wwPDB X-ray Structure Validation Summary Report (i)

Jun 22, 2024 – 06:38 PM EDT

PDB ID	:	6HYC
Title	:	The structure of full-length human phenylalanine hydroxylase in complex with the cofactor and negative regulator tetrahydrobiopterin
Authors	:	Alcorlo Pages, M.; Flydal, I.M.
Deposited on	:	2018-10-19
Resolution	:	3.18 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:

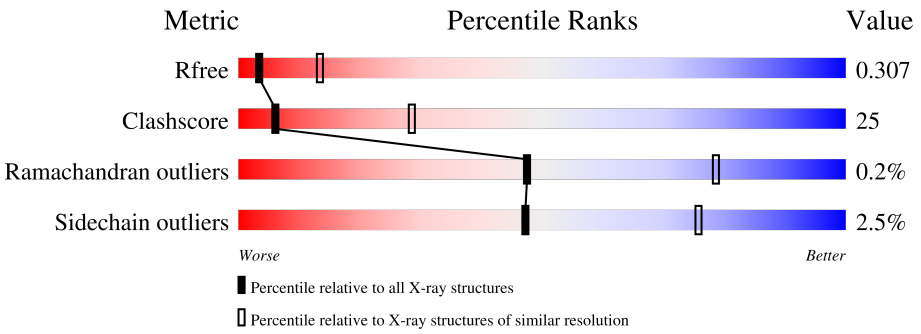
MolProbity	:	4.02b-467
Mogul	:	2022.3.0, CSD as543be (2022)
Xtriage (Phenix)	:	1.20.1
EDS	:	2.37.1
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.37.1

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 3.18 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	1467 (3.20-3.16)
Clashscore	141614	1599 (3.20-3.16)
Ramachandran outliers	138981	1574 (3.20-3.16)
Sidechain outliers	138945	1573 (3.20-3.16)

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	452	56% 36% • 6%
1	B	452	54% 39% • 6%
1	C	452	50% 42% • 6%
1	D	452	49% 42% • 6%

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
2	H4B	D	501	-	-	X	-

2 Entry composition (i)

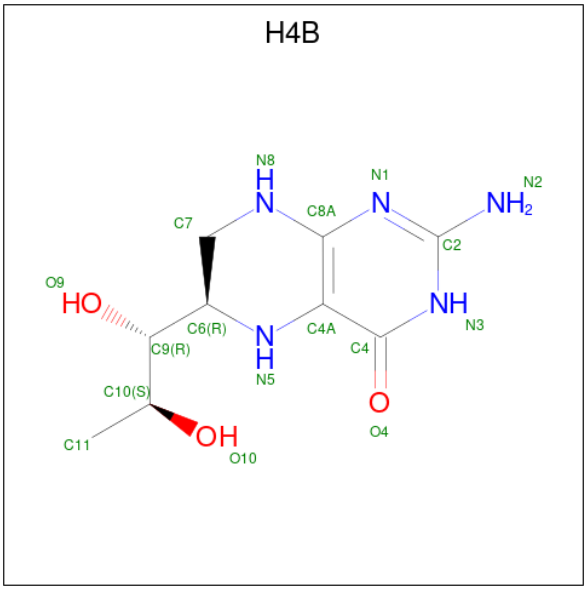
There are 3 unique types of molecules in this entry. The entry contains 13879 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Phenylalanine-4-hydroxylase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	D	423	3477	2237	587	642	11	0	4	0
1	A	423	3445	2213	583	638	11	0	0	0
1	C	423	3446	2214	583	638	11	0	0	0
1	B	424	3455	2219	584	641	11	0	0	0

- Molecule 2 is 5,6,7,8-TETRAHYDROBIOPTERIN (three-letter code: H4B) (formula: $C_9H_{15}N_5O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	D	1	17	9	5	3	0	0
2	C	1	17	9	5	3	0	0

Continued on next page...

Continued from previous page...

Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
2	B	1	17	9	5	3	0	0

- Molecule 3 is water.

