



# wwPDB X-ray Structure Validation Summary Report ⓘ

Mar 9, 2018 – 06:38 pm GMT

PDB ID : 5CTR  
Title : Crystal structure of human SART3 HAT-C domain-human USP4 DUSP-UBL domain complex  
Authors : Park, J.K.; Kim, E.E.  
Deposited on : 2015-07-24  
Resolution : 3.01 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  
Xtriage (Phenix) : 1.13  
EDS : trunk30967  
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) : trunk30967

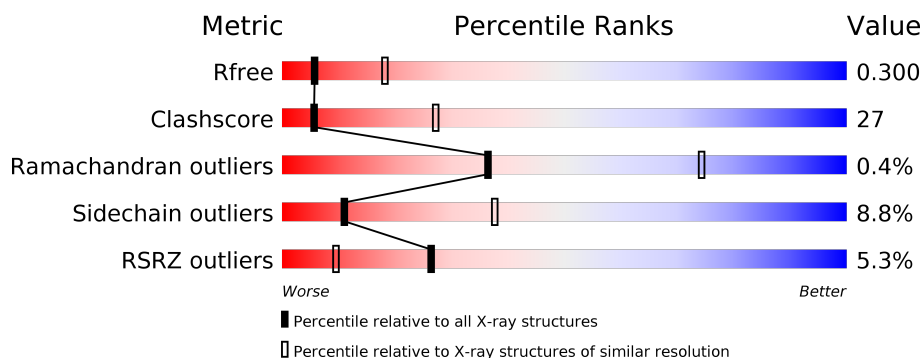
# 1 Overall quality at a glance ⓘ

The following experimental techniques were used to determine the structure:

## *X-RAY DIFFRACTION*

The reported resolution of this entry is 3.01 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	2110 (3.04-3.00)
Clashscore	122126	2436 (3.04-3.00)
Ramachandran outliers	120053	2362 (3.04-3.00)
Sidechain outliers	120020	2365 (3.04-3.00)
RSRZ outliers	108989	2001 (3.04-3.00)

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	338	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 1%, green 64%, yellow 24%, orange 2%, grey 9%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>%</span> <span>64%</span> <span>24%</span> <span>•• 9%</span> </div> </div>
1	B	338	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, green 57%, yellow 29%, orange 10%, grey 4%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>57%</span> <span>29%</span> <span>• 10%</span> </div> </div>
2	C	233	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 14%, green 52%, yellow 31%, orange 5%, grey 12%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>14%</span> <span>52%</span> <span>31%</span> <span>5%</span> <span>12%</span> </div> </div>
2	D	233	<div> <div style="width: 100%; height: 10px; background: linear-gradient(to right, red 9%, green 49%, yellow 35%, orange 6%, grey 9%);"></div> <div style="display: flex; justify-content: space-between; width: 100%;"> <span>9%</span> <span>49%</span> <span>35%</span> <span>6%</span> <span>9%</span> </div> </div>

## 2 Entry composition [i](#)

There are 3 unique types of molecules in this entry. The entry contains 8577 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 Squamous cell carcinoma antigen recognized by T-cells 3.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
1	A	308	Total	C	N	O	S	0	0	0
			2563	1615	457	476	15			
1	B	305	Total	C	N	O	S	0	0	0
			2536	1599	453	469	15			

There are 8 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
A	274	GLY	-	expression tag	UNP Q15020
A	275	SER	-	expression tag	UNP Q15020
A	276	HIS	-	expression tag	UNP Q15020
A	277	MET	-	expression tag	UNP Q15020
B	274	GLY	-	expression tag	UNP Q15020
B	275	SER	-	expression tag	UNP Q15020
B	276	HIS	-	expression tag	UNP Q15020
B	277	MET	-	expression tag	UNP Q15020

- Molecule 2 is a protein called Ubiquitin carboxyl-terminal hydrolase 4.

Mol	Chain	Residues	Atoms					ZeroOcc	AltConf	Trace
2	C	206	Total	C	N	O	S	0	0	0
			1694	1085	284	317	8			
2	D	212	Total	C	N	O	S	0	0	0
			1745	1114	294	328	9			

There are 6 discrepancies between the modelled and reference sequences:

Chain	Residue	Modelled	Actual	Comment	Reference
C	-2	GLY	-	expression tag	UNP Q13107
C	-1	SER	-	expression tag	UNP Q13107
C	0	HIS	-	expression tag	UNP Q13107
D	-2	GLY	-	expression tag	UNP Q13107

