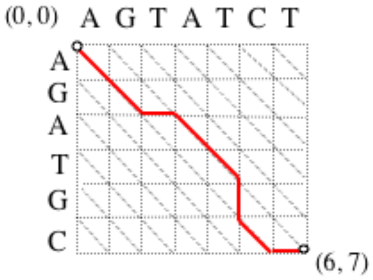
4. Sequence comparison

- How to predict gene/protein functions?
- Why gene/protein sequences may be similar?
- Measuring sequence similarity
- Aligning sequences is an optimization problem
- A sequence alignment as a path
- **A path is optimal if all its sub-paths are optimal**
- Alignment costs
- A recursive algorithm
- Recursion can be avoided: an iterative version
- How efficient is this algorithm?

A proof can be obtained

- If a path of length L is optimal, the path of length L-1 is also optimal
- Proof *ad absurdum*



(a)	A	G	–	A	T	G	C	–
	A	G	T	A	T	–	C	T

The proof can be obtained by recurrence

- If a path of length L is optimal, the path of length L-1 is also optimal
- Proof *ad absurdum*

- An optimal path is made up of optimal sub-paths

Toward a “dynamic programming” algorithm

- There are only 3 unitary paths ending on the last node

- Let’s suppose the costs of the optimal paths ending on these 3 nodes are known

- Then the optimal unitary last path is easily computed

Computation of the cost on the last node