EMBnet.news

Editorial

New design, new front cover, new articles, new editorial board, it seems that nothing was left over from the previous EMBnet.news!

Although we made a lot of changes, we tried to stick to the spirit of the journal. Since its beginnings in July 1994, our journal targets the EMBnet user community providing information about the nodes and descriptions of useful bioinformatics tools. In this issue and in the future, we will provide articles about EMBnet nodes (e.g. Focus on UMBER), technical tutorials (e.g. MacOSX and Bioinformatics), and more scientific articles (e.g. DNA Microarray technology). We are open to publish articles from our user community and encourage our readers to submit their manuscripts. Of course we will favor articles dealing with topics in bioinformatics or to the use of bioinformatics for successful science.

To broaden our scope and starting with this issue of EMBnet.news, we are happy to announce a new collaboration between EMBnet.news and the SWISS-PROT team publishing the "Protein Spotlight" review. This monthly review selects proteins of interest to the general public. From anthrax toxins to prion proteins and ricin, from antifreeze glycoprotein to defensins and cocaine esterase, all of these proteins are "famous" in the media, but not really well described by our newspapers. The goal of "Protein Spotlight" is to popularize science by providing people with short and accurate information about proteins. We hope that you will enjoy this valuable addition to our journal.

The editorial board:
Erik Bongcam-Rudloff
Laurent Falquet
Pedro Fernandes
Gonçalo Guimaraes Pereira

Cover picture: DNA sculpture Bror Marklund 1973, Linnaeus Centre, Uppsala, [© Erik Bongcam-Rudloff]

Contents

Editorial Board:

Erik Bongcam-Rudloff, "The Linnaeus Centre for Bioinformatics" LCB, SLU/UU, SE
Email: erik.bongcam@bmc.uu.se
Tel: +46-(0)18-4716696
Fax: +46-(0)18-4714525

Laurent Falquet, SIB, Lausanne, CH
Email: laurent.falquet@isb-sib.ch
Tel: +4121 6925954
Fax: +4121 6925945

Pedro Fernandes, Instituto Gulbenkian, PT
Email: pfern@igc.gulbenkian.pt
Tel: +315 214407912
Fax: +315 214407970

Gonçalo Guimaraes Pereira, UNICAMP, BR
Email: goncalo@unicamp.br
Tel: +55-19-37886237/6238
Fax: +55-19-37886235

Focus on UMBER
University of Manchester Bioinformatics Education & Research

T.K. Attwood

School of Biological Sciences & Department of Computer Science
University of Manchester
Oxford Road
Manchester M13 9PT, UK

Introduction

In March 1999, the Specialist Node at UCL transferred to the University of Manchester. Although that was only 5 minutes ago, incredibly, 4 years have passed in the meantime! Much has happened during this time, so what follows is an update on UMBER's bioinformatics research and education activities.

Our work has centred for some years around the provision of the PRINTS protein fingerprint database [1] and associated tools for searching and visualising protein sequence information: e.g., the Expert Analysis SYstem, EASY, which combines protein sequence and family database searches to provide a consensus diagnosis of the identity of a protein in the context of the family to which it belongs [2]; FingerPRINTScan, the custom-built tool for searches against fingerprints in the current version of PRINTS [3]; BLAST PRINTS, for searches of sequences matched in the current release [4]; and the Colour INteractive Editor for Multiple Alignments, CINEMA [5]). Another major theme of the group has been the development of on-line materials for education, notably BioActivity, our interactive bioinformatics practical [6].
Some of these activities have been supported either directly from EMBnet's coffers, or through collaboration with other EMBnet Nodes. In the sections that follow, we outline our progress in these and related areas.

PRINTS and its automatic supplement, prePRINTS

PRINTS is a database of 'fingerprints' – i.e., groups of conserved motifs excised from sequence alignments that provide diagnostic signatures for particular protein families. PRINTS can be used to make family and tentative functional assignments for newly determined sequences, offering greater diagnostic reliability than single-motif approaches, such as those embodied in PROSITE [7]. A limitation of using databases like PRINTS and PROSITE to infer function is that the success of a search depends on the database size – if a family isn't encoded in the resource, a search will never find it. Unfortunately, both PROSITE and PRINTS are compromised by the fact that they are maintained and annotated manually; they therefore lag behind resources that are generated by largely automatic means (e.g., Pfam).

