wwPDB X-ray Structure Validation Summary Report (i)

Feb 1, 2016 – 02:15 AM GMT

PDB ID	:	2GBG
Title	:	rat DPP-IV with alkynyl cyanopyrrolidine #2
Authors	:	Longenecker, K.L.; Jakob, C.G.; Fry, E.H.; Wilk, S.
Deposited on	:	2006-03-10
Resolution	:	3.00 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:

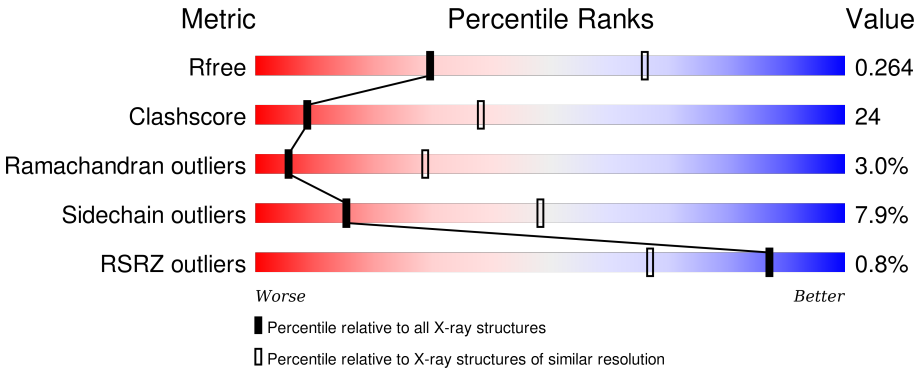
MolProbity	:	4.02b-467
Mogul	:	1.7 (RC4), CSD as536be (2015)
Xtriage (Phenix)	:	1.9-1692
EDS	:	rb-20026688
Percentile statistics	:	20151230.v01 (using entries in the PDB archive December 30th 2015)
Refmac	:	5.8.0135
CCP4	:	6.5.0
Ideal geometry (proteins)	:	Engh & Huber (2001)
Ideal geometry (DNA, RNA)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

1 Overall quality at a glance (i)

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

The reported resolution of this entry is 3.00 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}	91344	1578 (3.00-3.00)
Clashscore	102246	1912 (3.00-3.00)
Ramachandran outliers	100387	1853 (3.00-3.00)
Sidechain outliers	100360	1856 (3.00-3.00)
RSRZ outliers	91569	1592 (3.00-3.00)

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	730	% 58% 35% 6%
1	B	730	% 58% 35% 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	SO4	A	900	-	-	-	X

2 Entry composition (i)

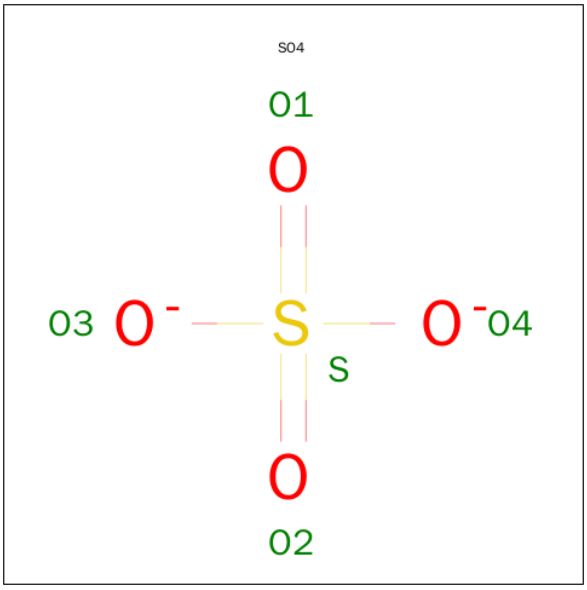
There are 3 unique types of molecules in this entry. The entry contains 11864 atoms, of which 0 are hydrogens and 0 are deuteriums.

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 Dipeptidyl peptidase 4.