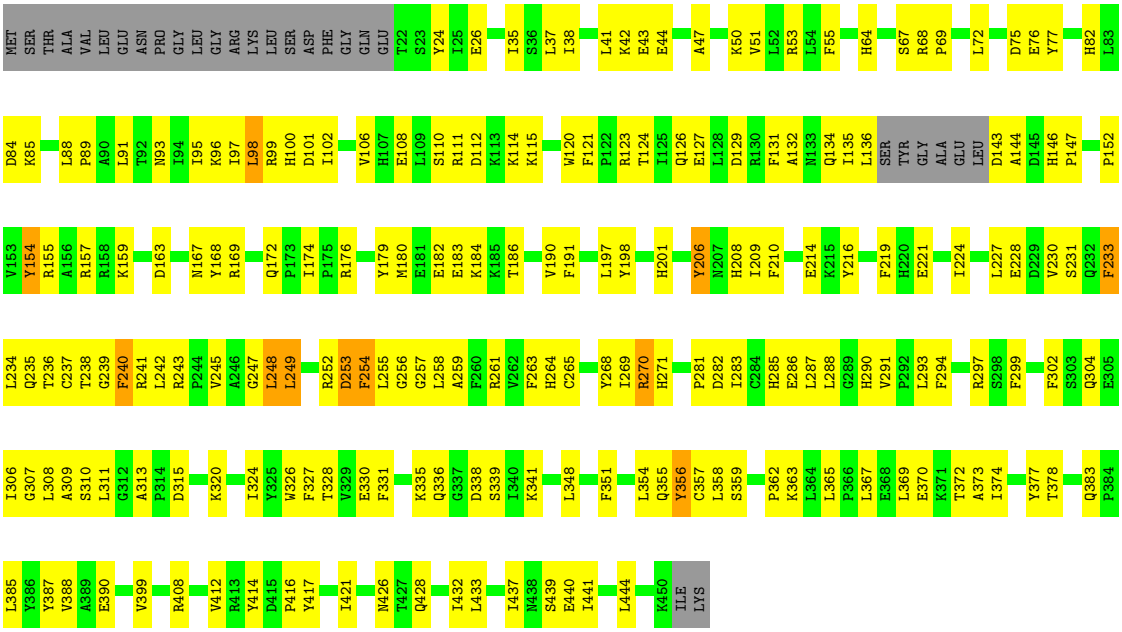
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
3	A	1	1	1	0	0
			Total	O		
3	C	3	3	3	0	0
			Total	O		
3	B	1	1	1	0	0

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. 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.

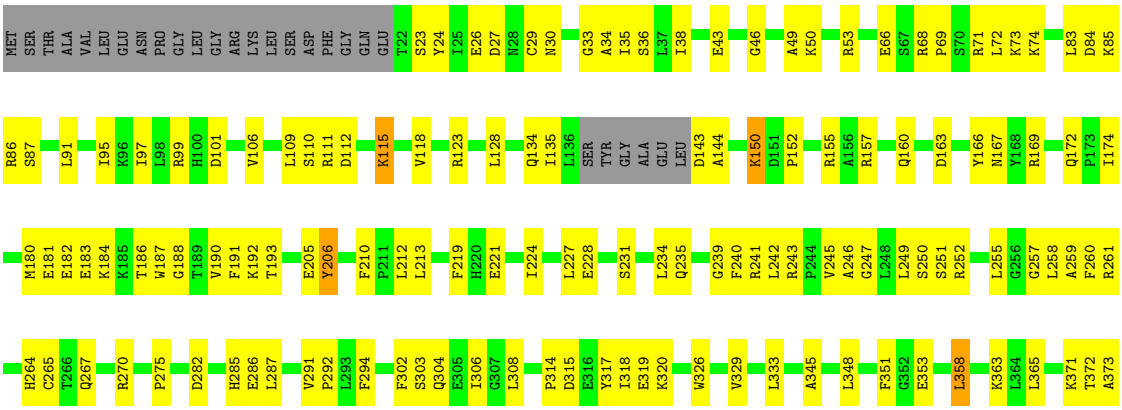
- Molecule 1: Phenylalanine-4-hydroxylase

Chain D:



- Molecule 1: Phenylalanine-4-hydroxylase

Chain A:



I374 Y377 E381 F382 Q383 P384 L385 Y386 Y387 V388 E397 N401 F402 T405 F410 R413 Y414 D415 P416 R420 I421 E422 V423 L424 D425 Q428 Q429 L430 K431 I432 L433 I437 I443 L444 L448 Q449 K450 ILE LYS

- Molecule 1: Phenylalanine-4-hydroxylase

Chain C: 50% 42% • 6%

MET SER THR ALA VAL LEU GLU ASN PRO GLY LEU GLY ARG LYS LEU SER ASP PHE GLY GLN GLU T22 S23 Y24 D27 A34 I35 I38 L41 K42 G46 A47 L48 A49 K50 V51 L54 F55 N61 L62 T63 H64 I65 E66 S67 R68 P69 S70 R71 L72 D75

F79 H82 L83 D84 K85 R86 S87 L88 P89 A90 L91 I95 K96 I97 L98 R99 A104 L109 S110 R111 D112 K113 K114 K115 D116 T117 V118 P119 P122 R123 T124 E127 L128 D129 R130 F131 Q134 I135 L136 SER TYR GLY ALA GLU LEU D143 P147 Y154 R155 A156 R157