To address this problem, we developed a suite of tools to reduce the manual burdens inherent in PRINTS' maintenance. In so doing, we hoped to be able to increase the rate of growth of the resource and hence improve its diagnostic utility. It was not our intention to try to automate PRINTS' maintenance entirely, as the amount of human knowledge required to produce complete, fully annotated fingerprints is considerable, and is not readily captured computationally. Nevertheless, we devised an automatic pipeline in which sequence alignments are generated, motifs are excised, the source sequence database (a Swiss-Prot/TrEMBL composite) is searched for related family members, the scanning process is repeated until no further family members are identified, and the results are collated into 'naked' templates. By applying this pipeline to a public source of seed alignments (ProDom), we were able to generate large numbers of naked fingerprints in batch mode; we then developed a database to store the results, which we termed prePRINTS [8].

The next challenge was to develop a tool capable of injecting annotation into the templates. For pragmatic reasons, we decided to use Swiss-Prot rather than the literature: by contrast with the free-texts available in, say, MEDLINE abstracts, Swiss-Prot annotation is semi-structured and has already been culled from the literature by

human curators. Accordingly, we designed a tool to produce protein family reports from information in Swiss-Prot (see below), which we then fed into the prePRINTS pipeline. Thus, prePRINTS now provides an automatic, annotated supplement to PRINTS, from which entries can be migrated into PRINTS itself with relative ease.

PRECIS

There have been several endeavours to address the problem of annotating sequence data computationally, but the task is non-trivial and few tools have emerged that gather useful information on a given sequence or set of sequences in a simple and convenient manner. As more genome projects bear fruit, the mass of uncharacterised sequence data accumulating in public repositories grows ever larger, and with it the need for tools to facilitate their analysis and annotation. With this in mind, we developed a tool that automatically creates Protein Reports Engineered from Concise Information in Swiss-Prot (PRECIS) [9,10]. PRECIS was initially developed to help with the tedious task of annotating PRINTS entries (hence, in part, its reliance on Swiss-Prot). However, to make the software more generally useful, we implemented a BLAST interface, which allows it to be run with a single starting sequence and an E-value cut-off, below which Swiss-Prot matches are to be allowed into the annotation-culling process.

PRECIS processes the Swiss-Prot entries of matched sequences using a number of rules and filters to determine whether a protein superfamily, a discrete functional family or domain family is being analysed. It then collates the results in the form of a structured report detailing (where known) the function and structure of the protein, the disease(s) with which it is associated, the family to which the protein belongs, database and literature references, and a list of keywords. Figure 1 illustrates the report generated for the rhodopsin family, having BLASTed Swiss-Prot with TrEMBL entry Q9PUZ5, a triggerfish rhodopsin. The search returned 107 significant matches (with E-values <1e-60), for each of which the full entry was retrieved from Swiss-Prot. The

complete data-set, occupying 204 pages of text, was analysed and processed, and the results distilled into a 1.5-page report (Fig. 1).

The principal limitation of PRECIS is that the current implementation is entirely dependent on the quality of pre-existing annotation in Swiss-Prot (if Swiss-Prot annotation is poor, incorrect or absent, PRECIS will inherit its defects); its value, however, is in moving beyond simply providing lists of keywords (as do tools such as the Protein Annotator's Assistant), endeavouring as it does to construct concise protein reports that are English-like. PRECIS is thus a useful annotation tool.

BioMinT

As mentioned above, using Swiss-Prot alone as a source of annotation is problematic, as any errors or deficiencies it has are simply inherited by PRECIS. To address this problem, we are re-visiting on-line texts, to try to refine our protein reports with information extracted from the literature – PRECIS was originally configured to collect MEDLINE abstracts referenced in each Swiss-Prot entry analysed, but we had neither the time nor in-house expertise to properly make use of this information. However, we now lead a European project, BioMinT, for biological text mining and database annotation, which involves 6 partners, including the Swiss Institute of Bioinformatics and the University of Geneva, through which we hope to develop the next generation of annotation tools for Swiss-Prot and PRINTS.

