wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 09:29 AM GMT

PDB ID	:	3IKM
Title	:	Crystal structure of human mitochondrial DNA polymerase holoenzyme
Authors	:	Lee, Y-S.; Kennedy, W.D.; Yin, Y.W.
Deposited on	:	2009-08-06
Resolution	:	3.24 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:

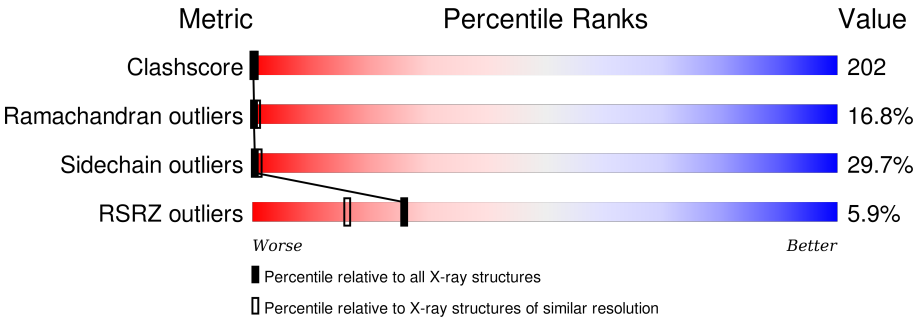
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)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

1 Overall quality at a glance

The following experimental techniques were used to determine the structure:
X-RAY DIFFRACTION

The reported resolution of this entry is 3.24 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))
Clashscore	102246	1227 (3.28-3.20)
Ramachandran outliers	100387	1204 (3.28-3.20)
Sidechain outliers	100360	1203 (3.28-3.20)
RSRZ outliers	91569	1097 (3.28-3.20)

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 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	1172	5% 5% 45% 34% 9% 7%
1	D	1172	5% 5% 43% 34% 11% 7%
2	B	427	7% 11% 50% 17% 5% 17%
2	C	427	5% 6% 54% 25% 8% 7%
2	E	427	7% 6% 55% 19% • 17%
2	F	427	5% 5% 54% 25% 8% 7%

2 Entry composition

There are 2 unique types of molecules in this entry. The entry contains 29480 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 DNA polymerase subunit gamma-1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	1093	Total	C	N	O	S	0	0	0
			8681	5505	1537	1589	50			
1	D	1094	Total	C	N	O	S	0	0	0
			8695	5515	1540	1590	50			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	354	ASN	-	SEE REMARK 999	UNP P54098
A	355	SER	-	SEE REMARK 999	UNP P54098
D	354	ASN	-	SEE REMARK 999	UNP P54098
D	354A	SER	-	SEE REMARK 999	UNP P54098

- Molecule 2 is a protein called DNA polymerase subunit gamma-2.

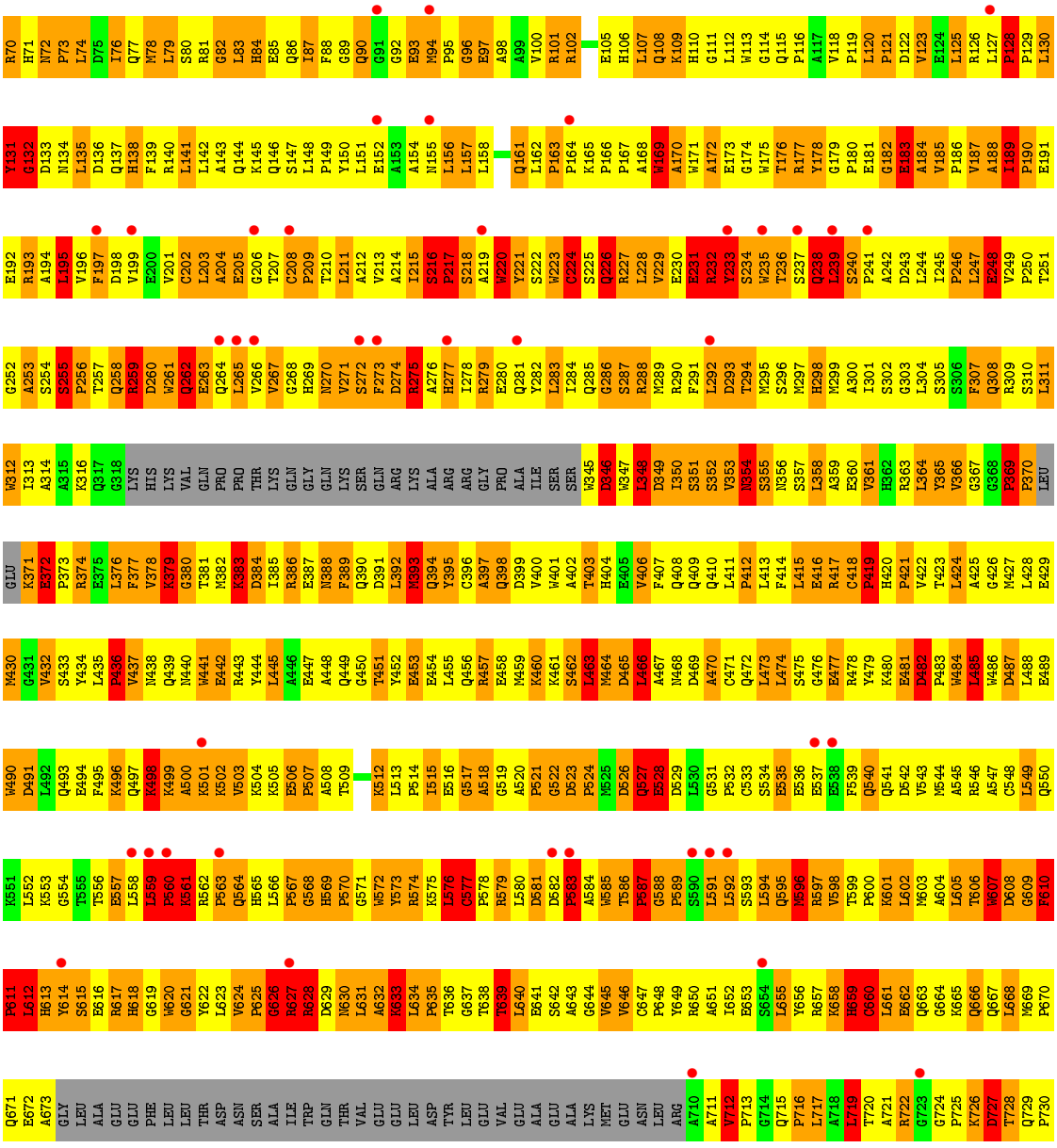
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	355	Total	C	N	O	S	0	0	0
			2871	1840	504	511	16			
2	C	396	Total	C	N	O	S	0	0	0
			3181	2031	563	571	16			
2	E	355	Total	C	N	O	S	0	0	0
			2871	1840	504	511	16			
2	F	396	Total	C	N	O	S	0	0	0
			3181	2031	563	571	16			

