WORLDWIDE PDB PROTEIN DATA BANK

wwPDB X-ray Structure Validation Summary Report (i)

Aug 20, 2023 – 12:03 PM EDT

PDB ID	:	2ONO
Title	:	Arg475Gln Mutant of Mitochondrial Aldehyde Dehydrogenase, apo form, pseudo-merohedrally twinned
Authors	:	Larson, H.N.; Hurley, T.D.
Deposited on	:	2007-01-24
Resolution	:	2.15 Å(reported)

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

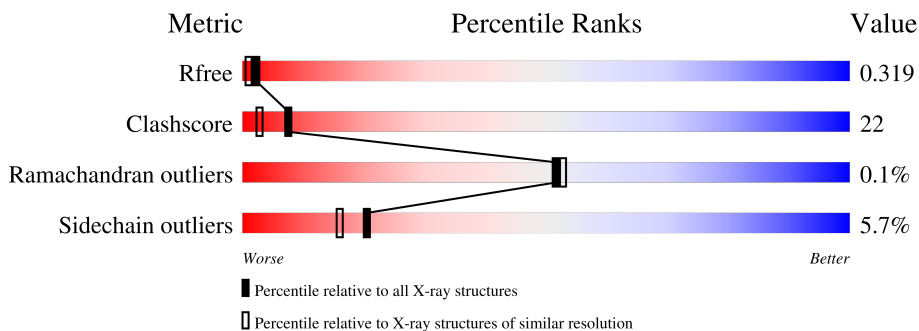|		
MolProbity	:	4.02b-467
Xtriage (Phenix)	:	1.13
EDS	:	2.35
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation 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 2.15 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		0.319
Clashscore		22
Ramachandran outliers		0.1%
Sidechain outliers		5.7%

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	1479 (2.16-2.16)
Clashscore	141614	1585 (2.16-2.16)
Ramachandran outliers	138981	1560 (2.16-2.16)
Sidechain outliers	138945	1559 (2.16-2.16)

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%

Mol	Chain	Length	Quality of chain
1	A	500	51% 43% ··
1	B	500	57% 39% ··
1	C	500	65% 32% ··
1	D	500	55% 41% ··
1	E	500	57% 39% ··
1	F	500	61% 35% ···
1	G	500	58% 38% ··

Continued on next page...

Continued from previous page...

Mol	Chain	Length	Quality of chain
1	H	500	58% 39%

2   Entry composition ⓘ

There are 2 unique types of molecules in this entry. The entry contains 31388 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 Aldehyde dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	494	Total	C	N	O	S	0	0	0
			3796	2414	646	718	18			
1	B	494	Total	C	N	O	S	0	0	0
			3796	2414	646	718	18			
1	C	494	Total	C	N	O	S	0	0	0
			3796	2414	646	718	18			
1	D	494	Total	C	N	O	S	0	0	0
			3796	2414	646	718	18			
1	E	494	Total	C	N	O	S	0	0	0
			3796	2414	646	718	18			
1	F	494	Total	C	N	O	S	0	0	0
			3796	2414	646	718	18			
1	G	494	Total	C	N	O	S	0	0	0
			3796	2414	646	718	18			
1	H	494	Total	C	N	O	S	0	0	0
			3796	2414	646	718	18			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	475	GLN	ARG	engineered mutation	UNP P05091
B	475	GLN	ARG	engineered mutation	UNP P05091
C	475	GLN	ARG	engineered mutation	UNP P05091
D	475	GLN	ARG	engineered mutation	UNP P05091
E	475	GLN	ARG	engineered mutation	UNP P05091
F	475	GLN	ARG	engineered mutation	UNP P05091
G	475	GLN	ARG	engineered mutation	UNP P05091
H	475	GLN	ARG	engineered mutation	UNP P05091

