wwPDB X-ray Structure Validation Summary Report (i)

Jan 13, 2024 – 09:08 pm GMT

PDB ID	:	6S6W
Title	:	Crystal Structure of human ALDH1A3 in complex with 2,6-diphenylimidazo[1,2-a]pyridine (compound GA11) and NAD+
Authors	:	Gelardi, E.L.M.; Quattrini, L.; LaMotta, C.; Ferraris, D.M.; Garavaglia, S.
Deposited on	:	2019-07-03
Resolution	:	3.25 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:

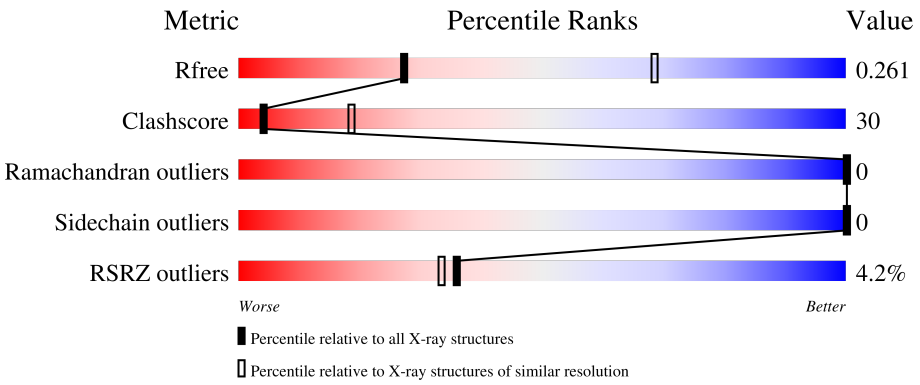
MolProbity	:	4.02b-467
Mogul	:	1.8.4, CSD as541be (2020)
Xtriage (Phenix)	:	1.13
EDS	:	2.36
buster-report	:	1.1.7 (2018)
Percentile statistics	:	20191225.v01 (using entries in the PDB archive December 25th 2019)
Refmac	:	5.8.0158
CCP4	:	7.0.044 (Gargrove)
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)	:	Parkinson et al. (1996)
Validation Pipeline (wwPDB-VP)	:	2.36

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 3.25 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	1191 (3.30-3.22)
Clashscore	141614	1251 (3.30-3.22)
Ramachandran outliers	138981	1229 (3.30-3.22)
Sidechain outliers	138945	1228 (3.30-3.22)
RSRZ outliers	127900	1154 (3.30-3.22)

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	489	3% 51% 46% .
1	B	489	5% 50% 50%

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
4	KXT	A	603	-	-	-	X

2 Entry composition (i)

There are 5 unique types of molecules in this entry. The entry contains 7620 atoms, of which 30 are hydrogens and 0 are deuteriums.

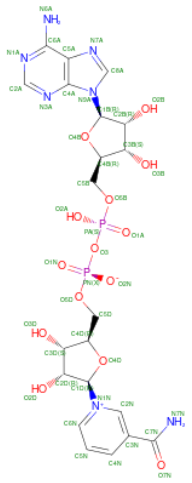
In the tables below, the ZeroOcc column contains the number of atoms modelled with zero occupancy, the AltConf column contains the number of residues with at least one atom in alternate conformation and the Trace column contains the number of residues modelled with at most 2 atoms.

- Molecule 1 is a protein called Aldehyde dehydrogenase family 1 member A3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	476	3679	2344	628	687	20	0	0	0
1	B	489	3779	2409	646	704	20	0	0	0

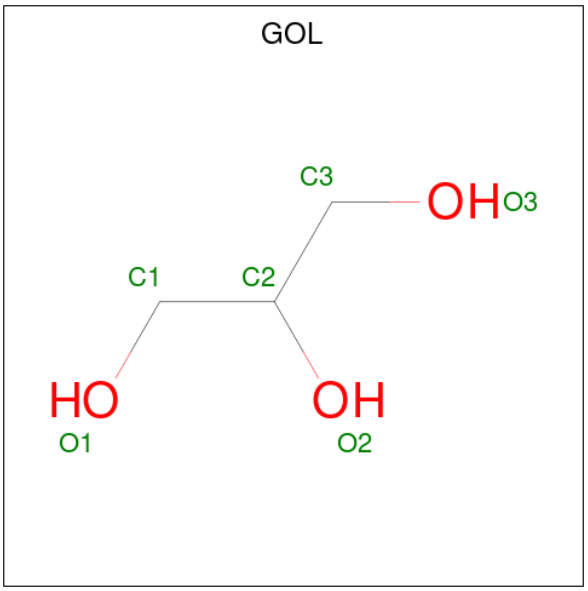
- Molecule 2 is NICOTINAMIDE-ADENINE-DINUCLEOTIDE (three-letter code: NAD) (formula: $C_{21}H_{27}N_7O_{14}P_2$).