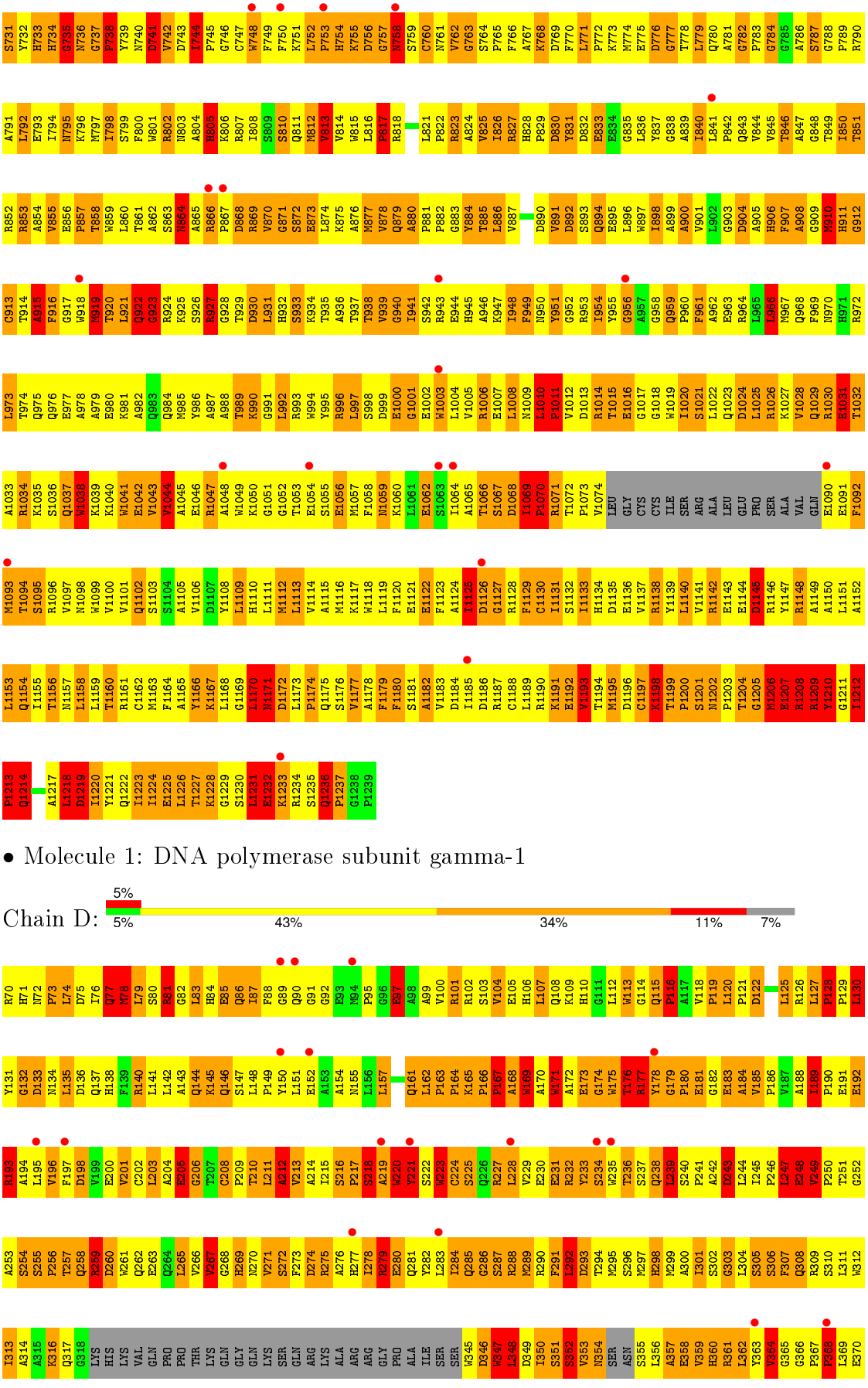
3 Residue-property plots (i)

