Figure 2 - Pseudo code for Pro-LEADER algorithm

**Input:** A training set $D$, $D = \{O_i\}_{j=1..n}$; $n$ is the size of $D$

**Initialize:** LeaderList = $\varnothing$

1. Select the first sequence, $L$, as a leader;
2. LeaderList = LeaderList $\cup L$;
3. For each $j \in [2..n]$ do
   - Compute the similarity score of $O_j$ with all leaders in LeaderList using Smith Waterman algorithm;
   - Find in LeaderList the nearest leader $R_i$ to $O_j$;
   - If Score ($R_i$, $O_j$) > Threshold then
     - Assign $O_j$ to the set of the leader $R_i$;
   - Else
     - LeaderList = LeaderList $\cup O_j$;
4. Compute $f(V)$;

**Output:** LeaderList; LeaderList is the best partition of $D$ into $K$ clusters; each cluster is defined by a Leader $R_i$