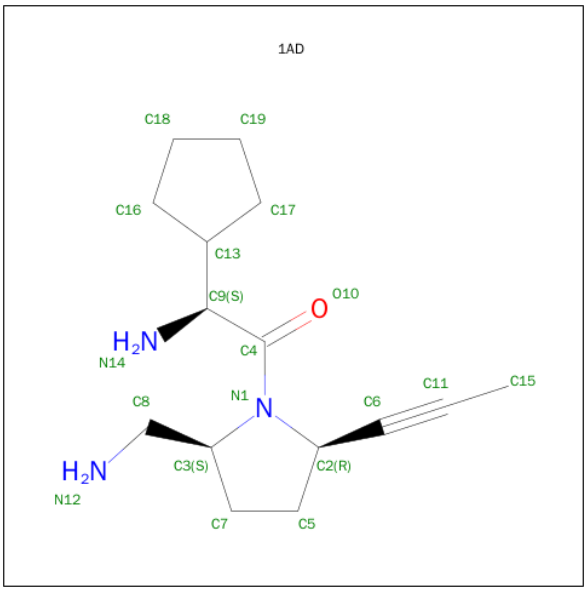
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	730	Total	C	N	O	S	0	0	0
			5920	3789	981	1124	26			
1	B	730	Total	C	N	O	S	0	0	0
			5920	3789	981	1124	26			

- Molecule 2 is SULFATE ION (three-letter code: SO4) (formula: O_4S).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	O	S	0	0
			5	4	1		

- Molecule 3 is (1S)-2-[(2S,5R)-2-(AMINOMETHYL)-5-PROP-1-YN-1-YLPYRROLIDIN-1-YL]-1-CYCLOPENTYL-2-OXOETHANAMINE (three-letter code: 1AD) (formula: $C_{15}H_{25}N_3O$).



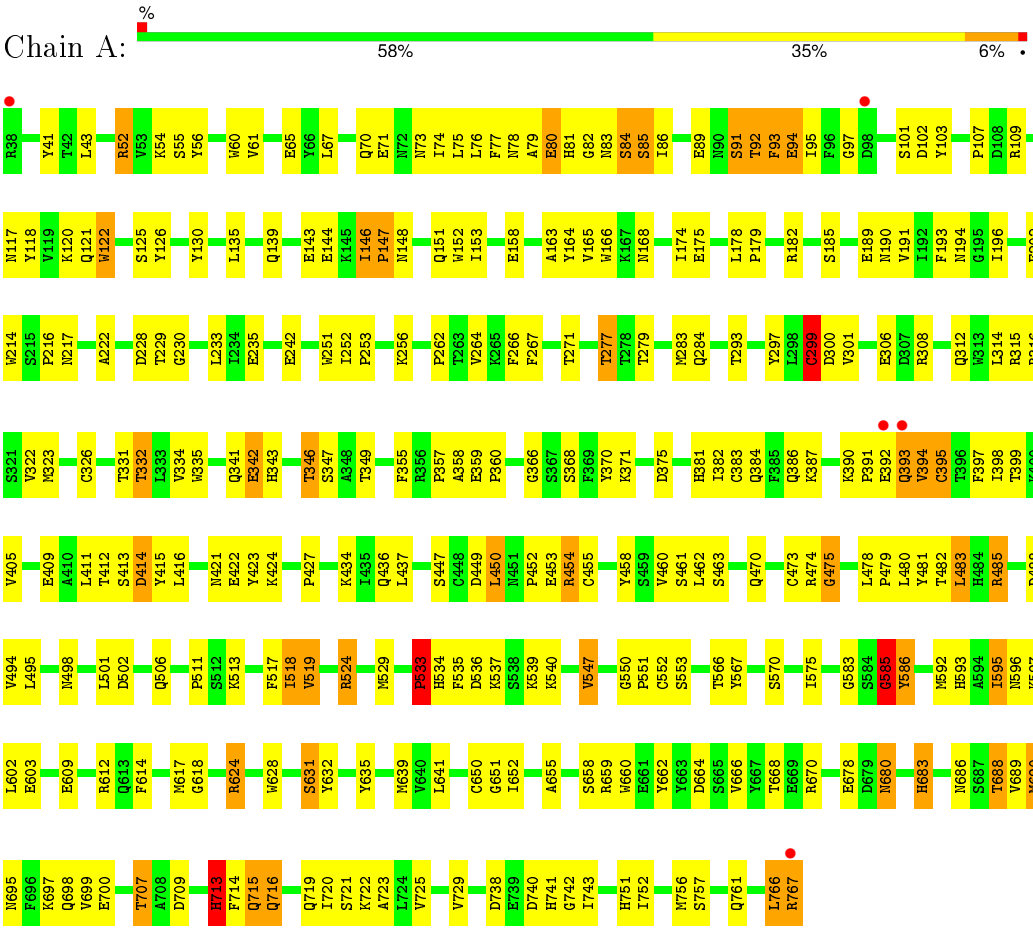
Mol	Chain	Residues	Atoms				ZeroOcc	AltConf
			Total	C	N	O		
3	A	1	19	15	3	1	0	0