- Molecule 2 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
2	A	134	Total 134	O 134	0	0
2	B	131	Total 131	O 131	0	0
2	C	145	Total 145	O 145	0	0
2	D	133	Total 133	O 133	0	0
2	E	140	Total 140	O 140	0	0
2	F	114	Total 114	O 114	0	0
2	G	118	Total 118	O 118	0	0
2	H	105	Total 105	O 105	0	0

3 Residue-property plots ⓘ

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. 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. 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: Aldehyde dehydrogenase

Chain A:

- Molecule 1: Aldehyde dehydrogenase

Chain B:

Q300 C301 C302 S306 R307 T308 Q311 E312 D313 I314 Y315 D316 E317 F318 V319 E320 R321 S322 V323 A326 K327 S328 R329 V330 V331 P334 T339 E340 Q341 V345 D346 E347 T348 Q349 F350 K351 K352 I353 I357 N358 T359 G360 K361 Q362 E363 K366 L367 L368 C369 G370 D376

R377 G378 Q382 V385 V389 M393 T394 K397 E398 E399 I400 F401 V404 I407 L408 K409 F410 I413 V416 V417 G418 N422 Y425 G426 L427 V431 D435 K438 Y441 L442 S443 Q444 G449 T450 V451 W452 C455 F459 G460 A461 Q462 S463

P464 F465 Y468 K469 M470 S471 G472 S473 G474 Q475 E476 L477 G478 E479 Y480 G481 L482 Q483 A484 E487 V488 K489 T490 V493 K498 N499 S500

• Molecule 1: Aldehyde dehydrogenase

Chain C: 65% 32%

SER ALA ALA ALA THR GLN A7 P11 N12 Q13 Q14 P15 Q21 I22 F23 I24 N25 N26 E27 W28 H29 D30 A31 V32 S33 R34 K35 T36 F37 P38 T39 V40 N41 P42 S43 I48 C49 Q50 E53 V63 F70 M79 R84 L87 R90 R99 T100 A103

E106 T107 L108 D109 N110 Y114 Y118 L119 V120 D121 L122 D123 M124 V125 L126 K127 Y131 G134 W135 K138 K142 T143 I144 P145 I146 D147 G148 T154 G160 Q164 P171 L172 L173 M174 Q175 K178 T185 K192 Q196 T197 P198 A201 L202

A205 P214 P215 G216 V217 V221 P222 H235 K240 V241 S246 T247 E248 I249 G250 R251 Q254 A257 S260 N261 L262 K263 R264 V265 E268 L269 G270 G271 K272 S273 P274 D280 V287 L294 N297 Q300 C301 C302 S306 D313 I314 Y315 D316 E317