NAD



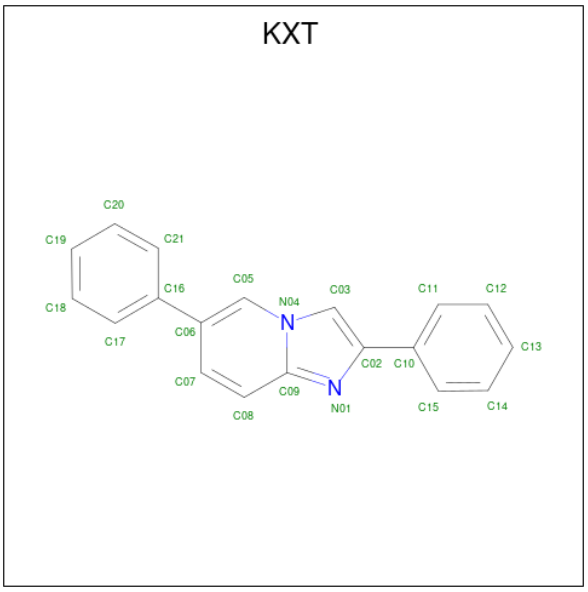
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
			Total	C	N	O	P		
2	A	1	44	21	7	14	2	0	0
2	B	1	44	21	7	14	2	0	0

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: $C_3H_8O_3$).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	O		
3	A	1	14	3	8	3	0	0
			Total	C	H	O		
3	A	1	14	3	8	3	0	0

- Molecule 4 is 2,6-diphenylimidazo[1,2-a]pyridine (three-letter code: KXT) (formula: $C_{19}H_{14}N_2$) (labeled as "Ligand of Interest" by depositor).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	H	N		
4	A	1	35	19	14	2	0	0

- Molecule 5 is water.

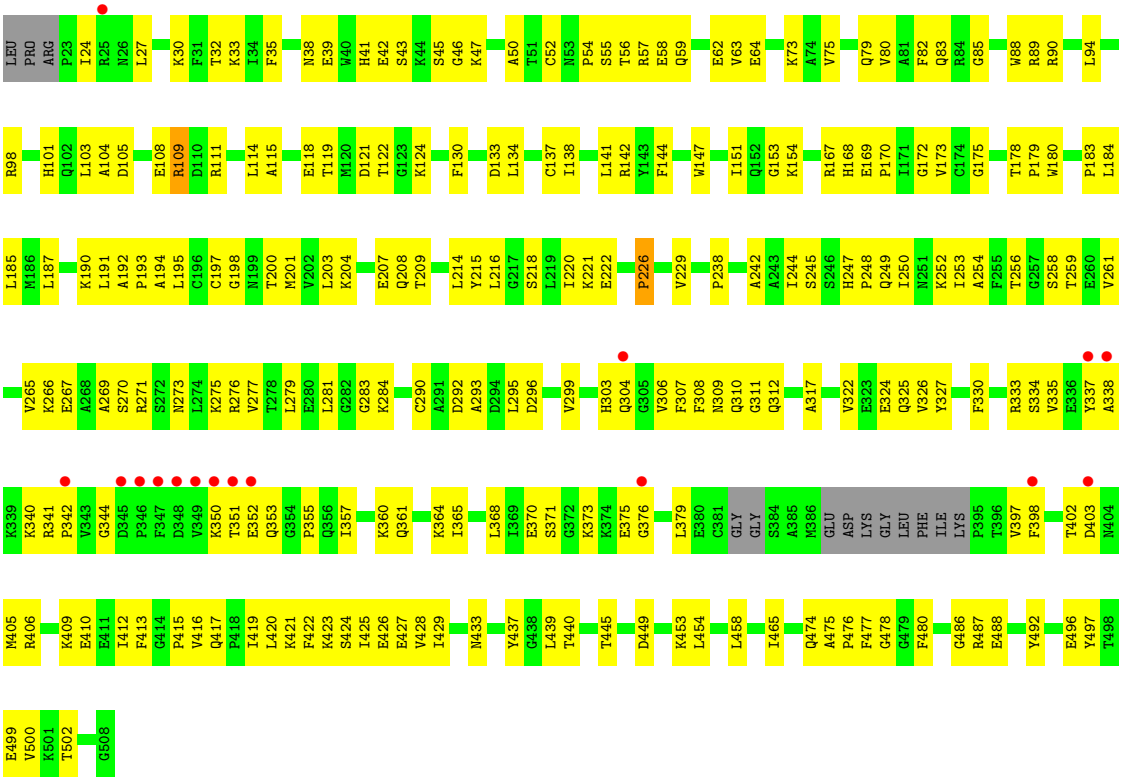
Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
			Total	O		
5	A	5	5	5	0	0
			Total	O		
5	B	6	6	6	0	0

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 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.

