wwPDB X-ray Structure Validation Summary Report (i)

Sep 20, 2023 – 05:25 PM EDT

PDB ID	:	5EXR
Title	:	Crystal structure of human primosome
Authors	:	Tahirov, T.H.; Baranovskiy, A.G.; Babayeva, N.D.
Deposited on	:	2015-11-24
Resolution	:	3.60 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 types of validation reports are described at
http://www.wwpdb.org/validation/2017/FAQs#types.

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

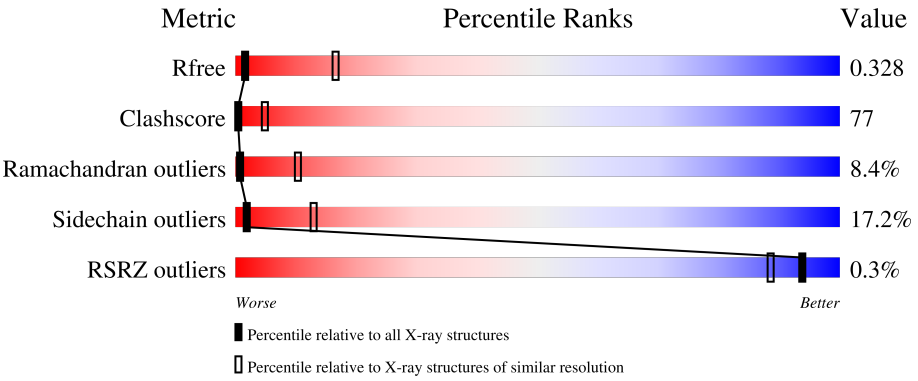
MolProbity	:	4.02b-467
Mogul	:	1.8.5 (274361), CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.35.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.35.1

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 3.60 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	1257 (3.70-3.50)
Clashscore	141614	1353 (3.70-3.50)
Ramachandran outliers	138981	1307 (3.70-3.50)
Sidechain outliers	138945	1307 (3.70-3.50)
RSRZ outliers	127900	1161 (3.70-3.50)

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	420	3% 21% 63% 9% 7%
1	E	420	23% 62% 8% 7%
2	B	509	20% 48% 16% • 15%
2	F	509	17% 50% 17% • 15%
3	C	1128	21% 54% 18% • 6%

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
3	G	1128	21% 53% 18% 6%
4	D	597	21% 38% 13% 26%
4	H	597	18% 41% 14% 26%

The following table lists non-polymeric compounds, carbohydrate monomers and non-standard residues in protein, DNA, RNA chains that are outliers for geometric or electron-density-fit criteria:

Mol	Type	Chain	Res	Chirality	Geometry	Clashes	Electron density
6	SF4	B	601	-	-	X	-

2 Entry composition (i)

There are 6 unique types of molecules in this entry. The entry contains 37658 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 primase small subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	389	Total	C	N	O	S	0	0	0
			3261	2099	564	583	15			
1	E	389	Total	C	N	O	S	0	0	0
			3261	2099	564	583	15			

- Molecule 2 is a protein called DNA primase large subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	B	434	Total	C	N	O	S	0	0	0
			3562	2280	616	653	13			
2	F	434	Total	C	N	O	S	0	0	0
			3562	2280	616	653	13			

- Molecule 3 is a protein called DNA polymerase alpha catalytic subunit.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
3	C	1057	Total	C	N	O	S	0	0	0
			8544	5477	1433	1578	56			
3	G	1057	Total	C	N	O	S	0	0	0
			8544	5477	1433	1578	56			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	516	ALA	VAL	engineered mutation	UNP P09884
G	516	ALA	VAL	engineered mutation	UNP P09884

- Molecule 4 is a protein called DNA polymerase alpha subunit B.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	D	444	Total	C	N	O	S	0	0	0
			3451	2194	576	666	15			

Continued on next page...

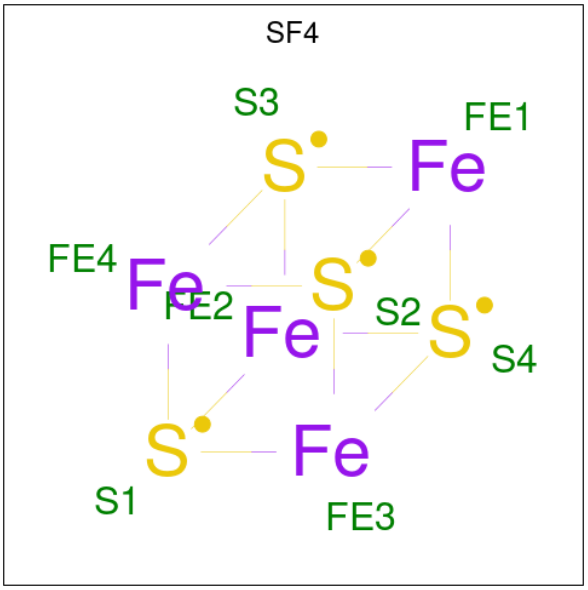
Continued from previous page...

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
4	H	444	Total	C	N	O	S	0	0	0
			3451	2194	576	666	15			

- Molecule 5 is ZINC ION (three-letter code: ZN) (formula: Zn).

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	1	Total	Zn	0	0
			1	1		
5	C	2	Total	Zn	0	0
			2	2		
5	E	1	Total	Zn	0	0
			1	1		
5	G	2	Total	Zn	0	0
			2	2		

- Molecule 6 is IRON/SULFUR CLUSTER (three-letter code: SF4) (formula: Fe_4S_4).