*Continued on next page...*

*Continued from previous page...*

Chain	Residue	Modelled	Actual	Comment	Reference
D	-1	SER	-	expression tag	UNP Q13107
D	0	HIS	-	expression tag	UNP Q13107

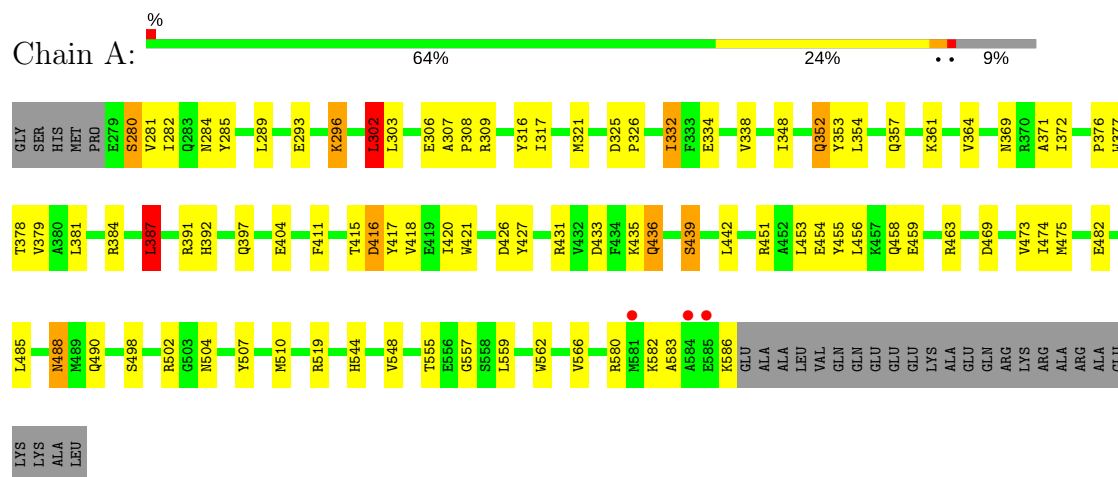
- Molecule 3 is water.

Mol	Chain	Residues	Atoms	ZeroOcc	AltConf
3	A	7	Total O 7 7	0	0
3	B	21	Total O 21 21	0	0
3	C	4	Total O 4 4	0	0
3	D	7	Total O 7 7	0	0

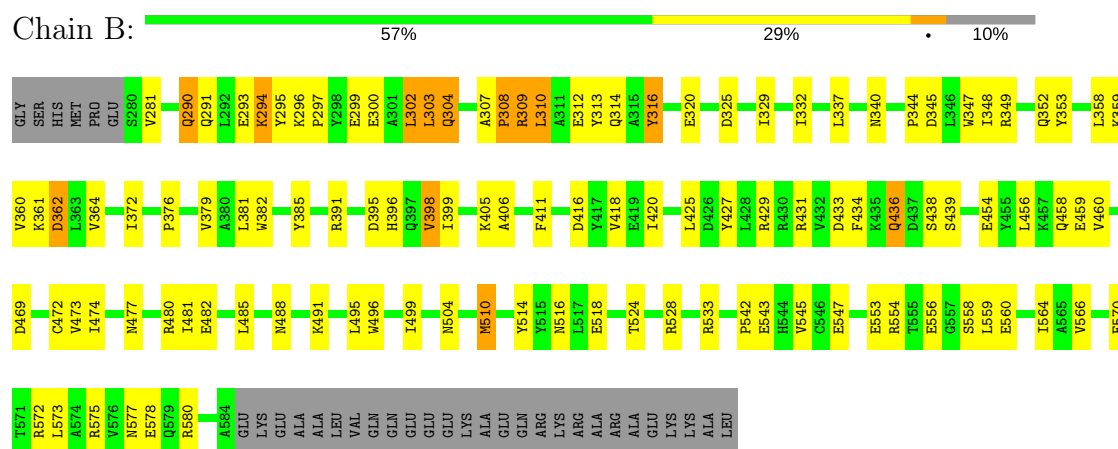
### 3 Residue-property plots

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.