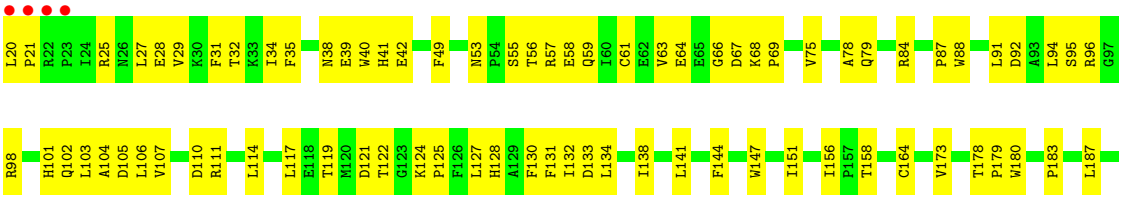
• Molecule 1: Aldehyde dehydrogenase family 1 member A3

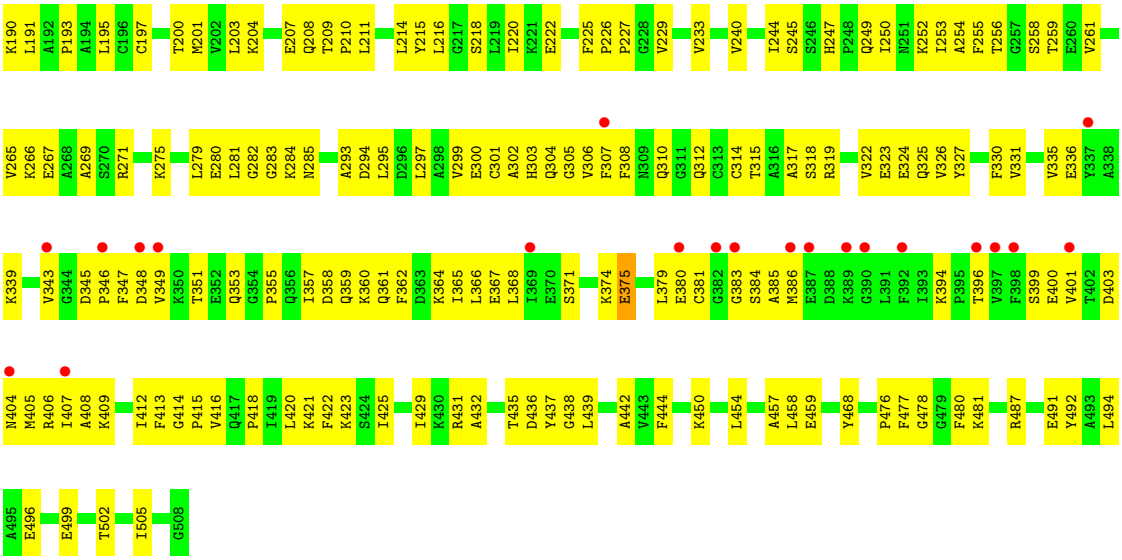
Chain A:



• Molecule 1: Aldehyde dehydrogenase family 1 member A3

Chain B:




K190 L191 A192 P193 A194 L195 C196 C197
T200 M201 V202 L203 K204
E207 Q208 T209 P210 L211
L214 Y215 L216 G217 S218 L219 I220 K221 E222
F225 P226 P227 G228 V229
V233
V240
I244 S245 S246 H247 P248 Q249 I250 N251 K252 I253 A254 F255 T256 G257 S258 T259 E260 V261
V265 K266 E267 A268 A269 S270 R271
K275
L279 E280 L281 G282 G283 K284 N285
A293 D294 L295 D296 L297 A298 V299 E300 C301 A302 H303 Q304 G305 V306 F307 F308 N309 Q310 G311 Q312 C313 C314 T315 A316 A317 S318 R319
V322 E323 E324 Q325 V326 Y327
F330 V331
V335 E336 Y337 A338
K339
V343 G344 D345 P346 F347 D348 V349 K350 T351 E352 Q353 G354 P355 Q356 I357 D358 Q359 K360 Q361 F362 D363 K364 I365 L366 E367 L368 I369 E370 S371
K374 E375
L379 E380 C381 G382 G383 S384 A385 M386 E387 D388 K389 G390 L391 F392 I393 K394 P395 T396 V397 F398 S399 E400 V401 T402 D403
N404 M405 R406 I407 A408 K409
I412 F413 G414 P415 V416 Q417 P418 I419 L420 K421 F422 K423 S424 I425
I429 K430 R431 A432
T435 D436 Y437 G438 L439
A442 V443 F444
K450
L454
A457 L458 E459
Y468
P476 F477 G478 G479 F480 K481
R487
E491 Y492 A493 L494
A495 E496
E499
T502
I505
G508

