



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 1, 2016 – 02:32 AM GMT

PDB ID : 2HK9  
Title : Crystal structure of shikimate dehydrogenase from aquifex aeolicus in complex with shikimate and NADP+ at 2.2 angstrom resolution  
Authors : Gan, J.H.; Prabakaran, P.; Gu, Y.J.; Andrykovitch, M.; Li, Y.; Liu, H.H.; Yan, H.; Ji, X.  
Deposited on : 2006-07-03  
Resolution : 2.20 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](mailto:validation@mail.wwpdb.org)  
A user guide is available at  
<http://wwpdb.org/validation/2016/XrayValidationReportHelp>  
with specific help available everywhere you see the ⓘ symbol.

---

The following versions of software and data (see [references ⓘ](#)) 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



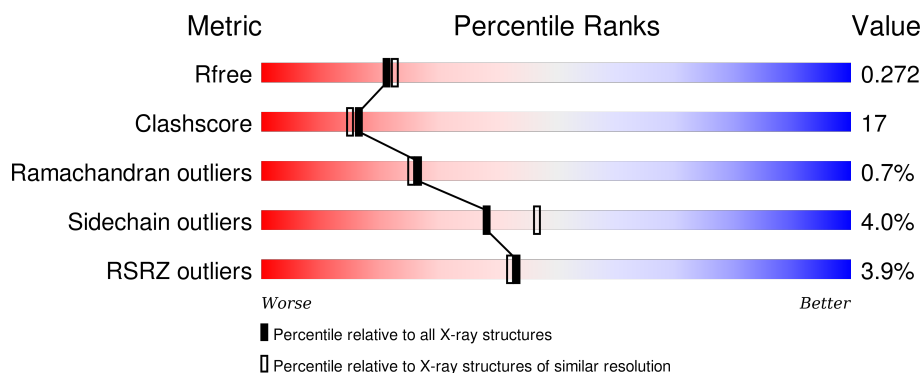
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.20 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	3774 (2.20-2.20)
Clashscore	102246	4477 (2.20-2.20)
Ramachandran outliers	100387	4404 (2.20-2.20)
Sidechain outliers	100360	4405 (2.20-2.20)
RSRZ outliers	91569	3781 (2.20-2.20)

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  $\geq 3$ , 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  $\leq 5\%$ . 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	275	<div> <div>3%</div> <div>68%</div> <div>27%</div> <div>..</div> </div>
1	B	275	<div> <div>3%</div> <div>63%</div> <div>33%</div> <div>..</div> </div>
1	C	275	<div> <div>7%</div> <div>57%</div> <div>38%</div> <div>..</div> </div>
1	D	275	<div> <div>3%</div> <div>56%</div> <div>40%</div> <div>..</div> </div>



## 2 Entry composition

There are 5 unique types of molecules in this entry. The entry contains 8871 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 Shikimate dehydrogenase.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	269	Total	C	N	O	S	0	0	0
			2127	1392	346	386	3			
1	B	267	Total	C	N	O	S	0	0	0
			2111	1384	341	383	3			
1	C	266	Total	C	N	O	S	0	0	0
			2103	1378	340	382	3			
1	D	266	Total	C	N	O	S	0	0	0
			2103	1378	340	382	3			

There are 32 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	HIS	-	EXPRESSION TAG	UNP O67049
A	-4	HIS	-	EXPRESSION TAG	UNP O67049
A	-3	HIS	-	EXPRESSION TAG	UNP O67049
A	-2	HIS	-	EXPRESSION TAG	UNP O67049
A	-1	HIS	-	EXPRESSION TAG	UNP O67049
A	0	HIS	-	EXPRESSION TAG	UNP O67049
A	195	GLU	LYS	SEE REMARK 999	UNP O67049
A	233	LEU	PHE	SEE REMARK 999	UNP O67049
B	-5	HIS	-	EXPRESSION TAG	UNP O67049
B	-4	HIS	-	EXPRESSION TAG	UNP O67049
B	-3	HIS	-	EXPRESSION TAG	UNP O67049
B	-2	HIS	-	EXPRESSION TAG	UNP O67049
B	-1	HIS	-	EXPRESSION TAG	UNP O67049
B	0	HIS	-	EXPRESSION TAG	UNP O67049
B	195	GLU	LYS	SEE REMARK 999	UNP O67049
B	233	LEU	PHE	SEE REMARK 999	UNP O67049
C	-5	HIS	-	EXPRESSION TAG	UNP O67049
C	-4	HIS	-	EXPRESSION TAG	UNP O67049
C	-3	HIS	-	EXPRESSION TAG	UNP O67049
C	-2	HIS	-	EXPRESSION TAG	UNP O67049
C	-1	HIS	-	EXPRESSION TAG	UNP O67049

