wwPDB X-ray Structure Validation Summary Report (i)

Feb 19, 2016 – 11:36 PM GMT

|		
PDB ID	:	5D98
Title	:	Influenza C Virus RNA-dependent RNA Polymerase - Space group P43212
Authors	:	Hengrung, N.; El Omari, K.; Serna Martin, I.; Vreede, F.T.; Cusack, S.; Rambo, R.P.; Vonrhein, C.; Bricogne, G.; Stuart, D.I.; Grimes, J.M.; Fodor, E.
Deposited on	:	2015-08-18
Resolution	:	3.90 Å(reported)

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 3.90 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.
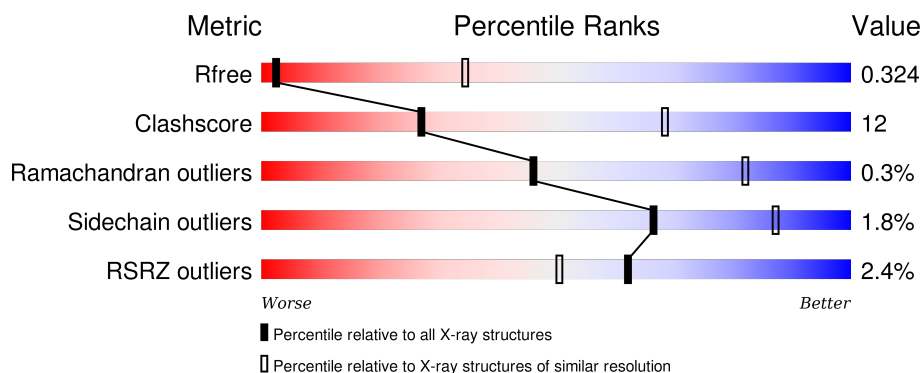
Metric	Whole archive (#Entries)	Similar resolution (#Entries, resolution range(Å))
R_{free}	91344	1014 (4.28-3.52)
Clashscore	102246	1031 (4.24-3.56)
Ramachandran outliers	100387	1012 (4.26-3.54)
Sidechain outliers	100360	1004 (4.26-3.54)
RSRZ outliers	91569	1018 (4.28-3.52)

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	709	2% / 67% / 30% / ••
1	D	709	% / 68% / 29% / ••
2	B	754	2% / 74% / 19% / • 6%
2	E	754	3% / 73% / 20% / • 6%
3	C	782	3% / 70% / 26% / ••

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	F	782	

2 Entry composition (i)

There are 4 unique types of molecules in this entry. The entry contains 34720 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 Polymerase acidic protein.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	693	Total 5630	C 3589	N 954	O 1043	S 44	0	0	0
1	D	693	Total 5630	C 3589	N 954	O 1043	S 44	0	0	0

- Molecule 2 is a protein called RNA-directed RNA polymerase catalytic subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	711	Total 5652	C 3587	N 956	O 1056	S 53	0	0	0
2	E	711	Total 5652	C 3587	N 956	O 1056	S 53	0	0	0

- Molecule 3 is a protein called Polymerase basic protein 2.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	762	Total 6076	C 3845	N 1066	O 1128	S 37	0	0	0
3	F	762	Total 6076	C 3845	N 1066	O 1128	S 37	0	0	0

There are 16 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	775	ALA	-	expression tag	UNP Q9IMP3
C	776	ARG	-	expression tag	UNP Q9IMP3
C	777	GLU	-	expression tag	UNP Q9IMP3
C	778	ASN	-	expression tag	UNP Q9IMP3
C	779	LEU	-	expression tag	UNP Q9IMP3
C	780	TYR	-	expression tag	UNP Q9IMP3
C	781	PHE	-	expression tag	UNP Q9IMP3
C	782	GLN	-	expression tag	UNP Q9IMP3

Continued on next page...

Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
F	775	ALA	-	expression tag	UNP Q9IMP3
F	776	ARG	-	expression tag	UNP Q9IMP3
F	777	GLU	-	expression tag	UNP Q9IMP3
F	778	ASN	-	expression tag	UNP Q9IMP3
F	779	LEU	-	expression tag	UNP Q9IMP3
F	780	TYR	-	expression tag	UNP Q9IMP3
F	781	PHE	-	expression tag	UNP Q9IMP3
F	782	GLN	-	expression tag	UNP Q9IMP3

- Molecule 4 is MAGNESIUM ION (three-letter code: MG) (formula: Mg).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	2	Total 2	Mg 2	0	0
4	D	2	Total 2	Mg 2	0	0

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 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: Polymerase acidic protein