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 2 21	Depositor
Cell constants a, b, c, α, β, γ	81.56Å 89.36Å 159.04Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	48.00 – 3.25 48.02 – 3.25	Depositor EDS
% Data completeness (in resolution range)	99.8 (48.00-3.25) 99.6 (48.02-3.25)	Depositor EDS
R_{merge}	(Not available)	Depositor
R_{sym}	0.13	Depositor
$< I/\sigma(I) >$ [1]	2.09 (at 3.25Å)	Xtriage
Refinement program	PHENIX	Depositor
R, R_{free}	0.177 , 0.256 0.200 , 0.261	Depositor DCC
R_{free} test set	912 reflections (4.84%)	wwPDB-VP
Wilson B-factor (Å^2)	82.2	Xtriage
Anisotropy	0.432	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 67.2	EDS
L-test for twinning[2]	$< \|L\| > = 0.49$, $< L^2 > = 0.32$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
F_o,F_c correlation	0.94	EDS
Total number of atoms	7620	wwPDB-VP
Average B, all atoms (Å^2)	78.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.10% 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: NAD, GOL, KXT

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	1/3753 (0.0%)	0.51	1/5076 (0.0%)
1	B	0.27	0/3857	0.49	3/5219 (0.1%)
All	All	0.29	1/7610 (0.0%)	0.50	4/10295 (0.0%)

All (1) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	226	PRO	C-N	8.16	1.49	1.34

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	A	109	ARG	CB-CA-C	-6.65	97.10	110.40
1	B	375	GLU	N-CA-CB	-6.28	99.30	110.60
1	B	375	GLU	N-CA-C	6.09	127.45	111.00
1	B	374	LYS	CB-CA-C	5.16	120.71	110.40

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	3679	0	3702	226	0

Continued on next page...

Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
1	B	3779	0	3811	238	0
2	A	44	0	23	5	0
2	B	44	0	24	2	0
3	A	12	16	16	2	0
4	A	21	14	0	0	0
5	A	5	0	0	1	0
5	B	6	0	0	0	0
All	All	7590	30	7576	450	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 30.

The worst 5 of 450 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:A:371:SER:HB2	1:A:406:ARG:HD3	1.42	1.00
1:B:306:VAL:HG22	1:B:317:ALA:HB3	1.40	0.99
1:B:283:GLY:HA3	1:B:437:TYR:HB3	1.43	0.98
1:A:290:CYS:HB2	1:A:445:THR:HG22	1.42	0.97
1:A:341:ARG:HH21	1:A:352:GLU:HB3	1.29	0.97

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	470/489 (96%)	427 (91%)	43 (9%)	0	100	100
1	B	487/489 (100%)	441 (91%)	46 (9%)	0	100	100
All	All	957/978 (98%)	868 (91%)	89 (9%)	0	100	100

There are no Ramachandran outliers to report.

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	395/405 (98%)	395 (100%)	0	100	100
1	B	405/405 (100%)	405 (100%)	0	100	100
All	All	800/810 (99%)	800 (100%)	0	100	100

There are no protein residues with a non-rotameric sidechain to report.

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	312	GLN
1	B	310	GLN
1	B	353	GLN

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)

5 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	NAD	A	601	1	42,48,48	5.46	18 (42%)	50,73,73	1.88	7 (14%)
3	GOL	A	604	-	5,5,5	0.90	0	5,5,5	1.05	0
3	GOL	A	602	-	5,5,5	1.15	0	5,5,5	1.42	1 (20%)
4	KXT	A	603	-	21,24,24	1.87	3 (14%)	25,33,33	1.03	1 (4%)
2	NAD	B	601	-	42,48,48	5.44	17 (40%)	50,73,73	1.80	6 (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
2	NAD	A	601	1	-	2/26/62/62	0/5/5/5
3	GOL	A	604	-	-	0/4/4/4	-
3	GOL	A	602	-	-	0/4/4/4	-
4	KXT	A	603	-	-	8/8/8/8	0/4/4/4
2	NAD	B	601	-	-	5/26/62/62	0/5/5/5

The worst 5 of 38 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	601	NAD	C2D-C1D	-16.76	1.28	1.53
2	B	601	NAD	C2D-C1D	-16.45	1.28	1.53
2	B	601	NAD	C2B-C1B	-16.11	1.29	1.53
2	A	601	NAD	C2B-C1B	-15.89	1.29	1.53
2	A	601	NAD	O4D-C1D	15.49	1.62	1.41

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

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
2	B	601	NAD	C5A-C6A-N6A	7.63	131.94	120.35
2	A	601	NAD	C5A-C6A-N6A	7.48	131.72	120.35
2	A	601	NAD	N3A-C2A-N1A	-5.29	120.41	128.68
2	B	601	NAD	N3A-C2A-N1A	-5.28	120.42	128.68
2	B	601	NAD	N6A-C6A-N1A	-5.02	108.15	118.57

There are no chirality outliers.