CINEMA/UTOPIA

At the end of its last grant period, EMBnet provided funds to extend the CINEMA alignment editor and render it compatible with EMBOSS (in fact, the project started the day before the EMBnet core grant terminated!). This project was fraught with difficulties from the outset, primarily as a result of problems acquiring and retaining suitable staff. A new version of CINEMA was eventually produced, termed CINEMA-MX [11], but the package was a poor replacement for the original package and lacked much of its functionality.

```
Rhodopsin

PRINTS; PR00237 GPCRRHODOPSN; PR00238 OPSIN; PR00579 RHODOPSIN
PROSITE; PS00237 G_PROTEIN_RECEP_F1_1; PS00238 OPSIN; PS50262 G_PROTEIN_RECEP_F1_2
PFAM; PF00001 7tm_1
INTERPRO; IPR000276; IPR001760
PDB; 1BOJ; 1BOK; 1EDS; 1EDV; 1EDW; 1EDX; 1F88; 1FDF
SCOP; 1BOJ; 1BOK; 1EDS; 1EDV; 1EDW; 1EDX; 1F88; 1FDF
CATH; 1BOJ; 1BOK; 1EDS; 1EDV; 1EDW; 1EDX; 1F88; 1FDF
MIM; 163500; 180380; 268000

1. FYHRQUIST, N., DONNER, K., HARGRAVE, P.A., MCDOWELL, J.H., POPP, M.P. AND SMITH, W.C.
Rhodopsins from three frog and toad species: sequences and functional comparisons.
EXP.EYE RES. 66 295-305 (1998).

2. ARCHER, S.N. AND HIRANO, J.
Opsin sequences of the rod visual pigments in two species of Poeciliid fish.
J.FISH BIOL. 51 215-219 (1997).

3. HUNT, D.M., FITZGIBBON, J., SLOBODYANYUK, S.J., BOWMAKER, J.K. AND DULAI, K.S.
Molecular evolution of the cottoid fish endemic to Lake Baikal deduced from nuclear DNA
evidence.
MOL.PHYLOGENET.EVOL. 8 415-422 (1997).

4. GALE, J.M., TOBEY, R.A., D'ANNA, A.
Localization and DNA sequence of a replication origin in the rhodopsin gene locus of
Chinese hamster cells.
J.MOL.BIOL. 224 343-358 (1992).

5. PALCZEWSKI, K., KUMASAKA, T., HORI, T., BEHNKE, C.A., MOTOSHIMA, H., FOX, B.A., LE
TRONG, I., TELLER, D.C., OKADA, T., STENKAMP, R.E., YAMAMOTO, M. AND MIYANO, M.
Crystal structure of rhodopsin: a G protein-coupled receptor.
SCIENCE 289 739-745 (2000).

6. YEAGLE, P.L., ALDERFER, J.L. AND ALBERT, A.D.
Structure of the third cytoplasmic loop of bovine rhodopsin.
BIOCHEMISTRY 34 14621-14625 (1995).

7. YEAGLE, P.L., SALLOUM, A., CHOPRA, A., BHAWSAR, N., ALI, L., KUZMANOVSKI, G.,
ALDERFER, J.L. AND ALBERT, A.D.
Structures of the intradiskal loops and amino terminus of the G-protein receptor,
rhodopsin.
J.PEPT.RES. 55 455-465 (2000).

Function:
Visual pigments are the light-absorbing molecules that mediate vision. They consist of an
apoprotein, opsin, covalently linked to cis-retinal.

Additional Info:
Integral membrane protein.

Disease:
Defects in rho are one of the causes of autosomal dominant retinitis pigmentosa (adrp).
Patients typically have night vision blindness and loss of midperipheral visual field; as
their condition progresses, they lose their far peripheral visual field and eventually
central vision as well. (OPSD_HUMAN).

Defects in rho are one of the causes of autosomal recessive retinitis pigmentosa (arrp).
(OPSD_HUMAN).

Defects in rho are also one of the causes of congenital stationary night blindness
(csnb4). (OPSD_HUMAN).

Family and structural information:
The structure has been determined, e.g., "Crystal structure of rhodopsin: a G protein-
coupled receptor" [5]; "Structure of the third cytoplasmic loop of bovine rhodopsin"
[6]; "Structures of the intradiskal loops and amino terminus of the G-protein receptor,
rhodopsin" [7].

Belongs to family 1 of g-protein coupled receptors. Opsin subfamily.

Keywords: Photoreceptor; Retinal protein; Transmembrane; Glycoprotein; Vision;
Phosphorylation; Lipoprotein; Palmitate; G-protein coupled receptor; Acetylation; 3D-
structure; Retinitis pigmentosa; Disease mutation.
```

Figure 1. Example PRECIS output, showing the report generated for the rhodopsin family using TrEMBL entry Q9PUZ5 (E-values <1e-60). From this single sequence, 107 significant matches were returned and distilled fully automatically into the concise report shown.

