



# wwPDB X-ray Structure Validation Summary Report ⓘ

Feb 27, 2019 – 01:36 PM EST

PDB ID : 2V1S  
Title : CRYSTAL STRUCTURE OF RAT TOM20-ALDH PRESEQUENCE COMPLEX  
Authors : Obita, T.; Igura, M.; Ose, T.; Endo, T.; Maenaka, K.; Kohda, D.  
Deposited on : 2007-05-29  
Resolution : 2.05 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

<https://www.wwpdb.org/validation/2017/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 : 1.7.3 (157068), CSD as539be (2018)  
Xtriage (Phenix) : 1.13  
EDS : rb-20031633  
Percentile statistics : 20171227.v01 (using entries in the PDB archive December 27th 2017)  
Refmac : 5.8.0158  
CCP4 : 7.0 (Gargrove)  
Ideal geometry (proteins) : Engh & Huber (2001)  
Ideal geometry (DNA, RNA) : Parkinson et al. (1996)  
Validation Pipeline (wwPDB-VP) : rb-20031633

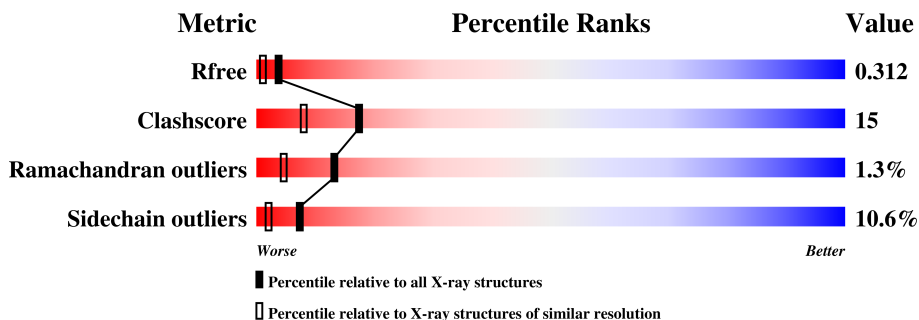
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 2.05 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))
$R_{free}$	111664	1449 (2.04-2.04)
Clashscore	122126	1524 (2.04-2.04)
Ramachandran outliers	120053	1512 (2.04-2.04)
Sidechain outliers	120020	1512 (2.04-2.04)

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  $\leq 5\%$ .

Mol	Chain	Length	Quality of chain
1	A	73	77% 11% • 10%
1	B	73	59% 23% 8% 10%
1	C	73	67% 23% • 5%
1	D	73	58% 27% 7% 8%
1	E	73	66% 25% 7% •
1	F	73	62% 25% • 12%
1	G	73	11% 7% • 81%

*Continued on next page...*

Continued from previous page...

Mol	Chain	Length	Quality of chain
2	H	13	<div><div></div><div></div><div></div><div></div></div> <div>54%38%8%</div>
2	I	13	<div><div></div><div></div><div></div><div></div></div> <div>54%23%8%15%</div>
2	J	13	<div><div></div><div></div><div></div><div></div></div> <div>69%31%</div>
2	K	13	<div><div></div><div></div><div></div><div></div></div> <div>46%38%15%</div>
2	L	13	<div><div></div><div></div><div></div><div></div></div> <div>54%31%15%</div>
2	M	13	<div><div></div><div></div><div></div><div></div></div> <div>31%15%8%46%</div>
2	N	13	<div><div></div><div></div><div></div><div></div></div> <div>46%15%8%31%</div>

## 2 Entry composition

There are 3 unique types of molecules in this entry. The entry contains 4232 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 MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM20 HOMOLOG.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	66	Total	C	N	O	S	0	0	0
			519	334	83	100	2			
1	B	66	Total	C	N	O	S	0	0	0
			519	334	83	100	2			
1	C	69	Total	C	N	O	S	0	0	0
			541	347	86	106	2			
1	D	67	Total	C	N	O	S	0	0	0
			520	333	83	102	2			
1	E	71	Total	C	N	O	S	0	0	0
			553	355	88	108	2			
1	F	64	Total	C	N	O	S	0	0	0
			502	324	80	96	2			
1	G	14	Total	C	N	O		0	0	0
			117	78	17	22				

- Molecule 2 is a protein called ALDEHYDE DEHYDROGENASE.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	H	12	Total	C	N	O	S	0	0	0
			93	58	19	15	1			
2	I	11	Total	C	N	O	S	0	0	0
			86	53	18	14	1			
2	J	13	Total	C	N	O	S	0	0	0
			97	60	20	16	1			
2	K	11	Total	C	N	O	S	0	0	0
			86	53	18	14	1			
2	L	13	Total	C	N	O	S	0	0	0
			97	60	20	16	1			
2	M	7	Total	C	N	O	S	0	0	0
			50	32	8	9	1			
2	N	9	Total	C	N	O		0	0	0
			75	48	15	12				