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.

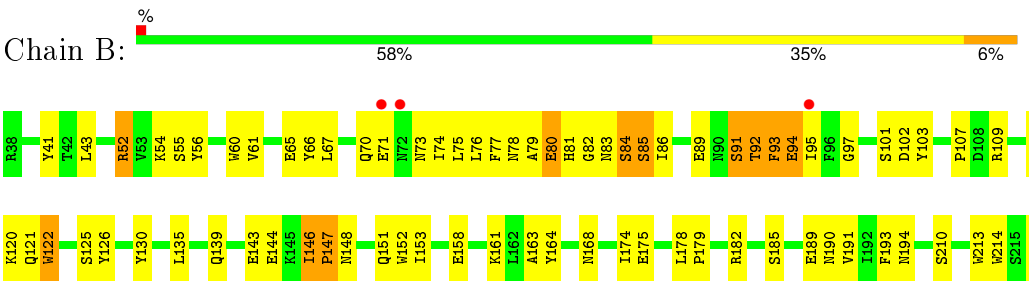
- Molecule 1: Dipeptidyl peptidase 4

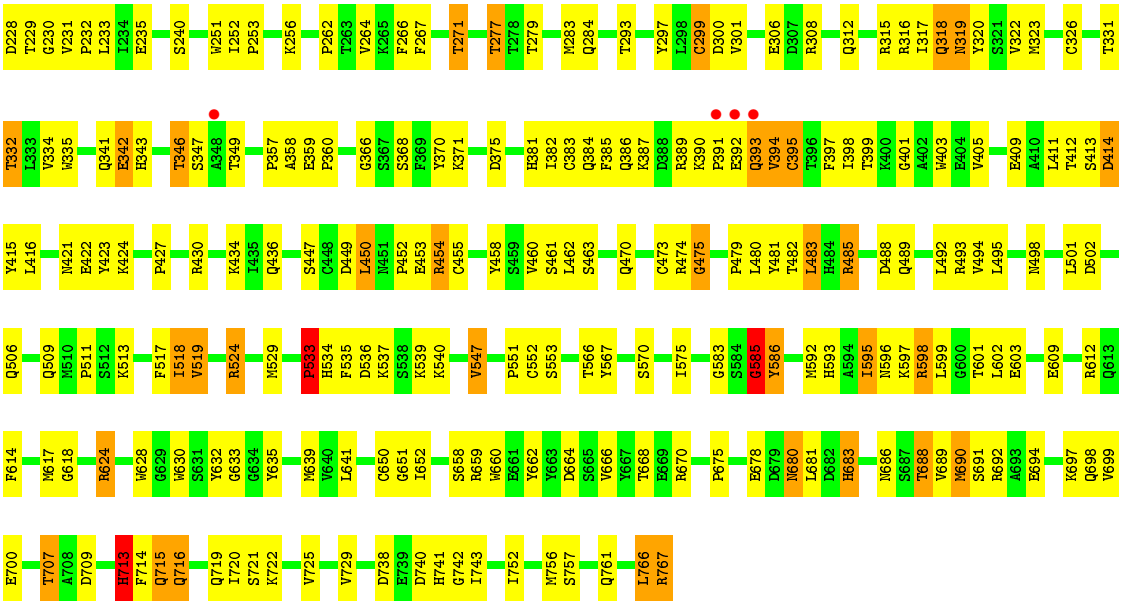
Chain A:



- Molecule 1: Dipeptidyl peptidase 4

Chain B:




D228 T229 G230 V231 P232 L233 I234 E235 S240 W251 I252 P253 K256 P262 T263 V264 K265 F266 F267 T271 T277 T278 T279 M283 Q284 T293 Y297 L298 C299 D300 V301 E306 D307 R308 Q312 R315 R316 I317 Q318 N319 Y320 S321 V322 M323 C326 T331
T332 L333 V334 W335 Q341 E342 H343 T346 S347 A348 T349 P357 A358 E359 P360 G366 S367 S368 F369 Y370 K371 D375 H381 I382 C383 Q384 Q384 F385 Q386 K387 D388 R389 K390 P391 E392 Q393 V394 C395 T396 F397 I398 T399 K400 G401 A402 W403 E404 V405 E409 A410 L411 T412 S413 D414
Y415 L416 N421 E422 Y423 K424 P427 R430 K434 I435 Q436 S447 C448 D449 L450 N451 P452 E453 R454 C455 Y458 S459 V460 S461 L462 S463 Q470 C473 R474 G475 P479 L480 Y481 T482 L483 H484 R485 D488 Q489 L492 R493 V494 L495 N498 L501 D502
Q506 Q509 M510 P511 S512 K513 F517 I518 V519 R524 M529 P533 H534 F535 D536 K537 S538 K539 K540 V547 P551 C552 S553 T566 Y567 S570 I575 G583 S584 Q585 Y586 M592 H593 A594 I595 N596 K597 R598 L599 G600 T601 L602 E603 E609 R612 Q613
F614 M617 G618 R624 W628 G629 W630 S631 Y632 G633 G634 Y635 M639 V640 L641 C650 G651 I652 S658 R659 W660 E661 Y662 Y663 D664 S665 V666 Y667 T668 E669 R670 P675 E678 D679 N680 L681 D682 H683 N686 S687 T688 V689 M690 S691 R692 A693 E694 K697 Q698 V699
E700 T707 A708 D709 H713 F714 Q715 Q716 Q719 I720 S721 K722 V725 V729 D738 E739 D740 H741 G742 I743 I752 M756 S757 Q761 L766 R767

4 Data and refinement statistics (i)

Property	Value	Source
Space group	P 21 3	Depositor
Cell constants a, b, c, α, β, γ	207.50Å 207.50Å 207.50Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	20.00 – 3.00 19.97 – 3.00	Depositor EDS
% Data completeness (in resolution range)	99.4 (20.00-3.00) 99.4 (19.97-3.00)	Depositor EDS
R_{merge}	0.11	Depositor
R_{sym}	(Not available)	Depositor
$< I/\sigma(I) >$ [1]	2.57 (at 2.98Å)	Xtriage
Refinement program	CNS	Depositor
R, R_{free}	0.235 , 0.269 0.230 , 0.264	Depositor DCC
R_{free} test set	2997 reflections (5.07%)	DCC
Wilson B-factor (Å^2)	61.7	Xtriage
Anisotropy	0.000	Xtriage
Bulk solvent k_{sol}(e/Å^3), B_{sol}(Å^2)	0.35 , 54.4	EDS
Estimated twinning fraction	0.012 for l,-k,h	Xtriage
L-test for twinning[2]	$< \|L\| > = 0.51$, $< L^2 > = 0.34$	Xtriage
Outliers	2 of 59077 reflections (0.003%)	Xtriage
F_o,F_c correlation	0.92	EDS
Total number of atoms	11864	wwPDB-VP
Average B, all atoms (Å^2)	57.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 1.92% 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

Bond lengths and bond angles in the following residue types are not validated in this section: 1AD, SO4

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	0.46	2/6088 (0.0%)	0.70	2/8278 (0.0%)
1	B	0.42	0/6088	0.69	2/8278 (0.0%)
All	All	0.44	2/12176 (0.0%)	0.69	4/16556 (0.0%)

