

**Algorithms in Genome Research**  
**Winter 2005/2006**

**Exercises**

**Sheet 1, Discussion: 03.11.2005**

1. Recall the different types of data that were mentioned in class: sequence data, structural data, mass spectra, 2D gels, ...
  - (a) To what scientific questions/problems may these kinds of data provide answers?
  - (b) Use EMBL Nucleotide Sequence Database, NCBI Database and Swiss2Dpage to find out:
    - i. How many genes are known for the human chromosome 4?
    - ii. At which chromosome band position is gene ALB located?
    - iii. For which protein does ALB code for?
    - iv. What is the function of this protein?
    - v. Which 2D page maps contain this protein?
    - vi. Click on any map location of our protein. Why is there shown a red box and a dashed red box?
2. Remember physical mapping by clone-probe hybridization.
  - (a) What are the main assumptions when the problem is modelled as the consecutive ones problem?
  - (b) Discuss experimental reasons why the assumptions do not hold in practice.
3. Solve the consecutive-ones problem for the following clone-probe hybridization matrix  $M$  (if possible).

$$M = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 1 & 1 & 1 & 0 \end{pmatrix}$$

4. For the following matrix  $M$  create the graph  $G(M)$  and solve the Travelling Salesman Pro-

blem.

$$M = \begin{pmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

- (a) What is the length of the shortest tour?
- (b) What is the probe order that this tour corresponds to?
- (c) What is the overall number of blocks of consecutive ones in this order?