

**Table S2. Formats for UniProt, NCBI and PDB IDs; FASTA Headers for Option C**

**A. Formats for UniProt IDs, NCBI IDs, and PDB IDs**

**UniProt IDs**

UniProtKB ID is 6 or 10 alphanumerical characters in the following formats:

1	2	3	4	5	6	7	8	9	10
[O,P,Q]	[0-9]	[A-Z,0-9]	[A-Z,0-9]	[A-Z,0-9]	[0-9]				
[A-N,R-Z]	[0-9]	[A-Z]	[A-Z,0-9]	[A-Z,0-9]	[0-9]				
[A-N,R-Z]	[0-9]	[A-Z]	[A-Z,0-9]	[A-Z,0-9]	[0-9]	[A-Z]	[A-Z,0-9]	[A-Z,0-9]	[0-9]

For example:

P11444  
T2HDW6  
A0A0A7PVN6

**NCBI RefSeq IDs**

An NCBI RefSeq ID is 2 letters followed by an underscore followed by a series of digits, a period, and one or more digits for the sequence version number, e.g.,

WP\_016501748.1  
NP\_708575.1  
YP\_002409124.1

**NCBI UniProt/Swiss-Prot IDs**

An NCBI UniProt/Swiss-Prot ID is the UniProt ID followed by a period and one or more digits for the sequence version number, e.g.,

Q31XL1.1  
B7LEJ8.1  
C4ZZT2.1

**NCBI GenBank IDs**

The format for NCBI GenBank IDs is 3 letters followed by five digits, a period, and one or more digits for the sequence version number, e.g.,

BAN56663.1  
AAC15504.1  
BAM38409.1

**PDB IDs**

The format for PDB IDs is one digit followed by two letters and a digit/letter:

1MDL  
1MRA  
3UXL

**NCBI GI Numbers**

An NCBI GI number (now retired) is a series of digits.

## B. Formats for FASTA headers for Option C

### UniProt (TrEMBL and SwissProt, respectively; from UniProt BLAST)

```
>tr|R9RJF1|R9RJF1_PSEAI Mandelate racemase OS=Pseudomonas aeruginosa PE=4
SV=1
>sp|P11444|MANR_PSEPU Mandelate racemase OS=Pseudomonas putida GN=mdlA PE=1
SV=1
```

### NCBI RefSeq (from NCBI BLAST)

```
>WP_016501748.1 mandelate racemase [Pseudomonas putida]
```

### NCBI UniProt/Swiss-Prot ID (from NCBI BLAST)

```
>Q0TE80.1 RecName: Full=Enolase; AltName: Full=2-phospho-D-glycerate hydro-
lyase; AltName: Full=2-phosphoglycerate dehydratase
```

### NCBI GenBank ID (from NCBI BLAST)

```
>AAA25887.1 mandelate racemase (EC 5.1.2.2) [Pseudomonas putida]
```

### NCBI PDB ID (from NCBI BLAST)

```
>pdb|1MDR|A Chain A, The Role Of Lysine 166 In The Mechanism Of Mandelate
Racemase From Pseudomonas Putida: Mechanistic And Crystallographic Evidence
For Stereospecific Alkylation By (r)-alpha-phenylglycidate
```

### NCBI GI Number (from NCBI BLAST; now retired)

```
>gi|347012980| 4-O-methyl-glucuronoyl methylesterase [Myceliophthora
thermophila ATCC 42464]
```

Option C also accepts FASTA headers in which the IDs (formats described in Option D) immediately follow the “>” symbol, e.g., the following headers abbreviated from those shown above:

### UniProt

```
>R9RJF1
>P11444
```

### NCBI RefSeq

```
>WP_016501748.1
```

### NCBI UniProt/Swiss-Prot ID

```
>Q0TE80.1
```

### NCBI GenBank ID

```
>AAA25887.1
```

### NCBI PDB ID

```
>1MDR
```

### NCBI GI Number (now retired)

```
>347012980
```