Today: Alignments 1

☐ LAST
  • 3D comparison

☐ TODAY
  • Alignments and “reach of comparative modeling”

☐ NEXT
  • alignment contd
Announcements

 Videos:  SciVe / www.rostlab.org

 THANKS:
  Tim Karl + ?

 Special lectures:
  • May 16, 23, 28: Andrea Schafferhans
  • June 27: Thomas Hopf
  • Jan 31: Marco De Vivo (ISS Geneva)

 No lecture:
  • May 9 Thu (Ascension)
  • May 14 Tue (Student assembly)
  • May 21 Tue (Whitsun break)
  • May 30 Thu (Corpus Christi)

 LAST lecture:  Jul 4

 Examen:  Jul 9, 13:15 (likely this room + others)
  • Makeup:  Oct 17 - morning

CONTACT: Marlena Drabik assistant@rostlab.org

Wednesday May 1, 2013
Notation: protein structure 1D, 2D, 3D

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>P</td>
<td>PP</td>
<td>P</td>
</tr>
<tr>
<td>Q</td>
<td>QQQQ</td>
<td>Q</td>
</tr>
<tr>
<td>I</td>
<td>FFQVI</td>
<td>I</td>
</tr>
<tr>
<td>T</td>
<td>SSIVR</td>
<td>T</td>
</tr>
<tr>
<td>L</td>
<td>LLSTL</td>
<td>L</td>
</tr>
<tr>
<td>W</td>
<td>WWQED</td>
<td>W</td>
</tr>
<tr>
<td>Q</td>
<td>RKQAK</td>
<td>Q</td>
</tr>
<tr>
<td>R</td>
<td>RRRPQ</td>
<td>R</td>
</tr>
<tr>
<td>P</td>
<td>PPPPP</td>
<td>P</td>
</tr>
<tr>
<td>L</td>
<td>VVTKF</td>
<td>L</td>
</tr>
<tr>
<td>V</td>
<td>VVLII</td>
<td>V</td>
</tr>
<tr>
<td>T</td>
<td>TTKEK</td>
<td>T</td>
</tr>
<tr>
<td>I</td>
<td>AALIV</td>
<td>I</td>
</tr>
<tr>
<td>K</td>
<td>HYKKF</td>
<td>K</td>
</tr>
<tr>
<td>I</td>
<td>IILVI</td>
<td>I</td>
</tr>
<tr>
<td>G</td>
<td>EENGG</td>
<td>G</td>
</tr>
<tr>
<td>G</td>
<td>GGGTG</td>
<td>G</td>
</tr>
<tr>
<td>Q</td>
<td>QQKRR</td>
<td>Q</td>
</tr>
<tr>
<td>L</td>
<td>PPLWW</td>
<td>L</td>
</tr>
<tr>
<td>K</td>
<td>VVFKV</td>
<td>K</td>
</tr>
<tr>
<td>E</td>
<td>EESKK</td>
<td>E</td>
</tr>
<tr>
<td>A</td>
<td>VVGLG</td>
<td>A</td>
</tr>
<tr>
<td>L</td>
<td>LLLLE</td>
<td>L</td>
</tr>
<tr>
<td>L</td>
<td>LLLVV</td>
<td>L</td>
</tr>
<tr>
<td>D</td>
<td>DDDDD</td>
<td>D</td>
</tr>
<tr>
<td>T</td>
<td>TTTTT</td>
<td>T</td>
</tr>
<tr>
<td>G</td>
<td>GGGGG</td>
<td>G</td>
</tr>
<tr>
<td>A</td>
<td>AAAAA</td>
<td>A</td>
</tr>
<tr>
<td>D</td>
<td>DDDDD</td>
<td>D</td>
</tr>
<tr>
<td>D</td>
<td>DDAKE</td>
<td>D</td>
</tr>
<tr>
<td>T</td>
<td>SSTTV</td>
<td>T</td>
</tr>
<tr>
<td>V</td>
<td>IIIVIV</td>
<td>V</td>
</tr>
<tr>
<td>L</td>
<td>VVIVL</td>
<td>L</td>
</tr>
</tbody>
</table>

kcal/mol

© Burkhard Rost (TU Munich)
Sequence comparisons: pairwise methods
Alignments answer:
How similar are proteins?
Alignments

PAPI

OPA
Alignments

.PAPI
OPA..
Alignments

.PAPI
OPA..
.PAPA
Alignments

.PAPI
OPA..
.PAPA
.PAPU
.PIPI
Alignments: Steps to do

☐ goal: align two sequences

    GGQLAKEEAL    EGQPVEVLP

☐ 2 do:
  • (1) Find optimal superposition of the two
  • (2) Define “optimal”
Alignment: Local vs. Global

- compare:
  - global: all residues aligned
    - PAPI
    - O.PA
  - GGQLAKEEAL
  - EGQPVEVLP
Alignment: Local vs. Global

- **compare:**
  - **global:** all residues aligned
    - GGQLAKEEAL
    - EGQPVEVLP
  - **local:** best matches
    - GGQLAKEEAL
    - EGQPVEVLP

© Burkhard Rost (TU Munich)
Alignment: Assumptions about meaning

- Biological meaning of insertion (".")?

