Using the Hidden Markov Model to
Assign SANC Residue Numbers

Note: HMMER3 (http://hmmer.org/) runs natively only on UNIX compatible operating systems, including Linux and macOS. Windows users will need to install additional software to run HMMER3. The developer provides an excellent user manual (http://eddylab.org/software/hmmer/Userguide.pdf) and your institution's IT department may be able to help with installation.

Searching the Model and Assigning SANC Residue Numbers

1. Save the HMM file and a text file containing the protein sequence of interest (In this example, "Your_sequence" in the file "Your_sequence.fa") in FASTA format in your directory of choice.
2. Navigate to the appropriate directory in the terminal and enter the following:

```
hmmsearch SANC_Numbering.hmm Your_sequence.fa
```

3. Look for the poriton of the output titled "Alignments for each domain:"

```
SANC_Numbering   4 kekelkavvdaaikpl...lkkqkipGmavavivdgkkhyfnyGvasketkkpvt 55
                   +++lka vdaa++p+   +k+++ipG++     +g++hyf+yG+aske++++vt+
 Your_sequence  30 PADRLKALVDAAVQPVQQQMKANDIPGLS----LKGEPHYFSYGLASKEDGRRVT 80
                   578899*******99866688899***86....69******************** PP
```

4. Use the numbers from either end of the "SANC_Numbering" lines as a starting point
    a. Dots in "SANC_Numbering" indicate insertions in "Your_sequence" compared to the SANC numbering scheme
        i. In this example, the three inserted Gln residues would be 19a, 19b, and 19c
    b. Dashes in "Your_seqeunce" indicate deletions in "Your_sequence" compared to the SANC numbering scheme
        i. In this example, the deleted residues are: 30, 31, 32, and 33

Important Considerations

- Numbering residues of novel AmpC enzymes in this manner helps to remove ambiguity as to the location of insertions and deletions under the SANC numbering scheme
- Our HMM does not determine the site of signal peptide cleavage and will "ignore" residues not likely part of the mature enzyme. Use SignalP 5.0 (http://www.cbs.dtu.dk/services/SignalP-5.0/) or another tool to determine signal peptides
- Our HMM will only assign numbers to residues 7 through 360 (this is a limitation of the process used to generate the model) and residues outside this range will need to be manually assigned (although exact alignment is generally less crucial in these regions)
    - Residues downstream of 361 are numbered in sequence
    - Residues upstream of 4 are numbered down to 0 and letters appended if needed