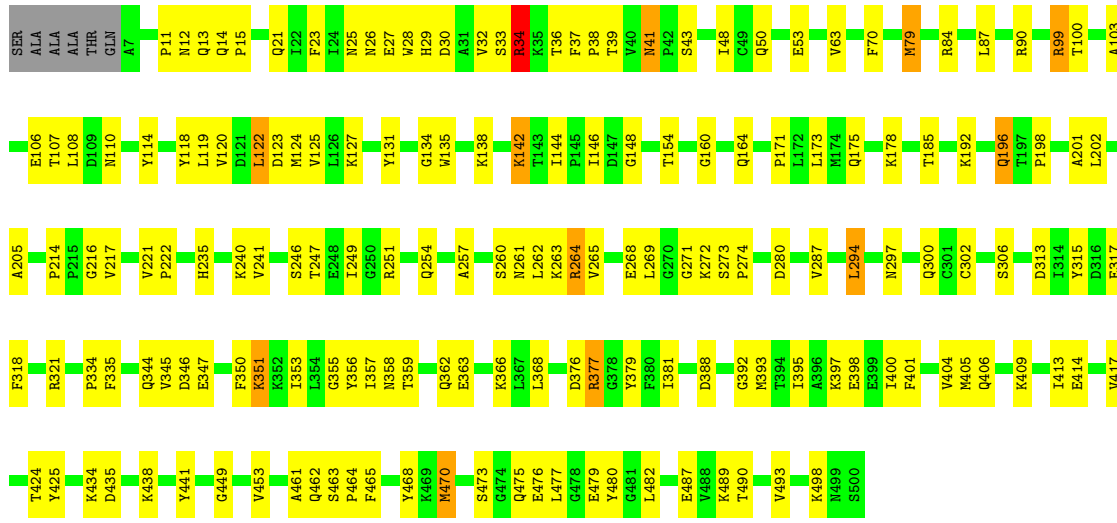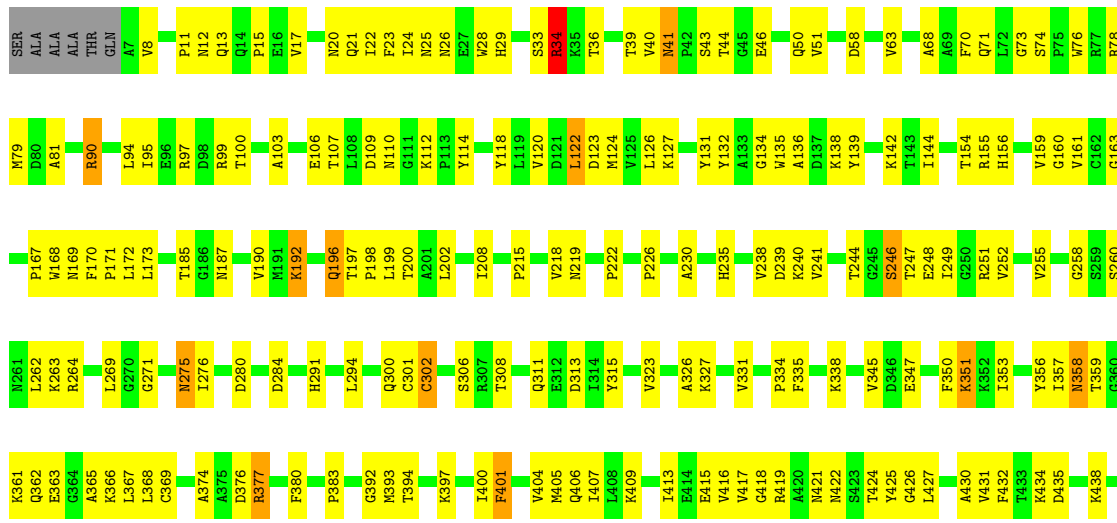F318 R321 P334 F335 Q344 V345 D346 E347 F350 K351 K352 I353 L354 G355 Y356 I357 N358 T359 Q362 E363 K366 L367 L368 D376 R377 G378 Y379 F380 I381 D388 G392 M393 T394 I395 A396 K397 E398 E399 I400 F401 V404 M405 Q406 K409 I413 E414 V417

T424 Y425 K434 D435 K438 Y441 G449 V453 A461 Q462 S463 P464 F465 Y468 K469 M470 S473 G474 Q475 E476 L477 G478 E479 Y480 G481 L482 E487 V488 K489 T490 V493 K498 N499 S500

• Molecule 1: Aldehyde dehydrogenase

Chain D: 55% 41%

SER ALA ALA ALA THR GLN A7 V8 P11 N12 Q13 Q14 P15 E16 V17 N20 Q21 I22 F23 I24 N25 N26 E27 W28 H29 S33 R34 K35 T36 T39 V40 N41 P42 S43 T44 G45 E46 Q50 V51 D58 V63 A68 A69 F70 Q71 L72 G73 S74 P75 W76 R77 R78

M79 D80 A81 R90 L94 I95 E96 R97 D98 R99 T100 A103 E106 T107 L108 D109 N110 G111 K112 P113 Y114 Y118 L119 V120 D121 L122 D123 M124 V125 L126 K127 Y131 Y132 A133 G134 W135 A136 D137 K138 Y139 K142 T143 I144 T154 R155 H156 V159 G160 V161 C162 G163