Determined not to give up, we embarked on a larger project funded by the e-Science Centre for the North West (ESNW), in which EMBnet is also a partner. Ambitiously referred to as UTOPIA (User-friendly Tools for OPerating Informatics Applications), this project aims to develop an easy-to-use bioinformatics analysis environment. It will provide a platform for building visualisation tools and applications capable of high performance data visualisation that is transparently accessible from Grid-enabled resources, and supporting real-time collaboration between geographically distant sites. It includes, at its heart, CINEMA3D (illustrated in Figure 2) and many additional sequence analysis, database search and visualisation tools. The first release of the CINEMA component is now available for download.

BioActivity/EMBER

BioActivity is an interactive bioinformatics Web practical with links to an introductory text-book [6, 12]. The book provides background material for the concepts, tools and resources used in the practical, setting the scene for one of the later chapters (a worked example through each of the practical steps); in turn, the practical provides on-line excerpts from the book, supplemented with colour pictures where appropriate.

In order to extend this idea and, crucially, update the Website, in 2001, the European Commission funded the development of a new European Multimedia Bioinformatics Educational Resource, EMBER [13,14]. EMBER was conceived as a mechanism both to bridge the skills gap in bioinformatics and to foster academic and private-sector collaboration in the development of standardised, multimedia courses in bioinformatics. There are 10 consortium partners, all of which are EMBnet Nodes.

Eventually, EMBER will offer a self-contained, interactive bioinformatics Web-tutorial, the equivalent stand-alone course on CD-ROM, and an accompanying introductory text-book. The decision to use conventional text, coupled with Web- and CD-based media, should ensure that students in European domains for whom Internet access is not optimal have access to the same educational materials. EMBER's path has not been smooth, and we hope to extend the project beyond its 2-year funding period to see it through to completion. Figure 3 illustrates the
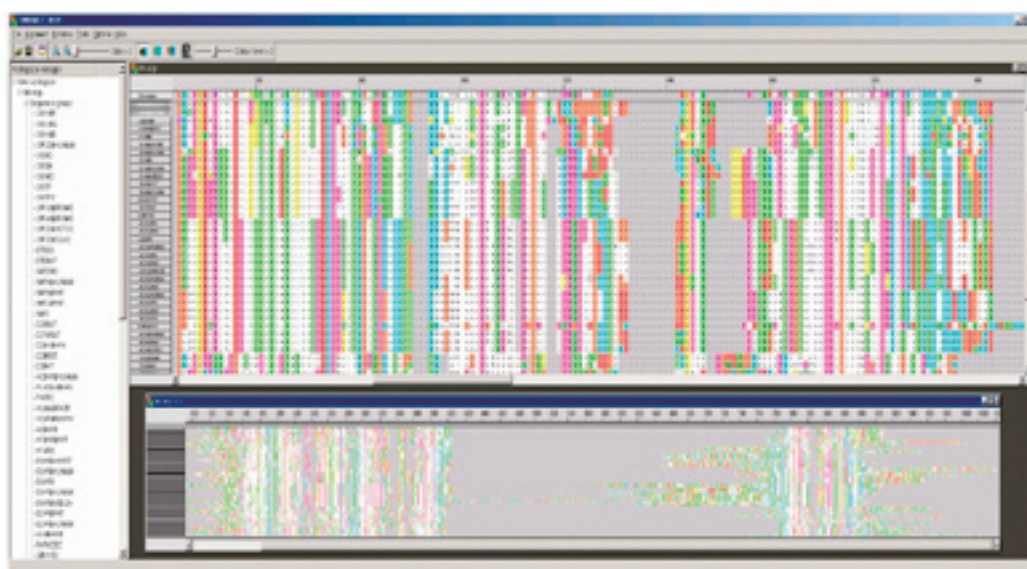

Figure 2. Screen-shot from the CINEMA3D module of UTOPIA illustrating part of a multiple alignment with an overview of the full alignment beneath.

prototype being developed at ETI, contrasting with the site available at UMBER for use in the meantime.

