**Course: Advanced Bioinformatics**

**Module title: Iterative Methods**

**Module no. : 41**

Iterative alignment methods begin by making an initial alignment of the sequences. These alignments are then revised to give a more reasonable result. The objective of this approach is to improve the overall alignment score.

MultAlin

PRRP

DIALIGN

#### Genetic Algorithms

The goal of genetic algorithms used in sequence alignment is to generate as many different multiple sequence alignments by rearrangements that simulate gaps and genetic recombination events. SAGA (Serial Alignment by Genetic Algorithm) is one such approach that yields very promising results, but becomes slow when more than 20 sequences are used.

**Steps of SAGA (Genetic Algorithm)**

1. Up to 20 different sequences are written in a row, allowing for overlaps of a random length. The ends of these sequences are then padded with gaps. Typically, upwards of 100 initial alignments are made.
2. The initial alignments are scored by the sum of pairs method. Standard amino acid scoring matrices and gap open, gap extension penalties are used.
3. Initial alignments are replaced to give another generation of multiple sequence alignments. One half of the multiple sequence alignments are chosen to proceed to the next generation unchanged (natural selection). This half is chosen by assigning probabilities to each sequence based on an inverse proportion of their SP scores (the best alignments, since the SP scores are weighted according to their distance from the parent). The other half of the alignments are sent to the next generation, but are first subject to mutation.
4. In the mutation process, gaps are inserted into the sequences subject to mutation and rearranged in an attempt to create a better scoring alignment. In this step, the sequences are split into two sets based on an estimated phylogenetic tree, and gaps of random lengths are inserted into random positions in the alignment.
5. Recombination of two parent alignments is accomplished???
6. The next generation is evaluated going back to step 2, and steps 2-5 are repeated a number (100-1000) times. The best scoring multiple sequence alignment is then obtained (note that it may not be the optimal scoring alignment).
7. The entire process is repeated several times, starting from a different initial alignment each time. The best scoring multiple sequence alignment is then chosen and reported to the user.

#### Simulated Annealing

Another approach to sequence alignment that works in a manner similar to genetic algorithms is simulated annealing. In these approaches, you begin with a heuristically determined multiple sequence alignment that is then changed using probabilistic models that identifies changes in the alignment that increase the alignment score. The drawback of simulated annealing approaches is that you can get stuck finding only the locally optimal alignment rather than the alignment score that is globally optimal.