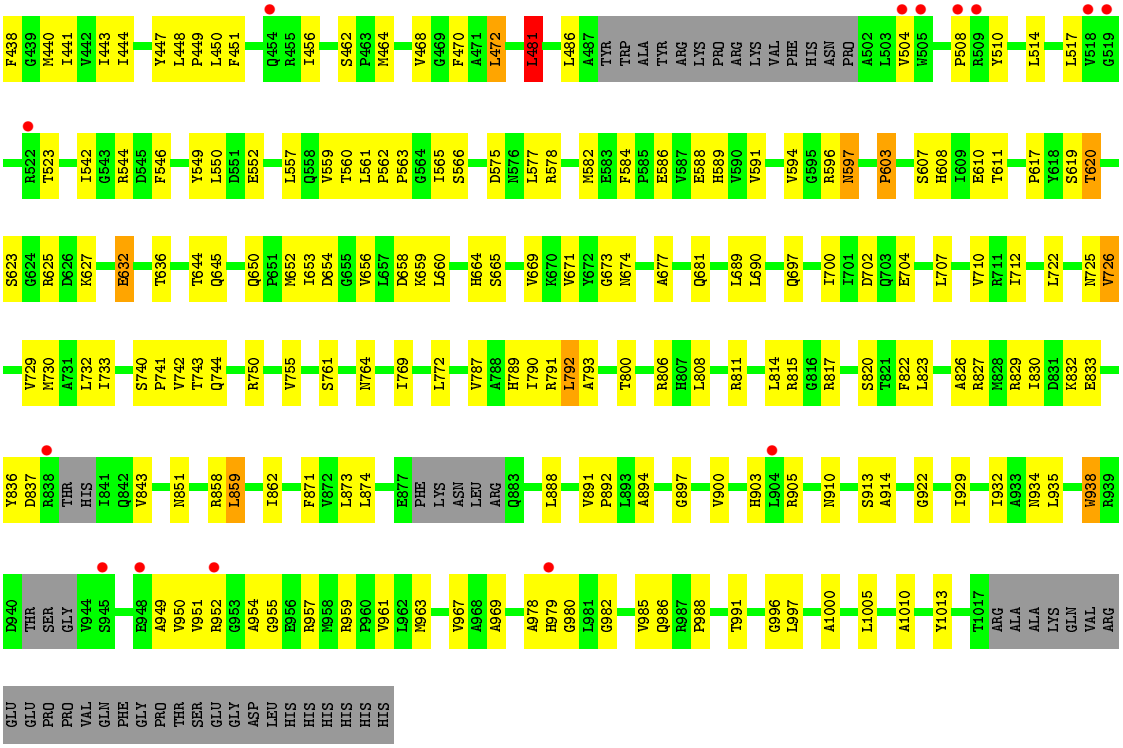
Chain B:



• Molecule 1: Heavy metal cation tricomponent efflux pump ZneA(CzcA-like)