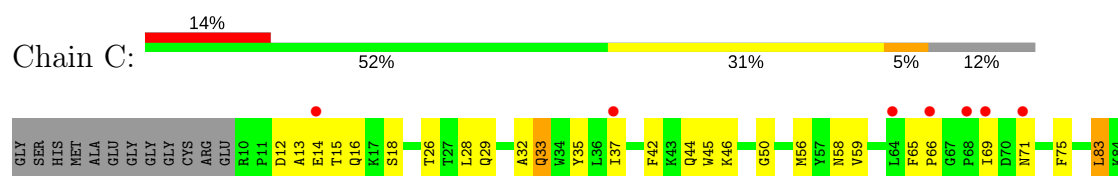
#### • Molecule 1: Squamous cell carcinoma antigen recognized by T-cells 3

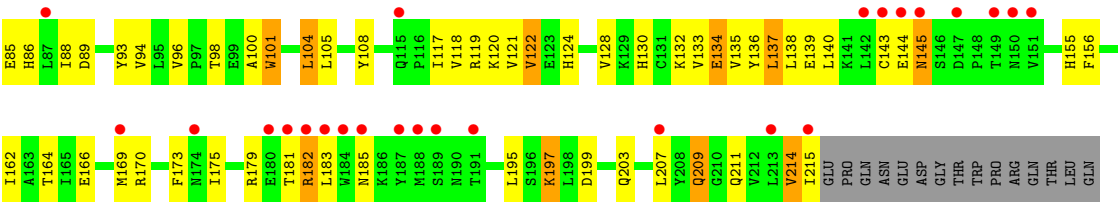


#### • Molecule 1: Squamous cell carcinoma antigen recognized by T-cells 3

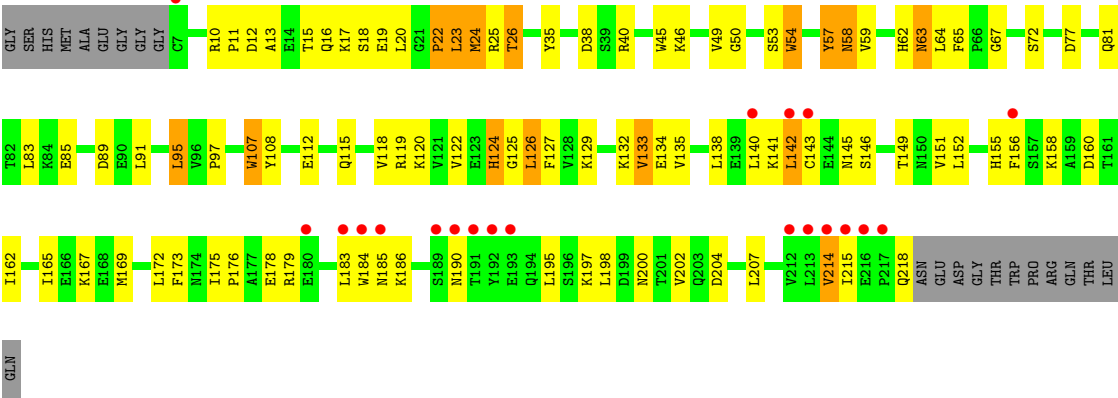


#### • Molecule 2: Ubiquitin carboxyl-terminal hydrolase 4





● Molecule 2: Ubiquitin carboxyl-terminal hydrolase 4



## 4 Data and refinement statistics

Property	Value	Source
Space group	P 43 21 2	Depositor
Cell constants a, b, c, $\alpha$ , $\beta$ , $\gamma$	115.20Å 115.20Å 306.14Å 90.00° 90.00° 90.00°	Depositor
Resolution (Å)	50.00 – 3.01 42.74 – 3.01	Depositor EDS
% Data completeness (in resolution range)	98.1 (50.00-3.01) 98.1 (42.74-3.01)	Depositor EDS
$R_{merge}$	0.11	Depositor
$R_{sym}$	(Not available)	Depositor
$\langle I/\sigma(I) \rangle$ <sup>1</sup>	2.11 (at 3.01Å)	Xtriage
Refinement program	PHENIX 1.7.3 _928	Depositor
R, $R_{free}$	0.309 , 0.334 0.299 , 0.300	Depositor DCC
$R_{free}$ test set	2064 reflections (5.04%)	wwPDB-VP
Wilson B-factor (Å <sup>2</sup> )	54.9	Xtriage
Anisotropy	0.131	Xtriage
Bulk solvent $k_{sol}$ (e/Å <sup>3</sup> ), $B_{sol}$ (Å <sup>2</sup> )	0.19 , -10.0	EDS
L-test for twinning <sup>2</sup>	$\langle  L  \rangle = 0.35$ , $\langle L^2 \rangle = 0.18$	Xtriage
Estimated twinning fraction	No twinning to report.	Xtriage
$F_o, F_c$ correlation	0.85	EDS
Total number of atoms	8577	wwPDB-VP
Average B, all atoms (Å <sup>2</sup> )	106.0	wwPDB-VP

Xtriage's analysis on translational NCS is as follows: *The largest off-origin peak in the Patterson function is 2.98% 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

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	0/2615	0.69	2/3535 (0.1%)
1	B	0.38	0/2588	0.67	0/3500
2	C	0.33	0/1737	0.53	0/2357
2	D	0.31	0/1789	0.49	0/2427
All	All	0.36	0/8729	0.62	2/11819 (0.0%)

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	2

There are no bond length outliers.

