A. Building a CM from a seed alignment:

```
Flexibacter    GUAUA.GCUC...AC
Agrobacterium  GUGCA.GCCC...AC
Escherichia    GUUUAcGGAC...AC
Chlamydia      GCU-A.GAAC...GC
Clostridium    GCG-A.GGCCauaGC
Arthrobacter   GCA-A.GAUC...GC
Thermus        GCCUA.-UGC...GC
Aquifex        GUCUA.GUGC...AC
#=GC SS_cons   <<<--.-->-...>>
               1 2 3 4 5   6 7 8 9   10 11
```

→ consensus structure

construction procedure → Covariance model

sequence + structure profile

Aquifex secondary structure

```
    A G
  U   U
  C•G
       C
  U•A
  G•C
```

consensus structure

```
   - -
  -   -
  <   >
       -
  <   >
  <   >
```

Model's consensus sequence and structure (Uppercase nt are highly conserved in the seed.)

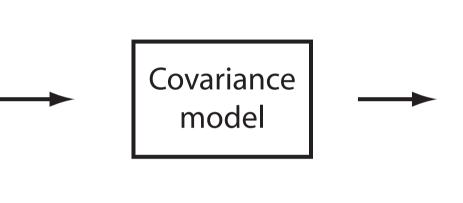```
   A G
  U   a
  a•u
       C
  c•g
  G•C
```

Seed alignment columns with < 80% gaps become consensus positions in the model. Other columns are considered inserts.

This model has 11 consensus positions, mapping to the 11 numbered positions in the seed.

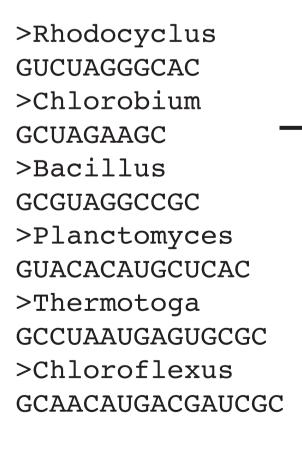## B. Aligning four target sequences with a CM:

```
>Rhodocyclus
GUCUAGGGCAC
>Chlorobium
GCUAGAAGC
>Bacillus
GCGUAGGCCGC
>Planctomyces
GUACACAUGCUCAC
```

→ Covariance model →

```
Rhodocyclus    GUCUA...GGGCAC
Chlorobium     GCU-A...GAA-GC
Bacillus       GCGUA...GGCCGC
Planctomyces   GUACAcauGCUCAC
#=GC SS_cons   <<<--..-->->>
#=GC RF        GcaUA...GauCgC
               1 2 3 4 5   6 7 8 9 10 11
```

Output alignments have a constant number of consensus columns (all of them) but a variable number of insert columns.

#=GC RF line annotates the model's consensus sequence. Gaps in RF are insert positions.

C. Aligning the four target sequences from (B), plus two more:

```
>Rhodocyclus
GUCUAGGGCAC
>Chlorobium
GCUAGAAGC
>Bacillus
GCGUAGGCCGC
>Planctomyces
GUACACAUGCUCAC
>Thermotoga
GCCUAAUGAGUGCGC
>Chloroflexus
GCAACAUGACGAUCGC
```

→ Covariance model →

```
Rhodocyclus    GUCUA......GGGCAC
Chlorobium     GCU-A......GAA-GC
Bacillus       GCGUA......GGCCGC
Planctomyces   GUACAc...auGCUCAC
Thermotoga     GCCUAau..gaGUGCGC
Chloroflexus   GCA-AcaugacGAUCGC
#=GC SS_cons   <<<--......-->->>
#=GC RF        GcaUA......GauCgC
               1 2 3 4 5        6 7 8 9 10 11
```

More target sequences yield more insert columns in output alignments.

Inserted nucleotides are not aligned between different sequences. They are simply split in half, the left half is flushed-left and the right half is flushed right.

```
Rhodocyclus    GUCUA......GGGCAC
Chlorobium     GCU-A......GAA-GC
Bacillus       GCGUA......GGCCGC
Planctomyces   GUACAcau...GCUCAC
Thermotoga     GCCUA.auga.GUGCGC
Chloroflexus   GCA-AcaugacGAUCGC
#=GC SS_cons   <<<--......-->->>
#=GC RF        GcaUA......GauCgC
               1 2 3 4 5        6 7 8 9 10 11
```

This is a better alignment with inserted nucleotides aligned more intuitively.

This alignment would NOT be generated by SSU-ALIGN.