



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 2, 2017 – 06:28 AM EDT

PDB ID : 2D06  
Title : Human Sult1A1 Complexed With Pap and estradiol  
Authors : Gamage, N.U.; Tsvetanov, S.; Duggleby, R.G.; McManus, M.E.; Martin, J.L.  
Deposited on : unknown  
Resolution : 2.30 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 : 1.7.2 (RC1), CSD as538be (2017)  
Xtriage (Phenix) : 1.9-1692  
EDS : rb-20030345  
Percentile statistics : 20161228.v01 (using entries in the PDB archive December 28th 2016)  
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) : rb-20030345

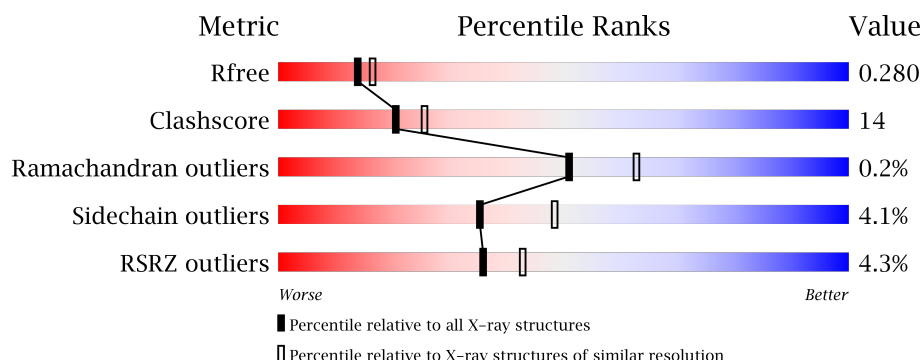
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.30 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))
$R_{free}$	100719	4130 (2.30-2.30)
Clashscore	112137	4751 (2.30-2.30)
Ramachandran outliers	110173	4705 (2.30-2.30)
Sidechain outliers	110143	4704 (2.30-2.30)
RSRZ outliers	101464	4156 (2.30-2.30)

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	295	<div> <div>2%</div> <div>74%</div> <div>22%</div> <div>..</div> </div>
1	B	295	<div> <div>6%</div> <div>63%</div> <div>33%</div> <div>..</div> </div>

## 2 Entry composition [i](#)

There are 4 unique types of molecules in this entry. The entry contains 4821 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.

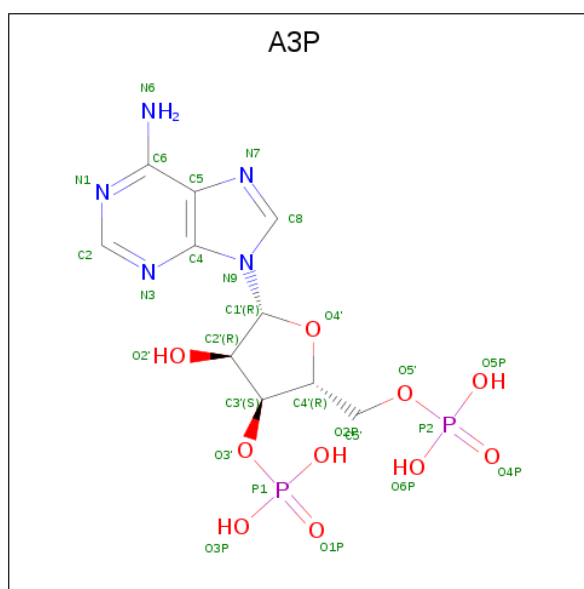
- Molecule 1 is a protein called Sulfotransferase 1A1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	288	Total	C	N	O	S	12	1	0
			2313	1502	377	419	15			
1	B	288	Total	C	N	O	S	26	2	0
			2295	1494	379	405	17			

There are 2 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	213	HIS	ARG	SEE REMARK 999	UNP P50225
B	213	HIS	ARG	SEE REMARK 999	UNP P50225

- Molecule 2 is ADENOSINE-3'-5'-DIPHOSPHATE (three-letter code: A3P) (formula:  $C_{10}H_{15}N_5O_{10}P_2$ ).



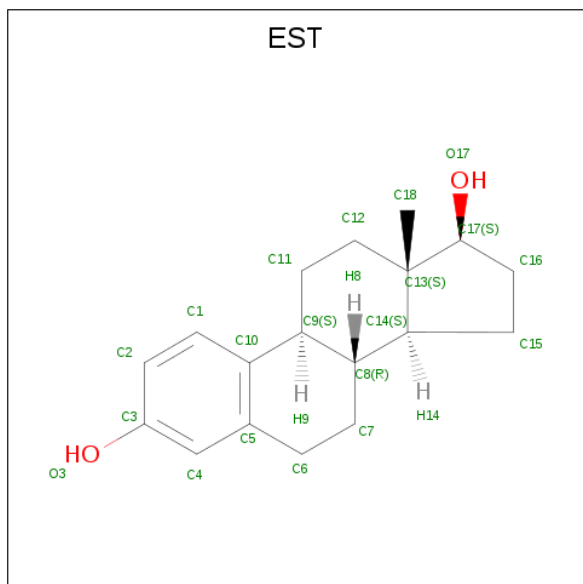
Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	A	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf
2	B	1	Total	C	N	O	P	0	0
			27	10	5	10	2		