- Molecule 1: Polymerase acidic protein

L355 D358 T362 E363 G364 L365 K366 Q367 S368 F378 P379 R380 W381 F382 I390 K391 T396 G397 W398 V399 A402 P405 N409 E410 E416 N419 K420 V425 L426 Q427 N430 C431 A432 R433 E434 M435 G436 K437 A444 E448 C449 R450 L451 F452

K457 P460 R464 E467 R468 K469 E473 G474 L475 P476 V477 P478 S479 E480 M481 D482 I487 C488 V489 K490 H494 LEU ASN LYS ASP D499 I504 I505 T506 F507 S510 P514 N515 K518 H519 V524 F525 H529 K536 GLY SER VAL ILE GLY R543

P546 L547 T554 K558 N561 D562 S565 K566 R569 S586 E591 N592 K596 T597 L598 I605 L617 L621 R622 V623 S637 Q646 I649 D656 K657 K658 M659 K660 F665 N666 P667 N681 C684 L685 F686 Q689 R690 L691 N692 V695

F705 F706 K707 T708 ASP

● Molecule 2: RNA-directed RNA polymerase catalytic subunit

Chain B:

2%
74%   19%   • 6%

M1 E2 I3 N4 P5 Y6 L7 M8 F9 I18 Y22 P23 Y24 T25 G26 P27 R51 T52 S53 V54 E55 R66 K67 F68 E73 E77 L78 V79 I87 L90 L93 F106 S123 T131 W134 Q147 V150 K154 P159 L166 E167

F183 V190 ARG ARG GLU LYS ASP LYS ARG SER GLY LYS GLU ILE LYS THR LYS VAL PRO V208 I221 T225 K231 G236 K237 L238 R241 A244 R251 C265 L273 P274 E279 K280 K281 A282 V302 N303 I304 C314 E318 I327 K336

A342 P343 F346 K349 F350 I357 R358 L359 S360 N361 K362 R363 K364 T365 K366 E367 V368 I369 I370 K374 Y378 M382 Y386 L389 L393 Q398 K399 D400 V401 P405 G406 G407 M408 L409 M412 L416 S417 T418 V419 L420 G421 T424 Y427 MET ASP

GLU E431 L432 G436 C437 F438 W439 T440 G441 L442 Q443 L450 V453 A454 W457 I460 I464 F467 V470 I474 G475 I476 N477 L480 L487 P488 E489 L490 F491 E492 F493 M496 F497 F498 F502 V503 S504 N505 L506 A507 T514 G517 D523

K532 T533 N534 M535 N538 S539 L540 S543 T544 A545 L546 K570 M574 K575 I576 I577 F580 I584 L590 L598 N601 N627 P628 F632 THR ASN PHE ASP LYS THR ILE ASP ILE PHE ARG ALA HIS GLY PRO ILE ARG VAL GLU GLU ASN GLU A655 R663

M675 M679 Y685 Q686 E697 I701 A707 I710 E720 M724 S753 ARG

● Molecule 2: RNA-directed RNA polymerase catalytic subunit

Chain E:

3%
73%   20%   • 6%

M1 E2 I3 N4 L7 V14 I18 Y22 P23 Y24 T25 P29 H32 Y49 S50 R51 T52 S53 V54 E55 V60 F61 R66 K67 F68 C69 E73 E77 L78 V79 K80 P81 I87 S88 S89 L90 L93 M97 F105 F106 T131 W134

V150 K154 P159 L166 E167 F183 V190 ARG ARG GLU ASP LYS ARG SER GLY LYS GLU ILE LYS THR LYS VAL PRO V208 A230 K231 I232 G236 K237 L238 Q239 R240 R241 A242 I243 A244 R251 P252 F253 I256 I264 C265 L273 P274 E279 K280 K281

A282 R294 V302 N303 I304 C314 E318 I327 L334 M335 D337 S340 V341 A342 P343 N348 K349 F350 I357 R358 L359 S360 N361 K362 R363 V368 I369 I370 K374 M382 Y386 L389 L393 Q398 K399 D400 V401 L404 P405 G406 G407 M408

L409 M412 L416 S417 T418 V419 L420 G421 V422 S423 T424 L425 C426 Y427 MET ASP GLU E431 L432 G436 C437 F438 V439 T440 G441 L442 V449 F451 A452 V453 A454 W457 I460 I464 R465 R466 F467 V470 I474 G475 I476 N477 L480 G485 S486 L487 P488 E489

L490 F491 E492 F493 M496 F497 F498 F502 V503 S504 N505 L506 D523 L540 S541 P542 S543 R550 R557 Y560 H563 K570 M574 K575 I576 I577 N578 E579 I584 L590 N601 N627 P628 F632 THR ASN PHE ASP LYS THR ILE ASP ILE PHE ARG