Figure 3. Screen-shots illustrating the look-&-feel of the ETI prototype Website (top), and our interim solution while the ETI site is being developed.

Book

Finally, we are negotiating with Prentice Hall to write a second edition of Introduction to Bioinformatics, first published in 1999 by Addison Wesley Longman [12]. Meanwhile, our colleagues at the Chinese National Node have translated the book into… Chinese (!); a Spanish translation has also just been published, and a low-cost version is also now available for the Indian market.

Web Addresses

UMBER
http://www.bioinf.man.ac.uk/dbbrowser/
PRINTS
http://www.bioinf.man.ac.uk/dbbrowser/
PRINTS/
PrePRINTS
http://www.bioinf.man.ac.uk/dbbrowser/
preprints/
PRECIS
http://www.bioinf.man.ac.uk/cgi-bin/db-browser/precis/blast _ precis.cgi
BioMinT
http://www.bioinf.man.ac.uk/dbbrowser/
biomint/
CINEMA
http://www.bioinf.man.ac.uk/dbbrowser/
CINEMA2.1/
CINEMA3D
http://aig.cs.man.ac.uk/utopia/
BioActivity
http://www.bioinf.man.ac.uk/dbbrowser/
bioactivity/
EMBER
http://www.bioinf.man.ac.uk/dbbrowser/
ember

References

1. Attwood, TK, Beck, ME, Bleasby, AJ and Parry-Smith, DJ (1994). PRINTS - A database of protein motif fingerprints. Nucleic Acids Research, 22(17), 3590–3596.
2. Selley, JN and Attwood, TK (2001) EASY – an Expert Analysis SYstem. Bioinformatics, 17(1), 105-106.
3. Scordis, P, Flower, DR and Attwood, TK (1999). FingerPRINTScan: Intelligent searching of the PRINTS motif database. Bioinformatics, 15(10), 799-806.
4. Wright, W, Scordis, P and Attwood, TK (1999). BLAST PRINTS - An alternative perspective on sequence similarity. Bioinformatics, 15(6), 523-524.
5. Parry-Smith, DJ, Payne, AWR, Michie, AD and Attwood, TK (1998). CINEMA - A novel Colour INteractive Editor for Multiple Alignments. Gene, 221, GC57–63.
6. Attwood, TK (2000). An interactive practical at the interface of Web-based and conventional publishing. CAL-laborate. A collaborative publication on the use

of Computer Aided Learning for tertiary level physical sciences, 4, 1–6.

7. Attwood, TK (2002) The PRINTS database, a resource for identification of protein families. Briefings in Bioinformatics, 3(3), 252-263.

8. Attwood, TK, Bradley, P, Flower, DR, Gaulton, A, Maudling, N, Mitchell, AL, Moulton, G, Nordle, A, Paine, K, Taylor, P, Uddin, A and Zygouri, C (2003) PRINTS and its automatic supplement, pre-PRINTS. Nucleic Acids Research, 31(1), 400-402.

9. Reich, JR, Mitchell, A, Goble, CA and Attwood, TK (2001) PRECIS: Protein Reports Engineered from Concise Information in SWISS-PROT. IEEE Intelligent Systems, 16(6), 42-51.

10. Mitchell, AL, Reich, JR, and Attwood, TK (2003) PRECIS: Protein Reports Engineered from Concise Information in SWISS-PROT. Bioinformatics, accepted.

11. Lord, P and Attwood, TK (2002) CINEMA-MX – a modular multiple alignment editor. Bioinformatics, 18(10), 1402-1403.

12. Attwood, T.K. and Parry-Smith, D.J. (1999) Introduction to bioinformatics. Addison Wesley Longman, Harlow, Essex CM20 2JE, UK. ISBN 0 582 327881

13. Mabey, J.E. and Attwood, T.K. (2001) EMBER: a European Multimedia Bioinformatics Educational Resource. CAL-laborate. A collaborative publication on the use of Computer Aided Learning for tertiary level physical sciences, Vol. 6, pp.13-16.