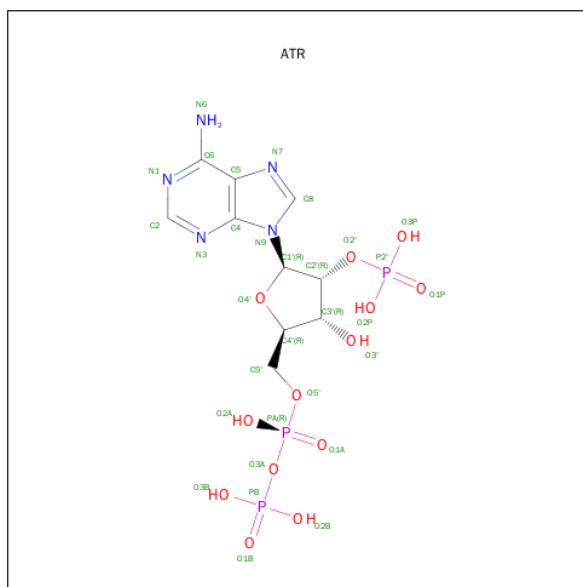
*Continued on next page...*



Continued from previous page...

Chain	Residue	Modelled	Actual	Comment	Reference
C	0	HIS	-	EXPRESSION TAG	UNP O67049
C	195	GLU	LYS	SEE REMARK 999	UNP O67049
C	233	LEU	PHE	SEE REMARK 999	UNP O67049
D	-5	HIS	-	EXPRESSION TAG	UNP O67049
D	-4	HIS	-	EXPRESSION TAG	UNP O67049
D	-3	HIS	-	EXPRESSION TAG	UNP O67049
D	-2	HIS	-	EXPRESSION TAG	UNP O67049
D	-1	HIS	-	EXPRESSION TAG	UNP O67049
D	0	HIS	-	EXPRESSION TAG	UNP O67049
D	195	GLU	LYS	SEE REMARK 999	UNP O67049
D	233	LEU	PHE	SEE REMARK 999	UNP O67049

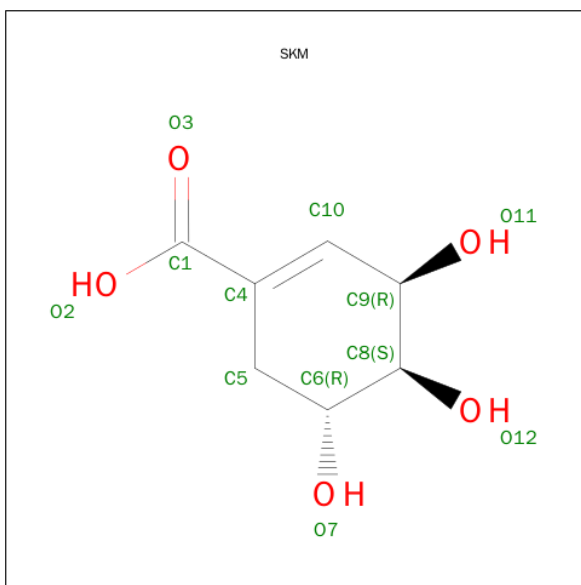
- Molecule 2 is 2'-MONOPHOSPHOADENOSINE-5'-DIPHOSPHATE (three-letter code: ATR) (formula: C<sub>10</sub>H<sub>16</sub>N<sub>5</sub>O<sub>13</sub>P<sub>3</sub>).



Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			31	10	5	13	3		

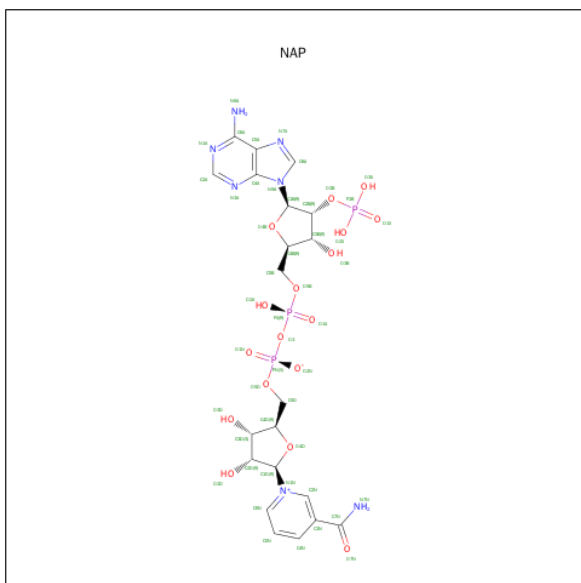
- Molecule 3 is (3R,4S,5R)-3,4,5-TRIHIDROXYCYCLOHEX-1-ENE-1-CARBOXYLIC ACID (three-letter code: SKM) (formula: C<sub>7</sub>H<sub>10</sub>O<sub>5</sub>).





Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			12	7	5		
3	B	1	Total	C	O	0	0
			12	7	5		
3	C	1	Total	C	O	0	0
			12	7	5		
3	D	1	Total	C	O	0	0
			12	7	5		

- Molecule 4 is NADP NICOTINAMIDE-ADENINE-DINUCLEOTIDE PHOSPHATE (three-letter code: NAP) (formula:  $C_{21}H_{28}N_7O_{17}P_3$ ).





Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
4	B	1	Total 48	C 21	N 7	O 17	P 3	0	0
4	C	1	Total 48	C 21	N 7	O 17	P 3	0	0
4	D	1	Total 48	C 21	N 7	O 17	P 3	0	0

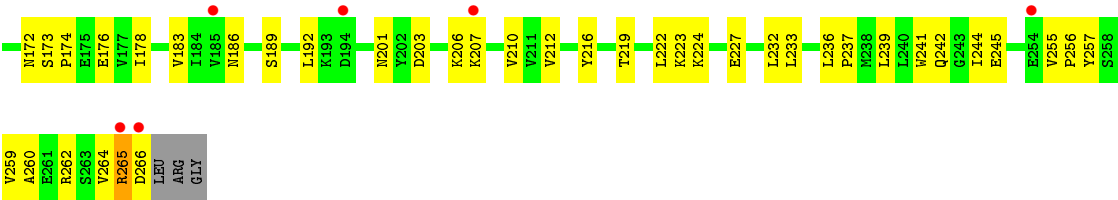
- Molecule 5 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
5	A	56	Total 56	O 56	0	0
5	B	46	Total 46	O 46	0	0
5	C	56	Total 56	O 56	0	0
5	D	46	Total 46	O 46	0	0