5 of 15 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	B	601	NAD	PA-O3-PN-O5D
2	B	601	NAD	O4D-C1D-N1N-C6N
4	A	603	KXT	C03-C02-C10-C11
4	A	603	KXT	C03-C02-C10-C15
4	A	603	KXT	N01-C02-C10-C11

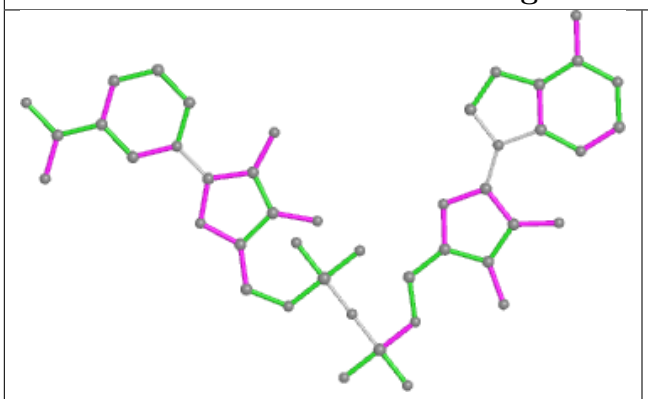
There are no ring outliers.

3 monomers are involved in 9 short contacts:

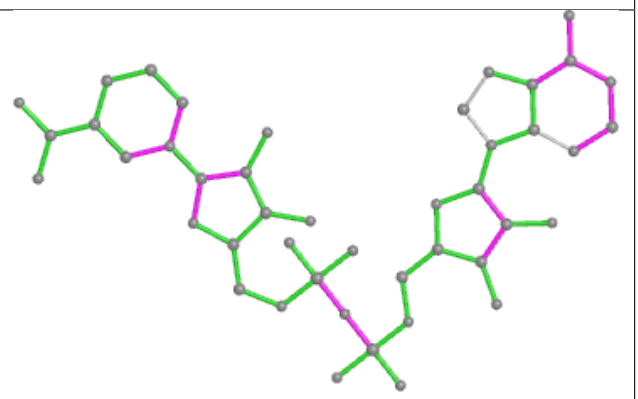
Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	601	NAD	5	0
3	A	604	GOL	2	0
2	B	601	NAD	2	0

The following is a two-dimensional graphical depiction of Mogul quality analysis of bond lengths, bond angles, torsion angles, and ring geometry for all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the validation Tables will also be included. For torsion angles, if less then 5% of the Mogul distribution of torsion angles is within 10 degrees of the torsion angle in question, then that torsion angle is considered an outlier. Any bond that is central to one or more torsion angles identified as an outlier by Mogul will be highlighted in the graph. For rings, the root-mean-square deviation (RMSD) between the ring in question and similar rings identified by Mogul is calculated over all ring torsion angles. If the average RMSD is greater than 60 degrees and the minimal RMSD between the ring in question and any Mogul-identified rings is also greater than 60 degrees, then that ring is considered an outlier. The outliers are highlighted in purple. The color gray indicates Mogul did not find sufficient equivalents in the CSD to analyse the geometry.