14. Attwood, TK (2002) EMBER – A European Multimedia Bioinformatics Educational Resource. EMBnet News, 8(1), 15-1

Bioinformatics courses in Colombia

Laurent Falquet
Swiss EMBnet node manager, Swiss Institute of Bioinformatics
155, ch. des Boveresses,
CH-1066 Epalinges,
SWITZERLAND
Laurent.Falquet@isb-sib.ch

Introduction

Over the past 4 years I had the opportunity to travel to Colombia to teach bioinformatics. Although this country is known mostly for its violence, guerrilla and drug production, it has a friendly population that is eager to learn and to have contacts with foreign people. This experience is an enrichment and proves, if necessary, the needs for high level courses in developing countries.

History

The first contact I had with Colombia was in Cali in 1999 where I was invited with two colleagues by Prof. Socrates Herrera (Univ. del Valle) to give an introductory course. During this course I met Prof. Emiliano Barreto (Univ. Nacional) who is now the Colombian EMBnet node manager[1] member of the Biotechnology Institute of the National University[2]. We organised three courses in Bogota, following the first one: introductory courses in 2000 and 2002, and advanced course in sequence analysis and introduction to microarrays data analysis in 2001.

The students

Each course is run over one full week with both morning and afternoon divided in theoretical and practical sessions. The students usually 30 per course, originated from several regions of Colombia (Bogota, Medellin, Cali, Cartagena) and several universities (Nacional, Los Andes, La Javeriana, Militar, CIAT, etc.). Most of them were biologists, pharmaceutical chemists, or computer scientists, and even some

local professors! We greatly appreciated their contact, smart questions and most importantly, their enthusiasm and willing to learn behavior.

Courses in 2001 (top) and 2002 (bottom)

The infrastructure

The teaching and the computer rooms are simple and equipped with recent PCs under Windows.

The network is sufficiently fast, but the bandwidth limits somehow the flow of pages. Before the first course in Bogota we helped them to set up the servers, a BLAST machine (dual proc Pentium under Linux) and a file server with users accounts and software, mainly the EMBOSS package plus several other tools and the databases (EMBL, SwissProt, TrEMBL). Currently they continue to maintain and enhance the services with a web portal and other tools.

The culture

Well my first effort was to try learn a bit of Spanish to improve the communication and year after year I enjoy to practice the spanish language. My first discovery came from the food and especially the fruits. They have an incredibly rich collection of tropical fruits that we don't even know about here in Europe! The few tourist places that we visited are really astonishing (e.g., the Gold Museum in Bogota or the Salt Cathedral of Zipaquira)[3,4], but the most interesting feature of the Colombian culture is the music and especially the Salsa dance, that you hear everywhere and that you can even learn in a moving bus "la chiva rumbera"! (see picture below)

Conclusion

Finally I can only encourage openminded people to visit this beautiful country, a paradise for biologists interested in biodiverstiy, and to meet fine colombian people. Of course you should listen to them and behave cautiously when they tell you.

References

1. Colombian EMBnet node (http://bioinf.ibun.unal.edu.co)
2. University Nacional (http://www.unal.edu.co)
3. El Museo del Oro (http://www.banrep.gov.co/museo)
4. Cathedral de sal (http://www.catedraldesal.com)

Computational gene expression analysis; correlating data and knowledge

Torgeir R. Hvidsten
The Linnaeus Centre for Bioinformatics
Uppsala University
Husargatan 3
SE-751 24 Uppsala
SWEDEN
Torgeir.Hvidsten@lcb.uu.se