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.

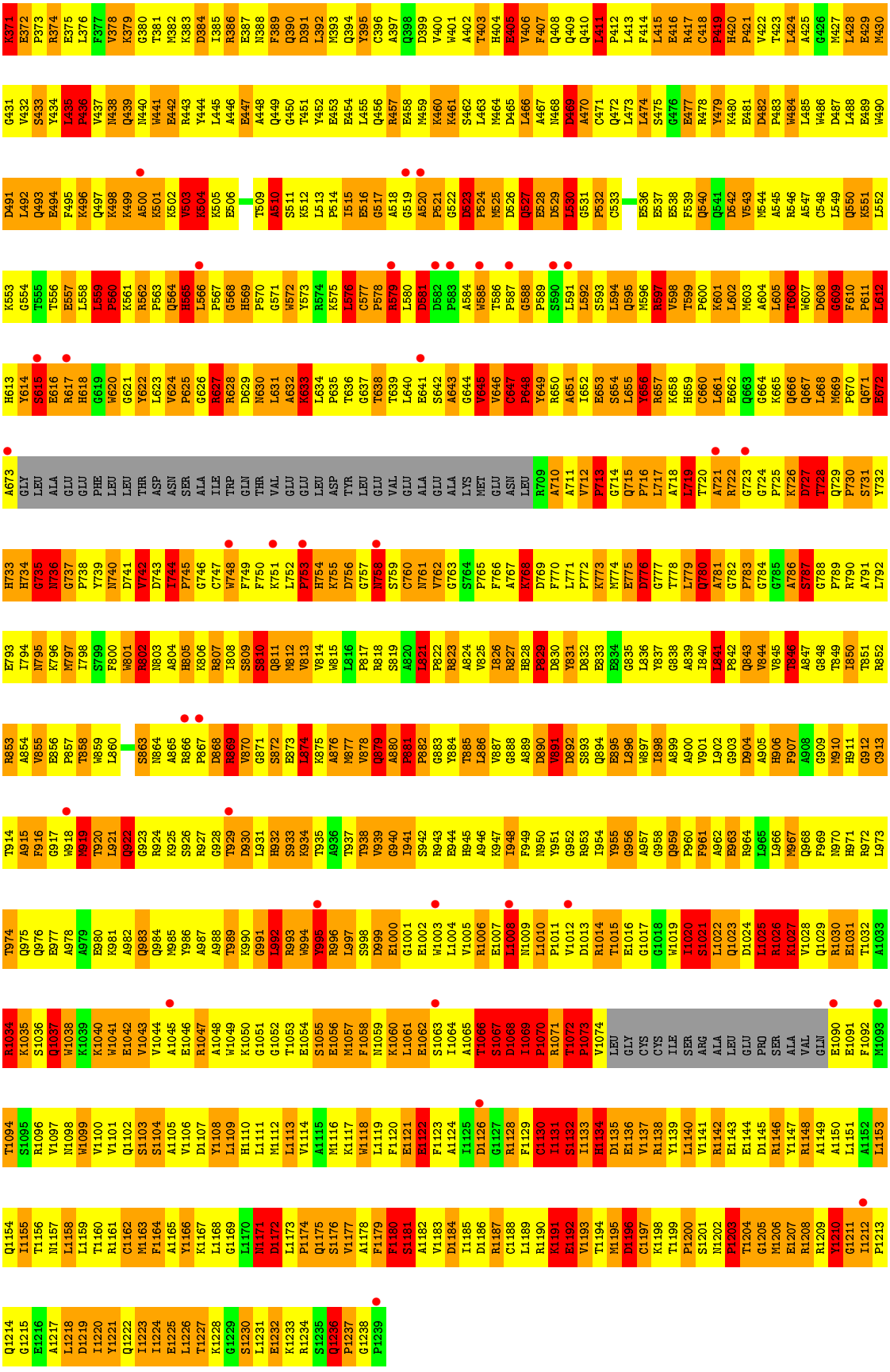
- Molecule 1: DNA polymerase subunit gamma-1