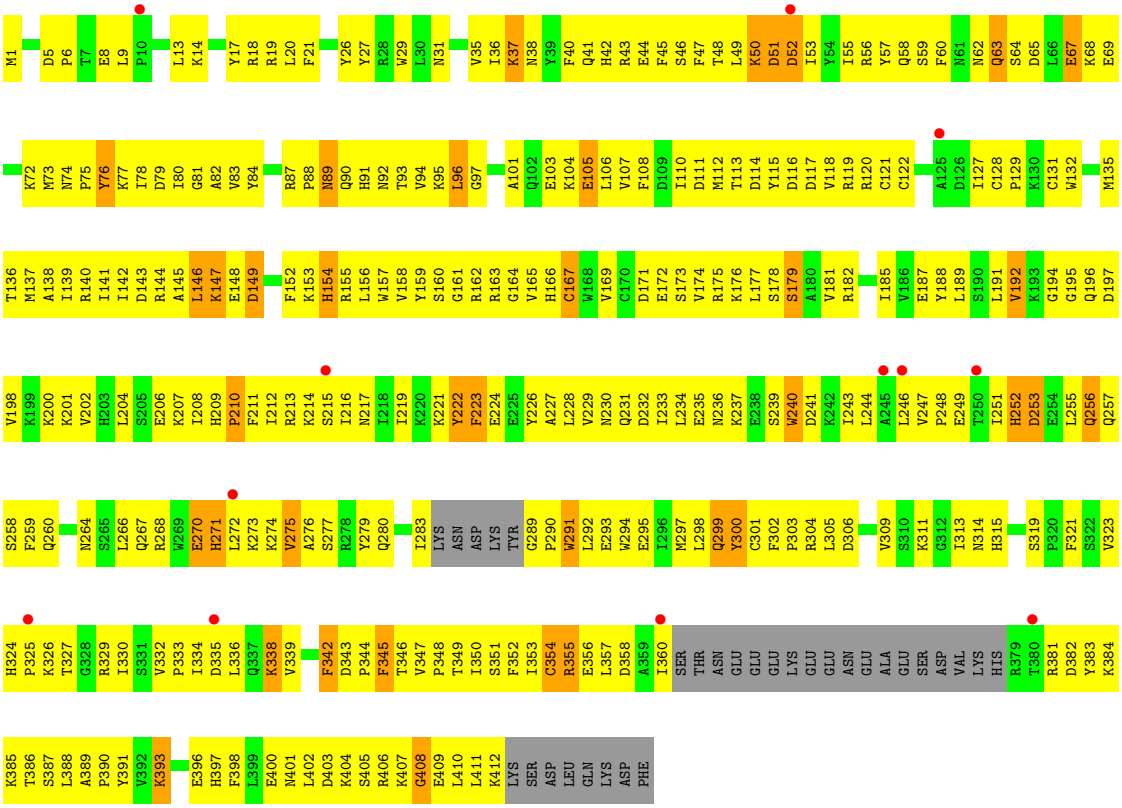
Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
6	B	1	Total	Fe	S	0	0
			8	4	4		
6	F	1	Total	Fe	S	0	0
			8	4	4		

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.

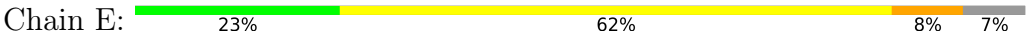
- Molecule 1: DNA primase small subunit

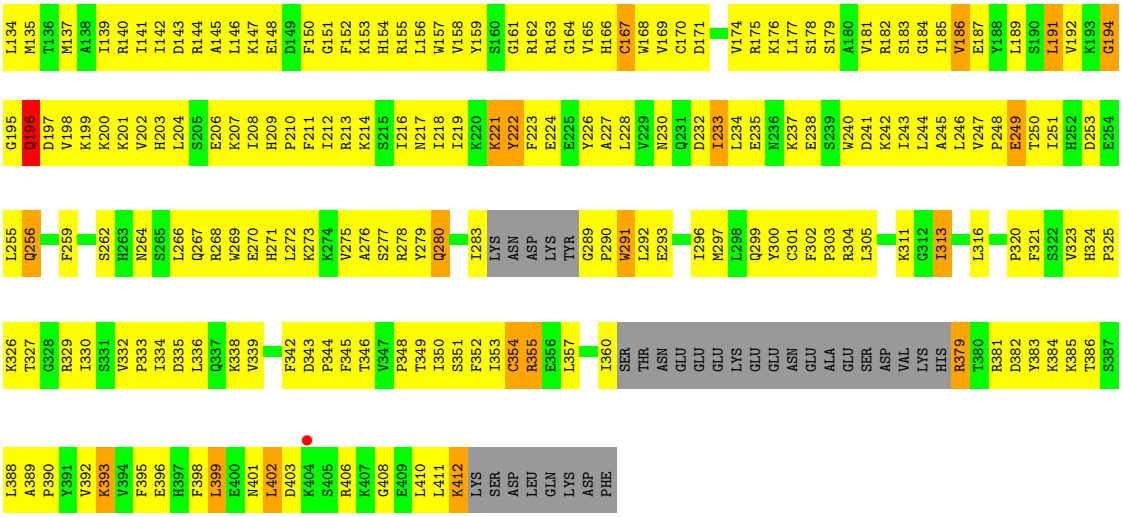
Chain A:



- Molecule 1: DNA primase small subunit

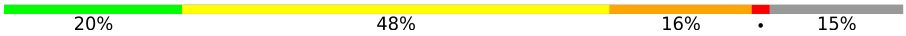
Chain E:

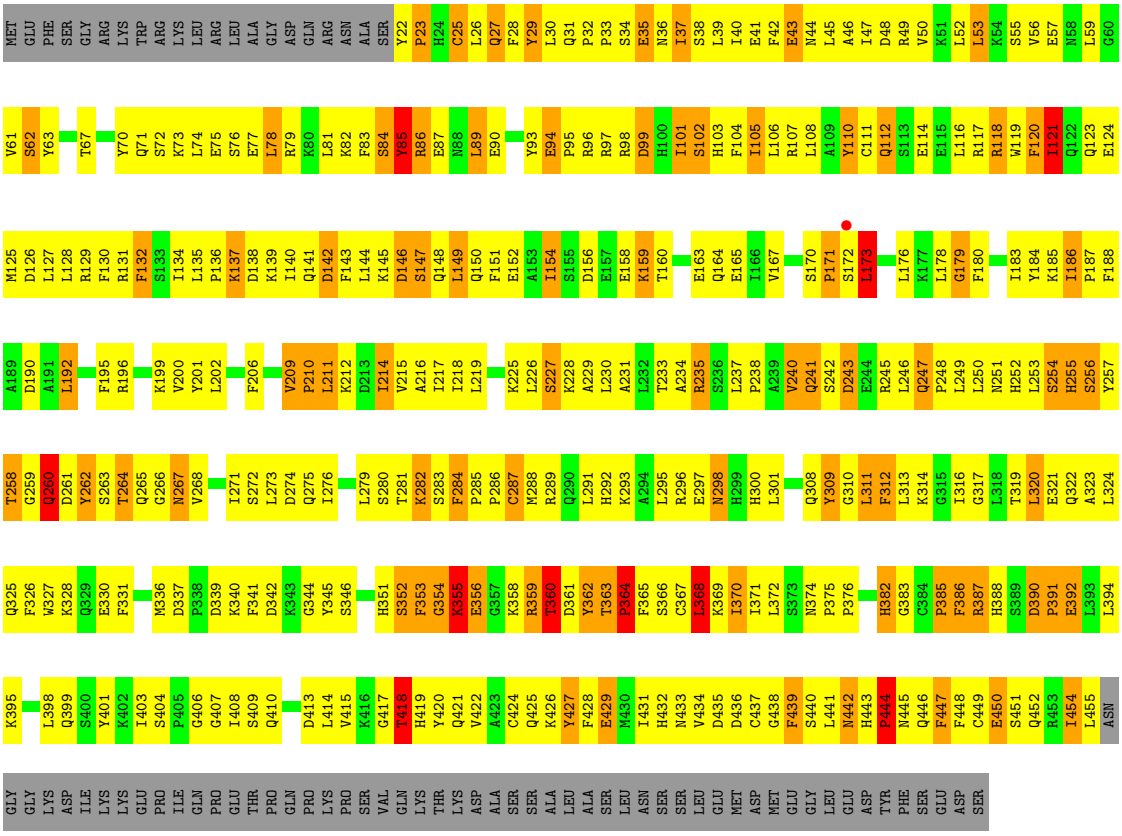




• Molecule 2: DNA primase large subunit

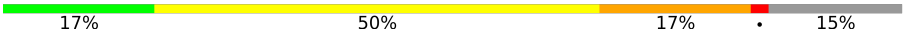
Chain B:

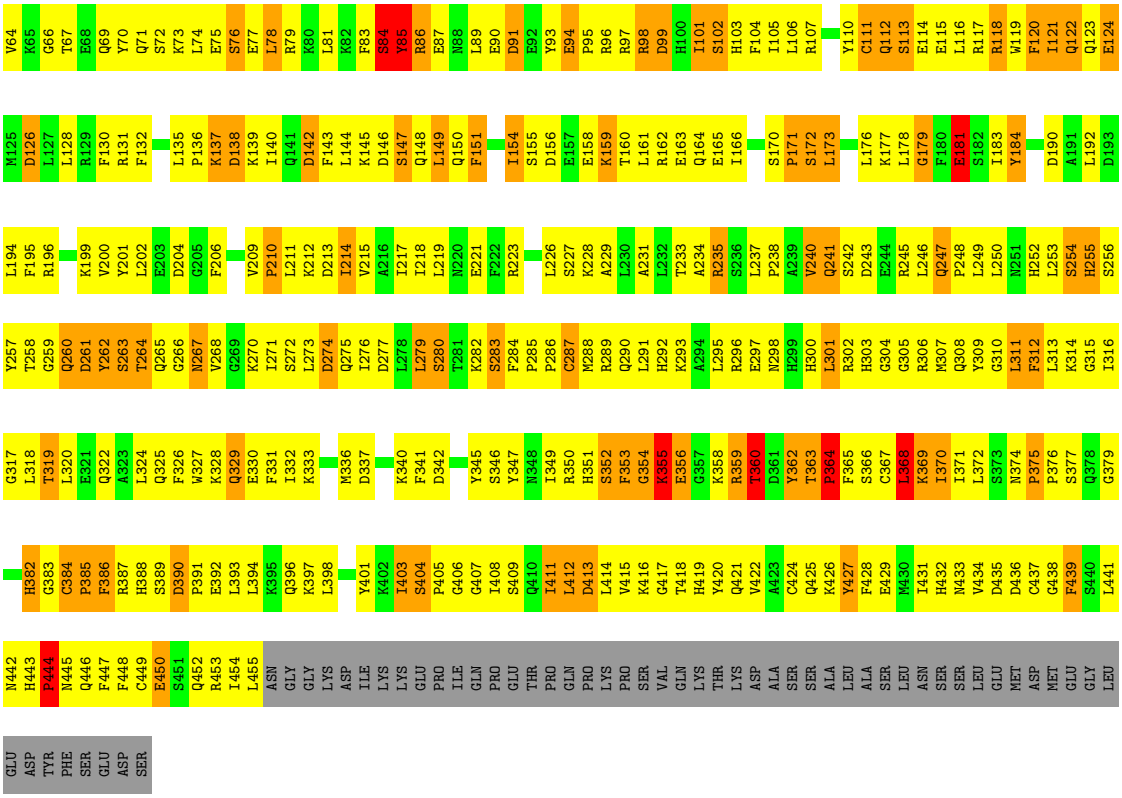




• Molecule 2: DNA primase large subunit

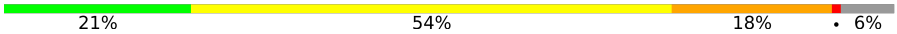
Chain F:

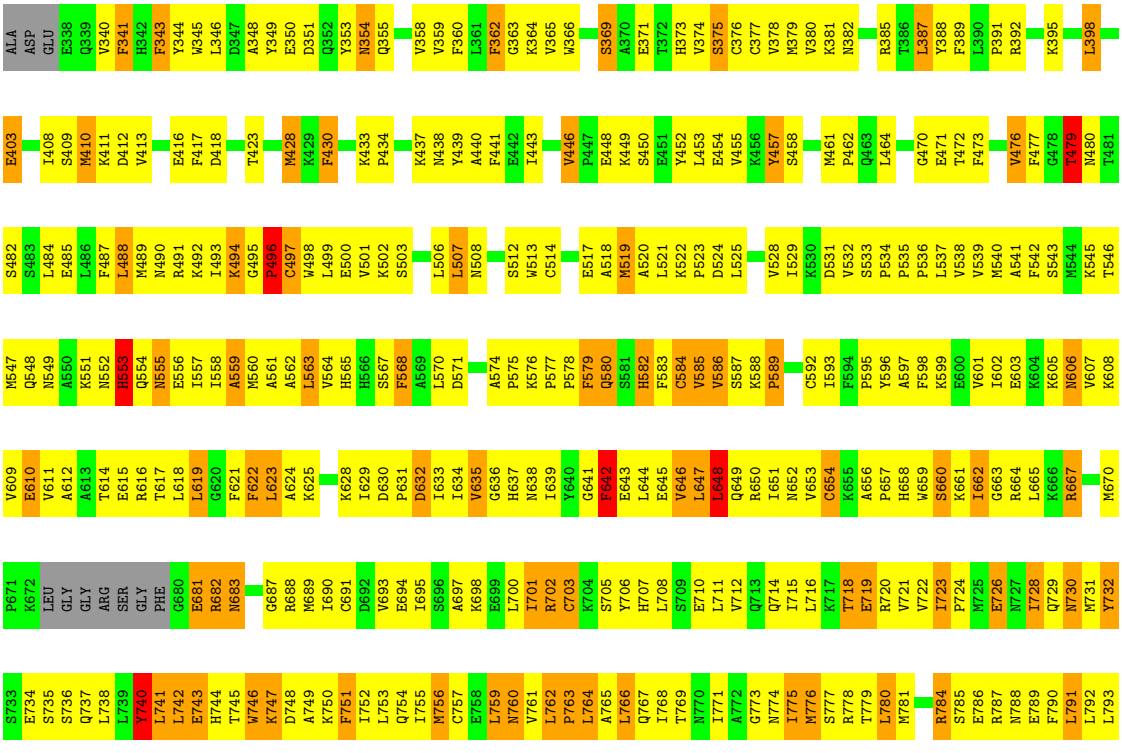


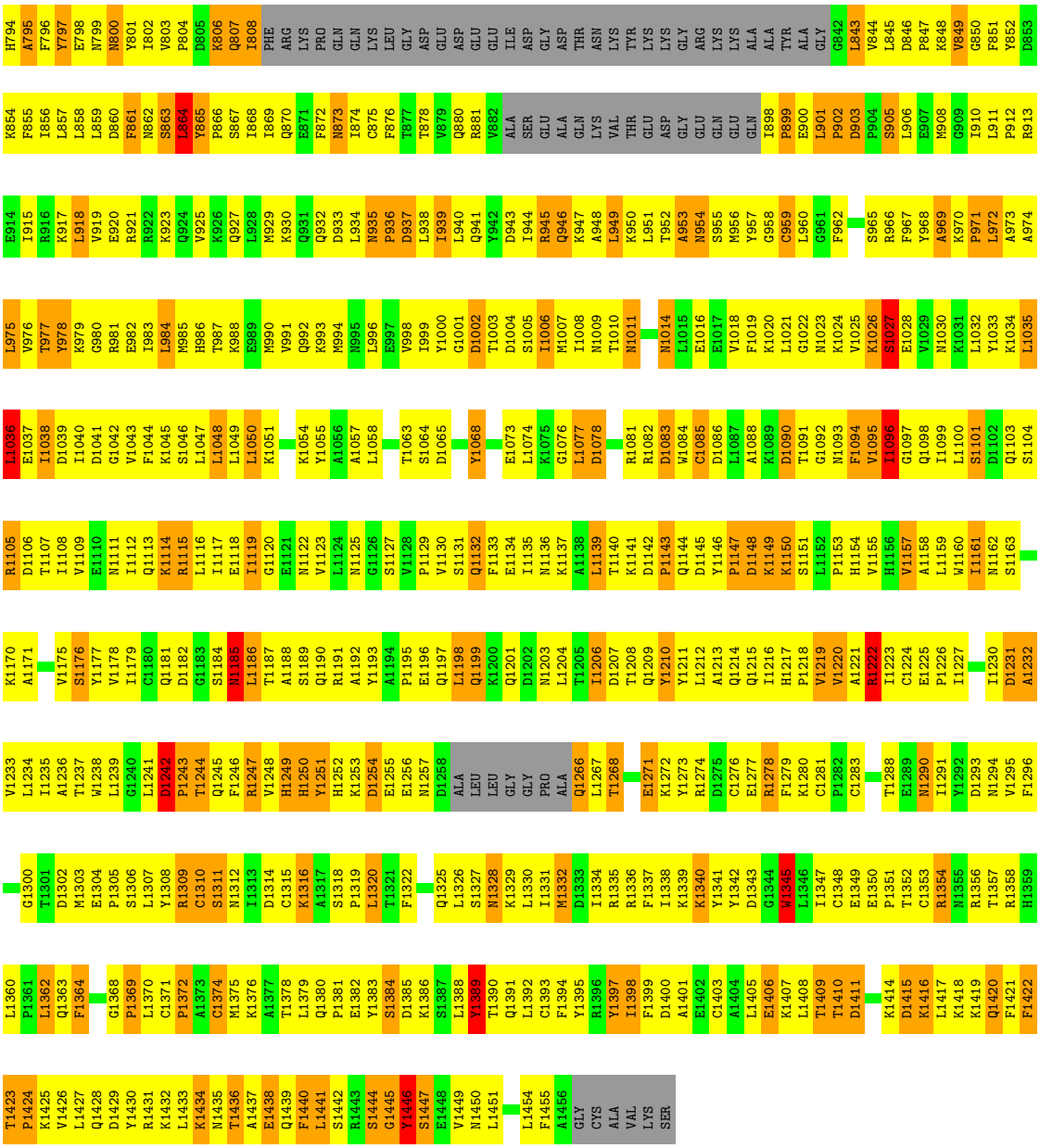


• Molecule 3: DNA polymerase alpha catalytic subunit

Chain C:

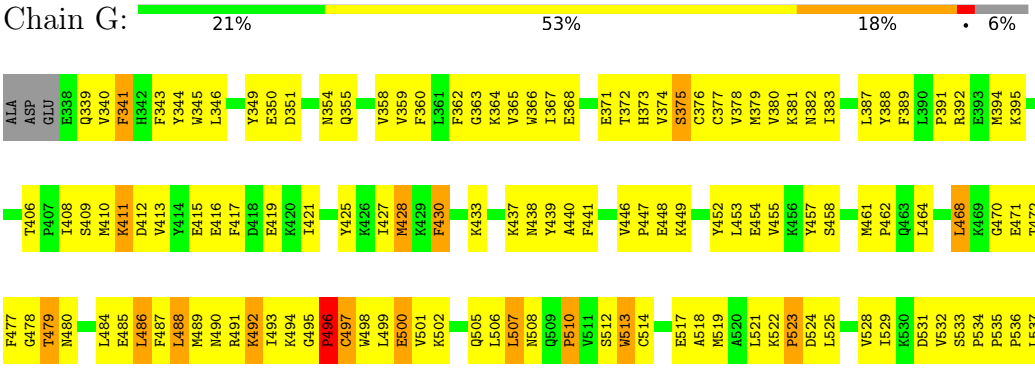


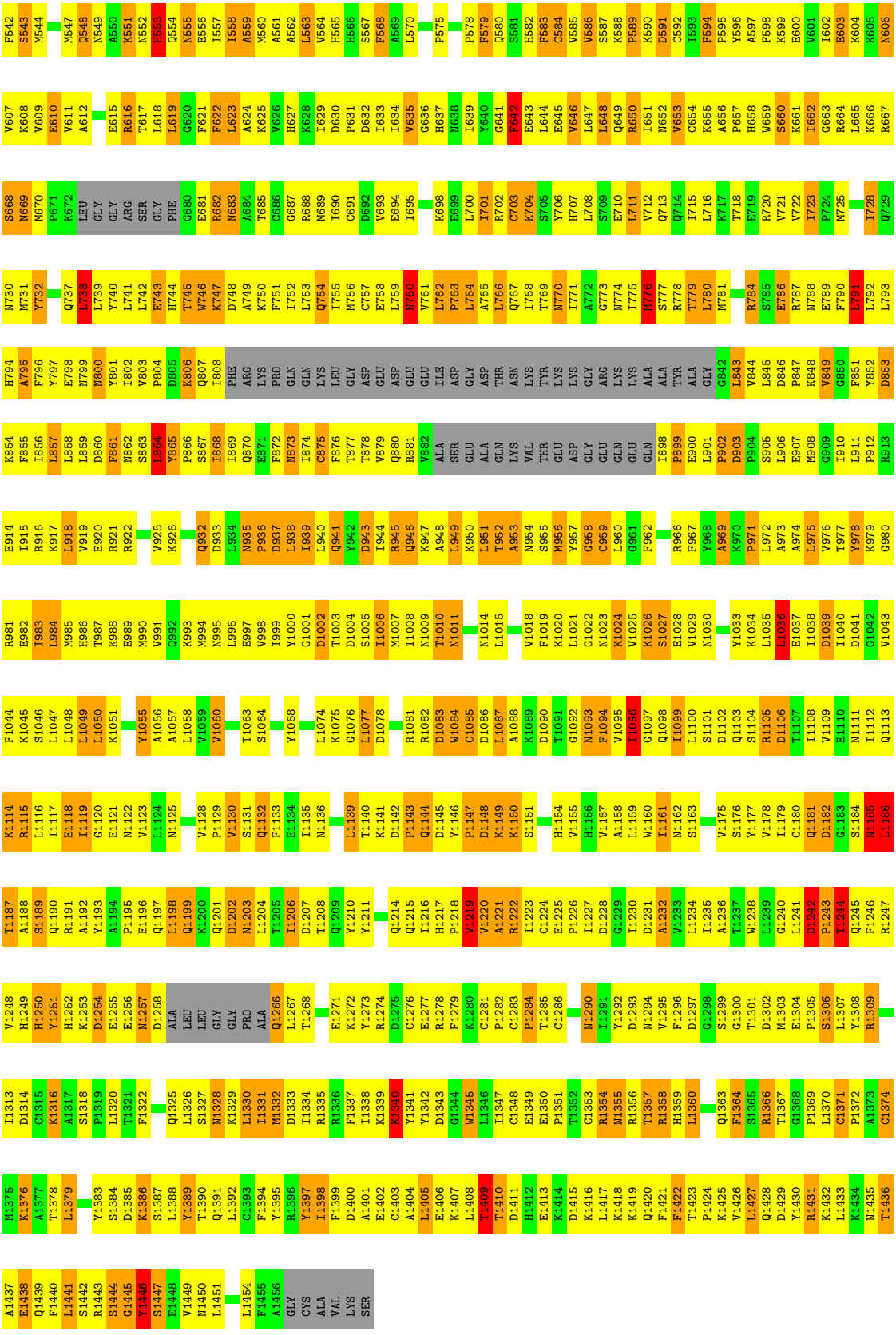




• Molecule 3: DNA polymerase alpha catalytic subunit