There are 21 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
H	21	TYR	ALA	engineered mutation	UNP P11884
H	23	GLY	ALA	engineered mutation	UNP P11884
H	24	CY3	THR	engineered mutation	UNP P11884
I	21	TYR	ALA	engineered mutation	UNP P11884
I	23	GLY	ALA	engineered mutation	UNP P11884
I	24	CY3	THR	engineered mutation	UNP P11884
J	21	TYR	ALA	engineered mutation	UNP P11884
J	23	GLY	ALA	engineered mutation	UNP P11884
J	24	CY3	THR	engineered mutation	UNP P11884
K	21	TYR	ALA	engineered mutation	UNP P11884
K	23	GLY	ALA	engineered mutation	UNP P11884
K	24	CY3	THR	engineered mutation	UNP P11884
L	21	TYR	ALA	engineered mutation	UNP P11884
L	23	GLY	ALA	engineered mutation	UNP P11884
L	24	CY3	THR	engineered mutation	UNP P11884
M	21	TYR	ALA	engineered mutation	UNP P11884
M	23	GLY	ALA	engineered mutation	UNP P11884
M	24	CY3	THR	engineered mutation	UNP P11884
N	21	TYR	ALA	engineered mutation	UNP P11884
N	23	GLY	ALA	engineered mutation	UNP P11884
N	24	CY3	THR	engineered mutation	UNP P11884

- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	70	Total O 70 70	0	0
3	B	50	Total O 50 50	0	0
3	C	57	Total O 57 57	0	0
3	D	41	Total O 41 41	0	0
3	E	69	Total O 69 69	0	0
3	F	45	Total O 45 45	0	0
3	G	5	Total O 5 5	0	0
3	H	7	Total O 7 7	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	I	3	Total 3	O 3	0	0
3	J	9	Total 9	O 9	0	0
3	K	8	Total 8	O 8	0	0
3	L	7	Total 7	O 7	0	0
3	M	2	Total 2	O 2	0	0
3	N	4	Total 4	O 4	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 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: MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM20 HOMOLOG

Chain A: 



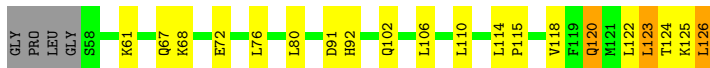
- Molecule 1: MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM20 HOMOLOG

Chain B: 



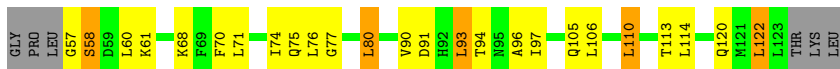
- Molecule 1: MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM20 HOMOLOG

Chain C: 



- Molecule 1: MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM20 HOMOLOG

Chain D: 



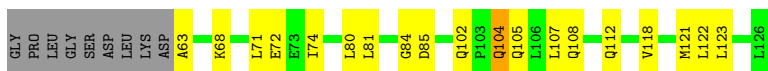
- Molecule 1: MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM20 HOMOLOG

Chain E: 



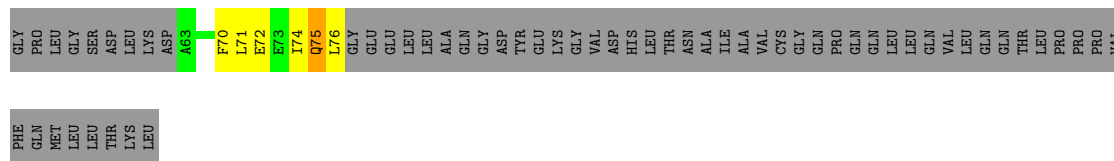
- Molecule 1: MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM20 HOMOLOG

Chain F: 



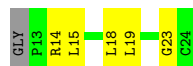
- Molecule 1: MITOCHONDRIAL IMPORT RECEPTOR SUBUNIT TOM20 HOMOLOG