Chain A:



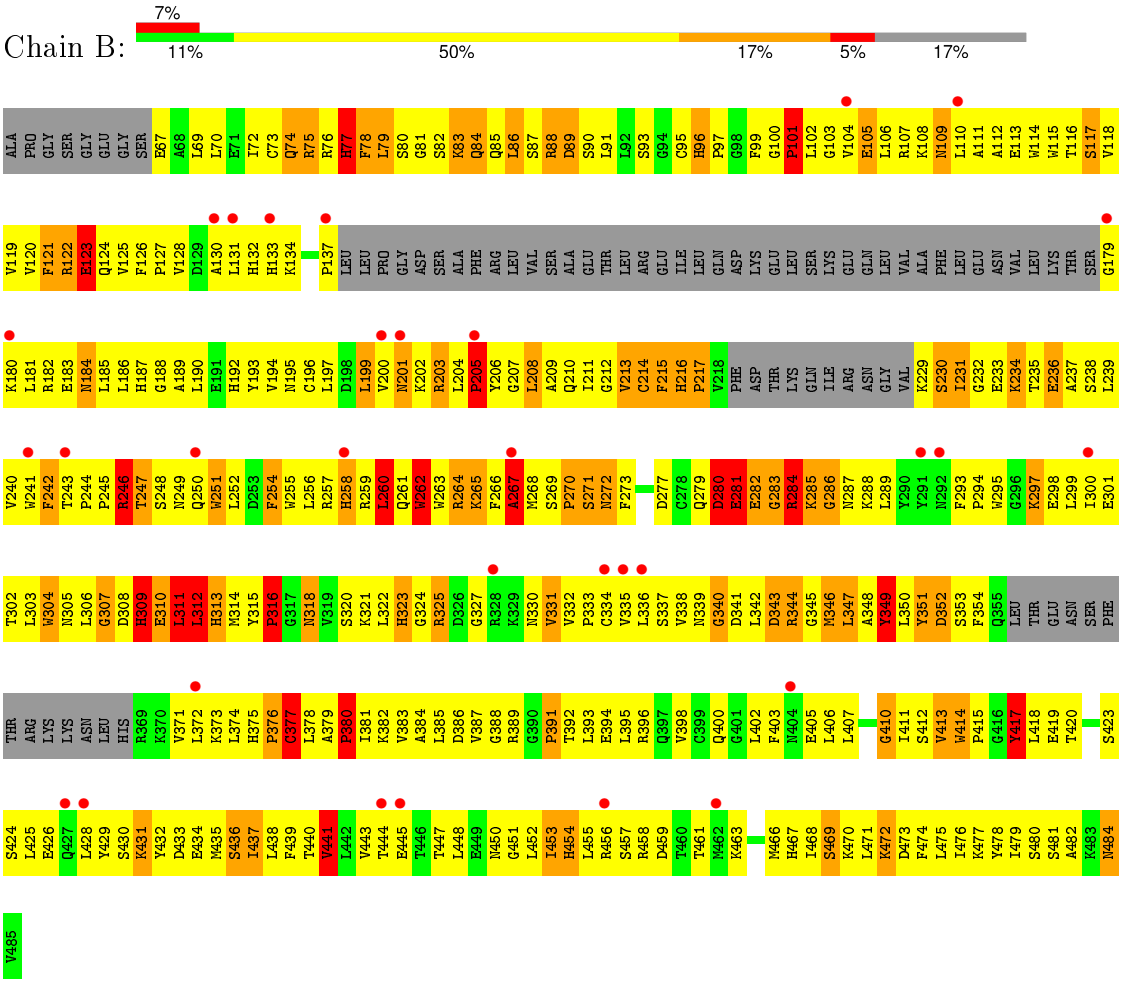


• Molecule 1: DNA polymerase subunit gamma-1



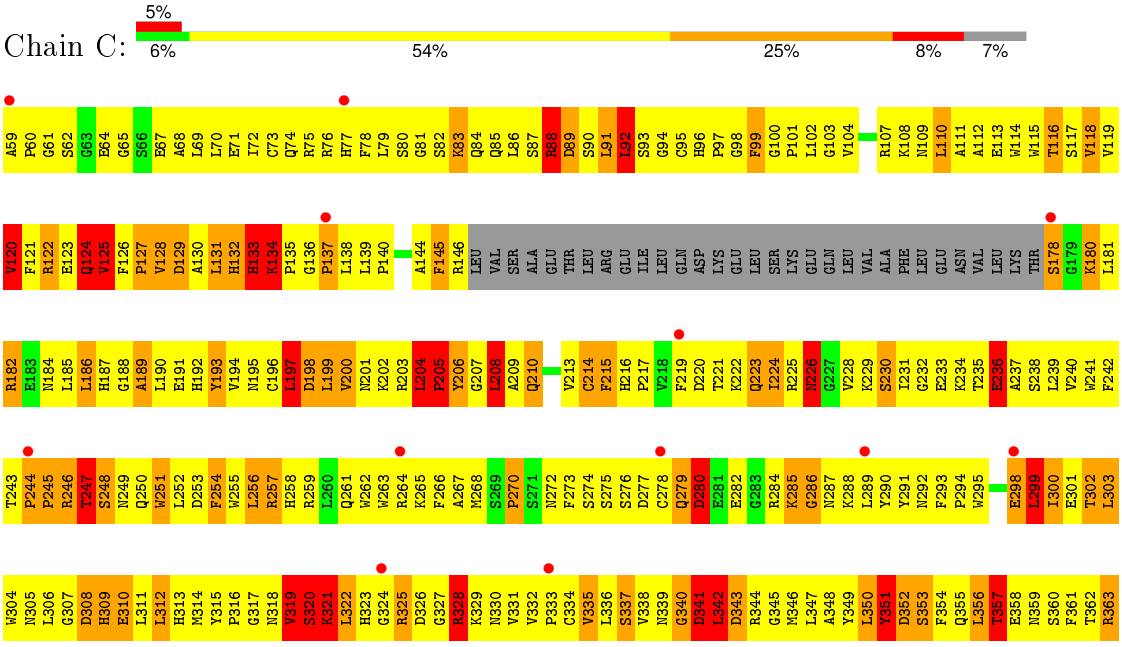
