Figure 4 - Pseudo code for Pro-CLARA algorithm

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

**Initialize:** $f(V)_{\text{max}} = 0$; $\text{iteration} = 0$;

**Repeat**

1. Draw a sample $S$ of $40 + 2K$ sequences randomly from $D$;

2. Call Pro-PAM algorithm to find $K$ medoids of $S$: $R_i (i \in \{1..K\})$;

3. **For each** $O_h \in D$ **do**

   - Compute the similarity score of $O_h$ with each medoid $R_i (i \in \{1..K\})$, using Smith Waterman algorithm;

   - Assign $O_h$ to the cluster with the nearest $R_i$;

4. **Compute** $f(V)$;

5. **If** $f(V) < f(V)_{\text{max}}$ **then**

   \hspace{1em} $iteration = iteration + 1$;

   **Else**

   \hspace{1em} $f(V)_{\text{max}} = f(V)$;

   **BestSets** = **CurrentSets**; (**CurrentSets** are Subsets obtained in this partition)

   Go back to Step 2;

**Until** $iteration = q$;

**End**

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