GGQLAKEEAL
EGQ.PVEVL
Alignment: Assumptions about meaning

- Biological meaning of insertion (".")?

- Historical interpretation
  
  (reason to start working on the problem):

  evolutionary relation
Meaning of “homology”?
phylogeny:
phylo: race/tribe/kind
genesis: birth
genos: origin
tree of life - Darwin 1859 Origin of Species
Phylogenetic Tree Of Life

© Wikipedia
Phylogenetic Tree Of Life

© Wikipedia
Phylogenetic Tree Of Life

© Wikipedia
Phylogenetic Tree of Life

Bacteria
- Spirochetes
- Gram positives
- Proteobacteria
- Cyanobacteria
- Rickettsiales
- Bacteroides
- Actinobacteria
- Thermotogae
- Aquifex

Archaea
- Methanomicrobia
- Methanosarcina
- Methanobacterium
- Halophiles
- Methanococcus
- T. maritima

Eucaryota
- Entamoebae
- Slime molds
- Animals
- Fungi
- Plants
- Ciliates
- Flagellates
- Trichomonads
- Microsporidia
- Diplomonads

© Wikipedia
Evolution: speciation

© http://evolution.berkeley.edu/evosite/evo101/
homology
etymology: Greek: ομόλογος (homologos):
homo (agreeing/same) + logos (word/reason/opinion) (wikipedia: relation)

originally (1843 Richard Owen): “same organ in different animals under every variety of form and function”

Genes:
- homologs originate from common ancestor
  - orthologs: speciation event
  - paralogs: duplication event
Evolution: speciation

© http://evolution.berkeley.edu/evosite/evo101/

species=?
Evolution: speciation

© http://evolution.berkeley.edu/evosite/evo101/

species = mating
Evolution: speciation

© http://evolution.berkeley.edu/evosite/evo101/

happily munching
Evolution: speciation

© http://evolution.berkeley.edu/evosite/evo101/

disaster strikes
Evolution: speciation

© http://evolution.berkeley.edu/evosite/evo101/

populations diverge
Evolution: speciation

© http://evolution.berkeley.edu/evosite/evo101/

rejoined - yet separated
Evolution: speciation


Evolution: speciation

Happy Face Spider *Theridion grallator* (same species-interbreed)

Carrion/Hooded crow (same species?)
homology definition

- **etymology:** Greek: ομόλογος (homologos):
  homo (agreeing/same) + logos (word/reason/opinion)
  (wikipedia: relation)

- originally (1843 Richard Owen):
  “same organ in different animals under every variety of form and function”

- **Genes:**
  homologs originate from common ancestor
  - orthologs: speciation event
  - paralogs: duplication event

- homolog structure => similar structure
Alignment: Assumptions about meaning

- Biological meaning of insertion (".")?

- Historical interpretation (reason to start working on the problem):
  evolutionary relation
  common ancestor
  homology

GGQLAKEEAL
EGQ.PVEVL
Dynamic programming: brute force
How to align sequences?

1. visually

GGQLAKEEAL
EGQ.PVEVL
How to align sequences?

- 1. Visually
- 2. Needleman-Wunsch: JMB
  1970 48:443-53

**A General Method Applicable to the Search for Similarities in the Amino Acid Sequence of Two Proteins**

Saul B. Needleman and Christian D. Wunsch
Department of Biochemistry, Northwestern University, and Nuclear Medicine Service, V. A. Research Hospital Chicago, Ill. 60611, U.S.A.

(Received 21 July 1969)

A computer adaptable method for finding similarities in the amino acid sequences of two proteins has been developed. From these findings it is possible to determine whether significant homology exists between the proteins. This information is used to trace their possible evolutionary development.

The maximum match is a number dependent upon the similarity of the sequences. One of its definitions is the largest number of amino acids of one protein that can be matched with those of a second protein allowing for all possible interruptions in either of the sequences. While the interruptions give rise to a very large number of comparisons, the method efficiently excludes from consideration those comparisons that cannot contribute to the maximum match.

