



# wwPDB X-ray Structure Validation Summary Report ⓘ

Aug 29, 2023 – 11:53 AM EDT

PDB ID : 3MSJ  
Title : Structure of bace (beta secretase) in complex with inhibitor  
Authors : Madden, J.; Kramer, J.; Smith, M.A.; Barker, J.; Godemann, R.  
Deposited on : 2010-04-29  
Resolution : 1.80 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 types of validation reports are described at

<http://www.wwpdb.org/validation/2017/FAQs#types>.

---

The following versions of software and data (see [references ⓘ](#)) were used in the production of this report:

MolProbity : 4.02b-467  
Mogul : 1.8.5 (274361), CSD as541be (2020)  
Xtrriage (Phenix) : 1.13  
EDS : 2.35  
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) : 2.35

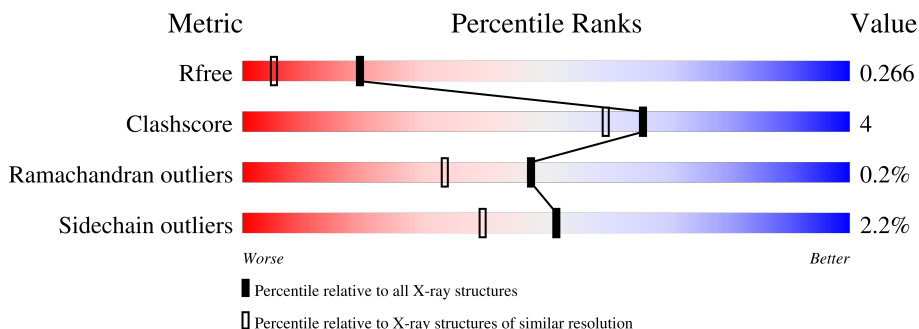
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

*X-RAY DIFFRACTION*

The reported resolution of this entry is 1.80 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	5950 (1.80-1.80)
Clashscore	141614	6793 (1.80-1.80)
Ramachandran outliers	138981	6697 (1.80-1.80)
Sidechain outliers	138945	6696 (1.80-1.80)

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  $\geq 3$ , 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  $\leq 5\%$

Mol	Chain	Length	Quality of chain
1	A	411	
1	B	411	
1	C	411	

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
3	GOL	B	394	-	-	X	-

## 2 Entry composition

There are 4 unique types of molecules in this entry. The entry contains 9702 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.

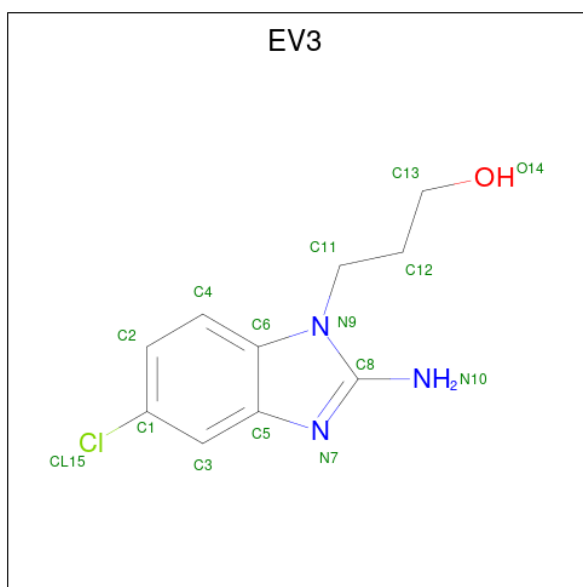
- Molecule 1 is a protein called BETA-SECRETASE 1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
			Total	C	N	O	S			
1	A	373	2962	1897	493	558	14	0	4	0
1	B	373	2956	1894	493	555	14	0	4	0
1	C	375	2987	1912	494	567	14	0	6	0

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	-5	LYS	ARG	engineered mutation	UNP P56817
A	-4	THR	ARG	engineered mutation	UNP P56817
B	-5	LYS	ARG	engineered mutation	UNP P56817
B	-4	THR	ARG	engineered mutation	UNP P56817
C	-5	LYS	ARG	engineered mutation	UNP P56817
C	-4	THR	ARG	engineered mutation	UNP P56817

- Molecule 2 is 3-(2-amino-5-chloro-1H-benzimidazol-1-yl)propan-1-ol (three-letter code: EV3) (formula: C<sub>10</sub>H<sub>12</sub>ClN<sub>3</sub>O).