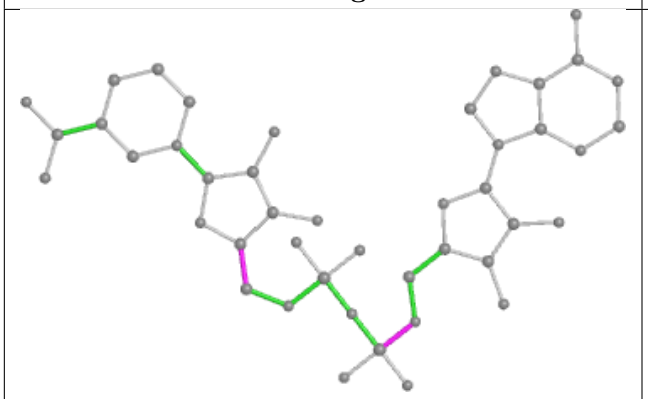
Ligand NAD A 601



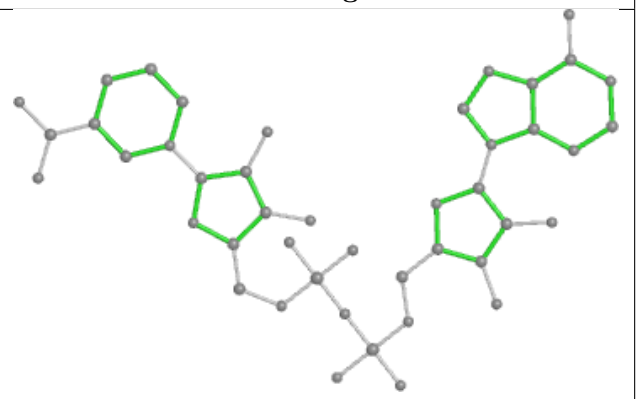
Bond lengths



Bond angles

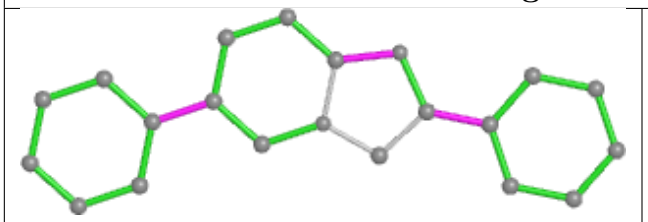


Torsions

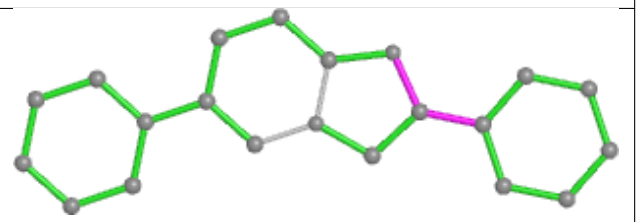


Rings

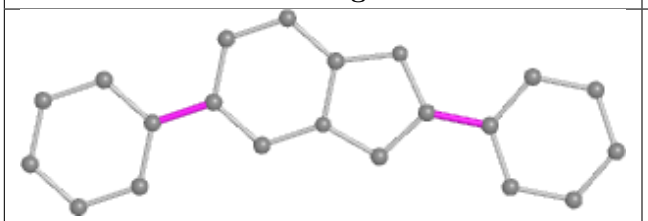
Ligand KXT A 603



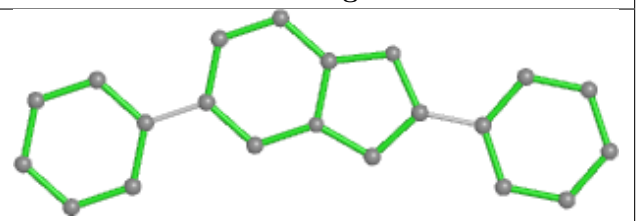
Bond lengths



Bond angles

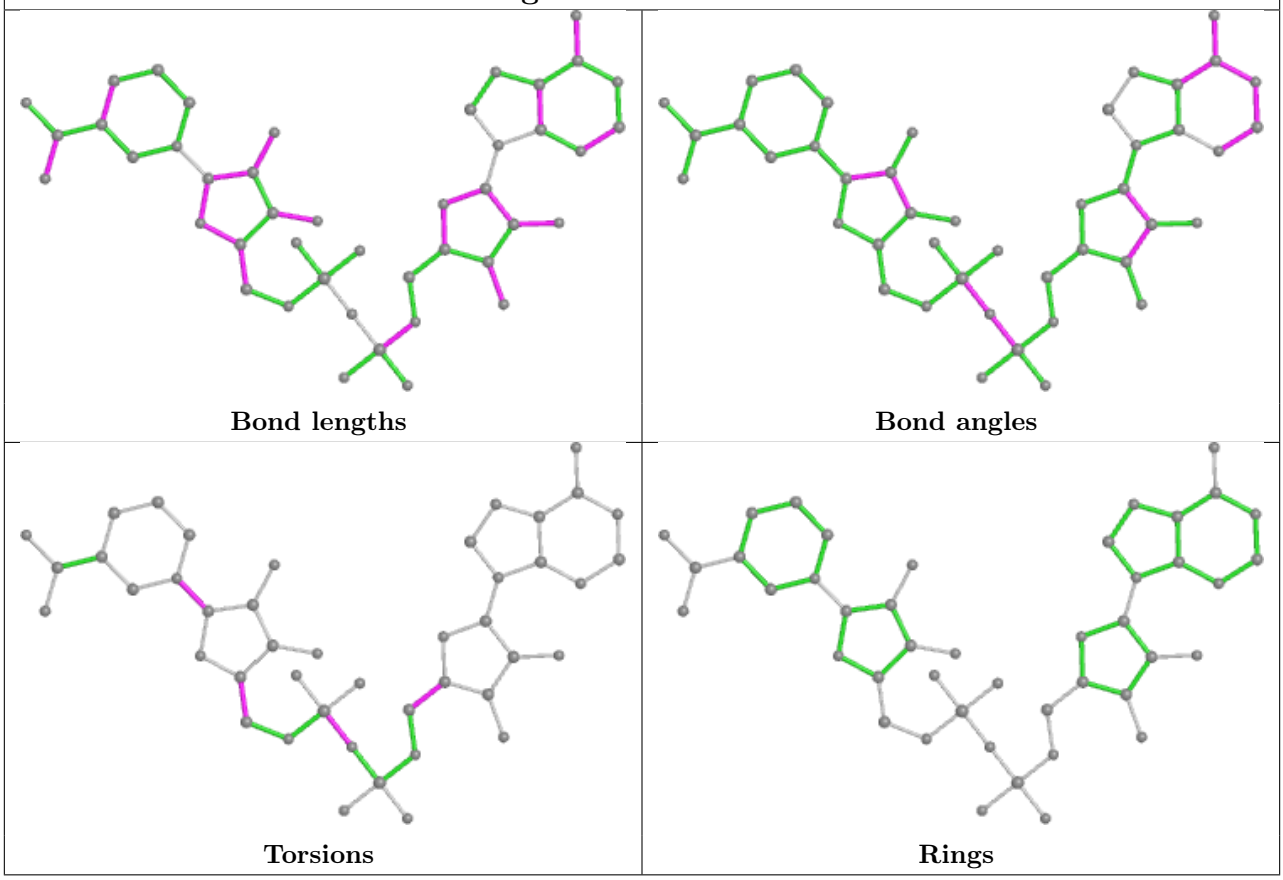


Torsions



Rings

Ligand NAD B 601



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(Å^2)	Q<0.9
1	A	476/489 (97%)	0.17	16 (3%)	45	42	43, 69, 131, 153	0
1	B	489/489 (100%)	0.15	25 (5%)	28	26	39, 70, 116, 144	0
All	All	965/978 (98%)	0.16	41 (4%)	36	33	39, 70, 125, 153	0

The worst 5 of 41 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	349	VAL	4.9
1	B	21	PRO	4.2
1	B	386	MET	4.1
1	A	348	ASP	4.0
1	A	347	PHE	3.9

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 monosaccharides in this entry.

6.4 Ligands (i)

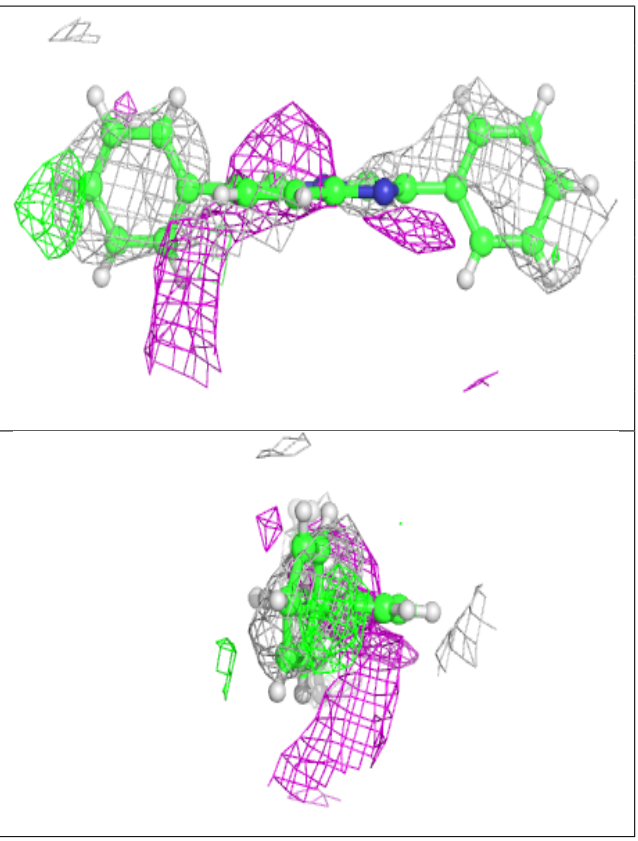
In the following table, the Atoms column lists the number of modelled atoms in the group and the number defined in the chemical component dictionary. The B-factors column lists the minimum, median, 95th percentile and maximum values of B factors of atoms in the group. The column labelled 'Q< 0.9' lists the number of atoms with occupancy less than 0.9.

Mol	Type	Chain	Res	Atoms	RSCC	RSR	B-factors(Å^2)	Q<0.9
4	KXT	A	603	21/21	0.63	0.67	79,122,154,160	0
3	GOL	A	604	6/6	0.78	0.34	61,80,97,97	0
3	GOL	A	602	6/6	0.85	0.31	62,81,100,100	0
2	NAD	A	601	44/44	0.89	0.23	50,76,133,136	0
2	NAD	B	601	44/44	0.93	0.23	64,83,142,143	0

The following is a graphical depiction of the model fit to experimental electron density of all instances of the Ligand of Interest. In addition, ligands with molecular weight > 250 and outliers as shown on the geometry validation Tables will also be included. Each fit is shown from different orientation to approximate a three-dimensional view.