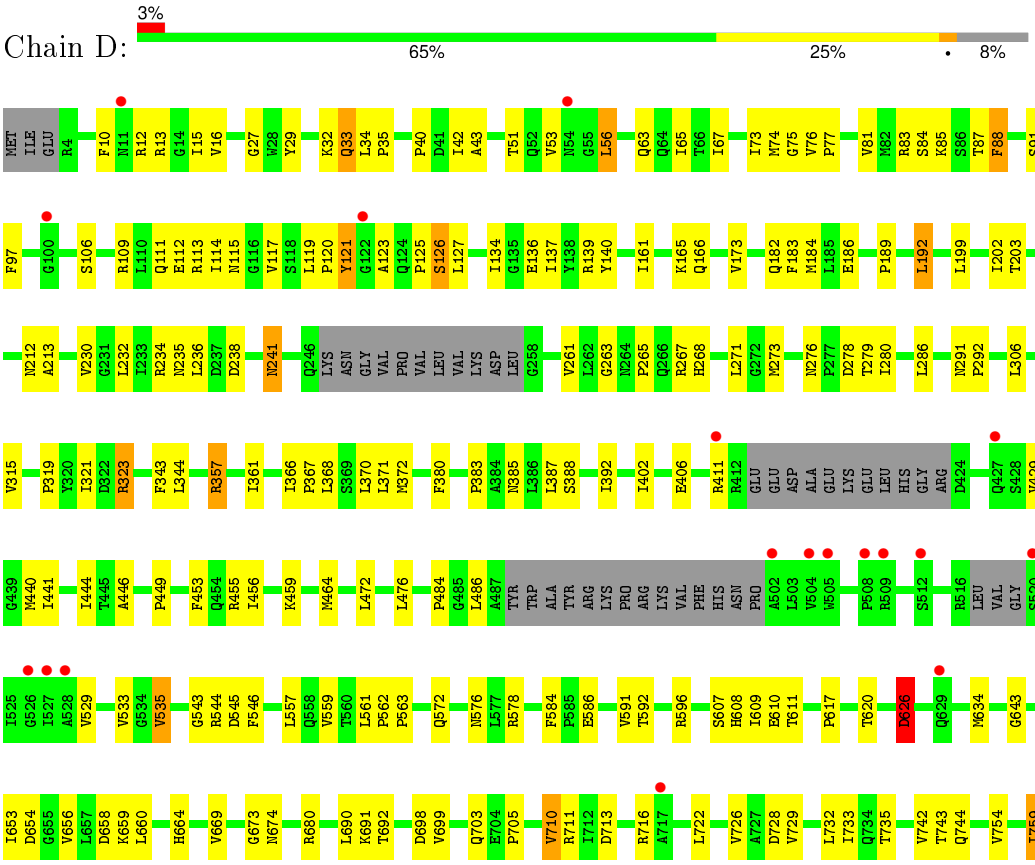
Chain C:

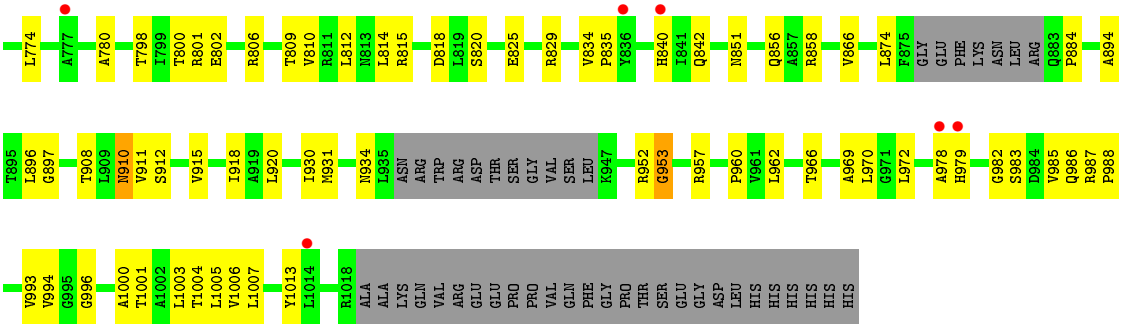




• Molecule 1: Heavy metal cation tricomponent efflux pump ZneA(CzcA-like)