Mol	Chain	Residues	Atoms				ZeroOcc	AltConf	
			Total	C	Cl	N			O
2	A	1	Total	C	Cl	N	O	0	0
			15	10	1	3	1		
2	B	1	Total	C	Cl	N	O	0	0
			15	10	1	3	1		
2	C	1	Total	C	Cl	N	O	0	0
			15	10	1	3	1		

- Molecule 3 is GLYCEROL (three-letter code: GOL) (formula: C<sub>3</sub>H<sub>8</sub>O<sub>3</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
			Total	C	O		
3	A	1	Total	C	O	0	0
			6	3	3		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0
3	A	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0
3	B	1	Total C O 6 3 3	0	0
3	C	1	Total C O 6 3 3	0	0
3	C	1	Total C O 6 3 3	0	0


- Molecule 4 is water.

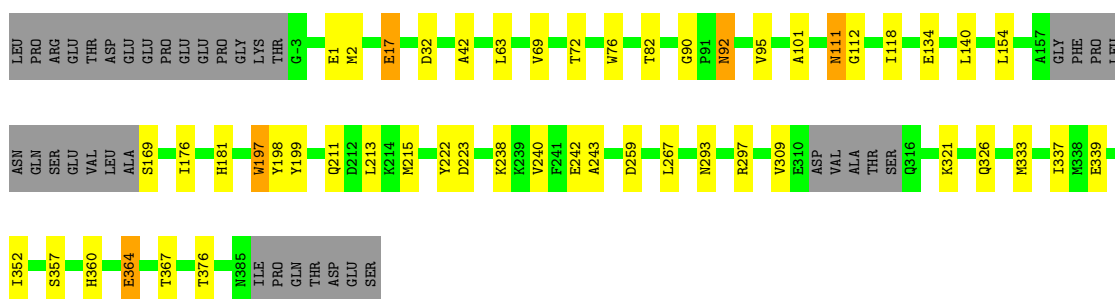
Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
4	A	236	Total O 236 236	0	0
4	B	243	Total O 246 246	0	3
4	C	167	Total O 168 168	0	1

### 3 Residue-property plots


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.

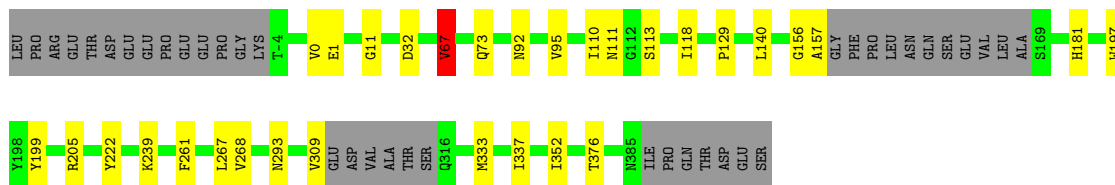
- Molecule 1: BETA-SECRETASE 1

Chain A: 




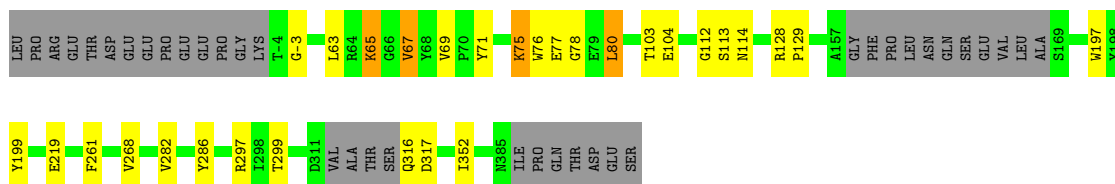
- Molecule 1: BETA-SECRETASE 1

Chain B: 



- Molecule 1: BETA-SECRETASE 1

Chain C: 



