



# wwPDB X-ray Structure Validation Summary Report ⓘ

Oct 9, 2017 – 02:52 PM EDT

PDB ID : 2H84  
Title : Crystal Structure of the C-terminal Type III Polyketide Synthase (PKS III) Domain of 'Steely1' (a Type I/III PKS Hybrid from Dictyostelium)  
Authors : Austin, M.B.; Saito, T.; Bowman, M.E.; Haydock, S.; Kato, A.; Moore, B.S.; Kay, R.R.; Noel, J.P.  
Deposited on : unknown  
Resolution : 2.90 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

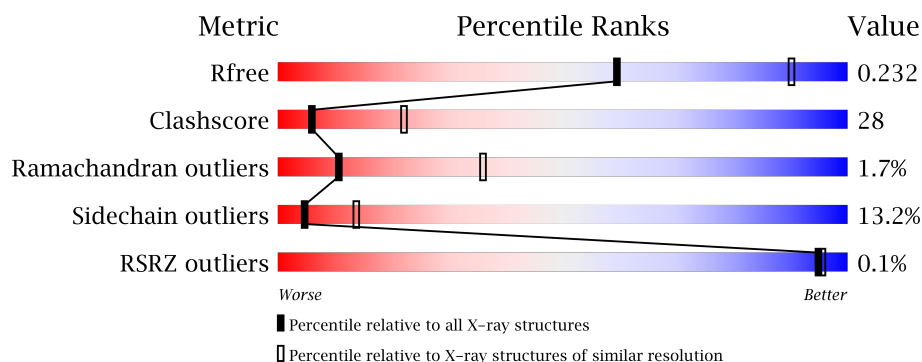
# 1 Overall quality at a glance

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 2.90 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	1586 (2.90-2.90)
Clashscore	112137	1807 (2.90-2.90)
Ramachandran outliers	110173	1768 (2.90-2.90)
Sidechain outliers	110143	1770 (2.90-2.90)
RSRZ outliers	101464	1596 (2.90-2.90)

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	374	
1	B	374	

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	P6G	A	1000	-	-	-	X

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 5949 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.

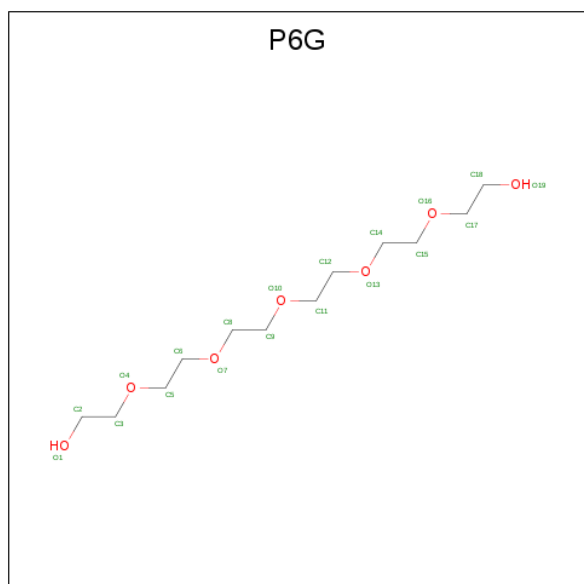
- Molecule 1 is a protein called Steely1.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	363	Total	C	N	O	S	0	0	0
			2782	1754	473	540	15			
1	B	363	Total	C	N	O	S	0	0	0
			2782	1754	473	540	15			

There are 4 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	2774	GLY	-	CLONING ARTIFACT	UNP Q55E72
A	2775	SER	-	CLONING ARTIFACT	UNP Q55E72
B	2774	GLY	-	CLONING ARTIFACT	UNP Q55E72
B	2775	SER	-	CLONING ARTIFACT	UNP Q55E72

- Molecule 2 is HEXAETHYLENE GLYCOL (three-letter code: P6G) (formula: C<sub>12</sub>H<sub>26</sub>O<sub>7</sub>).