Comparisons are made from the smallest unit of significance, a pair of amino acids, one from each protein. All possible pairs are represented by a two-dimensional array, and all possible comparisons are represented by pathways through the array. For this maximum match only certain of the possible pathways must be evaluated. A numerical value, one in this case, is assigned to every cell in the array representing like amino acids. The maximum match is the largest number that would result from summing the cell values of every pathway.

1. Introduction
Dynamic programming: Global

GGQLAKEEAL
EGQPVEVL
Dynamic programming: Global

GGQLAKEEAL
EGQPVEVL

GGQLAKEEAL
EGQPVEVL
Dynamic programming: Global

GGQLAKKEEAL
EGQPVEVL

GGQLAKKEEAL
EGQPVEVL

GGQLAKKEEAL
EGQPVEVL

© Burkhard Rost (TU Munich)
### Dynamic programming concept

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>G</th>
<th>Q</th>
<th>L</th>
<th>A</th>
<th>K</th>
<th>E</th>
<th>E</th>
<th>A</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Q</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

© Burkhard Rost (TU Munich)
allowing for gaps
Dynamic programming concept ins

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>G</th>
<th>Q</th>
<th>L</th>
<th>A</th>
<th>K</th>
<th>E</th>
<th>E</th>
<th>A</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td>2</td>
<td>1</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td>2</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dynamic programming concept ins

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>G</th>
<th>Q</th>
<th>L</th>
<th>A</th>
<th>K</th>
<th>E</th>
<th>E</th>
<th>A</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Wednesday May 1, 2013
Dynamic programming concept ins

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>G</th>
<th>Q</th>
<th>L</th>
<th>A</th>
<th>K</th>
<th>E</th>
<th>E</th>
<th>A</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Dynamic programming concept ins
### Dynamic programming concept

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>G</th>
<th>Q</th>
<th>L</th>
<th>A</th>
<th>K</th>
<th>E</th>
<th>E</th>
<th>A</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GGQLAKEEAL
GQP.VEVL

GGQLAKEEAL
G.QP.VE.VL
what is best?

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>G</th>
<th>Q</th>
<th>L</th>
<th>A</th>
<th>K</th>
<th>E</th>
<th>E</th>
<th>A</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
2 of 2 better than 4 of 8?
Gap/Insertions: penalty

- linear gap penalty:

  N gaps cost N

  \[
  \text{EGQADQPDVEVLP} \leftrightarrow \text{E.G.Q.P.V.EVLP}
  \]
Gap/Insertions: penalty

- **linear gap penalty:**
  
  \[
  N \text{ gaps cost } N
  \]

  \[
  \text{EGQ.....PVEVLP} \leftrightarrow \text{E.G.Q.P.V.EVLP}
  \]

- **idea of BLOCKS:**

How to enforce blocks in alignments?
Gap/Insertions: penalty

- linear gap penalty:
  
  N gaps cost N

  \[
  \text{EGQ......PVEVLP} \leftrightarrow \text{E.G.Q.P.V.EVLP}
  \]

- affine gap penalty:
  
  gap open

  gap extension (elongation)

  typical go=10 x ge
Dynamic programming: optimal alignment

• Global/no gap:
  SB Needleman and CD Wunsch 1970 J Mol Biol 48, 443-53
• Local/Gap:
  TF Smith and MS Waterman 1981 J Mol Biol 147, 195-197
Dynamic programming: optimal alignment

- Global/no gap:
  SB Needleman and CD Wunsch 1970 J Mol Biol 48, 443-53
- Local/Gap:
  TF Smith and MS Waterman 1981 J Mol Biol 147, 195-197

\[
\text{SW} = \sum_{k=l}^{L_{\text{ali}}} M_{U_k T_k} - G_o \cdot N_{\text{gap}} - G_e \cdot (L_{\text{gap}} - N_{\text{gap}})
\]
Identity the best criterion?
Substitution matrices
Alignments in brief

>1bl8_A  mol:protein length:97  Potassium Channel Protein
ALHWRAAGAATVLVLLLAGSYLAVLAERGAPGAQLIT
YPRALWWSVETATTGTYGDLYPVTLWGRVCAVVVVMVA
GITSFGLVTAAALATWFVFGREQ

