

C C A T T 0 1 0 0 0
G A G G A 0 1 1 0 1
G A A T T 0 0 1 1 0
A C A A G 0 0 1 0 0
T A C C A 0 0 1 1 0
T T A C A 0 1 0 0 0
A C C T C 0 0 0 1 0
A A G G A 0 0 0 0 0
G A T G A 0 1 1 0 0
T A G A T 0 0 1 0 0
G A T G A 1 0 1 0 0
T G T A G 1 0 0 0 0
T A G T A 0 0 0 0 0
G A T A T 1 0 0 0 0
G A G T G 1 1 0 0 0
A G A T T 1 1 0 0 0
G A G T A 1 1 0 0 0
T G A T G 1 1 0 0 0
A T T A G 1 1 0 0 0
T A G A T 1 1 0 0 0
T A G T G 1 1 0 0 0
G A G A A 1 1 0 0 0
G T A A A 1 1 0 0 0
G A T A A 1 1 0 0 0
T A G A A 1 1 0 0 0
A G A A A 1 1 0 0 0
G A G A A 1 1 0 0 0
A A A A A 1 1 0 0 0

Case Study

Bioinformatics in Molecular Parasitology - A case study on malaria vaccine research

October 4, 2006



Bioinformatics in Molecular Parasitology - A case study on malaria vaccine research

Bioinformatics, or tools for DNA and protein sequence search and analysis, have developed rapidly during the past 5 to 10 years. Using bioinformatics software solutions for sequence search and sequence analysis, in combination with other techniques, constitute a timesaving and costeffective way to obtain important data on gene level as well as protein level - information not easily obtainable by other techniques.

CLC bio has leveraged on this development and created advanced bioinformatics Workbenches that can be used for a number of scientific purposes.

One important research area is molecular parasitology as infections by parasites constitute a lethal threat to more than one billion humans worldwide.

Below is a description of the nature of malaria and of how CLC Combined Workbench can help in discovering the secrets of malaria's scientifically fascinating - but unfortunately often lethal - nature.

About Malaria

The World Health Organization (WHO) estimated in 1996 that the malaria parasite had infected more than 300,000,000 people and was responsible for more than 1,000,000 deaths annually.

Malaria is a parasitic disease transmitted through the blood meal of infected mosquitoes, which transmit infecting cells, sporozoites, into the mammalian host. Within minutes, the sporozoites invade the liver (hepatocytes) and develop into what is called "merozoites" within each cell.

The merozoite-infected cells then burst, and the merozoites invade the red blood cells, producing the various symptoms associated with malaria. The life-cycle is completed when germ cells, gametocytes, are produced from the infected blood cells. When another mosquito stings, the gametocyte is transferred through the blood meal and infects this new mosquito (see figure 1). Hereafter the malaria life cycle repeats.

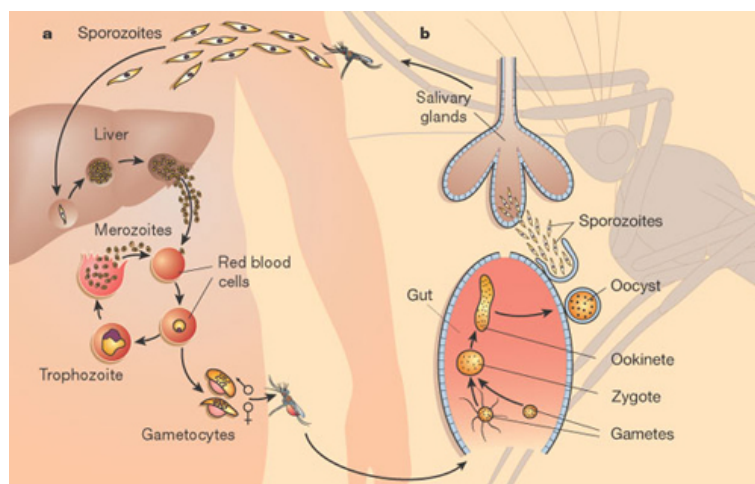


Figure 1: Ménard, Robert, 2005: Knockout malaria vaccine? (Nature 433, 113-114).

Upon repeated malaria infections, partial antibody-dependent immunity directed against the erythrocytic (red blood cell) stage is elicited in humans.

One of the major research foci today is development of vaccine to prevent life cycle dependent parasitic penetration of target organs e.g. liver or red blood cells.

Bioinformatics in malaria vaccine development

The complete genome of the malaria parasite, Plasmodium, has been sequenced and consists of about 14.5 megabases of DNA bearing more than 6000 genes. Over 500 genes have already been predicted from direct genome sequences, and an additional 2800 unique examples have been found from processing expressed sequence tags (ESTs).