All (2) bond angle outliers are listed below:

Mol	Chain	Res	Type	Atoms	Z	Observed( $^{\circ}$ )	Ideal( $^{\circ}$ )
1	A	302	LEU	CA-CB-CG	8.52	134.89	115.30
1	A	387	LEU	CA-CB-CG	6.67	130.63	115.30

There are no chirality outliers.

All (2) planarity outliers are listed below:

Mol	Chain	Res	Type	Group
1	A	416	ASP	Sidechain
1	A	488	ASN	Mainchain



## 5.2 Too-close contacts ⓘ

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	2563	0	2514	82	2
1	B	2536	0	2489	168	1
2	C	1694	0	1654	137	1
2	D	1745	0	1702	83	0
3	A	7	0	0	0	0
3	B	21	0	0	0	0
3	C	4	0	0	0	0
3	D	7	0	0	0	0
All	All	8577	0	8359	462	2

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 27.

The worst 5 of 462 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:320:GLU:HG3	1:B:332:ILE:CD1	1.33	1.56
1:A:317:ILE:CD1	1:A:332:ILE:HD12	1.45	1.45
2:C:138:LEU:HG	2:C:209:GLN:CD	1.36	1.44
2:C:137:LEU:HD11	2:C:156:PHE:C	1.38	1.39
2:D:175:ILE:HG23	2:D:179:ARG:NH1	1.33	1.38

All (2) symmetry-related close contacts are listed below. The label for Atom-2 includes the symmetry operator and encoded unit-cell translations to be applied.

Atom-1	Atom-2	Interatomic distance (Å)	Clash overlap (Å)
1:A:436:GLN:OE1	1:B:361:LYS:CE[5_544]	1.73	0.47
1:A:280:SER:OG	2:C:18:SER:OG[5_544]	2.08	0.12

## 5.3 Torsion angles ⓘ

### 5.3.1 Protein backbone ⓘ

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	306/338 (90%)	295 (96%)	11 (4%)	0	100	100
1	B	303/338 (90%)	295 (97%)	7 (2%)	1 (0%)	43	78
2	C	204/233 (88%)	184 (90%)	18 (9%)	2 (1%)	17	54
2	D	210/233 (90%)	199 (95%)	10 (5%)	1 (0%)	31	70
All	All	1023/1142 (90%)	973 (95%)	46 (4%)	4 (0%)	36	74

All (4) Ramachandran outliers are listed below:

Mol	Chain	Res	Type
1	B	308	PRO
2	C	59	VAL
2	C	26	THR
2	D	22	PRO

### 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	271/294 (92%)	261 (96%)	10 (4%)	37	73
1	B	268/294 (91%)	245 (91%)	23 (9%)	11	39
2	C	188/209 (90%)	170 (90%)	18 (10%)	9	33
2	D	194/209 (93%)	164 (84%)	30 (16%)	3	13
All	All	921/1006 (92%)	840 (91%)	81 (9%)	11	38

5 of 81 residues with a non-rotameric sidechain are listed below:

Mol	Chain	Res	Type
2	C	58	ASN
2	C	164	THR
2	D	141	LYS
2	C	83	LEU
2	C	122	VAL

Some sidechains can be flipped to improve hydrogen bonding and reduce clashes. All (3) such sidechains are listed below:

Mol	Chain	Res	Type
2	C	44	GLN
2	C	130	HIS
2	D	124	HIS

### 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 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 [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	308/338 (91%)	-0.07	3 (0%) 82 58	24, 74, 127, 201	0
1	B	305/338 (90%)	0.05	0 100 100	24, 89, 149, 191	0
2	C	206/233 (88%)	0.63	32 (15%) 2 0	24, 147, 215, 238	0
2	D	212/233 (90%)	0.40	20 (9%) 8 3	24, 111, 203, 227	0
All	All	1031/1142 (90%)	0.20	55 (5%) 26 9	24, 96, 196, 238	0

The worst 5 of 55 RSRZ outliers are listed below:

Mol	Chain	Res	Type	RSRZ
2	D	217	PRO	7.8
2	D	213	LEU	6.8
2	C	151	VAL	6.2
2	C	213	LEU	6.1
2	D	184	TRP	6.0

### 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](#)

There are no ligands in this entry.

## 6.5 Other polymers [i](#)

There are no such residues in this entry.