P167 W168 N169 F170 P171 L172 L173 T185 G186 N187 V190 M191 K192 Q196 T197 P198 L199 T200 A201 L202 I208 P215 V218 N219 P222 P226 A230 H235 V238 D239 K240 V241 T244 G245 S246 T247 E248 I249 G250 R251 V252 V255 G258 S259 S260

N261 L262 K263 R264 L269 G270 G271 N275 I276 D280 D284 H291 L294 Q300 C301 C302 S306 R307 T308 Q311 E312 D313 I314 Y315 V323 A326 K327 V331 P334 F335 K338 V345 D346 E347 F350 K351 K352 I353 Y356 I357 N358 T359 G360

K361 Q362 E363 G364 A365 K366 L367 L368 C369 A374 A375 D376 R377 F380 P383 G392 M393 T394 K397 I400 F401 V404 M405 Q406 I407 L408 K409 I413 E414 E415 V416 V417 G418 R419 A420 N421 N422 S423 T424 Y425 G426 L427 A430 V431 F432 T433 K434 D435 K438

L442 A445 G449 W452 V453 G460 A461 Q462 S463 P464 F465 G466 G467 Y468 K469 M470 S471 G472 S473 G474 Q475 E476 L477 Y480 G481 L482 Q483 A484 Y485 T486 K489 T490 V493 K498 N499 S500

• Molecule 1: Aldehyde dehydrogenase

Chain E:

57%   39%   • •

SER ALA ALA THR GLN A7 P11 N12 P15 F18 Q21 I22 F23 I24 N25 N26 E27 W28 H29 D30 A31 V32 S33 R34 F37 P38 N41 P42 S43 Q50 V51 K56 D60 K61 K64 R67 F70 Q71 W76 R77 R78 M79 D80 H83 R84

R90 R99 T100 Y101 L102 E106 T107 L108 D109 N110 G111 K112 P113 Y114 V115 L119 V120 D121 L122 D123 M124 K127 Y131 Y132 W135 K138 K142 P145 I146 T154 R155 H156 E157 P158 V159 G160 V161 C162 I166 F170 L173 M174 Q175 K178

N187 K192 Q196 T197 P198 A201 L202 Y203 V204 A205 I208 F213 V217 V218 P222 A233 S234 K240 V241 A242 F243 T244 G245 S246 T247 E248 I249 G250 R251 V252 I253 Q254 V255 A256 A257 S260 N261 L262 K263 R264 V265 T266 L267 E268 L269 G270 G271 K272 S273 P274

N275 A281 D282 M283 H291 L294 N297 Q300 C301 C302 C303 A304 G305 S306 R307 E312 E317 E320 R321 R325 V330 K338 P343 Q344 V345 D346 E347 T348 Q349 F350 K351 K352 I353 L354 G355 Y356 I357 N358 K361 Q362 E363 K366 L367 L368 I373

D376 R377 G378 Y379 F380 I381 Q390 M393 T394 K397 E398 E399 I400 F401 G402 P403 V404 M405 Q406 F410 K411 V416 V417 G418 N422 S423 T424 Y425 G426 L427 A428 A429 A430 V431 F432 D435 L436 D437 K438 A439 N440 Y441 L442 S443 Q444 G449 W452 V453 N454 C455

Q462 S463 P464 F465 G466 G467 Y468 K469 M470 S471 G472 S473 G474 Q475 E476 L477 Y480 G481 L482 Y485 T486 E487 V488 K489 T490 K494 K498 N499 S500

• Molecule 1: Aldehyde dehydrogenase

Chain F:

61%   35%   • • •

SER ALA ALA THR GLN A7 V8 P11 N12 P15 Q21 I22 F23 I24 N25 N26 E27 W28 H29 S33 R34 K35 T36 T39 V40 N41 P42 S43 T44 G45 E46 V47 I48 C49 Q50 D55 V59 D60 V63 K64 L87 R90 L94 I95 E96 R97 L102