Chain G:  11% 7% 81%



- Molecule 2: ALDEHYDE DEHYDROGENASE

Chain H:  54% 38% 8%



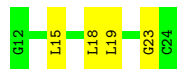
- Molecule 2: ALDEHYDE DEHYDROGENASE

Chain I:  54% 23% 8% 15%



- Molecule 2: ALDEHYDE DEHYDROGENASE

Chain J:  69% 31%



- Molecule 2: ALDEHYDE DEHYDROGENASE

Chain K:  46% 38% 15%

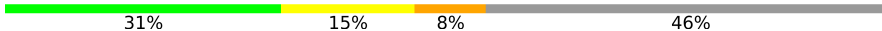


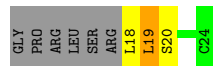
- Molecule 2: ALDEHYDE DEHYDROGENASE

Chain L:  54% 31% 15%



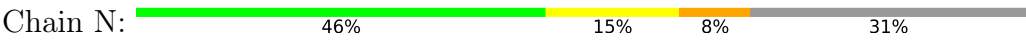
- Molecule 2: ALDEHYDE DEHYDROGENASE

Chain M:  31% 15% 8% 46%



- Molecule 2: ALDEHYDE DEHYDROGENASE





## 4 Data and refinement statistics

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	151.78Å 64.15Å 68.02Å 90.00° 94.70° 90.00°	Depositor
Resolution (Å)	20.00 – 2.05 33.89 – 2.05	Depositor EDS
% Data completeness (in resolution range)	93.1 (20.00-2.05) 93.1 (33.89-2.05)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.95 (at 2.05Å)	Xtriage
Refinement program	REFMAC 5.2.0019	Depositor
R, $R_{free}$	0.248 , 0.308 0.251 , 0.312	Depositor DCC
$R_{free}$ test set	3635 reflections (9.54%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	42.0	Xtriage
Anisotropy	0.420	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.29 , 53.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.33$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.90	EDS
Total number of atoms	4232	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	54.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 5.00% 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, 0.333 respectively for untwinned datasets, and 0.375, 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: CY3

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	0.54	0/527	0.57	0/713
1	B	0.53	0/527	0.61	0/713
1	C	0.52	0/549	0.62	0/743
1	D	0.47	0/528	0.53	0/716
1	E	0.56	0/561	0.66	0/759
1	F	0.48	0/510	0.59	0/691
1	G	0.44	0/118	0.53	0/157
2	H	0.48	0/87	0.70	0/115
2	I	0.50	0/79	0.74	0/104
2	J	0.56	0/91	0.52	0/121
2	K	0.49	0/79	0.69	0/104
2	L	0.53	0/91	0.67	0/121
2	M	0.60	0/43	0.71	0/57
2	N	0.41	0/75	0.63	0/99
All	All	0.52	0/3865	0.61	0/5213

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
2	H	0	1
2	I	0	2
2	J	0	1
2	K	0	1
2	L	0	2
All	All	0	7

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

5 of 7 planarity outliers are listed below:

Mol	Chain	Res	Type	Group
2	H	23	GLY	Mainchain
2	I	23	GLY	Mainchain,Peptide
2	J	23	GLY	Mainchain
2	K	23	GLY	Peptide
2	L	23	GLY	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	519	0	526	6	0
1	B	519	0	526	22	0
1	C	541	0	546	20	0
1	D	520	0	518	20	0
1	E	553	0	560	18	0
1	F	502	0	509	16	0
1	G	117	0	116	5	0
2	H	93	0	98	2	0
2	I	86	0	90	3	0
2	J	97	0	100	2	0
2	K	86	0	90	5	0
2	L	97	0	100	4	0
2	M	50	0	48	4	0
2	N	75	0	82	4	0
3	A	70	0	0	2	0
3	B	50	0	0	2	0
3	C	57	0	0	3	0
3	D	41	0	0	1	0
3	E	69	0	0	1	0
3	F	45	0	0	3	0
3	G	5	0	0	1	0
3	H	7	0	0	0	0
3	I	3	0	0	0	0
3	J	9	0	0	0	0

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
3	K	8	0	0	0	0
3	L	7	0	0	0	0
3	M	2	0	0	0	0
3	N	4	0	0	0	0
All	All	4232	0	3909	115	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 15.

The worst 5 of 115 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:F:74:ILE:HG23	2:N:15:LEU:HD23	1.48	0.94
1:E:72:GLU:O	1:E:76:LEU:HD13	1.72	0.90
1:B:76:LEU:CD1	1:B:80:LEU:HD22	2.01	0.89
1:A:67:GLN:NE2	3:A:2009:HOH:O	2.13	0.82
2:M:18:LEU:HD12	2:M:19:LEU:HD12	1.63	0.81

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	64/73 (88%)	60 (94%)	3 (5%)	1 (2%)	11	3
1	B	64/73 (88%)	63 (98%)	1 (2%)	0	100	100
1	C	67/73 (92%)	64 (96%)	3 (4%)	0	100	100
1	D	65/73 (89%)	61 (94%)	2 (3%)	2 (3%)	4	0
1	E	69/73 (94%)	67 (97%)	2 (3%)	0	100	100
1	F	62/73 (85%)	60 (97%)	2 (3%)	0	100	100

*Continued on next page...*

Continued from previous page...