- Molecule 3 is ESTRADIOL (three-letter code: EST) (formula:  $C_{18}H_{24}O_2$ ).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
3	A	1	Total	C	O	0	0
			20	18	2		
3	B	1	Total	C	O	0	0
			20	18	2		

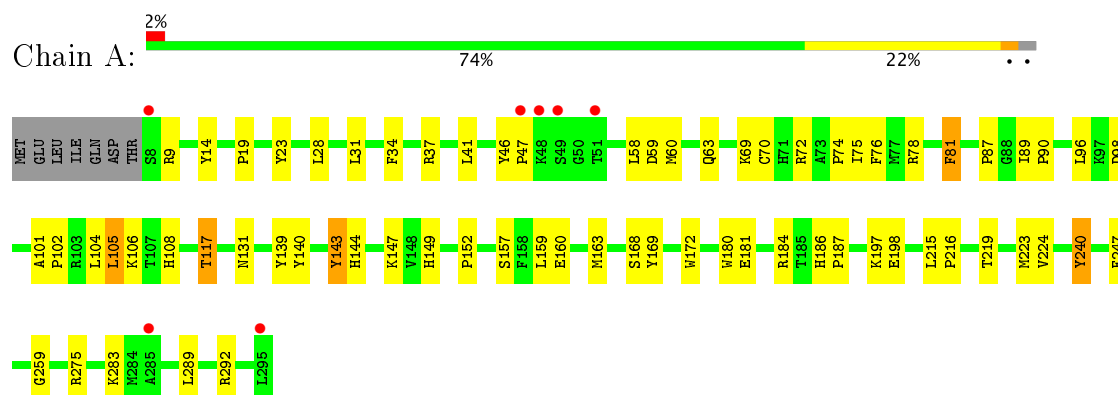
- Molecule 4 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
4	A	64	Total	O	0	0
			64	64		
4	B	55	Total	O	0	0
			55	55		

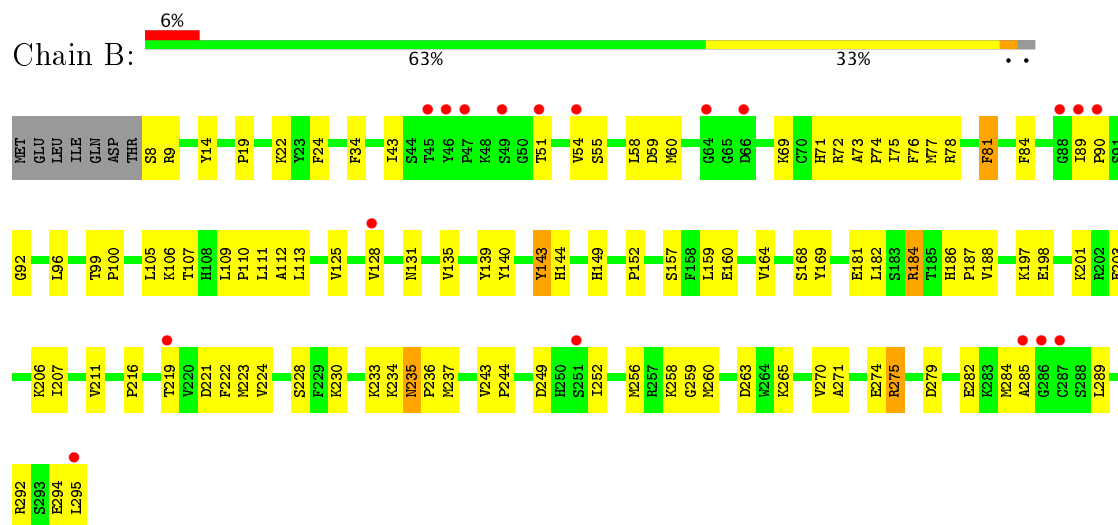
### 3 Residue-property plots [i](#)

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 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: Sulfotransferase 1A1



#### • Molecule 1: Sulfotransferase 1A1



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 2 <sub>1</sub> 2 <sub>1</sub> 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	123.64 Å   86.07 Å   72.95 Å 90.00°   90.00°   90.00°	Depositor
Resolution (Å)	29.77 – 2.30 31.41 – 2.30	Depositor EDS
% Data completeness (in resolution range)	99.6 (29.77-2.30) 99.5 (31.41-2.30)	Depositor EDS
$R_{merge}$	0.09	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	1.69 (at 2.29 Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.217   ,   0.283 0.217   ,   0.280	Depositor DCC
$R_{free}$ test set	3508 reflections (9.99%)	DCC
Wilson B-factor (Å <sup>2</sup> )	40.5	Xtriage
Anisotropy	0.442	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.27 , 61.7	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.50$ , $\langle L^2 \rangle = 0.34$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.96	EDS
Total number of atoms	4821	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	51.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The analyses of the Patterson function reveals a significant off-origin peak that is 29.91 % of the origin peak, indicating pseudo translational symmetry. The chance of finding a peak of this or larger height randomly in a structure without pseudo translational symmetry is equal to 1.4306e-03. The detected translational NCS is most likely also responsible for the elevated intensity ratio.*

<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: A3P, EST

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	0.57	0/2387	0.74	0/3248
1	B	0.53	0/2369	0.70	0/3221
All	All	0.55	0/4756	0.72	0/6469

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
1	A	0	1

There are no bond length outliers.