E106 T107 L108 D109 N110 Y114 Y118 L119 V120 D121 L122 D123 M124 V125 R130 Y131 Y132 A133 G134 W135 K138 G141 K142 I146 F150 R155 V159 G160 V161 Q164 I165 I166 P167 W168 N169 F170 L173 M174 Q175 K178 N187 V190 M191 K192

Q196 L199 L202 Y203 V204 A205 I208 K209 F213 P214 P215 V218 P222 H235 K240 V241 A242 F243 T244 G245 S246 T247 E248 I249 G250 R251 A257 G258 L262 K263 R264 V265 T266 L267 E268 L269 G270 G271 I276 D280 E288 H291 F292 A293 L294 F295

F296 N297 Q298 G299 Q300 C301 C302 S306 Q311 E312 D313 I314 Y315 V319 E320 V323 K327 S328 R329 G332 N333 P334 K338 Q344 E347 T348 Q349 F350 K351 K352 I353 I357 N358 T359 G360 K361 Q362 E363 K366 L367 L368 C369 I373 R377 F380

D388 M393 T394 I395 A396 K397 E398 E399 I400 F401 G402 P403 V404 M405 Q406 K409 I413 E414 E415 V416 V417 G418 R419 N422 S423 T424 Y425 G426 L427 K434 D435 K438 G460 A461 Q462 S463 P464 F465 K469 M470 S473 G474 Q475 E476 L477 Y480 T490

V495 P496 Q497 K498 N499 S500

• Molecule 1: Aldehyde dehydrogenase

Chain G:

• Molecule 1: Aldehyde dehydrogenase

Chain H:

4 Data and refinement statistics ⓘ

Property	Value	Source
Space group	P 1 21 1	Depositor
Cell constants a, b, c, α, β, γ	101.47Å 175.85Å 101.46Å 90.00° 94.79° 90.00°	Depositor
Resolution (Å)	10.00 – 2.15 24.89 – 2.15	Depositor EDS
% Data completeness (in resolution range)	92.5 (10.00-2.15) 96.5 (24.89-2.15)	Depositor EDS
R_{merge}	0.06	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	3.93 (at 2.15Å)	Xtriage
Refinement program	SHELXL-97	Depositor
R, R_{free}	0.251 , 0.314 0.264 , 0.319	Depositor DCC
R_{free} test set	9468 reflections (5.06%)	wwPDB-VP
Wilson B-factor (Å^2)	18.8	Xtriage
Anisotropy	0.340	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 24.9	EDS
L-test for twinning[2]	$< |L| > = 0.39$, $< L^2 > = 0.21$	Xtriage
Estimated twinning fraction	0.467 for l,-k,h	Xtriage
F_o,F_c correlation	0.93	EDS
Total number of atoms	31388	wwPDB-VP
Average B, all atoms (Å^2)	22.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.45% 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 ⓘ

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	0.38	3/3880 (0.1%)	0.87	10/5265 (0.2%)
1	B	0.28	0/3880	0.67	7/5265 (0.1%)
1	C	0.36	2/3880 (0.1%)	0.95	10/5265 (0.2%)
1	D	0.29	0/3880	0.67	6/5265 (0.1%)
1	E	0.36	2/3880 (0.1%)	1.01	11/5265 (0.2%)
1	F	0.30	0/3880	0.68	8/5265 (0.2%)
1	G	0.33	2/3880 (0.1%)	0.81	8/5265 (0.2%)
1	H	0.31	1/3880 (0.0%)	0.68	7/5265 (0.1%)
All	All	0.33	10/31040 (0.0%)	0.80	67/42120 (0.2%)

The worst 5 of 10 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	E	248	GLU	CD-OE1	10.60	1.37	1.25
1	G	361	LYS	CD-CE	9.20	1.74	1.51
1	A	115	VAL	CB-CG1	9.17	1.72	1.52
1	C	268	GLU	CG-CD	8.86	1.65	1.51
1	C	268	GLU	CD-OE2	8.37	1.34	1.25