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	631	SER	C-O	8.39	1.39	1.23
1	A	299	CYS	CB-SG	-5.43	1.73	1.81

All (4) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
1	B	713	HIS	N-CA-C	7.38	130.91	111.00
1	A	713	HIS	N-CA-C	7.22	130.49	111.00
1	B	585	GLY	N-CA-C	5.25	126.22	113.10
1	A	585	GLY	N-CA-C	5.06	125.76	113.10

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	5920	0	5631	274	0
1	B	5920	0	5632	274	0
2	A	5	0	0	0	0
3	A	19	0	23	1	0
All	All	11864	0	11286	544	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 24.

The worst 5 of 544 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:B:688:THR:HG22	1:B:691:SER:H	1.21	1.03
1:A:688:THR:HG22	1:A:691:SER:H	1.21	1.00
1:B:349:THR:HB	1:B:593:HIS:HD2	1.27	0.99
1:A:349:THR:HB	1:A:593:HIS:HD2	1.27	0.97
1:A:52:ARG:HH11	1:A:52:ARG:HB2	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	728/730 (100%)	635 (87%)	71 (10%)	22 (3%)	5 29
1	B	728/730 (100%)	634 (87%)	73 (10%)	21 (3%)	6 29
All	All	1456/1460 (100%)	1269 (87%)	144 (10%)	43 (3%)	5 29

5 of 43 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	80	GLU

Continued on next page...

Continued from previous page...

Mol	Chain	Res	Type
1	A	91	SER
1	A	92	THR
1	A	143	GLU
1	A	450	LEU

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	643/651 (99%)	592 (92%)	51 (8%)	15 48
1	B	643/651 (99%)	592 (92%)	51 (8%)	15 48
All	All	1286/1302 (99%)	1184 (92%)	102 (8%)	15 48

5 of 102 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	707	THR
1	B	148	ASN
1	B	688	THR
1	A	715	GLN
1	B	52	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 39 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	719	GLN
1	B	151	GLN
1	B	716	GLN
1	A	749	HIS
1	B	90	ASN

5.3.3 RNA

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 carbohydrates in this entry.

5.6 Ligand geometry (i)

2 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	$\#\|Z\| > 2$
3	1AD	A	768	1	17,20,20	1.63	2 (11%)	15,27,27	1.66	3 (20%)
2	SO4	A	900	-	4,4,4	1.15	0	6,6,6	0.23	0

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
3	1AD	A	768	1	-	0/14/37/37	0/2/2/2
2	SO4	A	900	-	-	0/0/0/0	0/0/0/0

All (2) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	768	1AD	C8-N12	-5.46	1.33	1.47
3	A	768	1AD	C4-N1	3.29	1.41	1.35

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(^{o})	Ideal(^{o})
3	A	768	1AD	O10-C4-C9	-2.28	115.63	120.12
3	A	768	1AD	C13-C9-C4	3.10	117.51	110.86
3	A	768	1AD	C7-C3-N1	3.52	105.64	101.93

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 1 short contact:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	A	768	1AD	1	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	730/730 (100%)	-0.41	5 (0%)	89	70	21, 49, 84, 143	0
1	B	730/730 (100%)	-0.28	7 (0%)	84	60	27, 58, 96, 149	0
All	All	1460/1460 (100%)	-0.35	12 (0%)	87	67	21, 53, 91, 149	0

The worst 5 of 12 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	767	ARG	6.2
1	A	393	GLN	3.7
1	B	393	GLN	3.1
1	B	392	GLU	3.1
1	B	391	PRO	2.9

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

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. LLDF column lists the quality of electron density of the group with respect to its neighbouring residues in protein, DNA or RNA chains. 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	LLDF	B-factors(Å^2)	Q<0.9
2	SO4	A	900	5/5	0.95	0.39	4.00	54,54,54,54	0
3	1AD	A	768	19/19	0.93	0.20	0.66	50,50,50,50	0

6.5 Other polymers (i)

There are no such residues in this entry.