Mol	Chain	Analysed	Favoured	Allowed	Outliers	Percentiles	
1	G	12/73 (16%)	10 (83%)	1 (8%)	1 (8%)	1	0
2	H	10/13 (77%)	10 (100%)	0	0	100	100
2	I	9/13 (69%)	9 (100%)	0	0	100	100
2	J	11/13 (85%)	10 (91%)	1 (9%)	0	100	100
2	K	9/13 (69%)	9 (100%)	0	0	100	100
2	L	11/13 (85%)	10 (91%)	0	1 (9%)	1	0
2	M	5/13 (38%)	4 (80%)	0	1 (20%)	0	0
2	N	7/13 (54%)	7 (100%)	0	0	100	100
All	All	465/602 (77%)	444 (96%)	15 (3%)	6 (1%)	13	4

5 of 6 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	63	ALA
1	D	58	SER
1	G	75	GLN
2	L	13	PRO
2	M	19	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	57/62 (92%)	54 (95%)	3 (5%)	25	15
1	B	57/62 (92%)	49 (86%)	8 (14%)	4	1
1	C	60/62 (97%)	55 (92%)	5 (8%)	12	5
1	D	57/62 (92%)	49 (86%)	8 (14%)	4	1
1	E	61/62 (98%)	54 (88%)	7 (12%)	6	1
1	F	55/62 (89%)	48 (87%)	7 (13%)	5	1
1	G	12/62 (19%)	11 (92%)	1 (8%)	12	5
2	H	9/9 (100%)	8 (89%)	1 (11%)	7	2

Continued on next page...

*Continued from previous page...*

Mol	Chain	Analysed	Rotameric	Outliers	Percentiles	
2	I	8/9 (89%)	6 (75%)	2 (25%)	0	0
2	J	9/9 (100%)	9 (100%)	0	100	100
2	K	8/9 (89%)	8 (100%)	0	100	100
2	L	9/9 (100%)	8 (89%)	1 (11%)	7	2
2	M	4/9 (44%)	4 (100%)	0	100	100
2	N	8/9 (89%)	7 (88%)	1 (12%)	5	1
All	All	414/497 (83%)	370 (89%)	44 (11%)	7	2

5 of 44 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	D	80	LEU
1	E	81	LEU
2	I	14	ARG
1	D	93	LEU
1	D	110	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 7 such sidechains are listed below:

Mol	Chain	Res	Type
1	C	102	GLN
1	G	67	GLN
1	E	75	GLN
1	B	108	GLN
1	E	108	GLN

### 5.3.3 RNA ⓘ

There are no RNA molecules in this entry.

## 5.4 Non-standard residues in protein, DNA, RNA chains ⓘ

6 non-standard protein/DNA/RNA residues 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	CY3	H	24	1,2	6,6,6	0.55	0	6,7,7	0.94	0
2	CY3	I	24	1,2	6,6,6	0.51	0	6,7,7	1.11	0
2	CY3	J	24	1,2	6,6,6	0.77	0	6,7,7	1.02	0
2	CY3	K	24	1,2	6,6,6	0.54	0	6,7,7	0.73	0
2	CY3	L	24	1,2	6,6,6	0.43	0	6,7,7	1.62	1 (16%)
2	CY3	M	24	1,2	6,6,6	0.36	0	6,7,7	0.36	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
2	CY3	H	24	1,2	-	0/5/6/6	0/0/0/0
2	CY3	I	24	1,2	-	0/5/6/6	0/0/0/0
2	CY3	J	24	1,2	-	0/5/6/6	0/0/0/0
2	CY3	K	24	1,2	-	0/5/6/6	0/0/0/0
2	CY3	L	24	1,2	-	0/5/6/6	0/0/0/0
2	CY3	M	24	1,2	-	0/5/6/6	0/0/0/0

There are no bond length outliers.

All (1) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	L	24	CY3	CB-CA-C	3.18	116.35	109.64

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

2 monomers are involved in 4 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	K	24	CY3	3	0
2	L	24	CY3	1	0



## 5.5 Carbohydrates [i](#)

There are no carbohydrates in this entry.

## 5.6 Ligand geometry [i](#)

There are no ligands in this entry.

## 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

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.