- Molecule 2: DNA polymerase subunit gamma-2

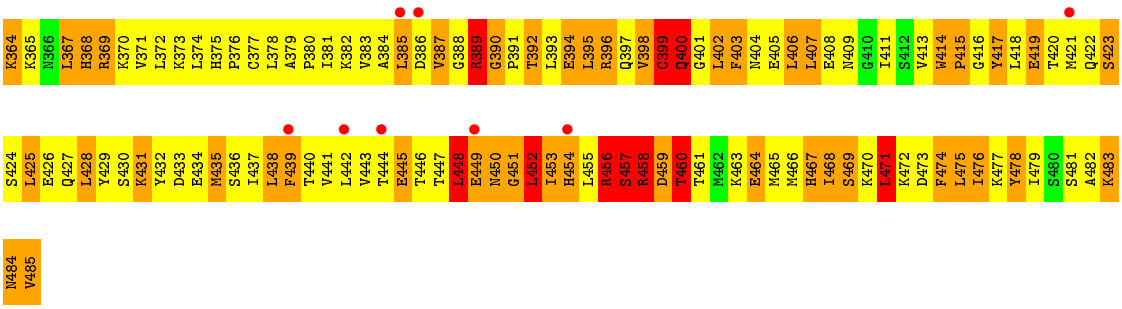
Chain B:



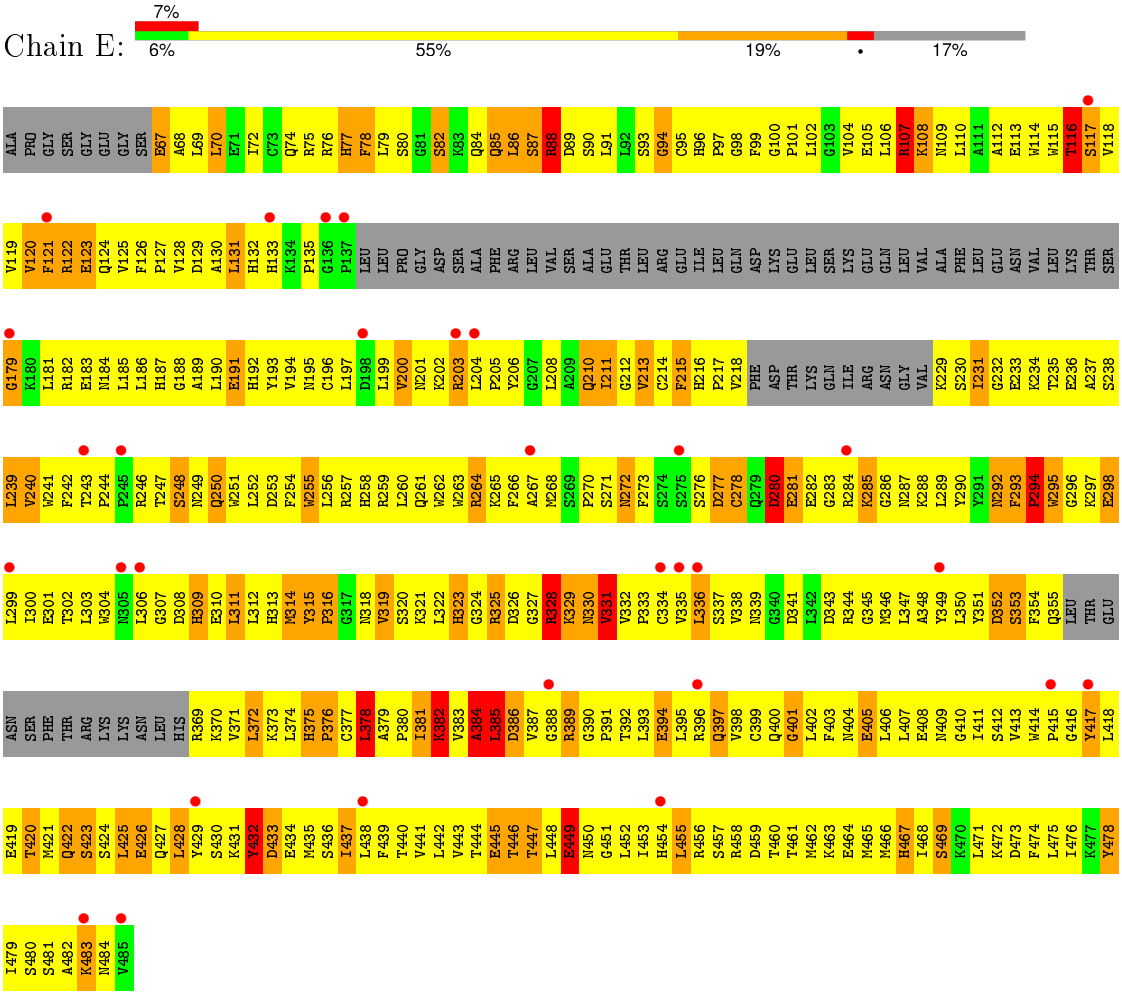
• Molecule 2: DNA polymerase subunit gamma-2