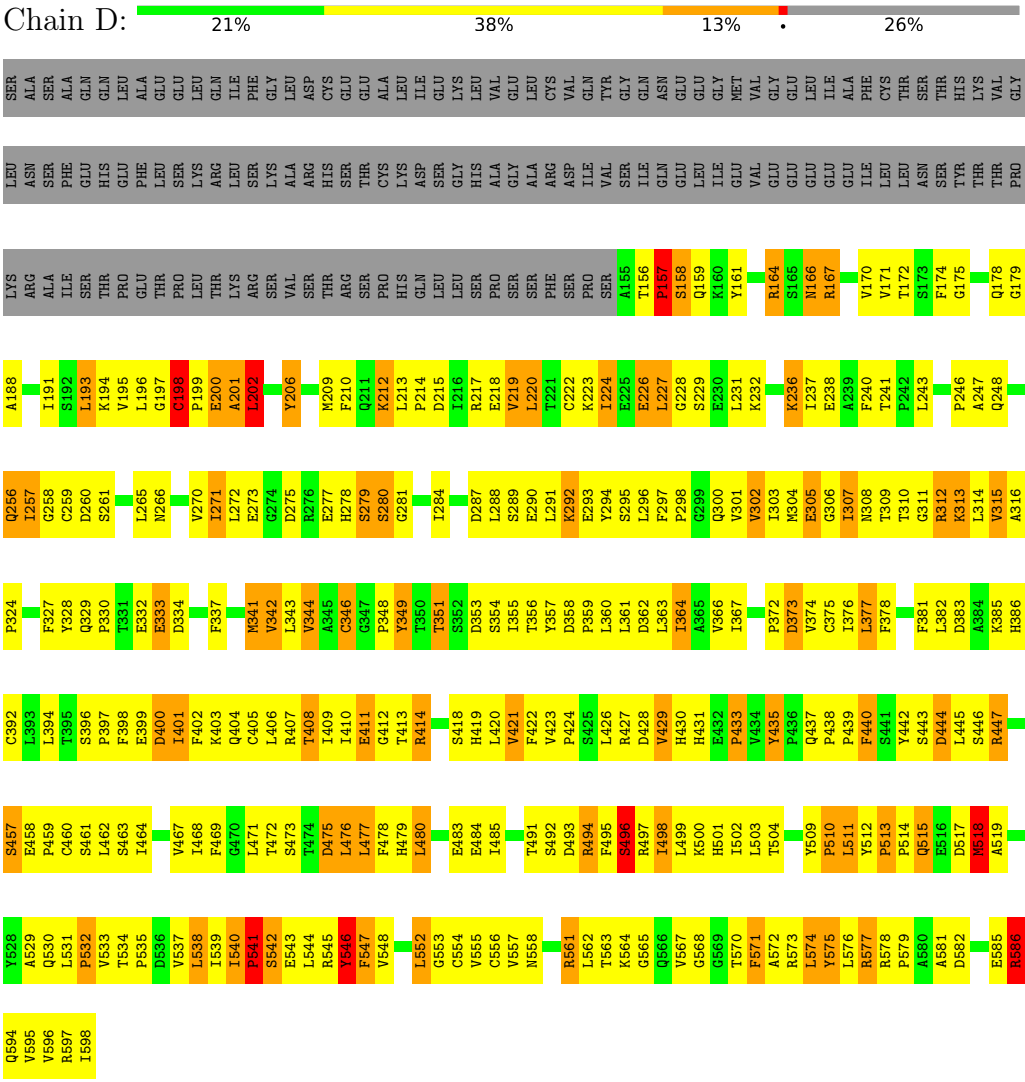
Chain G:





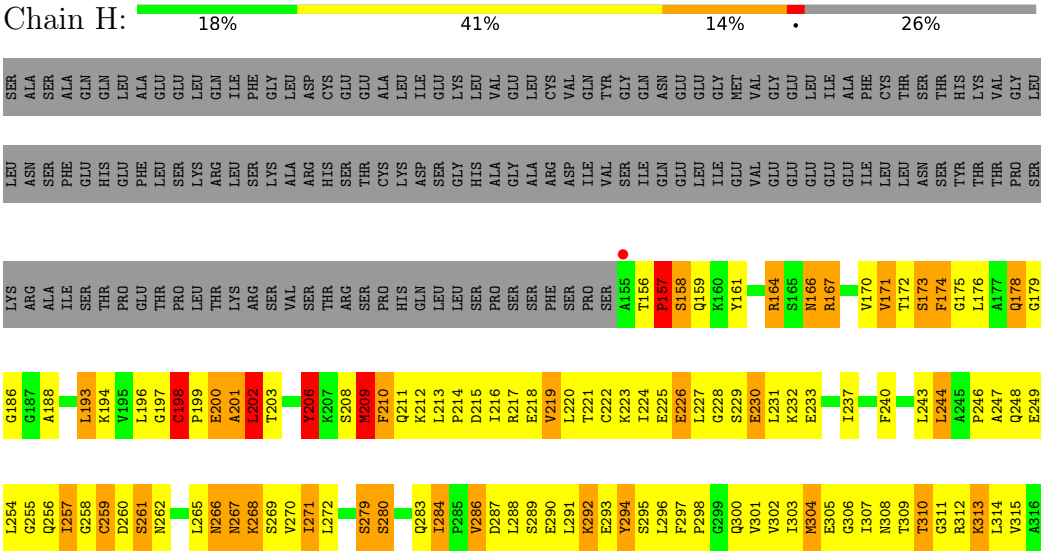
• Molecule 4: DNA polymerase alpha subunit B