Using sequence analysis to study the genomic setup and the identity and functionality of glycoproteins expressed on the surface of the different stage of the parasite has increase the detailed molecular understanding of how the malaria disease develops (pathogenesis).

In addition to identifying potential drug targets, further application of bioinformatics can provide information about virulence, antigenicity, evolution, and gene and protein interactions. These genomic sequence data constitute the basis of vaccine research today.

Bioinformatics in malaria vaccine development is used as a tool for

- Identification of target antigen
- Immunogenic analysis of target antigen
- Selection of target antigen

How to use CLC bio’s software for vaccine research

An example of how bioinformatics can be applied in vaccine research is described below.

The sporozoites (see figure 1) express several proteins on the surface. Some of them have interest as target for vaccine development (see figure 2).

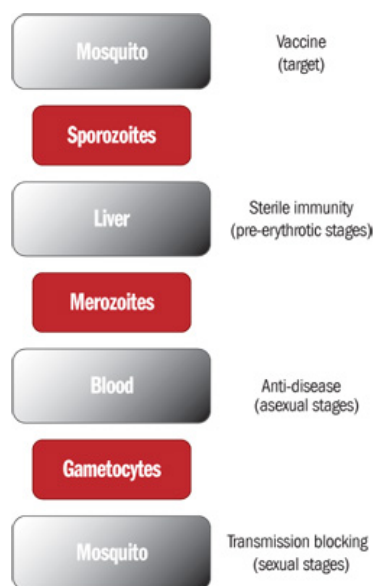


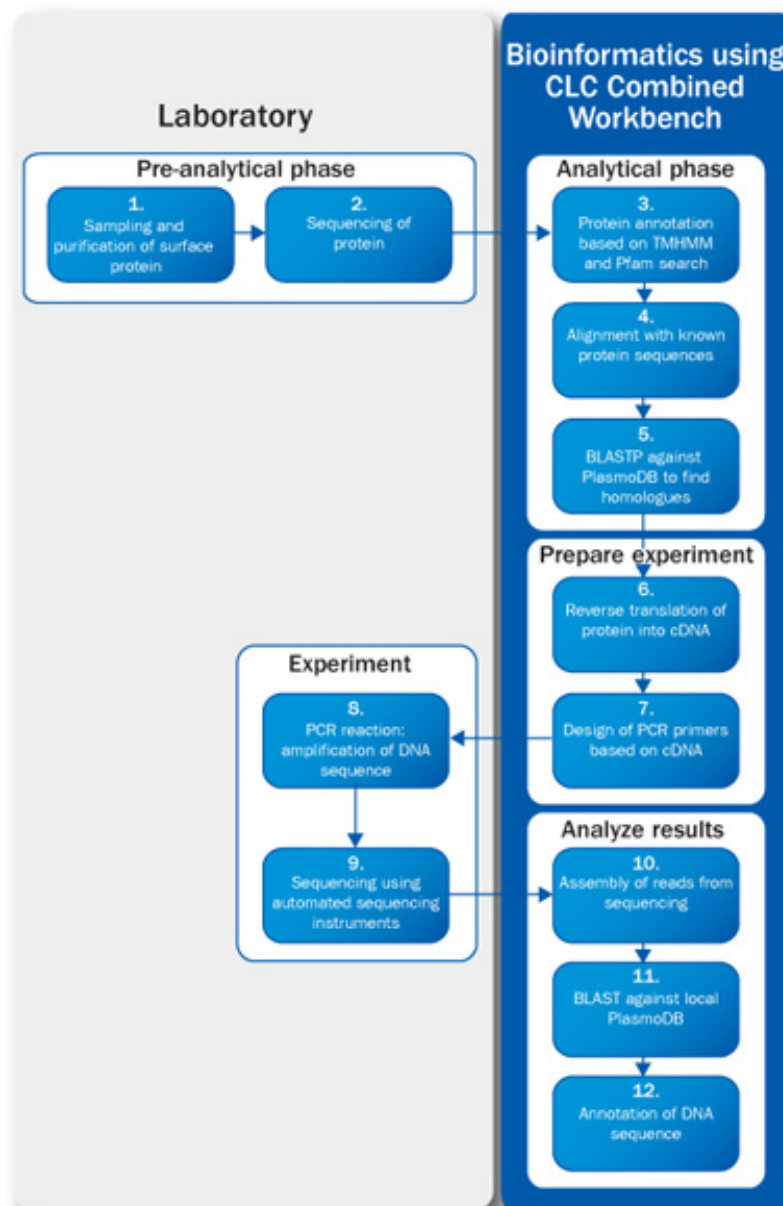
Figure 2:



In the laboratory, the protein of interest initially has to be isolated, purified, sequenced and subsequently cloned.