Chain C:

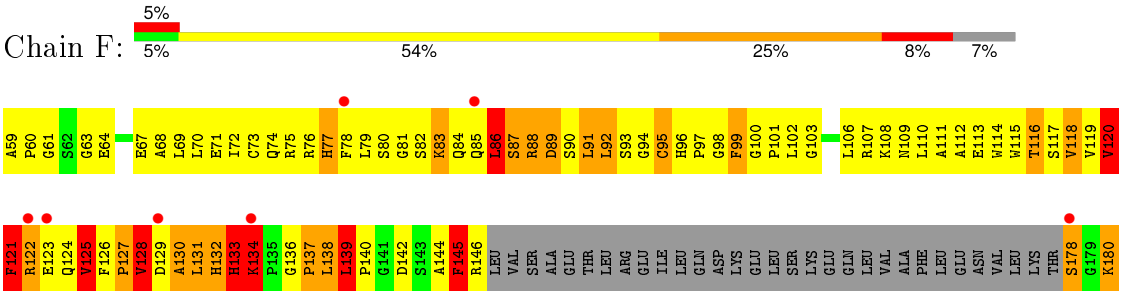


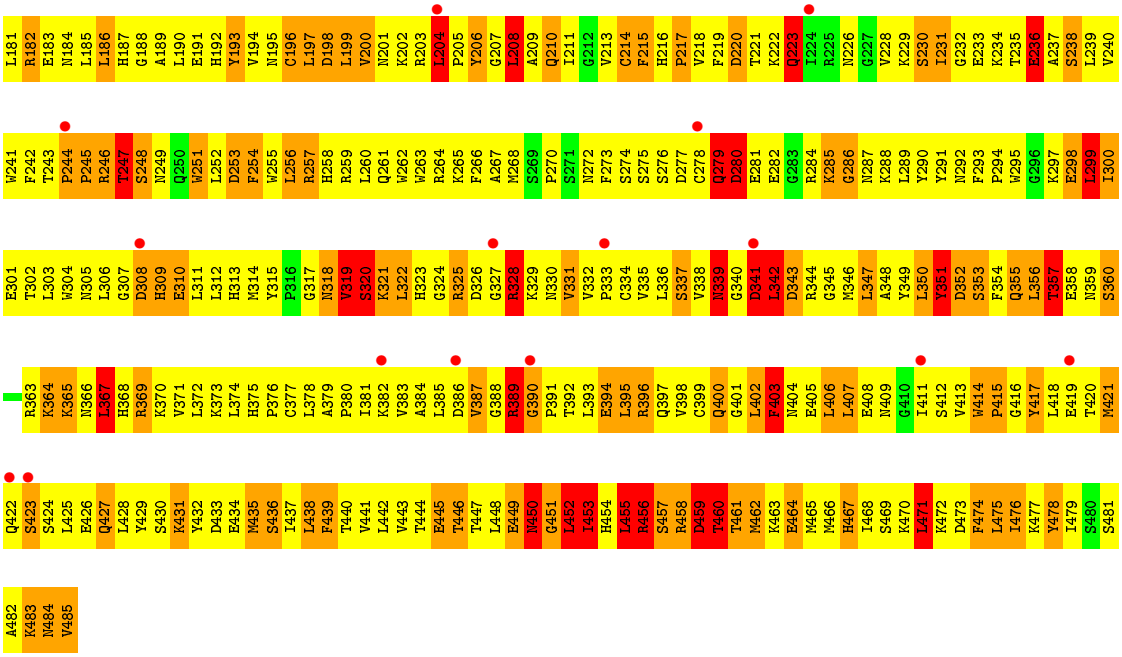


• Molecule 2: DNA polymerase subunit gamma-2



• Molecule 2: DNA polymerase subunit gamma-2




L181 R182 E183 N184 L185 L186 H187 G188 A189 L190 E191 H192 Y193 Y194 N195 G196 L197 D198 L199 V200 N201 K202 R203 L204 P205 Y206 G207 L208 A209 Q210 I211 G212 V213 C214 F215 H216 P217 V218 F219 D220 T221 K222 Q223 I224 R225 N226 G227 V228 K229 S230 I231 G232 E233 K234 T235 E236 A237 S238 L239 V240
W241 F242 T243 P244 P245 R246 T247 S248 N249 Q250 W251 L252 D253 F254 W255 L256 R257 H258 R259 L260 Q261 W262 W263 R264 K265 F266 A267 M268 S269 P270 S271 N272 F273 S274 S275 S276 D277 C278 Q279 D280 E281 E282 G283 R284 K285 G286 N287 K288 L289 Y290 Y291 N292 F293 P294 W295 G296 K297 E298 L299 I300
E301 T302 L303 W304 N305 L306 G307 D308 H309 E310 L311 L312 H313 M314 Y315 P316 G317 N318 V319 S320 K321 L322 H323 G324 R325 D326 G327 R328 K329 N330 V331 V332 P333 C334 V335 L336 S337 V338 N339 G340 D341 L342 D343 R344 G345 M346 L347 A348 Y349 L350 Y351 D352 S353 F354 Q355 L356 T357 E358 N359 S360
R363 K364 K365 N366 L367 H368 R369 K370 V371 L372 K373 L374 H375 P376 C377 L378 A379 P380 I381 K382 V383 A384 L385 D386 V387 G388 R389 G390 P391 T392 L393 E394 L395 R396 Q397 V398 C399 Q400 G401 L402 F403 N404 E405 L406 L407 E408 N409 G410 I411 S412 V413 W414 P415 G416 Y417 L418 E419 T420 M421
Q422 S423 S424 L425 E426 Q427 L428 Y429 S430 K431 Y432 D433 E434 M435 S436 I437 L438 F439 T440 V441 L442 V443 T444 E445 T446 T447 L448 E449 N450 G451 L452 I453 H454 L455 R456 S457 R458 D459 T460 T461 M462 K463 E464 M465 M466 H467 I468 S469 K470 L471 K472 D473 P474 L475 I476 K477 Y478 I479 S480 S481
A482 K483 N484 V485