ALA HIS GLY PRO ILE ARG VAL GLU GLU ASN GLU A655 V656 V657 H660 M675 M679 A680 K683 Q686 I701 I705 A749 S753 ARG

● Molecule 3: Polymerase basic protein 2

Chain C:

M1 L4 T28 V29 T34 T39 T40 S41 P47 S57 K58 F59 P60 I61 I62 E76 V80 A81 L82 E87 ASP VAL SER K91 V95 S100 N103 F107 C111 Y122 K123 S124 R128 R131 I135 K138 E139 L144 R147

Q148 R149 R150 R151 V152 D153 T154 V157 E158 Q159 R160 L161 E165 I166 Q170 L174 F175 E176 D177 N189 E195 M196 R197 S198 K199 F200 L205 E208 A211 H212 M213 F218 V227 F228 G229 A230 R235 M236 E237 I238 I239 G243 G244 W247 E250

T253 K263 N264 D265 I266 R267 A268 V269 K272 A277 S280 I281 M282 N283 A284 K287 R298 I299 G300 E301 T302 E308 L309 I310 S317 V320 C323 G328 K333 T334 L335 M340 L341 L342 K343 S346 G349 V350 K351 T355 I358 GLN

GLY VAL ARG ALA VAL Q365 F366 E367 Y368 W369 S370 E371 E374 S381 A384 S387 E390 R391 S392 L393 E394 W395 I396 T397 I398 D405 R406 K407 R408 L409 L410 M414 I415 F416 C417 R418 D419 Y422 A426 P427 A428 T431 K438 L439 G440 R441 E442 I443 P444

Y445 V448 M449 M450 N451 W452 I453 L463 R467 G468 I469 S479 L493 T497 I498 Q499 I503 D504 S517 E521 A522 F523 A531 T532 I533 L534 D535 L536 P537 K541 V542 T543 V547 Q550 H551 P552 D553 L554 T565 R574 L575 G578 I582

P583 Y591 L594 V600 I607 R611 K612 G613 V614 R620 Q626 V627 Y632 S633 L634 L637 I638 K639 P642 F643 T644 G651 F658 N670 F674 S675 W676 S681 I684 Y685 C686 H687 G688 I691 R692 V693 P694 L695 V696 V697 G698 D699 E700

H701 M702 D703 T704 A707 L708 L709 S713 V714 C715 E716 N717 D718 P719 R720 A721 T725 F734 K737 V738 R739 T749 F750 K751 R752 T753 V763 V767 K768 K769 I770 K771 MET SER ASN ALA ARG GLU ASN LEU TYR PHE GLN

● Molecule 3: Polymerase basic protein 2

Chain F:

M1 L4 Q16 T28 V29 T39 L49 R52 W53 S56 S57 K58 F59 P60 E76 H77 V80 A81 L82 E87 ASP VAL SER K91 I102 N103 N106 C111 V117 V121 Y122 K123 S124 R125 F126 G127 R128 K129 E130 R131 I135 M136 W137 K138

E139 L140 L144 Q148 R149 V152 V157 E158 Q159 R160 L161 E165 I166 Q170 L174 F175 E176 D177 N189 E195 M196 R197 S198 K199 N202 K203 P204 L205 E208 A211 H212 M213 F218 F228 G229 A230 E234 R235 M236 E237 L238 I239 W247 I248

Q249 E250 T253 A254 K263 N264 D265 I266 R267 A268 V269 C270 R271 K272 V273 C274 A276 A277 S280 N283 K287 L288 I292 R298 I299 G300 E308 L309 I310 S317 P318 E319 V320 T321 L322 C323 K324 S325 Q330 L331 G332 K333 M340 K343 K344 I345 S346

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, α, β, γ	185.66Å 185.66Å 598.22Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.01 – 3.90 100.57 – 3.90	Depositor EDS
% Data completeness (in resolution range)	98.8 (50.01-3.90) 98.8 (100.57-3.90)	Depositor EDS
R_{merge}	0.20	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.92 (at 3.89Å)	Xtriage
Refinement program	REFMAC 5.8.0103	Depositor
R, R_{free}	0.286 , 0.326 0.286 , 0.324	Depositor DCC
R_{free} test set	4770 reflections (5.28%)	DCC
Wilson B-factor (Å^2)	161.2	Xtriage
Anisotropy	0.644	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.27 , 132.8	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning[2]	$< |L| > = 0.42, < L^2 > = 0.24$	Xtriage
Outliers	0 of 95152 reflections	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	34720	wwPDB-VP
Average B, all atoms (Å^2)	207.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.59% 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 (i)

5.1 Standard geometry (i)

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

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	0.41	0/5746	0.57	1/7717 (0.0%)
1	D	0.43	0/5746	0.57	1/7717 (0.0%)
2	B	0.40	0/5749	0.56	0/7723
2	E	0.41	0/5749	0.57	0/7723
3	C	0.41	0/6185	0.59	0/8322
3	F	0.42	0/6185	0.59	0/8322
All	All	0.41	0/35360	0.58	2/47524 (0.0%)

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	219	ASN	C-N-CD	5.11	139.14	128.40
1	D	219	ASN	C-N-CD	5.11	139.12	128.40

There are no chirality outliers.