Hereafter the lab work is to isolate the DNA from overnight cultures, and then to sequence the isolated cDNA clones.

When both the protein and the DNA sequence are known, the bioinformatics tools come into play. Using database searches on both National Center for Biotechnology Information (NCBI) and the PlasmoDB, which can be downloaded and searched using CLC Combined Workbench, homologue sequences are discovered. Sequence analyses and alignment with other sequences is performed in desktop programs like CLC Combined Workbench, and these analyses make up the basis for a better understanding of the characteristics of the protein. This understanding, in turn, makes it possible to define a target for development of vaccines against malaria.



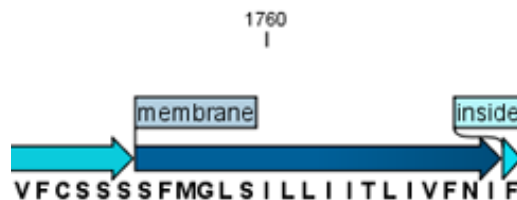
Zooming in on the analytical phase

Step 3. Automated protein annotation based on TMHMM and PFAM search

The protein sequence is imported into CLC Combined Workbench, and two types of analyses are performed:

- Transmembrane Helix Prediction (TMHMM)
- Finding PFAM domains

The Workbench automatically annotates the sequence. The annotations are shown graphically on the screen and the sequence is saved for further analysis.



Step 4. Alignment with proteins from previous research

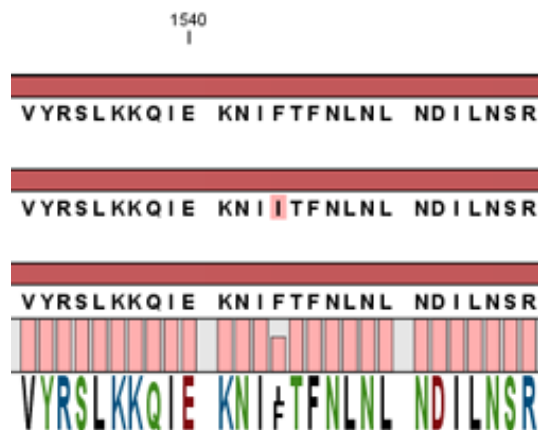
The protein is now aligned with other known and relevant proteins that the user has worked with in earlier research.

Sequence similarities between the known proteins and the newly sequenced protein are investigated.

Many of the known proteins already have functional annotations in regions that are similar to regions in the unknown protein.

This is a strong indication of similar function in the unknown protein. And through CLC Combined Workbench the annotations of interest are smoothly transferred to the protein under investigation.

The protein sequence, now including a number of functional annotations is saved for further analysis.





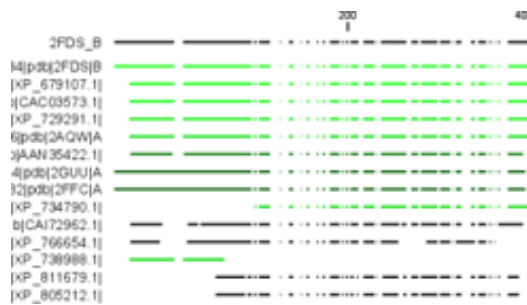
Step 5. BLASTP against PlasmoDB

A database including plasmodium proteins is downloaded from <http://www.plasmodb.org/>.

The data is in a FASTA-file which the user converts to a local BLAST database through CLC Combined Workbench. This is done with a few mouse-clicks.

CLC Combined Workbench is hereafter used for performing a BLASTP search against the database to find plasmodium-specific homologues. The speed is high as the BLASTP search is performed on the user's own computer - not through a web browser.

The BLASTP results are further analyzed in order to analyze the protein, and additional functional annotations and other types of annotations are added to the protein sequence.



Abbreviations

BLAST: Basic Local Alignment Search Tool PlasmoDB: specific sequence database for genomic search in Plasmodium falciparum. Erythrocytes: red blood cells TMHMM: One of the best algorithms for transmembrane helix prediction

Creative Commons License

All CLC bio's scientific articles are licensed under a Creative Commons Attribution-NonCommercial-NoDerivs 2.5 License. You are free to copy, distribute, display, and use the work for educational purposes, under the following conditions: You must attribute the work in it's original form and "CLC bio" has to be clearly labeled as author and provider of the work. You may not use this work for commercial purposes. You may not alter, transform, or build upon this work.

See <http://creativecommons.org/licenses/by-nc-nd/2.5/> for more about how you may use the contents.