>1orq_C  mol:protein length:223  Potassium Channel
IGDVMEHPLVELGVSYALLSVIVVVVECTMQLSGEYL
RLYLVDLILVILWADYAYRAYKSGDPAGYVKKTLYEI
PALVPAGLLALEGHLAGLGLFLRIVRLRFLRILLHSRG
SKFLSALADADKIRFYLFGAVMLTLYGAFAIYIVEY
PDPNSSIKSFDALWWAVVTATTGTYGDVVPATPIGVK
IGIAVMLTGISALTLLIGTVSNMFQKILV

Query= 1bl8_A  mol:protein length:97  Potassium Channel Protein
(97 letters)

>1orq_C  mol:protein length:223  Potassium Channel
Length = 223

Score = 58.5 bits (140), Expect = 4e-14
Identities = 26/72 (36%), Positives = 43/72 (59%)

Query: 21  GSYIALAERGAPGACLITYPRALWWSVETATTGTYGDLYPVTLWGRVCAVVVVMAGITS 80
G++ L+ + E  P + + +  ALWW+V TATTGTYGD+ P T  G+ + + VM+ GI++
Sbjct: 147  GAAFLIVVEYPDPSNSSIKSFDALWWAVVTATTGTYGDVVPATPIGVKIGIAVMLTGISA 206

Query: 81  FGLVTAALATWF 92
L + ++ F
Sbjct: 207  LTLLIGTVSNMF 218
Alignments in brief

Query= 1bl8_A  mol:protein length:97  Potassium Channel Protein  (97 letters)

>1orq_C  mol:protein length:223  Potassium Channel  
Length = 223

Score = 58.5 bits (140), Expect = 4e-14
Identities = 26/72 (36%), Positives = 43/72 (59%)

Query: 21  GSYLAVLAERGAPGACLITYPRLNWSSVETATTGYGDLYPVTLLWGRCVAVVVMVAGITS 80
       G++ E P + +  + ALW+V TATTGYGD+ P T + + VM+ GI++
Sbjct: 147  GAFAIYIVEYPDPNSSIKSVFDALWWAVVTATTGYGDVVPATPIGKVIGIAVMLTGISA 206

Query: 81  FGLVTAALATWF 92
L+ ++ F
Sbjct: 207  LTLLIGTVSNMF 218

Scoring matrix

Algorithm to optimize score
Alignment: Substitution Matrix: PAM

- **PAM**: Point Accepted Mutations (substitution matrix)

  Margaret Oakley Dayhoff (1925-1983)

  introduced 1978 based on 1572 observed mutations in 71 families of closely related (85% PIDE) proteins
  Protein Atlas (Atlas of Protein Sequence and Structure)
  (1st sequence: insulin: 1955 Fred Sanger, 1st structures: 1958: Perutz et al. (hemoglobin) / Kendrew et al. (myoglobin) )


- **PAM\textsuperscript{1}: 1 point mutation per hundred comparisons**

  \[ \text{PAM}^n = (\text{PAM}^1)^n \]
BLOSUM

- BLOcks of amino acid SUbstitution Matrices
  Align only conserved regions
- compile log-odd ratios

\[ S_{i,j} = \log \frac{p_i \cdot M_{i,j}}{p_i \cdot p_j} = \log \frac{M_{i,j}}{p_j} = \log \frac{\text{observed frequency}}{\text{expected frequency}} \]

- BLOSUM\(n\)=threshold at \(n\)% pairwise sequence identity

Alignments: scoring matrix

Scoring Matrix

© Burkhard Rost (TU Munich)
Alignments: scoring matrix

<table>
<thead>
<tr>
<th>C</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>2</td>
</tr>
<tr>
<td>T</td>
<td>1</td>
</tr>
<tr>
<td>P</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
</tr>
<tr>
<td>G</td>
<td>1</td>
</tr>
<tr>
<td>N</td>
<td>1</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
</tr>
<tr>
<td>Q</td>
<td>0</td>
</tr>
<tr>
<td>H</td>
<td>-1</td>
</tr>
<tr>
<td>R</td>
<td>-1</td>
</tr>
<tr>
<td>K</td>
<td>0</td>
</tr>
<tr>
<td>M</td>
<td>-2</td>
</tr>
<tr>
<td>I</td>
<td>-1</td>
</tr>
<tr>
<td>L</td>
<td>-3</td>
</tr>
<tr>
<td>V</td>
<td>-2</td>
</tr>
<tr>
<td>F</td>
<td>-4</td>
</tr>
<tr>
<td>Y</td>
<td>-3</td>
</tr>
<tr>
<td>W</td>
<td>-8</td>
</tr>
</tbody>
</table>

