1. **Semi-global alignment [25 pts].**

In this question we consider a variant of the global alignment problem, where we impose no penalty on gaps at the ends of one of the sequences. Consider the following two alternative alignments:

```
sequence1: CAGCA-CTTGGATTCTCGG
sequence2: ---CAGCGTGG--------
sequence1: CAGCACTTGGATTCTCGG
sequence2: CAGC-----G-T----GG
```

Under the simplest scoring scheme of +1 for match, -1 for mismatch, and -1 for a gap the second alignment is preferred, despite our intuition that the first alignment is more biologically relevant. If the gaps on the ends of `sequence2` are not penalized, then the first alignment scores higher. This approach is called *semi-global* alignment. Note that in this approach the end gaps in one of the sequences in the alignment

```
sequence1: ACGTCAT---
sequence2: ---TCATGCA
```

will be penalized. Show how to modify the Needleman-Wunsch algorithm to compute a semi-global alignment (including the initialization of the matrix, and the traceback operations). Illustrate your algorithm on the sequences `ACAGATAC` and `AGTT` using the above simple scoring scheme.

2. **Homology [10 pts].**

Is homology transitive? (i.e. if A and B are homologous, and B and C are homologous, are A and C homologous?)

3. **Global alignment with a limited number of gaps [20 pts].**

For a parameter \( k \), suggest a dynamic programming algorithm that finds an optimal global alignment between two sequences subject to the constraint that the alignment contains at most \( k \) blocks of consecutive indels.

4. **Local alignment [50 pts].**

Implement the Smith-Waterman local alignment algorithm, using fixed gap costs, and a fixed mismatch penalty (the same for all amino acids/nucleotides). Use your program to align the following DNA sequences:
Explore the parameter space of the algorithm, and comment on the alignments you obtain (e.g., what kind of parameter values give a global alignment, vs parameter values that give a local alignment). Show what you think is the best alignment for the two sequences. In addition, use your program to align random DNA sequences of the same length and repeat the same exploration of parameter space. In view of these experiments, do you think seq1 and seq2 are related? (In other words, do you think the level of similarity between them can be attributed to chance).

Email your program to the instructor. For those in search of a greater challenge, implement the affine gap version.