R158 K159 Q160 F161 A162 D163 I164 I174 V177 E178 Y179 M180 E181 E182 E183 K184 K185 T186 W187 V190 F191 K192 T193 L194 Y198 H201 A202 C203 Y206 N207 H208 I209 F210 P211 L212 L213 E214 K215 Y216 F219 H220 E221 I224 P225 Q226 L227 E228 S231 Q232 F233 L234

Q235 T236 C237 T238 G239 F240 R241 L242 R243 P244 V245 A246 G247 L248 L249 R252 D253 F254 L255 G256 G257 L258 A259 F260 R261 H264 C265 T266 I269 R270 H271 G272 S273 P281 D282 I283 C284 H285 E286 L287 H290 V291 P292 L293 F294 R297 S298 F299 F302 S303 Q304 E305 I306

G307 L308 L311 P314 D315 E316 Y317 I318 E319 K320 L321 I324 Y325 V329 L333 Q336 S339 I340 K341 G344 L348 E353 L354 Q355 Y356 C357 L358 S359 P362 L369 E370 A373 I374 V379 F382 V388 K396 F402 A403 A404 T405 I406

P407 R408 Y414 D415 P416 Y417 I421 E422 V423 L424 Q428 Q429 L430 K431 I432 L433 A434 D435 S436 I437 N438 I441 L444 C445 L448 Q449 K450 ILE LYS

- Molecule 1: Phenylalanine-4-hydroxylase

Chain B: 54% 39% • 6%

MET SER THR ALA VAL LEU GLU ASN PRO GLY LEU GLY ARG LYS LEU SER ASP PHE GLY GLN E21 T22 S23 D27 N28 N32 I35 S36 L37 I38 F39 S40 L41 K42 E43 E44 K50 R53 L54 E57 N58 N61 L62 E66 S67 R68 P69 S70 R71 L72 K73 K74

D75 E76 Y77 E78 F79 F80 T81 D84 K85 L88 T92 I95 K96 I97 L98 R99 H100 D101 V106 H107 E108 L109 K113 W120 F121 P122 R123 T124 I125 Q126 E127 I135 L136 SER TYR GLY ALA GLU LEU D143 H146 F149 R155 A156 R157 R158 K159 Q160

D163 Y168 Q172 F173 I174 E178 Y179 M180 K184 K185 T186 W187 G188 T189 V190 F191 K192 L197 Y198 H201 Y204 E205 Y206 N207 H208 I209 F210 P211 L212 L213 E214 K215 Y216 F219 H220 E221 S231 Q232 F233 L234 Q235 T236 C237 T238 G239 F240 R241 L242 R243 P244

G247 L248 L249 F254 F260 R261 H264 C265 T266 I269 R270 T278 P279 E280 P281 D282 I283 C284 H285 E286 L287 V291 P292 S295 D296 R297 A300 Q301 F302 S303 Q304 E305 I306 G307 L308 L311 D315 I318 Y325 W326 F331 K335 Q336 I340

L347 L348 F351 G352 E353 L354 Q355 Y356 C357 L358 K363 L369 E370 K371 I374 Y377 T378 F382 Q383 P384 L385 A395 K398 V399 R400 N401 F402 T405 Y414 D415 P416 Y417 T418 Q419 R420 I421 E422 N426 T427 Q428 Q429 L430 K431 I432 S436 I437 N438 S439

E440 I441 G442 I443 L444 C445 L448 Q449 K450 ILE LYS

4 Data and refinement statistics (i)

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, α, β, γ	101.94Å 101.37Å 203.54Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	33.92 – 3.18 49.18 – 2.90	Depositor EDS
% Data completeness (in resolution range)	97.8 (33.92-3.18) 94.4 (49.18-2.90)	Depositor EDS
R_{merge}	0.18	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	1.43 (at 2.91Å)	Xtriage
Refinement program	PHENIX (1.13_2998: ???)	Depositor
R, R_{free}	0.262 , 0.311 0.262 , 0.307	Depositor DCC
R_{free} test set	2070 reflections (4.74%)	wwPDB-VP
Wilson B-factor (Å^2)	58.3	Xtriage
Anisotropy	0.121	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.29 , -48.0	EDS
L-test for twinning[2]	$< \|L\| > = 0.40$, $< L^2 > = 0.23$	Xtriage
Estimated twinning fraction	0.039 for -k,-h,-l 0.030 for k,h,-l 0.087 for -h,-k,l	Xtriage
F_o,F_c correlation	0.89	EDS
Total number of atoms	13879	wwPDB-VP
Average B, all atoms (Å^2)	81.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 6.94% 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 (i)

5.1 Standard geometry (i)

Bond lengths and bond angles in the following residue types are not validated in this section: H4B

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	0.31	0/3531	0.56	0/4778
1	B	0.32	0/3541	0.55	0/4792
1	C	0.31	0/3532	0.56	0/4780
1	D	0.33	0/3564	0.57	0/4823
All	All	0.32	0/14168	0.56	0/19173

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	D	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	D	254[A]	PHE	Mainchain

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	3445	0	3394	137	1
1	B	3455	0	3404	164	1
1	C	3446	0	3398	211	0
1	D	3477	0	3430	215	1
2	B	17	0	15	1	0
2	C	17	0	15	6	0
2	D	17	0	15	26	0
3	A	1	0	0	1	0
3	B	1	0	0	0	0
3	C	3	0	0	2	0
All	All	13879	0	13671	679	2

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 25.

