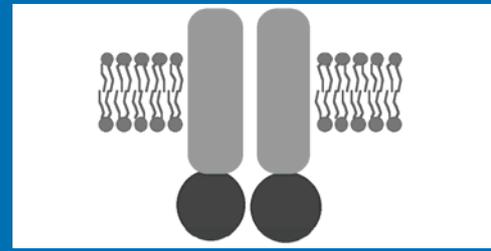


From gene to insilico structure

Projektleitung: Dr. Sander Smits & Benjamin Wenn



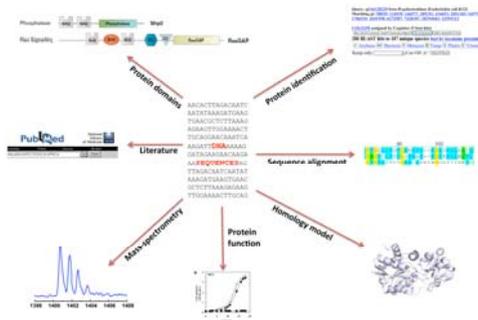
Die Idee

Das Geheimnis des Lebens liegt verborgen in der DNA. Diese DNA kodiert eine Proteinsequenz, die aus einer für jedes Protein spezifischen Funktionsweise resultiert. Mit steigender Anzahl von DNA-Genomen, deren Sequenzen bekannt sind, steigt auch die Anzahl von bioinformatischen Daten, die in verschiedenen Datenbanken im Internet auffindbar sind. Mit Hilfe der primären DNA Sequenz ist es möglich heraus zu finden:

- um welches Protein es sich handelt
- was die Funktion des Proteins ist
- in welchen anderen Organismen dieses Protein vorkommt
- welche biochemischen und physiologischen Eigenschaften eine Rolle spielen
- welche Aminosäuren an der Funktion beteiligt sind
- welche Mutationen entdeckt und bearbeitet wurden
- eventuell kann man die 3-dimensionale Struktur des Proteins vorhersagen.

Das Modul "From gene to insilico structure" zeigt und erklärt die Vielfalt von Bioinformatik-Programmen im Internet, die jeder Biochemiker, Biologe, Molekularbiologe, Chemiker, Biophysiker in seinem späteren Beruf täglich nutzen wird. Mithilfe dieser Programme ist es möglich, eine Vielzahl von Informationen zu gewinnen, deren Ergebnisse später in den Laborexperimenten eingesetzt werden.

Die Möglichkeiten



Die Webseiten

Eine Vielzahl von Webseiten sind im Internet kostenlos zu Benutzen. Deren Layout, Methode, Hintergrund, und Benutzungsweise sind jedoch unterschiedlicher Natur.

Um die Programme im Internet richtig anzuwenden und zu finden haben wir für das Modul eine Webseite erstellt, die sowohl die Programme theoretisch erklärt als auch mit praktischen Übungen zeigt, wie man diese anwenden sollte, um zu einem richtigen Ergebnis zu kommen.

Das Modul

Den Studierenden wird innerhalb des Moduls die Möglichkeit gegeben sich das Wissen über verschiedene Wege zu erarbeiten.

- eine Erklärung der Theorie mit Hintergrundinformationen (1)
- Übungen zur Lernziel Kontrolle (2)
- Weiterführende Literatur (3)
- Aufzeichnung der Vorlesung mit Lecturnity (4)
- Video-Erklärungen der verschiedenen Webseiten (erstellt mit Camtasia) (5)
- Podcast der Vorlesung für iPod und iPhone (6)



Exercise 1: Genome to ORF (2)

Genome search

Many genomes have been sequenced during the last two decades. The overwhelming amount of files is archived on the internet. You just have to find them. An fast way to searching the genome sequence, the genome fragments have to be put in right order. Searching will be the sequence pieces fit together forming one single piece of DNA. This piece of DNA is then taken as template to search for open reading frames or called "ORF".

On the following pages you will find 17 pieces of DNA fragments from a random sequenced genome using the data you received.

- Make a note of this sequence?
- Are there any gaps in the sequence. Do you think that the missing parts are essential. If yes how can we obtain that sequence?
- On the last four pages of the document another complete sequence is downloaded from the [www.expasy.org](http://expasy.org) website. Make an alignment with your coding. Does the sequence match this one?
- Find all genes in this sequence? Using the program gene finder at the NCBI website (<http://www.ncbi.nlm.nih.gov/blast/gene/>).

RESEARCH ARTICLE (3)

Draft Genome Sequence of the Sexually Transmitted Pathogen *Trichomonas vaginalis*

Trichomonas vaginalis is a common sexually transmitted pathogen that causes vaginitis in women and proctitis in men. The draft genome sequence of this organism is presented here. The genome is approximately 200 Mb in size and contains approximately 10,000 genes. The sequence is highly AT-rich and contains many repetitive elements. The draft genome sequence is available at <http://www.ncbi.nlm.nih.gov/genbank/tx/>.

modul 1: genome to orf (1)

1. A list of Genes (Sequences)

2. 17 Genes' Sequence in fasta

3. 17 Genes' Translation

4. 17 Genes' In silico ORF

5. 17 Genes' In silico ORF alignment

6. 17 Genes' In silico ORF alignment

7. 17 Genes' In silico ORF alignment

8. 17 Genes' In silico ORF alignment

9. 17 Genes' In silico ORF alignment

10. 17 Genes' In silico ORF alignment

11. 17 Genes' In silico ORF alignment

12. 17 Genes' In silico ORF alignment

13. 17 Genes' In silico ORF alignment

14. 17 Genes' In silico ORF alignment

15. 17 Genes' In silico ORF alignment

16. 17 Genes' In silico ORF alignment

17. 17 Genes' In silico ORF alignment

A History of Genome Sequencing

The sequencing of the human genome along with related organisms is one of the largest scientific endeavors in the history of mankind. The information gained from sequencing will provide the raw data for the emerging field of bioinformatics, which computer science and biology have combined to form. The large scale sequencing proposed by the Human Genome Project in 1990 could never have been a reality without modern computer facilities. Shortly, twenty years ago, complete genome sequencing was a daunting task in light of such a daunting amount and size of data. The sequencing of the human genome is a complex task that requires a lot of resources and a lot of money. The sequencing of the human genome is a complex task that requires a lot of resources and a lot of money. The sequencing of the human genome is a complex task that requires a lot of resources and a lot of money.

from gene to "in silico" structure 2010 (6)

Podcasts

The following podcasts are available for your iPod or iPhone:

- Podcast 1: DNA sequencing and genomes
- Podcast 2: DNA sequencing and genomes
- Podcast 3: DNA sequencing and genomes
- Podcast 4: DNA sequencing and genomes
- Podcast 5: DNA sequencing and genomes
- Podcast 6: DNA sequencing and genomes

How to translate DNA to protein sequence (5)

with the website

www.expasy.ch

Dr. Sander Smits
Institute for Biochemistry
University of Duisburg-Essen
www.expasy.ch

Module 1: (4)

DNA sequencing and genomes

Enter gene in "in silico" structure