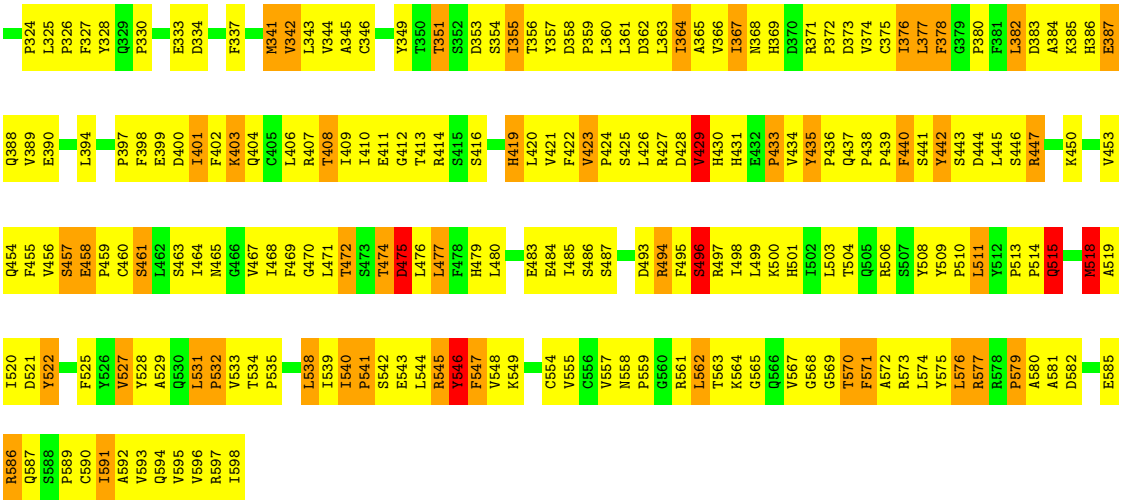
Chain D:



• Molecule 4: DNA polymerase alpha subunit B

Chain H:




P324 L325 P326 F327 Y328 Q329 P330 E333 D334 F337 M341 V342 L343 V344 A345 C346 Y349 T350 T351 S352 D353 S354 I355 T356 Y357 D358 P359 L360 L361 D362 L363 I364 A365 V366 I367 N368 H369 D370 R371 P372 D373 V374 C375 I376 L377 F378 G379 P380 F381 L382 D383 A384 K385 H386 E387
Q388 V389 E390 L394 P397 F398 E399 D400 I401 F402 K403 Q404 C405 L406 R407 T408 I409 I410 E411 G412 T413 R414 S415 S416 H419 L420 V421 F422 V423 P424 S425 L426 R427 D428 V429 H430 H431 E432 P433 V434 Y435 P436 Q437 P438 P439 F440 S441 Y442 S443 D444 L445 S446 R447 K450 V453
Q454 F455 V456 S457 E458 P459 C460 S461 L462 S463 I464 N465 G466 V467 I468 F469 G470 L471 T472 S473 T474 D475 L476 L477 F478 H479 L480 E483 E484 I485 S486 S487 D493 R494 F495 S496 R497 I498 L499 K500 H501 I502 L503 T504 Q505 R506 S507 Y508 Y509 P510 L511 Y512 P513 P514 Q515 M518 A519
I520 D521 Y522 F525 Y526 V527 Y528 A529 Q530 L531 P532 V533 T534 P535 L538 I539 I540 P541 S542 E543 L544 R545 Y546 F547 V548 K549 C554 V555 C556 V557 N558 P559 G560 R561 L562 T563 K564 G565 Q566 V567 G568 G569 T570 F571 A572 R573 L574 Y575 L576 R577 R578 P579 A580 A581 D582 E585
R586 Q587 S588 P589 C590 I591 A592 V593 Q594 V595 V596 R597 I598

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	113.10Å 210.16Å 172.56Å 90.00° 93.56° 90.00°	Depositor
Resolution (Å)	39.94 – 3.60 49.73 – 3.30	Depositor EDS
% Data completeness (in resolution range)	68.9 (39.94-3.60) 73.5 (49.73-3.30)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.36 (at 3.33Å)	Xtriage
Refinement program	CNS 1.1	Depositor
R, R_{free}	0.268 , 0.326 0.275 , 0.328	Depositor DCC
R_{free} test set	4621 reflections (5.01%)	wwPDB-VP
Wilson B-factor (Å^2)	78.1	Xtriage
Anisotropy	0.047	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , 69.5	EDS
L-test for twinning[2]	$< \|L\| > = 0.42$, $< L^2 > = 0.24$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.86	EDS
Total number of atoms	37658	wwPDB-VP
Average B, all atoms (Å^2)	61.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 44.59 % of the origin peak, indicating pseudo-translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo-translational symmetry is equal to 1.4935e-04. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

[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)

Bond lengths and bond angles in the following residue types are not validated in this section: ZN, SF4

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	0.51	0/3343	0.72	0/4508
1	E	0.48	0/3343	0.67	0/4508
2	B	0.57	0/3646	0.82	5/4908 (0.1%)
2	F	0.57	0/3646	0.80	3/4908 (0.1%)
3	C	0.58	0/8724	0.83	3/11788 (0.0%)
3	G	0.58	1/8724 (0.0%)	0.83	5/11788 (0.0%)
4	D	0.61	0/3529	0.86	2/4795 (0.0%)
4	H	0.60	0/3529	0.86	3/4795 (0.1%)
All	All	0.57	1/38484 (0.0%)	0.81	21/51998 (0.0%)

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
2	B	0	2
2	F	0	1
3	C	0	1
3	G	0	1
4	D	0	1
4	H	0	1
All	All	0	7

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	G	1371	CYS	CB-SG	-5.25	1.73	1.81