Mol	Chain	Residues	Atoms			ZeroOcc	AltConf
2	A	1	Total	C	O	0	0
			19	12	7		

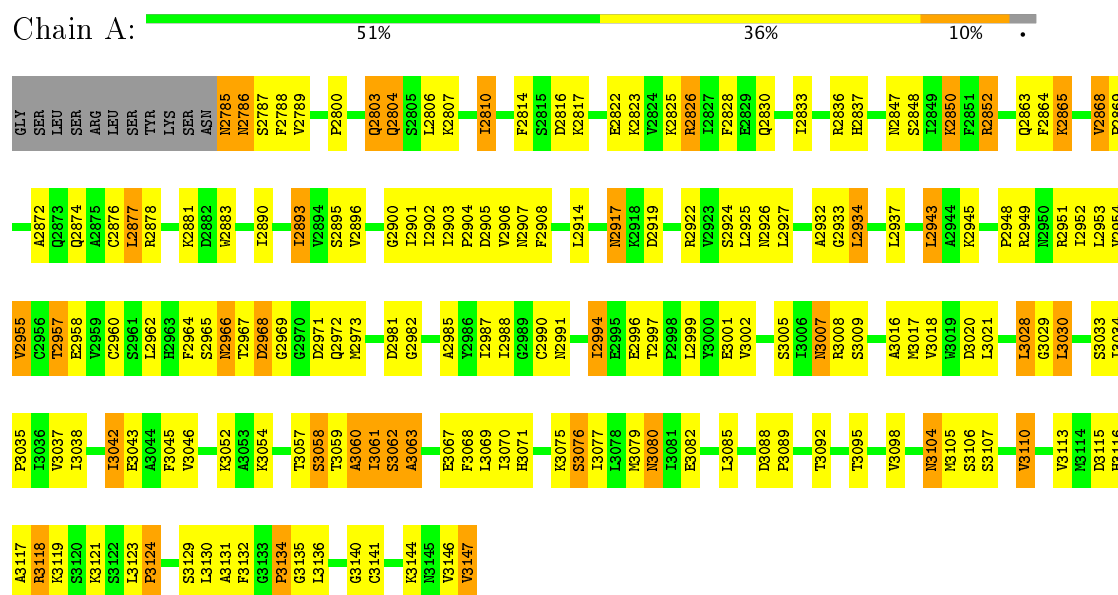
- Molecule 3 is water.

Mol	Chain	Residues	Atoms		ZeroOcc	AltConf
3	A	179	Total	O	0	0
			179	179		
3	B	187	Total	O	0	0
			187	187		

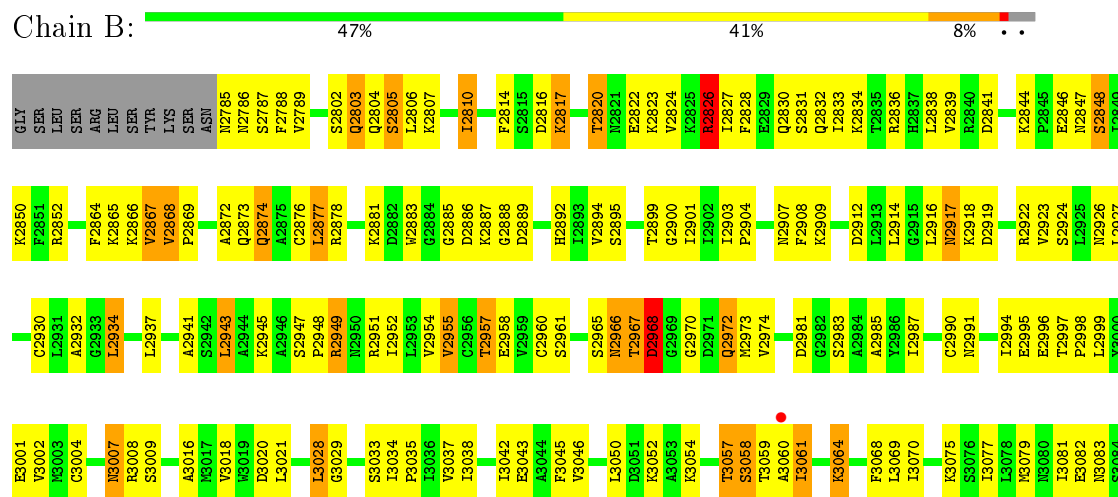
### 3 Residue-property plots

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: Steely1



#### • Molecule 1: Steely1



L3085	G3086	L3087	D3088	P3089	K3090	Q3091	T3092	K3093	N3094	T3095	V3098	M3104	M3105	S3106	S3107	V3110	V3113	M3114	D3115	H3116	A3117	R3118	K3119	S3120	K3121	S3122	L3123	P3124	L3130	A3131	F3132	G3133	P3134	E3139	L3143	K3144	R3145	V3146	V3147
-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------	-------

## 4 Data and refinement statistics

Property	Value	Source
Space group	P 21 21 21	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	81.96Å 83.31Å 114.30Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	39.13 – 2.90 39.14 – 2.91	Depositor EDS
% Data completeness (in resolution range)	97.4 (39.13-2.90) 98.8 (39.14-2.91)	Depositor EDS
$R_{merge}$	0.22	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	4.62 (at 2.90Å)	Xtriage
Refinement program	CNS	Depositor
R, $R_{free}$	0.199 , 0.233 0.198 , 0.232	Depositor DCC
$R_{free}$ test set	883 reflections (5.05%)	DCC
Wilson B-factor (Å <sup>2</sup> )	26.8	Xtriage
Anisotropy	0.415	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.33 , 15.3	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.47$ , $\langle L^2 \rangle = 0.30$	Xtriage
Estimated twinning fraction	0.066 for k,h,-l	Xtriage
$F_o, F_c$ correlation	0.91	EDS
Total number of atoms	5949	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	21.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 3.67% 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: P6G