There are no bond angle outliers.

There are no chirality outliers.

All (1) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	240	TYR	Sidechain

### 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	2313	0	2218	49	0
1	B	2295	0	2205	77	0
2	A	27	0	11	0	0
2	B	27	0	11	0	0
3	A	20	0	24	1	0
3	B	20	0	24	0	0
4	A	64	0	0	2	0
4	B	55	0	0	3	0
All	All	4821	0	4493	126	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 14.

The worst 5 of 126 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:197:LYS:HE3	1:B:259:GLY:O	1.75	0.86
1:B:8:SER:HB2	4:B:3089:HOH:O	1.80	0.81
1:A:76:PHE:HB3	1:A:240:TYR:OH	1.80	0.80
1:B:60:MET:HB3	1:B:223:MET:CE	2.11	0.80
1:B:89:ILE:HG23	1:B:90:PRO:HD2	1.69	0.74

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	287/295 (97%)	266 (93%)	20 (7%)	1 (0%)	44	55
1	B	288/295 (98%)	261 (91%)	27 (9%)	0	100	100
All	All	575/590 (98%)	527 (92%)	47 (8%)	1 (0%)	51	63

All (1) Ramachandran outliers are listed below:



Mol	Chain	Res	Type
1	A	147	LYS

### 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	247/265 (93%)	236 (96%)	11 (4%)	32	44
1	B	241/265 (91%)	232 (96%)	9 (4%)	39	53
All	All	488/530 (92%)	468 (96%)	20 (4%)	35	48

5 of 20 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	181	GLU
1	A	289	LEU
1	B	184	ARG
1	A	143	TYR
1	A	159	LEU

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (5) such sidechains are listed below:

Mol	Chain	Res	Type
1	A	63	GLN
1	A	71	HIS
1	A	144	HIS
1	B	144	HIS
1	B	235	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](#)

4 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	A3P	A	2001	-	26,29,29	1.73	4 (15%)	31,45,45	1.29	3 (9%)
3	EST	A	304	-	23,23,23	2.80	15 (65%)	36,36,36	1.20	4 (11%)
2	A3P	B	2002	-	26,29,29	1.99	5 (19%)	31,45,45	1.26	3 (9%)
3	EST	B	305	-	23,23,23	2.81	15 (65%)	36,36,36	1.19	4 (11%)

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	A3P	A	2001	-	-	0/11/31/31	0/3/3/3
3	EST	A	304	-	-	0/0/40/40	0/4/4/4
2	A3P	B	2002	-	-	0/11/31/31	0/3/3/3
3	EST	B	305	-	-	0/0/40/40	0/4/4/4

The worst 5 of 39 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	B	2002	A3P	C8-N7	-2.46	1.30	1.34
2	A	2001	A3P	P1-O2P	-2.06	1.46	1.54

*Continued on next page...*

*Continued from previous page...*

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
3	B	305	EST	C13-C17	2.05	1.57	1.54
3	A	304	EST	C7-C6	2.06	1.56	1.52
3	A	304	EST	C16-C17	2.10	1.57	1.53

The worst 5 of 14 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	2001	A3P	O4'-C4'-C5'	-3.05	99.11	109.40
3	B	305	EST	C11-C9-C8	-2.70	107.62	111.33
3	A	304	EST	C15-C14-C13	-2.65	100.56	103.83
3	A	304	EST	C11-C9-C8	-2.63	107.71	111.33
3	B	305	EST	C15-C14-C13	-2.63	100.59	103.83

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	304	EST	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 [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	288/295 (97%)	-0.19	7 (2%) 59 66	25, 47, 76, 99	8 (2%)
1	B	288/295 (97%)	0.02	18 (6%) 21 27	30, 51, 87, 107	12 (4%)
All	All	576/590 (97%)	-0.08	25 (4%) 36 43	25, 49, 82, 107	20 (3%)

The worst 5 of 25 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	88	GLY	5.1
1	B	285	ALA	4.3
1	B	286	GLY	4.1
1	B	45	THR	3.8
1	B	90	PRO	3.2

### 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
3	EST	A	304	20/20	0.89	0.26	1.29	48,62,70,82	0
3	EST	B	305	20/20	0.89	0.24	0.94	41,64,87,88	0
2	A3P	A	2001	27/27	0.98	0.19	0.18	21,40,52,55	0
2	A3P	B	2002	27/27	0.97	0.17	-0.26	29,48,59,67	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.