The worst 5 of 21 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	F	368	LEU	CA-CB-CG	-7.73	97.53	115.30
2	B	368	LEU	CA-CB-CG	-7.12	98.92	115.30
4	D	202	LEU	CA-CB-CG	6.87	131.10	115.30
4	H	202	LEU	CA-CB-CG	6.48	130.21	115.30
3	C	742	LEU	CA-CB-CG	-6.32	100.76	115.30

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	B	110	TYR	Sidechain
2	B	309	TYR	Sidechain
3	C	740	TYR	Sidechain
4	D	349	TYR	Sidechain
2	F	110	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	3261	0	3247	444	0
1	E	3261	0	3247	419	0
2	B	3562	0	3542	548	0
2	F	3562	0	3542	557	0
3	C	8544	0	8632	1431	0
3	G	8544	0	8634	1426	0
4	D	3451	0	3425	535	0
4	H	3451	0	3425	532	0
5	A	1	0	0	0	0
5	C	2	0	0	0	0
5	E	1	0	0	0	0
5	G	2	0	0	0	0
6	B	8	0	0	3	0
6	F	8	0	0	0	0
All	All	37658	0	37694	5765	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 77.

The worst 5 of 5765 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:G:858:LEU:HD13	3:G:1007:MET:HG3	1.23	1.21
2:B:209:VAL:HG12	2:B:210:PRO:HD2	1.25	1.19
3:C:730:ASN:H	3:C:730:ASN:ND2	1.28	1.19
4:D:476:LEU:HD11	4:D:502:ILE:HD11	1.22	1.14
1:E:20:LEU:HD21	1:E:357:LEU:HD22	1.16	1.13

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	383/420 (91%)	326 (85%)	45 (12%)	12 (3%)	4 32
1	E	383/420 (91%)	336 (88%)	36 (9%)	11 (3%)	4 33
2	B	432/509 (85%)	291 (67%)	93 (22%)	48 (11%)	0 6
2	F	432/509 (85%)	295 (68%)	84 (19%)	53 (12%)	0 5
3	C	1047/1128 (93%)	731 (70%)	226 (22%)	90 (9%)	1 10
3	G	1047/1128 (93%)	743 (71%)	209 (20%)	95 (9%)	1 9
4	D	442/597 (74%)	325 (74%)	79 (18%)	38 (9%)	1 10
4	H	442/597 (74%)	330 (75%)	73 (16%)	39 (9%)	1 9
All	All	4608/5308 (87%)	3377 (73%)	845 (18%)	386 (8%)	1 10

5 of 386 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	37	LYS
2	B	29	TYR
2	B	35	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	B	90	GLU
2	B	94	GLU

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	363/393 (92%)	322 (89%)	41 (11%)	6 30
1	E	363/393 (92%)	324 (89%)	39 (11%)	6 32
2	B	394/459 (86%)	326 (83%)	68 (17%)	2 12
2	F	394/459 (86%)	329 (84%)	65 (16%)	2 15
3	C	962/1013 (95%)	785 (82%)	177 (18%)	1 10
3	G	962/1013 (95%)	780 (81%)	182 (19%)	1 9
4	D	390/526 (74%)	314 (80%)	76 (20%)	1 9
4	H	390/526 (74%)	312 (80%)	78 (20%)	1 8
All	All	4218/4782 (88%)	3492 (83%)	726 (17%)	2 13

5 of 726 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	F	444	PRO
3	G	1083	ASP
3	G	430	PHE
2	F	427	TYR
3	G	732	TYR

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 180 such sidechains are listed below:

Mol	Chain	Res	Type
2	F	71	GLN
3	G	652	ASN
2	F	164	GLN

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
2	F	399	GLN
3	G	862	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

Of 8 ligands modelled in this entry, 6 are monoatomic - leaving 2 for Mogul analysis.

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	$\#\|Z\| > 2$
6	SF4	F	601	2	0,12,12	-	-	-		
6	SF4	B	601	2	0,12,12	-	-	-		

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
6	SF4	F	601	2	-	-	0/6/5/5
6	SF4	B	601	2	-	-	0/6/5/5

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.

1 monomer is involved in 3 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
6	B	601	SF4	3	0

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

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	389/420 (92%)	-0.09	12 (3%) 49 33	39, 101, 121, 137	0
1	E	389/420 (92%)	-0.40	1 (0%) 94 88	45, 98, 115, 136	0
2	B	434/509 (85%)	-0.72	1 (0%) 95 91	5, 61, 112, 135	0
2	F	434/509 (85%)	-0.66	0 100 100	4, 64, 115, 135	0
3	C	1057/1128 (93%)	-0.75	0 100 100	1, 51, 93, 123	0
3	G	1057/1128 (93%)	-0.73	0 100 100	2, 54, 96, 116	0
4	D	444/597 (74%)	-0.70	0 100 100	1, 44, 94, 111	0
4	H	444/597 (74%)	-0.68	1 (0%) 95 91	2, 47, 94, 121	0
All	All	4648/5308 (87%)	-0.64	15 (0%) 94 88	1, 60, 109, 137	0

The worst 5 of 15 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	125	ALA	3.6
1	A	360	ILE	2.9
1	A	272	LEU	2.9
1	A	215	SER	2.7
1	A	245	ALA	2.7

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, 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	B-factors(Å^2)	Q<0.9
5	ZN	A	501	1/1	0.91	0.04	123,123,123,123	0
5	ZN	E	501	1/1	0.94	0.10	91,91,91,91	0
5	ZN	G	1501	1/1	0.95	0.16	78,78,78,78	0
5	ZN	C	1502	1/1	0.97	0.15	26,26,26,26	0
6	SF4	F	601	8/8	0.97	0.18	1,1,8,18	0
6	SF4	B	601	8/8	0.98	0.18	1,1,2,9	0
5	ZN	C	1501	1/1	0.99	0.14	46,46,46,46	0
5	ZN	G	1502	1/1	0.99	0.12	1,1,1,1	0

6.5 Other polymers (i)

There are no such residues in this entry.