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 32	Depositor
Cell constants a, b, c, α, β, γ	138.39Å 138.39Å 226.31Å 90.00° 90.00° 120.00°	Depositor
Resolution (Å)	46.92 – 3.24 46.92 – 3.25	Depositor EDS
% Data completeness (in resolution range)	89.5 (46.92-3.24) 89.5 (46.92-3.25)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.12 (at 3.25Å)	Xtriage
Refinement program	CNS 1.2	Depositor
R, R_{free}	0.284 , 0.303 0.362 , (Not available)	Depositor DCC
R_{free} test set	No test flags present.	DCC
Wilson B-factor (Å^2)	93.9	Xtriage
Anisotropy	0.065	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.28 , 64.7	EDS
Estimated twinning fraction	0.008 for -h,-k,l 0.459 for h,-h-k,-l 0.009 for -k,-h,-l	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.52$, $< L^2 > = 0.36$	Xtriage
Outliers	0 of 76508 reflections	Xtriage
F_o,F_c correlation	0.83	EDS
Total number of atoms	29480	wwPDB-VP
Average B, all atoms (Å^2)	85.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.12% 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

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	0.72	10/8913 (0.1%)	1.33	149/12100 (1.2%)
1	D	0.81	18/8927 (0.2%)	1.45	161/12118 (1.3%)
2	B	0.60	2/2944 (0.1%)	0.99	16/3981 (0.4%)
2	C	0.68	0/3262	1.31	47/4411 (1.1%)
2	E	0.67	6/2944 (0.2%)	1.11	23/3981 (0.6%)
2	F	0.64	0/3262	1.28	37/4411 (0.8%)
All	All	0.72	36/30252 (0.1%)	1.31	433/41002 (1.1%)

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
1	A	0	4
1	D	1	3
2	B	0	1
2	C	0	1
2	E	1	2
2	F	0	1
All	All	2	12

The worst 5 of 36 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	D	358	GLU	CG-CD	12.87	1.71	1.51
2	E	385	LEU	CA-CB	-11.49	1.27	1.53
1	D	352	SER	N-CA	9.45	1.65	1.46
1	A	232	ARG	N-CA	-8.85	1.28	1.46
2	E	382	LYS	N-CA	-8.69	1.28	1.46