Scoring Matrix

© Burkhard Rost (TU Munich)
### Alignments: scoring matrix

#### Scoring Matrix

```
<table>
<thead>
<tr>
<th></th>
<th>C</th>
<th>S</th>
<th>T</th>
<th>P</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>0</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>-2</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>-3</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>1</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>N</td>
<td>-4</td>
<td>1</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>D</td>
<td>-5</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>E</td>
<td>-5</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Q</td>
<td>-5</td>
<td>-1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
<tr>
<td>H</td>
<td>-3</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>-2</td>
<td>2</td>
</tr>
<tr>
<td>R</td>
<td>-4</td>
<td>0</td>
<td>-1</td>
<td>0</td>
<td>-2</td>
<td>3</td>
</tr>
<tr>
<td>K</td>
<td>-5</td>
<td>0</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>2</td>
</tr>
<tr>
<td>M</td>
<td>-5</td>
<td>-2</td>
<td>-1</td>
<td>-1</td>
<td>-3</td>
<td>2</td>
</tr>
<tr>
<td>L</td>
<td>-6</td>
<td>-3</td>
<td>-2</td>
<td>-3</td>
<td>4</td>
<td>-4</td>
</tr>
<tr>
<td>V</td>
<td>-2</td>
<td>-1</td>
<td>0</td>
<td>-1</td>
<td>-1</td>
<td>2</td>
</tr>
<tr>
<td>F</td>
<td>-4</td>
<td>-3</td>
<td>-3</td>
<td>-5</td>
<td>-5</td>
<td>2</td>
</tr>
<tr>
<td>Y</td>
<td>-3</td>
<td>-3</td>
<td>-5</td>
<td>-7</td>
<td>4</td>
<td>2</td>
</tr>
<tr>
<td>W</td>
<td>-8</td>
<td>-2</td>
<td>-5</td>
<td>4</td>
<td>7</td>
<td>2</td>
</tr>
</tbody>
</table>
```

- The scoring matrix is used to calculate the similarity between sequences.
- The matrix entries represent the cost of aligning two characters: positive for matches, negative for mismatches, and 0 for gaps.
- The alignment is determined by the path with the lowest cost from the top left to the bottom right of the matrix.

© Burkhard Rost (TU Munich)
Other substitution matrices

- BLOSUM-62 took the day
- many more (sequence based):
  - GONNET
  - McLachlan
  - Claverie
- structure based
  - PHAT, SLIM
  - STROMA, SDM, HSDM, SM_SAUSAGE, AGAPE
- function based
- for threading/fold recognition
- asf. ....
Interactive software tool

Ignacio Ibarra & Francisco Melo:

Interactive software tool to comprehend the calculation of optimal sequence alignments with dynamic programming

Bioinformatics 2010, 26:1664-5

http:/melolab.org/sat
Dynamic programming

- time used?

- how to choose parameters?
Alignment: hashing (fast & dirty)
BLAST: fast matching of single ‘words’

TTYKLILNGKTLKGETTTEAVDAATAEKFVKQYANDNGVDGWTYDDATKTFVTTEK
TTYKLIIILLLLLLLLLLMMMLMLLLLAWTVEKAFKTFAAAAAAWTVEKAFKTFAAAAA

Default “word” size for “seeds” = 3
BLAST: fast matching of single ‘words’

Default “word” size for “seeds” = 3
BLAST: fast matching of single ‘words’

TTYKLILNLKTLGETTTEAVDAATAEKFQYANDNGVDGEWTVYDDATKTFTVTEK
TTYKLILLLLLLLLLLLLLLLLLLLLLAWTVEKAFKTFAAAAAAAAAAWTVEKAFKTFAAAAA

Default “word” size for “seeds” = 3
BLAST: fast matching of single ‘words’

TTYKLILNGKTLKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEK
TTYKLILLLLALLLLLLALLAWTVEKAFKTFAAAAAAAAWTVEKAFKTFAAAAAA

Default “word” size for “seeds” = 3
BLAST: fast matching of single ‘words’

Default “word” size for “seeds” = 3
Minds behind hashing

FASTA (extension of FASTN/FASTP)
DJ Lipman & WR Pearson (1985)
Science 227: 1435-41

BLAST
Basic Local Alignment Search Tool
SF Altschul, W Gish, W Miller, EW Myers, DJ Lipman (1990) JMB
215:403-10
BLAST: fast matching of single ‘words’

TTYKLILNGKTLKGETTTEAVDAATAEKEFKQYANDNGVDGEWTTYDDATKTFTVTEK
TTYKLILLLLLLLLLLLLAWTVEKAFKTFAAAABBBBBBBBBBBBBBBB

Default “word” size for “seeds” = 3
BLAST: fast matching of single ‘words’