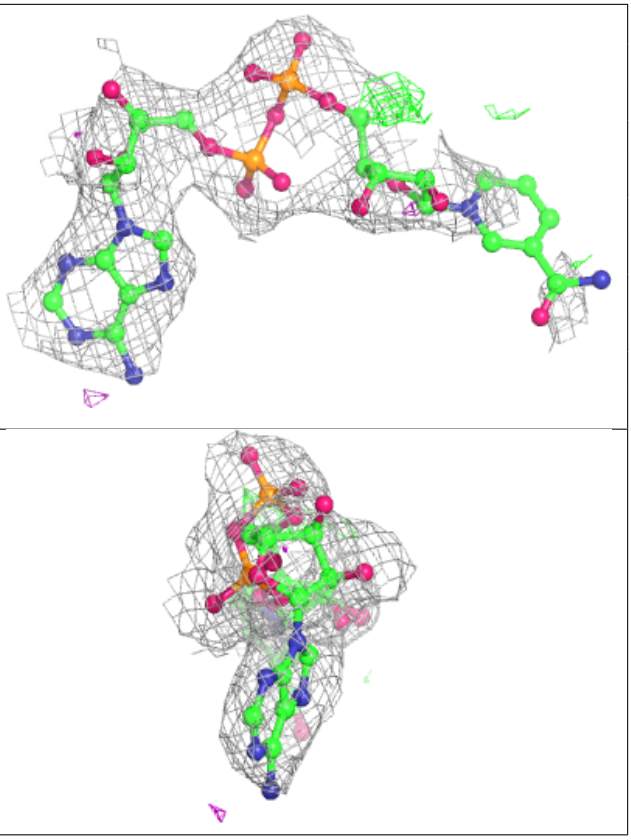
Electron density around KXT A 603:

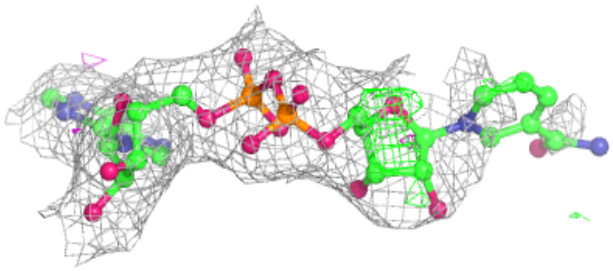
2mF$_o$-DF$_c$ (at 0.7 rmsd) in gray
mF$_o$-DF$_c$ (at 3 rmsd) in purple (negative)
and green (positive)



Electron density around NAD A 601:

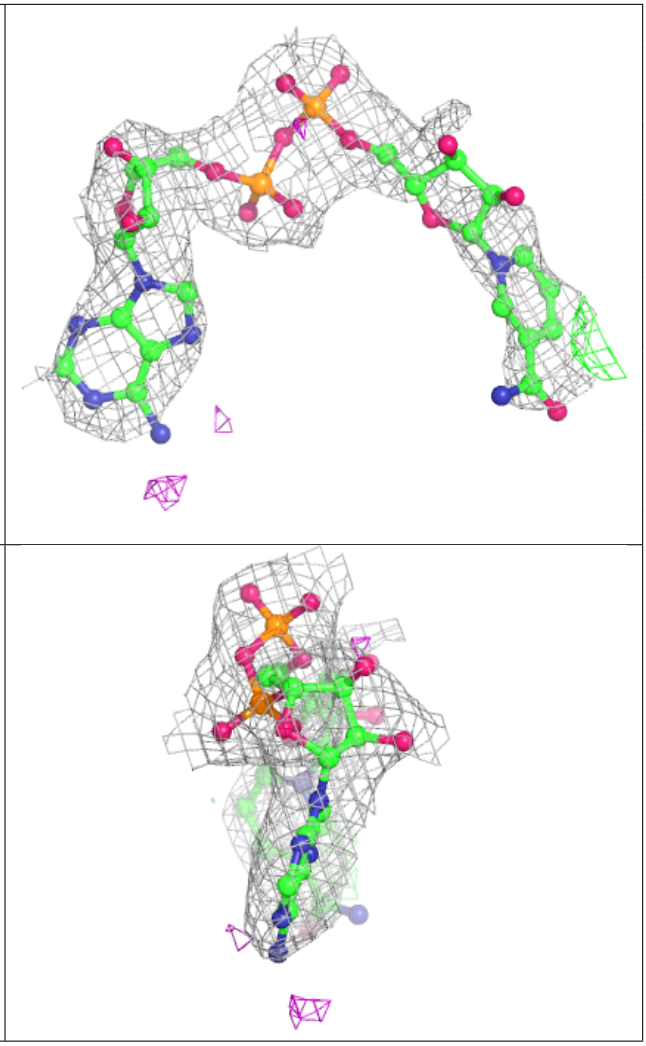
$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)





Electron density around NAD B 601:

$2mF_o$-DF_c (at 0.7 rmsd) in gray
mF_o-DF_c (at 3 rmsd) in purple (negative)
and green (positive)



6.5 Other polymers (i)

There are no such residues in this entry.