The worst 5 of 433 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	E	382	LYS	N-CA-C	16.56	155.72	111.00
2	E	385	LEU	N-CA-C	16.34	155.11	111.00
1	D	248	GLU	CA-C-N	-14.80	84.64	117.20
2	B	280	ASP	C-N-CA	14.32	157.51	121.70
1	A	1212	ILE	C-N-CD	-12.99	92.03	120.60

All (2) chirality outliers are listed below:

Mol	Chain	Res	Type	Atom
1	D	249	VAL	CA
2	E	385	LEU	CA

5 of 12 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	1179	PHE	Sidechain
1	A	221	TYR	Sidechain
1	A	239	LEU	Mainchain
1	A	395	TYR	Sidechain
2	B	417	TYR	Sidechain

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	8681	0	8521	3592	8
1	D	8695	0	8540	3900	7
2	B	2871	0	2862	809	0
2	C	3181	0	3168	1419	1
2	E	2871	0	2862	945	0
2	F	3181	0	3168	1467	4
All	All	29480	0	29121	11814	10

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 202.

The worst 5 of 11814 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:352:SER:C	1:D:352:SER:CA	1.74	1.55
1:D:460:LYS:HE2	1:D:460:LYS:N	1.20	1.48
2:E:432:TYR:HB2	2:E:437:ILE:CD1	1.51	1.39
1:D:352:SER:CA	1:D:358:GLU:OE2	1.73	1.36
1:D:914:THR:O	1:D:918:TRP:HB3	1.24	1.35

The worst 5 of 10 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)
1:A:673:ALA:O	2:F:318:ASN:O[2_734]	1.91	0.29
1:A:673:ALA:C	2:F:318:ASN:O[2_734]	1.98	0.22
1:A:1034:ARG:NE	2:F:450:ASN:ND2[1_545]	2.04	0.16
1:A:994:TRP:NE1	1:D:522:GLY:N[1_545]	2.06	0.14
1:D:1232:GLU:OE2	2:F:223:GLN:OE1[2_844]	2.11	0.09

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	1085/1172 (93%)	629 (58%)	254 (23%)	202 (19%)	0	0
1	D	1086/1172 (93%)	645 (59%)	234 (22%)	207 (19%)	0	0
2	B	349/427 (82%)	205 (59%)	94 (27%)	50 (14%)	0	1
2	C	394/427 (92%)	254 (64%)	84 (21%)	56 (14%)	0	1
2	E	349/427 (82%)	219 (63%)	79 (23%)	51 (15%)	0	1
2	F	394/427 (92%)	251 (64%)	93 (24%)	50 (13%)	0	2
All	All	3657/4052 (90%)	2203 (60%)	838 (23%)	616 (17%)	0	1

5 of 616 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	72	ASN
1	A	76	ILE
1	A	96	GLY
1	A	163	PRO
1	A	169	TRP

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	918/988 (93%)	611 (67%)	307 (33%)	0 0
1	D	919/988 (93%)	584 (64%)	335 (36%)	0 0
2	B	317/380 (83%)	259 (82%)	58 (18%)	2 10
2	C	350/380 (92%)	260 (74%)	90 (26%)	0 2
2	E	317/380 (83%)	270 (85%)	47 (15%)	4 17
2	F	350/380 (92%)	245 (70%)	105 (30%)	0 1
All	All	3171/3496 (91%)	2229 (70%)	942 (30%)	0 1

5 of 942 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	369	ARG
1	D	308	GLN
2	F	231	ILE
2	C	407	LEU
1	D	145	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 119 such sidechains are listed below:

Mol	Chain	Res	Type
2	C	454	HIS
1	D	394	GLN
2	F	84	GLN
1	D	86	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	D	155	ASN

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

There are no ligands in this entry.

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	1093/1172 (93%)	0.23	61 (5%)	28	18	27, 80, 112, 124	0
1	D	1094/1172 (93%)	0.15	53 (4%)	34	23	27, 81, 113, 124	0
2	B	355/427 (83%)	0.44	31 (8%)	13	8	81, 111, 129, 141	0
2	C	396/427 (92%)	0.31	20 (5%)	32	21	42, 85, 113, 127	1 (0%)
2	E	355/427 (83%)	0.41	30 (8%)	13	9	64, 108, 129, 144	0
2	F	396/427 (92%)	0.20	22 (5%)	28	18	42, 87, 114, 128	0
All	All	3689/4052 (91%)	0.25	217 (5%)	26	16	27, 88, 120, 144	1 (0%)

The worst 5 of 217 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	B	267	ALA	7.6
1	D	583	PRO	7.0
2	F	327	GLY	6.9
1	D	277	HIS	6.2
1	A	563	PRO	6.2

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

There are no ligands in this entry.

6.5 Other polymers (i)

There are no such residues in this entry.