## 4 Data and refinement statistics i

Property	Value	Source
Space group	C 1 2 1	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	229.20Å 98.83Å 63.13Å 90.00° 103.36° 90.00°	Depositor
Resolution (Å)	111.80 – 1.80 24.03 – 1.80	Depositor EDS
% Data completeness (in resolution range)	94.4 (111.80-1.80) 94.4 (24.03-1.80)	Depositor EDS
$R_{merge}$	0.05	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.39 (at 1.80Å)	Xtrriage
Refinement program	REFMAC 5.2.0005	Depositor
R, $R_{free}$	0.177 , 0.208 0.247 , 0.266	Depositor DCC
$R_{free}$ test set	1212 reflections (1.01%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	30.4	Xtrriage
Anisotropy	0.295	Xtrriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.38 , 44.1	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.49$ , $\langle L^2 \rangle = 0.32$	Xtrriage
Estimated twinning fraction	0.024 for -h-2*1,-k,l	Xtrriage
$F_o, F_c$ correlation	0.94	EDS
Total number of atoms	9702	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	26.0	wwPDB-VP

Xtrriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 4.45% 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 [i](#)

### 5.1 Standard geometry [i](#)

Bond lengths and bond angles in the following residue types are not validated in this section: EV3, GOL

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	0.55	0/3046	0.69	2/4136 (0.0%)
1	B	0.55	0/3039	0.71	1/4127 (0.0%)
1	C	0.50	0/3077	0.65	0/4179
All	All	0.53	0/9162	0.68	3/12442 (0.0%)

There are no bond length outliers.

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	B	67	VAL	CB-CA-C	-6.05	99.90	111.40
1	A	297	ARG	NE-CZ-NH1	5.74	123.17	120.30
1	A	297	ARG	NE-CZ-NH2	-5.10	117.75	120.30

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	2962	0	2875	38	0
1	B	2956	0	2882	21	0
1	C	2987	0	2897	15	0
2	A	15	0	12	3	0
2	B	15	0	12	1	0

*Continued on next page...*



Continued from previous page...

Mol	Chain	Non-H	H(model)	H(added)	Clashes	Symm-Clashes
2	C	15	0	12	0	0
3	A	42	0	56	2	0
3	B	48	0	64	7	0
3	C	12	0	16	0	0
4	A	236	0	0	3	0
4	B	246	0	0	0	0
4	C	168	0	0	0	0
All	All	9702	0	8826	72	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 4.

The worst 5 of 72 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:267:LEU:HD23	1:A:321:LYS:HG3	1.60	0.82
1:C:282:VAL:HG13	1:C:299:THR:HG23	1.68	0.75
1:A:215:MET:HE1	1:A:243:ALA:HB3	1.72	0.69
1:B:0:VAL:H	3:B:394:GOL:H31	1.63	0.63
1:A:309:VAL:HG21	1:A:321:LYS:HD2	1.83	0.61

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	371/411 (90%)	362 (98%)	9 (2%)	0	100 100
1	B	371/411 (90%)	364 (98%)	7 (2%)	0	100 100
1	C	375/411 (91%)	364 (97%)	9 (2%)	2 (0%)	29 15
All	All	1117/1233 (91%)	1090 (98%)	25 (2%)	2 (0%)	47 33

All (2) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	C	-3	GLY
1	C	113	SER

### 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	322/352 (92%)	313 (97%)	9 (3%)	43 30
1	B	322/352 (92%)	317 (98%)	5 (2%)	62 54
1	C	326/352 (93%)	319 (98%)	7 (2%)	53 42
All	All	970/1056 (92%)	949 (98%)	21 (2%)	52 39

5 of 21 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	C	65	LYS
1	C	80	LEU
1	C	317	ASP
1	C	197	TRP
1	C	75	LYS

Sometimes sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 20 such sidechains are listed below:

Mol	Chain	Res	Type
1	B	362	HIS
1	C	73	GLN
1	C	362	HIS
1	C	92	ASN
1	A	266	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](#)