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	0.82	15/2830 (0.5%)	0.86	6/3830 (0.2%)
1	B	0.76	11/2830 (0.4%)	0.81	7/3830 (0.2%)
All	All	0.79	26/5660 (0.5%)	0.84	13/7660 (0.2%)

The worst 5 of 26 bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
1	A	3147	VAL	C-OXT	14.68	1.51	1.23
1	A	3147	VAL	C-O	12.49	1.47	1.23
1	B	2968	ASP	CB-CG	-10.26	1.30	1.51
1	A	3135	GLY	N-CA	-9.06	1.32	1.46
1	A	2968	ASP	CB-CG	-8.09	1.34	1.51

The worst 5 of 13 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
1	A	3147	VAL	CA-C-O	24.29	171.11	120.10
1	A	2968	ASP	N-CA-C	7.46	131.13	111.00
1	B	3082	GLU	OE1-CD-OE2	-7.06	114.83	123.30
1	B	2949	ARG	NE-CZ-NH2	-6.49	117.06	120.30
1	A	3082	GLU	OE1-CD-OE2	-6.44	115.58	123.30

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	2782	0	2795	158	0
1	B	2782	0	2795	158	0
2	A	19	0	22	2	0
3	A	179	0	0	0	0
3	B	187	0	0	6	0
All	All	5949	0	5612	309	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 28.

The worst 5 of 309 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:3070:ILE:H	1:A:3095:THR:HG21	1.24	1.02
1:B:3070:ILE:H	1:B:3095:THR:HG21	1.24	1.01
1:B:2917:ASN:HD22	1:B:2919:ASP:H	1.06	0.98
1:B:2823:LYS:HG2	1:B:2973:MET:HE1	1.46	0.97
1:B:3095:THR:HG22	1:B:3113:VAL:HG11	1.47	0.96

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	361/374 (96%)	323 (90%)	34 (9%)	4 (1%)	17	48
1	B	361/374 (96%)	325 (90%)	28 (8%)	8 (2%)	8	29
All	All	722/748 (96%)	648 (90%)	62 (9%)	12 (2%)	11	36

5 of 12 Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	A	3124	PRO
1	B	2887	LYS
1	B	2968	ASP
1	B	3058	SER
1	A	3058	SER

### 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	314/324 (97%)	276 (88%)	38 (12%)	6	17
1	B	314/324 (97%)	269 (86%)	45 (14%)	4	11
All	All	628/648 (97%)	545 (87%)	83 (13%)	5	14

5 of 83 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
1	A	3134	PRO
1	B	2846	GLU
1	B	3115	ASP
1	B	2787	SER
1	B	2817	LYS

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. 5 of 33 such sidechains are listed below:

Mol	Chain	Res	Type
1	A	3083	ASN
1	B	2803	GLN
1	B	3056	GLN
1	A	3104	ASN
1	A	3145	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](#)

1 ligand is 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	P6G	A	1000	-	18,18,18	2.68	5 (27%)	17,17,17	2.37	8 (47%)

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	P6G	A	1000	-	-	0/16/16/16	0/0/0/0

All (5) bond length outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(Å)	Ideal(Å)
2	A	1000	P6G	C12-C11	-8.34	1.05	1.49
2	A	1000	P6G	O13-C14	-4.34	1.23	1.42
2	A	1000	P6G	O10-C9	-3.16	1.28	1.42
2	A	1000	P6G	O16-C15	-2.81	1.30	1.42
2	A	1000	P6G	C15-C14	2.64	1.62	1.49

The worst 5 of 8 bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed(°)	Ideal(°)
2	A	1000	P6G	O16-C17-C18	-2.47	98.74	110.15
2	A	1000	P6G	O13-C14-C15	-2.40	99.36	110.41
2	A	1000	P6G	O7-C8-C9	-2.31	99.78	110.41
2	A	1000	P6G	O10-C11-C12	2.10	120.04	110.41
2	A	1000	P6G	O13-C12-C11	2.26	120.80	110.41

There are no chirality outliers.

There are no torsion outliers.

There are no ring outliers.

1 monomer is involved in 2 short contacts:

Mol	Chain	Res	Type	Clashes	Symm-Clashes
2	A	1000	P6G	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	363/374 (97%)	-0.49	0	100   100	3, 19, 39, 62	0
1	B	363/374 (97%)	-0.49	1 (0%)	93   93	4, 20, 40, 60	0
All	All	726/748 (97%)	-0.49	1 (0%)	95   95	3, 20, 40, 62	0

All (1) RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
1	B	3060	ALA	2.3

### 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(Å <sup>2</sup> )	Q<0.9
2	P6G	A	1000	19/19	0.83	0.27	2.73	50,52,64,65	0

## 6.5 Other polymers [i](#)

There are no such residues in this entry.