TTYKLLNGKTLKGETTTEAVDAATAEKEKQVYANDNGVDGWEWTYDDATKTFTVTEK
TTYKLLLLLLLLLLLLLLLLLLLLLWAVEKAFKTFAAAAAAAAATVEKAFKTFAAAAA

Default “word” size for “seeds” = 3
BLAST: fast matching of single ‘words’

TTYKLILNGKTLKGETTTEAVDAATAEKFQYANDNGVDGEWTYDDATKTFTVTEK
TTYKLILLLLLLLLLLLLTTLLLLAWTVEKAFKTFAAAAAAAATVEKAFKTFAAAAA

Default “word” size for “seeds” = 3
BLAST: fast matching of single ‘words’

TTYKLILNGKTLKGETTTEAVDAATAEKFJKQYANDNGVDGEWTYDDATKFTFTVTEK
TTYKLILLLLLLLLLLLLLLLLLLAWTVEKAFKTFAAAAAAAWTVEKAFKTF

TTYKLIL
TTYKLIL

AATAEKFJKQYA
AWTVEKAFKTF

WTYDDATKTF
WTVEKAFKTF

Default “word” size for “seeds” = 3
BLAST: fast matching of single ‘words’

Default “word” size for “seeds” = 3
the major challenge for word search algorithms is to get the statistics right
Significance of match (e.g. BLAST E-values)

Background (database)

Homologues

Distribution (fraction of counts)

Score
Different approaches for background

- SEARCHSEQ
  - permutations of SEARCHSEQ

- FASTA

- pre-compile distribution for entire database

- BLAST
How accurate are pairwise alignments?
Annotation transfer

similar sequence \[\rightarrow\] similar function

similar sequence \[\rightarrow\] similar structure
How to assess alignment accuracy?
How to assess alignment accuracy?
All-vs-all: PDB

1D = sequence alignment
All-vs-all: PDB

3D = structural alignment

1D = sequence alignment
All-vs-all: PDB

3D = structural alignment
1D = sequence alignment

<0.2nm rmsd — SAME 3D
in between  — ignore
>0.5nm rmsd — DIFFER in 3D
PDB all-against-all ok?

proteins of known 3D structure (PDB)
Databases biased: MUST remove bias!
Sequence conservation of protein structure

B Rost 1999 Prot Engin 12, 85-94
Sequence conservation of protein structure

B Rost 1999 Prot Engin 12, 85-94
Structure prediction using homology


Wednesday May 1, 2013
Percentage of identical residues

Number of residues aligned

Score = 83.2 bits (205), Expect = 9e-17
Identities = 18/101 (X%), Positives = 36/101 (35%), Gaps = 2/101 (1%)

Query: 111 AAGGIAAKYLARKNSSVFGFIGCGTQAYFQLEALRRVFDIGEVKAYDVREKAAKKF 170
   + +  +      +G G       ++L     +  V   +   + A +
Sbjct: 153 AAVELAERELGSLHDKTVLVVGAGEMGKTVAKSLVD-RGVRAVLFANRTYERAVEL 211

Query: 171 EDRGISASVQPAEEASRCDVLVTTPSRKPVVKAEEVEECT 211
   +       +R DV+V+ T +  PV+ + V E
Sbjct: 212 GGEAVRFDE-LVDHLARSDDVVSAAPHPVIHVDDVERAEAL 251

Structure prediction using homology

P1

P2


Wednesday May 1, 2013

Structure prediction using homology

---

Score = 83.2 bits (205), Expect = 9e-17

Identities = 18/101 (X%), Positives = 36/101 (35%), Gaps = 2/101 (1%)

**Query:**

111 AAGGIAAKYLARKNSSVFGFIGCGTQAYFQLEALRRVFDIGEVKAYDVREKAKKF 170

**Sbjct:**

153 AAVELAERELGSLHDKTVLVVGAGEMGKTVAKSLVD-RGVRAVLVANRTYERAVEL 211

**Query:**

171 EDRGISASVQPAEEASCDVLVTTPSRKVAKAEWVEEGT 211

**Sbjct:**

212 GGEAVRFDE-LVDHLARSEAHPVIHVDDVREAL 251

---

Wednesday May 1, 2013
Structure prediction using homology


Percentage of identical residues
Number of residues aligned

Score = 83.2 bits (205), Expect = 9e-17
Identities = 18/101 (X%), Positives = 36/101 (35%), Gaps = 2/101 (1%)