There are no planarity outliers.

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	5630	0	5632	147	0
1	D	5630	0	5632	167	0
2	B	5652	0	5749	127	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	E	5652	0	5749	126	0
3	C	6076	0	6183	182	0
3	F	6076	0	6183	198	0
4	A	2	0	0	0	0
4	D	2	0	0	0	0
All	All	34720	0	35128	840	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 12.

The worst 5 of 840 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:138:LYS:HE2	3:F:250:GLU:CG	1.22	1.59
3:F:138:LYS:CE	3:F:250:GLU:HG2	1.34	1.48
3:F:138:LYS:CE	3:F:250:GLU:CG	1.83	1.46
3:F:138:LYS:CE	3:F:250:GLU:CD	1.86	1.41
3:F:138:LYS:NZ	3:F:250:GLU:HG2	1.33	1.36

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	685/709 (97%)	643 (94%)	42 (6%)	0	100	100
1	D	685/709 (97%)	645 (94%)	40 (6%)	0	100	100
2	B	703/754 (93%)	666 (95%)	35 (5%)	2 (0%)	46	82
2	E	703/754 (93%)	673 (96%)	27 (4%)	3 (0%)	39	79
3	C	756/782 (97%)	688 (91%)	63 (8%)	5 (1%)	26	70

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
3	F	756/782 (97%)	686 (91%)	67 (9%)	3 (0%)	39	79
All	All	4288/4490 (96%)	4001 (93%)	274 (6%)	13 (0%)	46	82

5 of 13 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
2	B	405	PRO
3	C	698	GLY
2	E	405	PRO
2	B	503	VAL
2	E	503	VAL

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	618/631 (98%)	603 (98%)	15 (2%)	57	82
1	D	618/631 (98%)	608 (98%)	10 (2%)	70	88
2	B	629/669 (94%)	620 (99%)	9 (1%)	74	89
2	E	629/669 (94%)	618 (98%)	11 (2%)	68	88
3	C	669/686 (98%)	660 (99%)	9 (1%)	76	89
3	F	669/686 (98%)	655 (98%)	14 (2%)	61	85
All	All	3832/3972 (96%)	3764 (98%)	68 (2%)	66	87

5 of 68 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
3	C	450	MET
1	D	366	LYS
3	F	445	TYR
3	C	594	LEU
1	D	48	PHE

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 39 such sidechains are listed below:

Mol	Chain	Res	Type
3	C	550	GLN
1	D	141	ASN
3	F	717	ASN
3	C	717	ASN
1	D	31	HIS

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)

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 (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(Å²)	Q<0.9
1	A	693/709 (97%)	-0.05	17 (2%)	61	48	132, 195, 269, 299	0
1	D	693/709 (97%)	-0.11	4 (0%)	90	86	129, 191, 258, 295	0
2	B	711/754 (94%)	-0.03	13 (1%)	71	61	131, 192, 256, 317	0
2	E	711/754 (94%)	0.06	23 (3%)	51	38	130, 184, 266, 315	0
3	C	762/782 (97%)	0.02	23 (3%)	54	40	149, 222, 285, 358	0
3	F	762/782 (97%)	0.13	26 (3%)	49	37	134, 224, 287, 343	0
All	All	4332/4490 (96%)	0.00	106 (2%)	62	50	129, 202, 275, 358	0

The worst 5 of 106 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	E	242	ALA	6.1
3	F	631	HIS	5.3
2	E	237	LYS	5.1
3	C	1	MET	4.8
3	C	696	VAL	4.7

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.

wwPDB X-ray Structure Validation Summary Report

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. 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, 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	LLDF	B-factors(Å^2)	Q<0.9
4	MG	A	801	1/1	0.93	0.10	-1.07	185,185,185,185	0
4	MG	D	800	1/1	0.94	0.09	-1.28	200,200,200,200	0
4	MG	A	800	1/1	0.89	0.07	-1.43	204,204,204,204	0
4	MG	D	801	1/1	0.90	0.16	-	182,182,182,182	0

6.5 Other polymers (i)

There are no such residues in this entry.