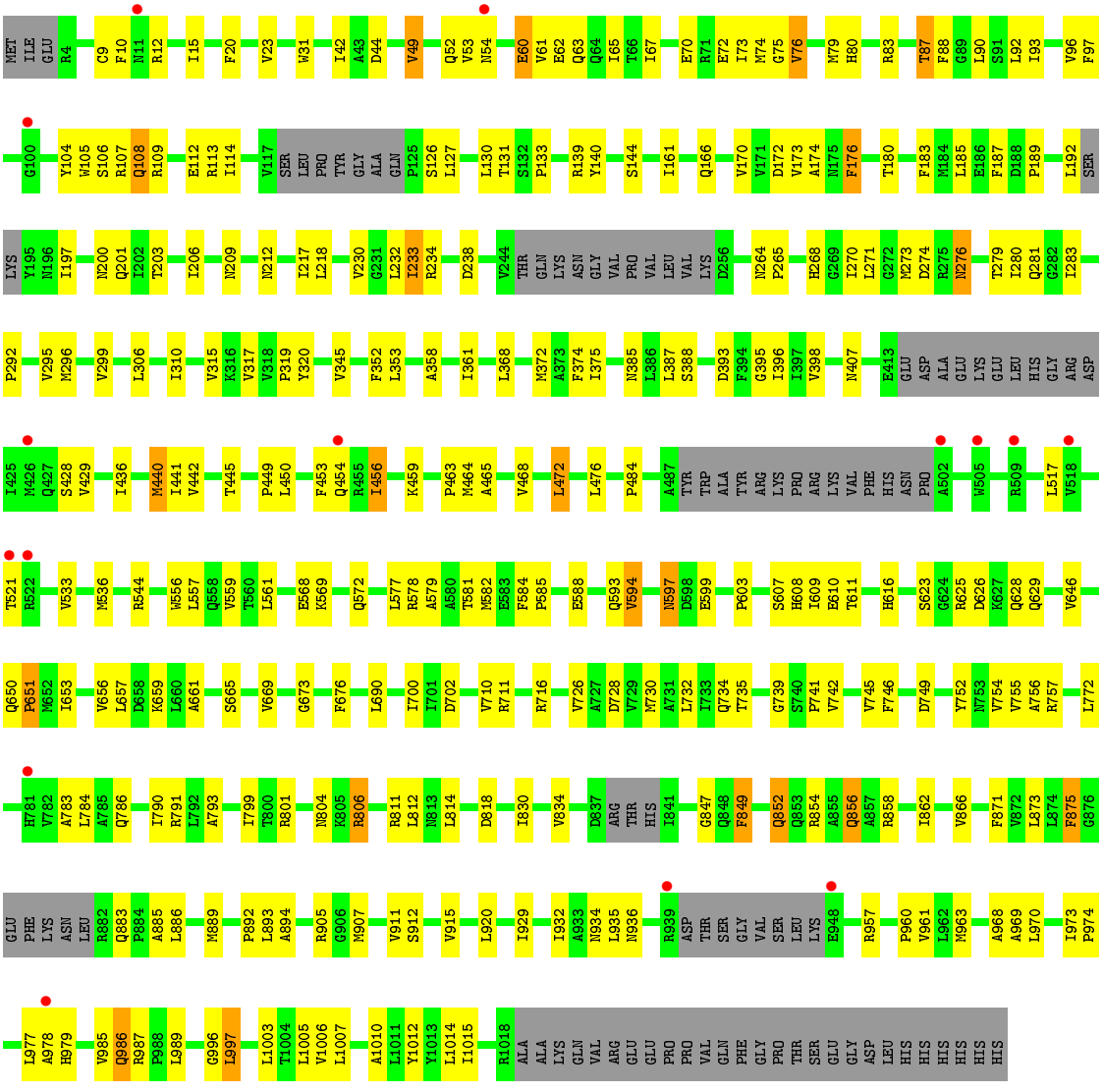
Chain D:





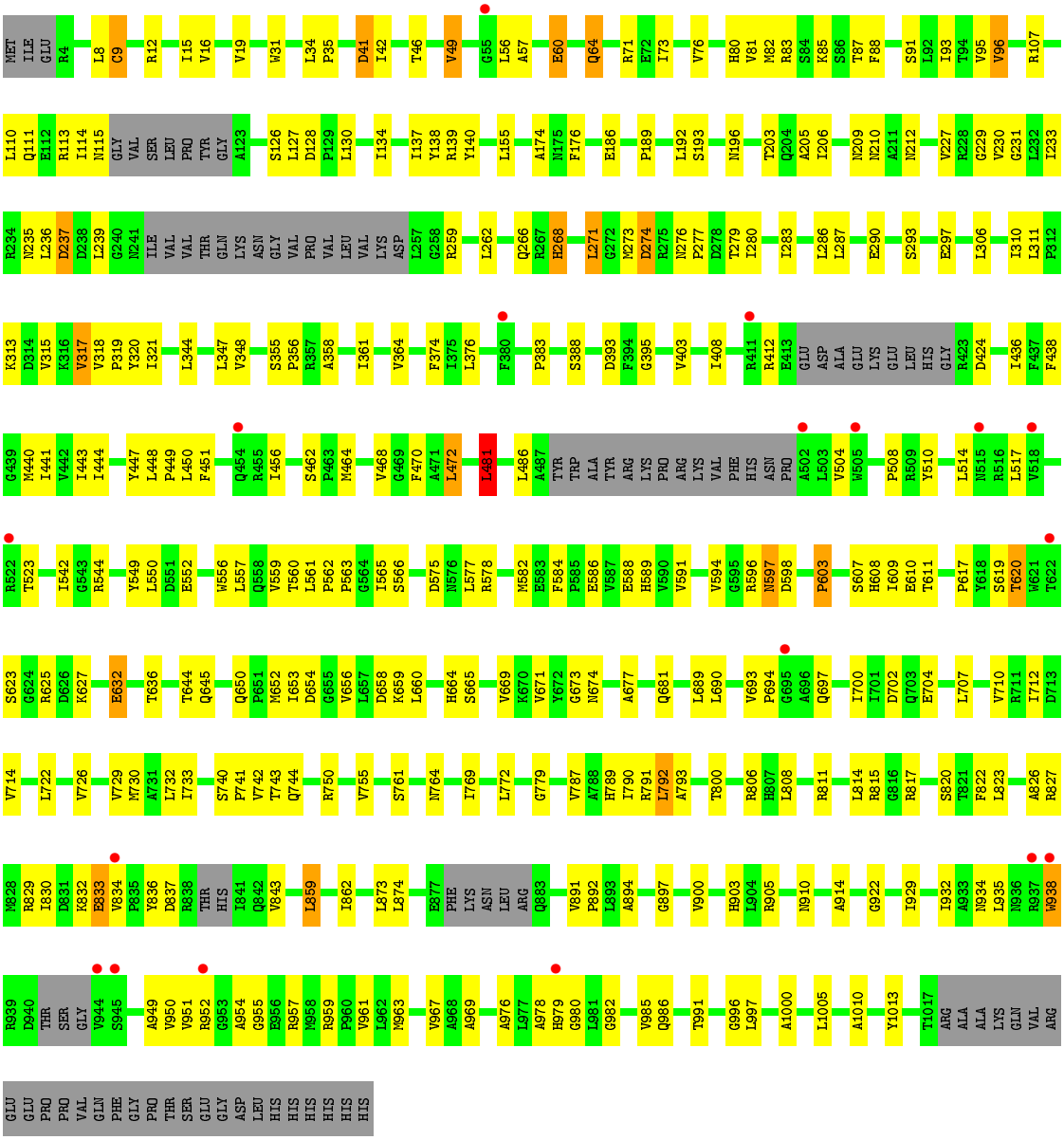
• Molecule 1: Heavy metal cation tricomponent efflux pump ZneA(CzcA-like)

Chain E:



• Molecule 1: Heavy metal cation tricomponent efflux pump ZneA(CzcA-like)

Chain F:


MET 4 ILE 8 GLU C9 R12 I15 V16 V19 W31 L34 P35 D41 I42 T46 V49 G55 L56 A57 E60 Q64 R71 E72 I73 V76 H80 V81 M82 R83 S84 K85 S86 T87 F88 S91 L92 I93 T94 V95 V96 R107
L110 Q111 E112 R113 I114 N115 GLY VAL SER LEU PRO TYR GLY A123 S126 L127 D128 P129 L130 I134 I137 Y138 R139 Y140 L155 A174 N175 F176 E186 P189 L192 S193 N196 T203 Q204 A205 I206 N209 N210 A211 N212 V227 R228 G229 V230 G231 L232 I233
R234 N235 L236 D237 D238 L239 G240 N241 ILE VAL VAL THR GLN LYS ASN GLY VAL PRO VAL LEU VAL LYS ASP L257 G258 R259 L262 Q266 R267 H268 L271 G272 M273 D274 R275 N276 P277 D278 T279 I280 I283 L286 L287 E290 S293 E297 I306 I310 L311 P312
K313 D314 V315 K316 V317 V318 P319 Y320 I321 L344 L347 V348 S355 P356 R357 A358 I361 V364 F374 I375 L376 F380 P383 S388 D393 F394 G395 V403 I408 R411 R412 E413 GLU ASP ALA GLU LYS GLU LEU HIS GLY R423 D424 I436 F437 F438
G439 M440 I441 V442 I443 I444 Y447 L448 P449 L450 F451 Q454 R455 I456 S462 P463 M464 V468 G469 F470 A471 L472 L481 L486 A487 TYR TRP ALA TYR ARG LYS PRO ARG LYS VAL PHE HIS ASN PRO A502 L503 V504 W505 P508 R509 Y510 L514 N515 R516 L517 V518
R522 T523 I542 G543 R544 Y549 L550 D551 E552 W556 L557 Q558 V559 T560 L561 P562 P563 G564 I565 S566 D575 N576 L577 R578 M582 E583 F584 P585 E586 V587 E588 H589 V590 V591 V594 G595 R596 N597 D598 P603 S607 H608 I609 E610 T611 P617 Y618 S619 T620 W621 T622
S623 G624 R625 D626 K627 E632 T636 T644 Q645 Q650 F651 M652 I653 D654 G655 V656 L657 D658 K659 L660 H664 S665 V669 K670 V671 Y672 G673 N674 A677 Q681 L689 L690 V693 P694 G695 A696 Q697 I700 I701 D702 Q703 E704 L707 V710 R711 I712 D713
V714 L722 V726 V729 M730 A731 L732 I733 S740 F741 V742 T743 Q744 R750 V755 S761 N764 I769 L772 G779 V787 A788 H789 I790 R791 L792 A793 T800 R806 H807 L808 R811 L814 R815 G816 R817 S820 T821 F822 L823 A826 R827
M828 R829 I830 D831 K832 E833 V834 P835 Y836 D837 R838 THR HIS I841 Q842 V843 L859 I862 L873 L874 E877 PHE LYS ASN LEU ARG Q883 V891 P892 L893 A894 G897 V900 H903 L904 R905 N910 A914 G922 I929 I932 A933 N934 L935 N936 R937 M938
R939 D940 THR SER GLY V944 S945 A949 V950 V951 R952 G953 A954 G955 E956 R957 M958 R959 P960 V961 L962 M963 V967 A968 A969 A976 L977 A978 H979 G980 L981 G982 V985 Q986 Y991 G996 L997 A1000 L1005 A1010 Y1013 T1017 ARG ALA ALA LYS GLN VAL ARG
GLU GLU PRO PRO VAL GLN PHE GLY PRO THR SER GLU GLY ASP LEU HIS HIS HIS HIS HIS HIS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	223.63Å 129.06Å 392.37Å 90.00° 94.62° 90.00°	Depositor
Resolution (Å)	19.98 – 3.00 19.98 – 2.99	Depositor EDS
% Data completeness (in resolution range)	97.3 (19.98-3.00) 95.9 (19.98-2.99)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.11	Depositor
$< I/\sigma(I) >$ [1]	2.24 (at 2.98Å)	Xtriage
Refinement program	PHENIX (phenix.refine: 1.8.2_1309)	Depositor
R, R_{free}	0.279 , 0.305 0.280 , 0.308	Depositor DCC
R_{free} test set	1878 reflections (0.88%)	DCC
Wilson B-factor (Å^2)	60.5	Xtriage
Anisotropy	0.086	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 73.1	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.34$, $< L^2 > = 0.16$	Xtriage
Outliers	8 of 219455 reflections (0.004%)	Xtriage
F_o,F_c correlation	0.84	EDS
Total number of atoms	44018	wwPDB-VP
Average B, all atoms (Å^2)	78.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.95% 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, 0.375 respectively for untwinned datasets, and 0.333, 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: ZN

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	0.25	0/7461	0.46	0/10149
1	B	0.26	0/7436	0.46	0/10111
1	C	0.26	0/7480	0.46	0/10170
1	D	0.26	0/7461	0.46	0/10149
1	E	0.26	0/7436	0.46	0/10111
1	F	0.26	0/7480	0.46	0/10170
All	All	0.26	0/44754	0.46	0/60860

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	7336	0	7607	168	0
1	B	7314	0	7588	155	0
1	C	7357	0	7626	173	0
1	D	7336	0	7607	161	0
1	E	7314	0	7588	167	0
1	F	7357	0	7626	175	0
2	A	1	0	0	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	B	1	0	0	0	0
2	D	1	0	0	0	0
2	E	1	0	0	0	0
All	All	44018	0	45642	966	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 11.