Query: 111 AAGGIAAKYLARKNSSVFGFIGCGTQAYFQLEALRRVFDIGEVKAYDREKAKKF 170
Sbjct: 153 AAVELAERELGSLHDKTVLVGAGEmKTVAKSLV-UGRAVILVRNRTYERAVEL 211

Query: 171 EDRGISASVQPAEASRCVDLVTTPSRKPVVKAEWVEEGT 211
Sbjct: 212 GGEAVRFDE-LVDHLARSDVVSATAAPHVIVHHDDVREAL 251


Wednesday May 1, 2013
Structure prediction using homology

Score = 33.9 bits (77), Expect = 0.068
Identities = 14/58 (y%), Positives = 28/58 (48%), Gaps = 2/58 (3%)

Query: 178 SVQPAEAESRCDELVTTPSRTPVVKAEBVEEGTHINAIGADGPKQELD-VEILKKA 234
    + EE ++ D+LV T + +VK EW++ G + G + ++ E ++A
Sbjct: 198 TAHLDEEKNKDILVVATAPQPE-MVKGEIKPAGVIDCGINYKVGDVAYDEAKERA 254

Percentage of identical residues

Score = 33.9 bits (77), Expect = 0.068
Identities = 14/58 (y%), Positives = 28/58 (48%), Gaps = 2/58 (3%)

Query: 178  SVQPAEASRCVTLTTPSRFPVKAEWVEETHNAIGADGPQELD-VEILKKA 234
+ EE ++ D+LV T + +VK EW++ G + G + ++ E ++A

Sbjct: 198  TMLDEEVNGDILVATGQPG-MVKGEWIKPGAIIDCGINKVVGDVAYDEAKERA 254

Sequence conservation of protein structure

B Rost 1999 Prot Engin 12, 85-94
Sequence conservation of protein structure

B Rost 1999 *Prot Engin* 12, 85-94
How to estimate performance from the curves?
Distance from new HSSP-curve

Sequence identity implies structural similarity!

Don't know region

Distance from curve = +10

Distance from curve = -10

Percentage sequence identity

Number of residues aligned

B Rost 1999 Prot Engin 12, 85-94

© Burkhard Rost (TU Munich)
Distance from new HSSP-curve

B Rost \textit{Prot Engin} 12, 85-94
Twilight zone = false positives explode!!

Number of protein pairs

Distance from HSSP threshold

Percentage sequence identity

Number of residues aligned

B Rost 1999 Prot Engin 12, 85-94
Twilight zone = false positives explode!!

B Rost 1999 Prot Engin 12, 85-94
Detecting true hits in Twilight zone

- Percentage of cumulative true positives
- Distance from threshold

**similarity-larger-than-identity**

**they-dont-know-what-they-do only sequence identity**

B Rost 1999 Prot Engin 12, 85-94

© Burkhard Rost (TU Munich)
Finding similar structures in Twilight zone

B Rost 1999 Prot Engin 12, 85-94

© Burkhard Rost (TU Munich)

Wednesday May 1, 2013
Accuracy vs. Coverage

ACCURACY = the number of correct proteins found

COVERAGE = how many of the proteins found are correct?

Accuracy vs. Coverage graph
BLAST is not enough ...

they-don't-know-what-they-do only sequence identity

B Rost 1999 Prot Engin 12, 85-94

© Burkhard Rost (TU Munich)
Threshold in percentage sequence identity

- PIDE
- >100, >30%
- new HSSP
Threshold in percentage sequence identity
So far: pairwise - anything more to do?
Triangle beyond reach

- Protein A
- Protein B
- Protein C
Sequence Space Hopping
Success through sequence space hopping

![Graph showing percentage sequence identity against distance from threshold with two curves labeled 'ide' and 'old'.]
Sequence comparisons: multiple alignment
Multiple alignments

Dynamic programming?
for 3 sequences: $O(N_1 \times N_2 \times N_3)$
NP-complete (L Wang & T Jiang (1994) JCB 1: 337-48)

<table>
<thead>
<tr>
<th></th>
<th>G</th>
<th>G</th>
<th>Q</th>
<th>L</th>
<th>A</th>
<th>K</th>
<th>E</th>
<th>E</th>
<th>A</th>
<th>L</th>
</tr>
</thead>
<tbody>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Q</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>P</td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>E</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>P</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Multiple alignments

- Dynamic programming?
  for 3 sequences: $O(N_1 \times N_2 \times N_3)$
  NP-complete (L Wang & T Jiang (1994) JCB 1: 337-48)

- hack 1:
  dynamic programming: pairwise, only space in vicinity of intersection searched n-wise
Multiple alignments