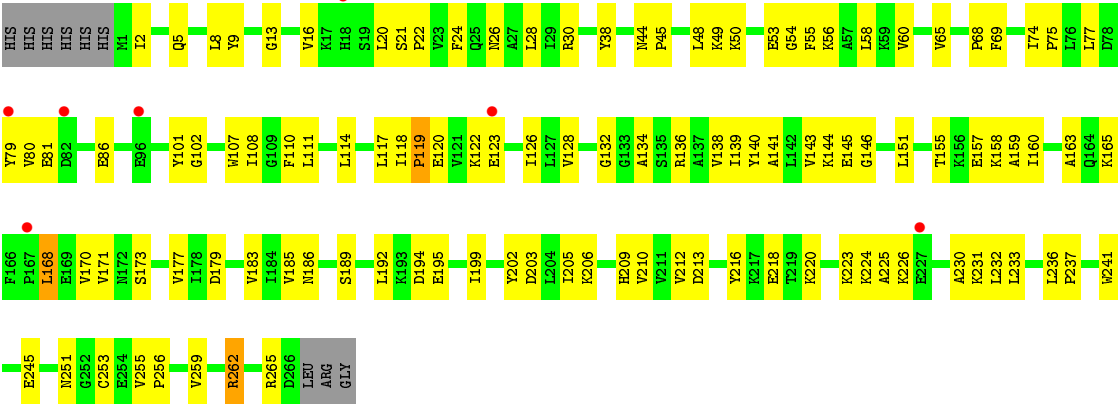








● Molecule 1: Shikimate dehydrogenase





## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	72.00Å 119.80Å 128.87Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	29.41 – 2.20 30.01 – 2.20	Depositor EDS
% Data completeness (in resolution range)	82.3 (29.41-2.20) 82.5 (30.01-2.20)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.97 (at 2.20Å)	Xtriage
Refinement program	CNS 1.0	Depositor
R, $R_{free}$	0.207 , 0.271 0.209 , 0.272	Depositor DCC
$R_{free}$ test set	2364 reflections (5.02%)	DCC
Wilson B-factor (Å <sup>2</sup> )	37.0	Xtriage
Anisotropy	0.411	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.34 , 60.9	EDS
Estimated twinning fraction	No twinning to report.	Xtriage
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.48$ , $\langle L^2 \rangle = 0.31$	Xtriage
Outliers	0 of 47934 reflections	Xtriage
$F_o, F_c$ correlation	0.95	EDS
Total number of atoms	8871	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	50.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.40% of the height of the origin peak. No significant pseudotranslation is detected.*

<sup>1</sup>Intensities estimated from amplitudes.

<sup>2</sup>Theoretical values of  $\langle |L| \rangle$ ,  $\langle L^2 \rangle$  for acentric reflections are 0.5, 0.375 respectively for untwinned datasets, and 0.333, 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: SKM, NAP, ATR

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	0/2169	0.63	1/2930 (0.0%)
1	B	0.38	0/2153	0.61	0/2911
1	C	0.38	0/2145	0.63	0/2900
1	D	0.37	0/2145	0.61	0/2900
All	All	0.38	0/8612	0.62	1/11641 (0.0%)

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	168	LEU	CA-CB-CG	5.04	126.89	115.30

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	2127	0	2222	67	0
1	B	2111	0	2206	80	0
1	C	2103	0	2195	79	0
1	D	2103	0	2195	81	0
2	A	31	0	11	1	0
3	A	12	0	9	0	0

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	B	12	0	9	0	0
3	C	12	0	9	0	0
3	D	12	0	9	0	0
4	B	48	0	25	2	0
4	C	48	0	25	0	0
4	D	48	0	25	2	0
5	A	56	0	0	0	0
5	B	46	0	0	2	0
5	C	56	0	0	3	0
5	D	46	0	0	2	0
All	All	8871	0	8940	293	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 17.

The worst 5 of 293 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:118:ILE:HD11	1:D:210:VAL:HG11	1.34	1.04
1:C:28:LEU:HD22	1:C:255:VAL:HG11	1.40	1.03
1:D:111:LEU:HD13	1:D:141:ALA:HB1	1.41	1.00
1:C:206:LYS:HG2	1:C:207:LYS:N	1.91	0.86
1:C:206:LYS:HG2	1:C:207:LYS:H	1.41	0.84

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	267/275 (97%)	257 (96%)	9 (3%)	1 (0%)	39	42
1	B	265/275 (96%)	248 (94%)	15 (6%)	2 (1%)	24	22

*Continued on next page...*



*Continued from previous page...*

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	C	264/275 (96%)	251 (95%)	11 (4%)	2 (1%)	24	22
1	D	264/275 (96%)	241 (91%)	21 (8%)	2 (1%)	24	22
All	All	1060/1100 (96%)	997 (94%)	56 (5%)	7 (1%)	26	25

5 of 7 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	209	HIS
1	C	265	ARG
1	D	119	PRO
1	A	16	VAL
1	B	16	VAL

### 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	229/235 (97%)	216 (94%)	13 (6%)	25	29
1	B	228/235 (97%)	221 (97%)	7 (3%)	47	59
1	C	227/235 (97%)	221 (97%)	6 (3%)	54	66
1	D	227/235 (97%)	217 (96%)	10 (4%)	35	42
All	All	911/940 (97%)	875 (96%)	36 (4%)	38	47

5 of 36 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	B	68	PRO
1	C	8	LEU
1	D	194	ASP
1	B	212	VAL
1	C	30	ARG

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 11 such sidechains are listed below:



Mol	Chain	Res	Type
1	B	5	GLN
1	B	201	ASN
1	C	242	GLN
1	A	251	ASN
1	C	201	ASN

### 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 carbohydrates in this entry.

### 5.6 Ligand geometry ⓘ

8 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	ATR	A	1101	-	26,33,33	2.03	7 (26%)	34,52,52	2.49	10 (29%)
3	SKM	A	1102	-	10,12,12	2.65	5 (50%)	9,17,17	1.37	2 (22%)
4	NAP	B	1201	-	42,52,52	1.36	5 (11%)	54,80,80	1.22	4 (7%)
3	SKM	B	1202	-	10,12,12	2.60	5 (50%)	9,17,17	1.36	2 (22%)
4	NAP	C	1301	-	42,52,52	1.33	5 (11%)	54,80,80	1.19	4 (7%)
3	SKM	C	1302	-	10,12,12	2.43	5 (50%)	9,17,17	1.35	2 (22%)
4	NAP	D	1401	-	42,52,52	1.38	5 (11%)	54,80,80	1.20	5 (9%)
3	SKM	D	1402	-	10,12,12	2.41	5 (50%)	9,17,17	1.31	2 (22%)



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	ATR	A	1101	-	-	0/17/37/37	0/3/3/3
3	SKM	A	1102	-	-	0/0/20/20	0/1/1/1
4	NAP	B	1201	-	-	0/27/67/67	0/5/5/5
3	SKM	B	1202	-	-	0/0/20/20	0/1/1/1
4	NAP	C	1301	-	-	0/27/67/67	0/5/5/5
3	SKM	C	1302	-	-	0/0/20/20	0/1/1/1
4	NAP	D	1401	-	-	0/27/67/67	0/5/5/5
3	SKM	D	1402	-	-	0/0/20/20	0/1/1/1

The worst 5 of 42 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	A	1102	SKM	C1-C4	-3.28	1.45	1.51
3	D	1402	SKM	C1-C4	-3.21	1.45	1.51
3	C	1302	SKM	C1-C4	-2.74	1.46	1.51
3	B	1202	SKM	C1-C4	-2.64	1.46	1.51
2	A	1101	ATR	C3'-C2'	-2.58	1.47	1.53

The worst 5 of 31 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1101	ATR	N3-C2-N1	-8.31	122.53	128.89
2	A	1101	ATR	C1'-N9-C4	-7.37	115.82	126.94
4	B	1201	NAP	N3A-C2A-N1A	-4.23	125.65	128.89
4	C	1301	NAP	N3A-C2A-N1A	-4.17	125.70	128.89
4	D	1401	NAP	N3A-C2A-N1A	-3.78	126.00	128.89

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

3 monomers are involved in 5 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1101	ATR	1	0
4	B	1201	NAP	2	0
4	D	1401	NAP	2	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](#)

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<sup>th</sup> 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 <sup>2</sup> )	Q<0.9
1	A	269/275 (97%)	0.03	8 (2%) 54 53	26, 43, 71, 89	0
1	B	267/275 (97%)	0.23	9 (3%) 49 47	26, 47, 77, 95	0
1	C	266/275 (96%)	0.27	18 (6%) 20 20	31, 46, 71, 99	0
1	D	266/275 (96%)	0.18	7 (2%) 59 58	29, 48, 76, 98	0
All	All	1068/1100 (97%)	0.18	42 (3%) 43 42	26, 46, 75, 99	0

The worst 5 of 42 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	A	269	GLY	4.4
1	B	160	ILE	4.4
1	D	79	TYR	4.2
1	C	265	ARG	3.9
1	C	82	ASP	3.8

### 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 carbohydrates 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. 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<sup>th</sup> 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( $\text{\AA}^2$ )	Q<0.9
4	NAP	C	1301	48/48	0.87	0.17	0.24	58,71,89,89	0
3	SKM	C	1302	12/12	0.91	0.16	0.14	32,41,52,53	0
3	SKM	D	1402	12/12	0.95	0.14	0.10	33,39,45,50	0
3	SKM	B	1202	12/12	0.92	0.17	0.05	32,49,55,58	0
3	SKM	A	1102	12/12	0.94	0.13	-0.22	47,53,56,60	0
4	NAP	B	1201	48/48	0.89	0.14	-0.23	48,58,72,77	0
2	ATR	A	1101	31/31	0.93	0.12	-0.31	34,54,89,93	0
4	NAP	D	1401	48/48	0.96	0.11	-0.77	37,46,57,59	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.