20 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	GOL	A	397	-	5,5,5	0.32	0	5,5,5	0.33	0
3	GOL	B	397	-	5,5,5	0.30	0	5,5,5	0.40	0
3	GOL	A	400	-	5,5,5	0.39	0	5,5,5	0.16	0
2	EV3	C	393	-	13,16,16	1.53	2 (15%)	15,22,22	0.94	0
3	GOL	A	399	-	5,5,5	0.40	0	5,5,5	0.73	0
3	GOL	B	401	-	5,5,5	0.35	0	5,5,5	0.49	0
3	GOL	B	395	-	5,5,5	0.35	0	5,5,5	0.29	0
3	GOL	B	396	-	5,5,5	0.43	0	5,5,5	0.54	0
3	GOL	A	398	-	5,5,5	0.37	0	5,5,5	0.36	0
2	EV3	A	393	-	13,16,16	1.50	2 (15%)	15,22,22	1.02	1 (6%)
2	EV3	B	393	-	13,16,16	1.53	2 (15%)	15,22,22	1.03	0
3	GOL	B	399	-	5,5,5	0.38	0	5,5,5	0.51	0
3	GOL	B	400	-	5,5,5	0.34	0	5,5,5	0.56	0
3	GOL	B	398	-	5,5,5	0.38	0	5,5,5	0.37	0
3	GOL	B	394	-	5,5,5	0.70	0	5,5,5	1.56	2 (40%)
3	GOL	A	394	-	5,5,5	0.38	0	5,5,5	0.24	0
3	GOL	A	395	-	5,5,5	0.63	0	5,5,5	0.45	0
3	GOL	C	394	-	5,5,5	0.43	0	5,5,5	0.46	0
3	GOL	C	395	-	5,5,5	0.41	0	5,5,5	0.33	0
3	GOL	A	396	-	5,5,5	0.53	0	5,5,5	0.57	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	GOL	A	397	-	-	0/4/4/4	-
3	GOL	B	397	-	-	2/4/4/4	-
3	GOL	A	400	-	-	0/4/4/4	-
2	EV3	C	393	-	-	1/4/4/4	0/2/2/2
3	GOL	A	399	-	-	2/4/4/4	-
3	GOL	B	401	-	-	0/4/4/4	-
3	GOL	B	395	-	-	2/4/4/4	-
3	GOL	B	396	-	-	2/4/4/4	-
3	GOL	A	398	-	-	3/4/4/4	-
2	EV3	A	393	-	-	1/4/4/4	0/2/2/2
2	EV3	B	393	-	-	1/4/4/4	0/2/2/2
3	GOL	B	399	-	-	2/4/4/4	-
3	GOL	B	400	-	-	2/4/4/4	-
3	GOL	B	398	-	-	4/4/4/4	-
3	GOL	B	394	-	-	2/4/4/4	-
3	GOL	A	394	-	-	3/4/4/4	-
3	GOL	A	395	-	-	4/4/4/4	-
3	GOL	C	394	-	-	2/4/4/4	-
3	GOL	C	395	-	-	1/4/4/4	-
3	GOL	A	396	-	-	0/4/4/4	-

The worst 5 of 6 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	393	EV3	C8-N10	3.35	1.40	1.33
2	C	393	EV3	C8-N10	3.25	1.40	1.33
2	A	393	EV3	C8-N10	3.07	1.40	1.33
2	A	393	EV3	C3-C5	-2.52	1.37	1.41
2	C	393	EV3	C3-C5	-2.40	1.38	1.41

All (3) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
3	B	394	GOL	O1-C1-C2	2.37	121.57	110.20
3	B	394	GOL	O3-C3-C2	2.16	120.56	110.20
2	A	393	EV3	C1-C3-C5	-2.10	117.73	119.50

There are no chirality outliers.

5 of 34 torsion outliers are listed below:

Mol	Chain	Res	Type	Atoms
2	C	393	EV3	N9-C11-C12-C13
3	A	395	GOL	O1-C1-C2-C3
3	A	399	GOL	C1-C2-C3-O3
3	B	394	GOL	O1-C1-C2-C3
3	B	396	GOL	O1-C1-C2-C3

There are no ring outliers.

7 monomers are involved in 11 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
3	B	397	GOL	1	0
3	B	395	GOL	1	0
3	A	398	GOL	2	0
2	A	393	EV3	3	0
2	B	393	EV3	1	0
3	B	400	GOL	1	0
3	B	394	GOL	4	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

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.