The worst 5 of 966 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:D:388:SER:HB2	1:D:464:MET:HB2	1.60	0.83
1:D:429:VAL:HG21	1:D:484:PRO:HB3	1.61	0.82
1:A:388:SER:HB2	1:A:464:MET:HB2	1.61	0.81
1:A:429:VAL:HG21	1:A:484:PRO:HB3	1.62	0.81
1:D:544:ARG:HH22	1:D:856:GLN:HG2	1.50	0.77

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	944/1045 (90%)	837 (89%)	92 (10%)	15 (2%)	12 48
1	B	936/1045 (90%)	858 (92%)	76 (8%)	2 (0%)	52 88
1	C	943/1045 (90%)	850 (90%)	82 (9%)	11 (1%)	16 56
1	D	944/1045 (90%)	836 (89%)	93 (10%)	15 (2%)	12 48
1	E	936/1045 (90%)	860 (92%)	74 (8%)	2 (0%)	52 88
1	F	943/1045 (90%)	851 (90%)	81 (9%)	11 (1%)	16 56
All	All	5646/6270 (90%)	5092 (90%)	498 (9%)	56 (1%)	19 61

5 of 56 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	88	PHE
1	C	979	HIS
1	D	88	PHE
1	F	979	HIS
1	A	74	MET

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	787/864 (91%)	755 (96%)	32 (4%)	37	76
1	B	784/864 (91%)	746 (95%)	38 (5%)	31	71
1	C	789/864 (91%)	746 (95%)	43 (5%)	27	66
1	D	787/864 (91%)	756 (96%)	31 (4%)	39	77
1	E	784/864 (91%)	747 (95%)	37 (5%)	32	72
1	F	789/864 (91%)	747 (95%)	42 (5%)	28	67
All	All	4720/5184 (91%)	4497 (95%)	223 (5%)	32	72

5 of 223 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	710	VAL
1	D	392	ILE
1	F	620	THR
1	C	750	ARG
1	D	33	GLN

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 68 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	764	ASN
1	D	664	HIS

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	F	289	ASN
1	C	804	ASN
1	D	33	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 carbohydrates in this entry.

5.6 Ligand geometry

Of 4 ligands modelled in this entry, 4 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	958/1045 (91%)	-0.37	24 (2%)	61	30	29, 77, 143, 215	0
1	B	954/1045 (91%)	-0.44	13 (1%)	78	51	27, 70, 135, 209	0
1	C	959/1045 (91%)	-0.39	21 (2%)	65	35	26, 71, 140, 241	0
1	D	958/1045 (91%)	-0.31	29 (3%)	54	25	29, 77, 142, 215	0
1	E	954/1045 (91%)	-0.44	15 (1%)	74	47	25, 70, 134, 209	0
1	F	959/1045 (91%)	-0.40	18 (1%)	70	41	26, 71, 139, 241	0
All	All	5742/6270 (91%)	-0.39	120 (2%)	67	36	25, 73, 140, 241	0

The worst 5 of 120 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	C	518	VAL	7.5
1	F	979	HIS	5.7
1	D	522	ARG	5.5
1	A	522	ARG	5.5
1	A	504	VAL	5.2

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 carbohydrates 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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
2	ZN	E	1101	1/1	0.88	0.09	-3.48	99,99,99,99	0
2	ZN	B	1101	1/1	0.96	0.04	-	108,108,108,108	0
2	ZN	D	1101	1/1	0.97	0.08	-	111,111,111,111	1
2	ZN	A	1101	1/1	0.96	0.08	-	112,112,112,112	0

6.5 Other polymers

There are no such residues in this entry.