- Dynamic programming?
  for 3 sequences: $O(N_1 \times N_2 \times N_3)$
  NP-complete (L Wang & T Jiang (1994) JCB 1: 337-48)

- hack 1:
  dynamic programming: pairwise, only space
  in vicinity of intersection searched n-wise

- hack 2:
  map to tree / pairwise
  Russell Doolittle, UCSD

Russell Doolittle
Multiple alignment: progressive 1

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90</td>
<td>80</td>
<td>70</td>
</tr>
<tr>
<td>B</td>
<td>90</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

GGQLAKEEAL
GGQLAKDEAL
GGQIAKDEAL
GGQIAKDEAI
Multiple alignment: progressive

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>90</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Step 1

GGQLAKEEAL
GGQLAKDEAL
GGQIAKDEAI

Wednesday May 1, 2013
Multiple alignment: progressive

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>90</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Step 1**
- GGQLAKEEAL
- GGQLAKDEAL
- ggqlakeeal

**Step 2**
- GGQIAKDEAL
- GGQIAKDEAI
- ggqiaakdeal
Multiple alignment: progressive 1

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>90</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Step 1
- GGQLAKEEAL
- GGQLAKDEAL
- ggqlakeeal

Step 2
- GGQIAKDEAL
- GGQIAKDEAI
- ggqiaakdeal

Step 3
- GGQIAKDEAL
- GGQIAKDEAI
- ggqiaakdeal
### Step 1

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td></td>
<td>90</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td></td>
<td>90</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Step 1**
- GGQLAKEEAL
- GGQLAKDEAL
- GGQIAKDEAL
- GGQIAKDEAI

**Multiple alignment: progressive 2**
Multiple alignment: progressive 2

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>90</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td></td>
<td>90</td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Step 1
- GGQLAKEEAL
- GGQLAKDEAL
- ggqlakeeal

Step 2
- GGQIAKDEAL
- GGQIAKDEAI
- ggqlakeeal
Multiple alignment: progressive 2

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90</td>
<td>80</td>
<td>70</td>
<td></td>
</tr>
<tr>
<td>B</td>
<td>90</td>
<td>80</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>90</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Step 1
- GGQLAKEEAL
- GGQLAKDEAL
- ggqlakeeal

Step 2
- GGQIAKDEAL
- ggqiakeal

Step 3
- GGQIAKDEAI
Today: Alignments

☐ LAST WEEKs
  • 3D comparison

☐ THIS WEEK
  • Alignments and “reach of comparative modeling”

☐ NEXT WEEK
  • alignment contd
Lecture plan (PP1: Structure)

- 01: 2013/04/16 Tue: welcome: who we are
- 02: 2013/04/18 Thu: no room, no lecture
- 03: 2013/04/23 Tue: intro I - acids/structure
- 05: 2013/04/30 Tue: intro IIa - 3D comparisons
- 06: 2013/05/02 Thu: alignment 1
- 07: 2013/05/07 Tue: alignment 2
- 08: 2013/05/09 Thu: holiday (Ascension Day)
- 09: 2013/05/14 Tue: no lecture: student assembly (true for 2013!!)
- 10: 2013/05/16 Thu: Andrea Schafferhans: 3D compare 2
- 11: 2013/05/21 Tue: Whitsun holiday - Andrea Schafferhans: Comparative Modeling 2
- 12: 2013/05/23 Thu: Andrea Schafferhans: Comparative Modeling 1
- 13: 2013/05/28 Tue: Andrea Schafferhans: Comparative Modeling 2
- 14: 2013/05/30 Thu: Fronleichnam=Corpus Christi - Alignment 2
- 15: 2013/06/04 Tue: 3D->1D: sec str
- 16: 2013/06/06 Thu: sec str pred 1 (white board)
- 17: 2013/06/11 Tue: sec str pred 2
- 18: 2013/06/13 Thu: sec str pred 3
- 19: 2013/06/18 Tue: transmembrane helix prediction
- 20: 2013/06/20 Thu: transmembrane strand prediction, solvent accessibility
- 21: 2013/06/25 Tue: 2D prediction, intro
- 22: 2013/06/27 Thu: 2D prediction - Thomas Hopf
- 23: 2013/07/02 Tue: 3D prediction
- 24: 2013/07/04 Thu: summary: what we do in our group
- 25: 2013/07/09 Tue: examen, no lecture
- 26: 2013/07/11 Thu: no lecture
- 27: 2013/07/16: no lecture
- 28: 2013/07/18: no lecture