The worst 5 of 67 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	E	90	ARG	CD-NE-CZ	37.79	176.51	123.60
1	E	99	ARG	NE-CZ-NH1	27.66	134.13	120.30
1	C	34	ARG	NE-CZ-NH1	27.45	134.03	120.30
1	A	90	ARG	CD-NE-CZ	26.80	161.12	123.60
1	E	99	ARG	NE-CZ-NH2	-25.64	107.48	120.30

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	3796	0	3740	221	0
1	B	3796	0	3740	171	0
1	C	3796	0	3740	140	0
1	D	3796	0	3740	182	0
1	E	3796	0	3740	191	0
1	F	3796	0	3740	171	0
1	G	3796	0	3740	182	0
1	H	3796	0	3740	173	0
2	A	134	0	0	27	0
2	B	131	0	0	9	0
2	C	145	0	0	10	0
2	D	133	0	0	11	0
2	E	140	0	0	13	0
2	F	114	0	0	11	0
2	G	118	0	0	18	0
2	H	105	0	0	18	0
All	All	31388	0	29920	1330	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 1330 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:249:ILE:CD1	1:D:249:ILE:CG1	1.83	1.52
1:E:424:THR:HG22	1:E:470:MET:HB2	1.20	1.11
1:A:262:LEU:HD21	1:B:251:ARG:HG2	1.41	1.02
1:E:424:THR:HG21	1:E:470:MET:SD	2.02	0.99
1:H:247:THR:HA	1:H:269:LEU:HD22	1.45	0.99

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	492/500 (98%)	475 (96%)	16 (3%)	1 (0%)	47	46
1	B	492/500 (98%)	476 (97%)	16 (3%)	0	100	100
1	C	492/500 (98%)	471 (96%)	21 (4%)	0	100	100
1	D	492/500 (98%)	473 (96%)	18 (4%)	1 (0%)	47	46
1	E	492/500 (98%)	472 (96%)	19 (4%)	1 (0%)	47	46
1	F	492/500 (98%)	476 (97%)	16 (3%)	0	100	100
1	G	492/500 (98%)	472 (96%)	20 (4%)	0	100	100
1	H	492/500 (98%)	477 (97%)	14 (3%)	1 (0%)	47	46
All	All	3936/4000 (98%)	3792 (96%)	140 (4%)	4 (0%)	51	53

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	E	270	GLY
1	H	426	GLY
1	A	426	GLY
1	D	426	GLY

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	399/402 (99%)	371 (93%)	28 (7%)	15	10

Continued on next page...

Continued from previous page...

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
1	B	399/402 (99%)	373 (94%)	26 (6%)	17	12
1	C	399/402 (99%)	378 (95%)	21 (5%)	22	19
1	D	399/402 (99%)	381 (96%)	18 (4%)	27	24
1	E	399/402 (99%)	377 (94%)	22 (6%)	21	17
1	F	399/402 (99%)	376 (94%)	23 (6%)	20	15
1	G	399/402 (99%)	378 (95%)	21 (5%)	22	19
1	H	399/402 (99%)	377 (94%)	22 (6%)	21	17
All	All	3192/3216 (99%)	3011 (94%)	181 (6%)	20	16

5 of 181 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	F	26	ASN
1	G	195	GLU
1	F	121	ASP
1	F	347	GLU
1	G	259	SER

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 76 such sidechains are listed below:

Mol	Chain	Res	Type
1	F	300	GLN
1	H	83	HIS
1	G	26	ASN
1	G	300	GLN
1	H	462	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 ⓘ

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 ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.2 Non-standard residues in protein, DNA, RNA chains ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.3 Carbohydrates ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.4 Ligands ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.

6.5 Other polymers ⓘ

Unable to reproduce the depositors R factor - this section is therefore empty.