The worst 5 of 679 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:254[A]:PHE:HE2	2:D:501:H4B:C2	1.20	1.55
1:D:254[A]:PHE:CE2	2:D:501:H4B:C4	2.02	1.41
1:D:254[A]:PHE:HE2	2:D:501:H4B:N3	1.20	1.35
1:C:178:GLU:CB	1:B:180:MET:HE2	1.56	1.35
1:D:254[A]:PHE:CE2	2:D:501:H4B:N3	1.94	1.34

All (2) 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:D:441:ILE:CG2	1:D:441:ILE:CG2[2_556]	2.01	0.19
1:A:115:LYS:NZ	1:B:374:ILE:O[3_555]	2.06	0.14

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	419/452 (93%)	388 (93%)	30 (7%)	1 (0%)	47	78
1	B	420/452 (93%)	385 (92%)	35 (8%)	0	100	100
1	C	419/452 (93%)	387 (92%)	31 (7%)	1 (0%)	47	78
1	D	423/452 (94%)	395 (93%)	26 (6%)	2 (0%)	29	66
All	All	1681/1808 (93%)	1555 (92%)	122 (7%)	4 (0%)	47	78

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	246	ALA
1	D	115	LYS
1	D	374	ILE
1	A	374	ILE

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	376/399 (94%)	369 (98%)	7 (2%)	57	80
1	B	377/399 (94%)	367 (97%)	10 (3%)	44	74
1	C	376/399 (94%)	369 (98%)	7 (2%)	57	80
1	D	379/399 (95%)	364 (96%)	15 (4%)	31	64
All	All	1508/1596 (94%)	1469 (97%)	39 (3%)	47	75

5 of 39 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	402	PHE
1	B	216	TYR
1	B	21	GLU
1	B	61	ASN
1	B	270	ARG

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 8 such

sidechains are listed below:

Mol	Chain	Res	Type
1	B	61	ASN
1	C	290	HIS
1	A	30	ASN
1	D	438	ASN
1	C	264	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 monosaccharides in this entry.

5.6 Ligand geometry (i)

3 ligands are modelled in this entry.

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	$\#\|Z\| > 2$
2	H4B	C	501	-	16,18,18	1.08	1 (6%)	14,26,26	1.99	5 (35%)
2	H4B	B	501	-	16,18,18	1.05	1 (6%)	14,26,26	2.00	5 (35%)
2	H4B	D	501	-	16,18,18	1.06	1 (6%)	14,26,26	2.00	5 (35%)

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
2	H4B	C	501	-	-	4/8/17/17	0/2/2/2
2	H4B	B	501	-	-	2/8/17/17	0/2/2/2
2	H4B	D	501	-	-	2/8/17/17	0/2/2/2

All (3) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	C	501	H4B	C4A-C8A	3.59	1.48	1.41
2	D	501	H4B	C4A-C8A	3.54	1.48	1.41
2	B	501	H4B	C4A-C8A	3.49	1.47	1.41

The worst 5 of 15 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	D	501	H4B	C2-N3-C4	4.51	122.23	115.96
2	C	501	H4B	C2-N3-C4	4.49	122.21	115.96
2	B	501	H4B	C2-N3-C4	4.47	122.18	115.96
2	C	501	H4B	C4-C4A-N5	2.96	122.80	118.57
2	D	501	H4B	C4-C4A-N5	2.95	122.79	118.57

There are no chirality outliers.

5 of 8 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	D	501	H4B	N5-C6-C9-O9
2	C	501	H4B	N5-C6-C9-O9
2	B	501	H4B	N5-C6-C9-O9
2	C	501	H4B	C7-C6-C9-O9
2	C	501	H4B	N5-C6-C9-C10

There are no ring outliers.

3 monomers are involved in 33 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	C	501	H4B	6	0
2	B	501	H4B	1	0
2	D	501	H4B	26	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 (i)

6.1 Protein, DNA and RNA chains (i)

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

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

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

6.3 Carbohydrates (i)

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

6.4 Ligands (i)

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

6.5 Other polymers (i)

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