Introduction

Physiological and pathophysiological responses are associated with specific changes in cellular gene expression and gaining insight into these specific patterns enable hypotheses about gene function, understanding of gene regulation and promising new possibilities for molecular medical diagnosis. Microarray technology is a fundamental research vehicle in this context, allowing simultaneous readouts of relative gene expression levels for thousands of pre-selected genes in a tissue sample (Figure 1). Conduction several experiment using samples from different time points, clinical states and/or different treatments constitute a powerful setup for studying the genome-wide changes in cellular gene expression. However, this setup also raises non-trivial computational challenges not only related to storage and analysis of data, but also to evaluation and interpretation of the results according to current biological knowledge.
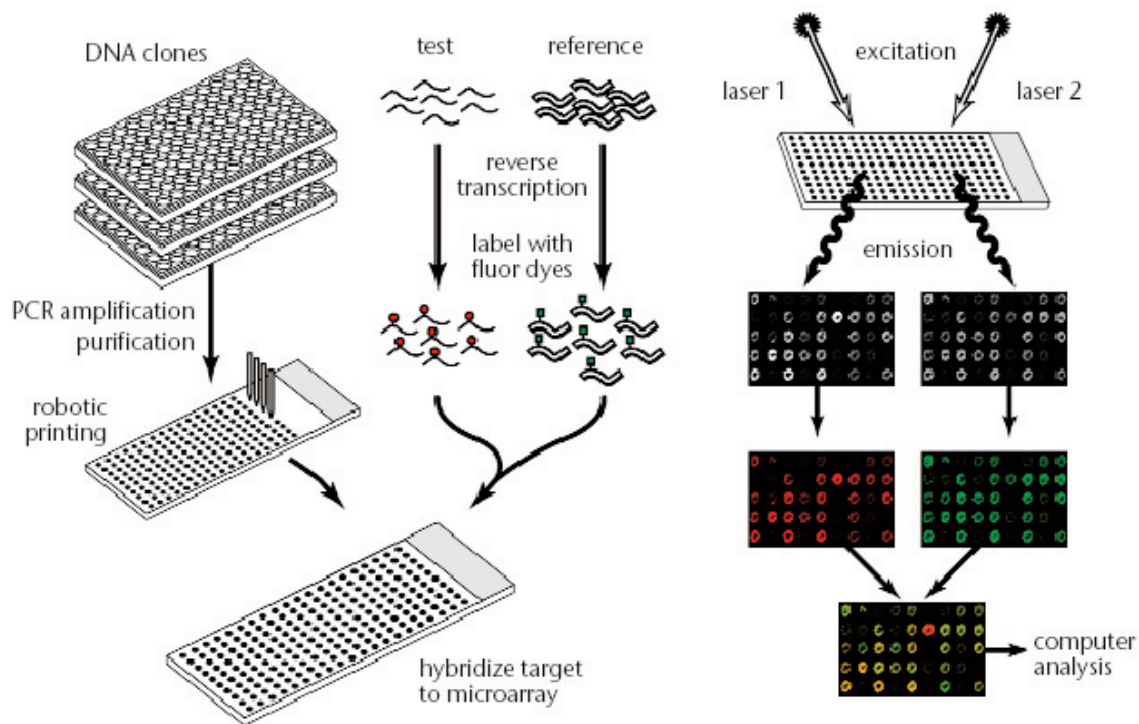

Figure 1 Microarray experiment. Gene probes of interest are printed on glass microscope slides. RNA from both test and reference sample are first fluorescently labeled with Cye3 (red) or Cye5 (green) and then hybridized to the clones on the array. The arrays are scanned and the resulting images are pseudo-colored and merged. Data from hybridization experiment is commonly viewed as normalized ratios Cye3/Cye5, indicating increased (>1), decreased (<1) or unchanged (=1) expression level in the test sample relative to the reference sample. Conduction several experiment using samples from different time points, clinical states and/or different treatments constitute a powerful setup for studying the genome-wide changes in cellular gene expression in living systems (D. J. Duggan et al., Nature Genetics 21, 10-14, 1999).

Gene expression data analysis

The typical intermediate result of a microarray study is the genes ◊ samples – matrix where each row holds the expression profile of one gene over all samples and each column hold the expression profile of one sample over all genes. Common for most data analysis methods is that they are based on representing genes/samples as points in a multi-dimensional expression spaced spanned by the samples/genes (obviously, the exception is studies in which samples are taken from different time point and genes can be represented as functions of time in two dimensions).

Data analyses strategies typically try to reduce the complexity by assume that there are underlying, known or unknown, classes of genes or samples that corresponds to biological or medical phenomena. Such classes can, for examples, be groups of genes that code for proteins with the same molecular function or are involved in the same biological process, or groups of tissue samples that are taken from patients with the same disease or that underwent the same medical treatment. There are two conceptually different analysis strategies. The first strategy is called class discovery or clustering (1,2). Such methods try to discover biological related genes or samples by assuming that they will appear close to each other in the expression space. Hierarchical clustering is one commonly used clustering method in the context of gene expression analysis (Figure 2). The second strategy take advantage of already well studied examples and try to learn models that can be used both to gain insight into the parameters discriminating different classes and for prediction/diagnosis. These methods are called supervised learning or machine learning (3). Models can range from simply identifying genes that are differentially expressed in, for example, samples from different cancer types, through IF-THEN rules combining such discriminatory genes, to complex illegible models such as artificial neural networks and hyper-planes from support vector machines.
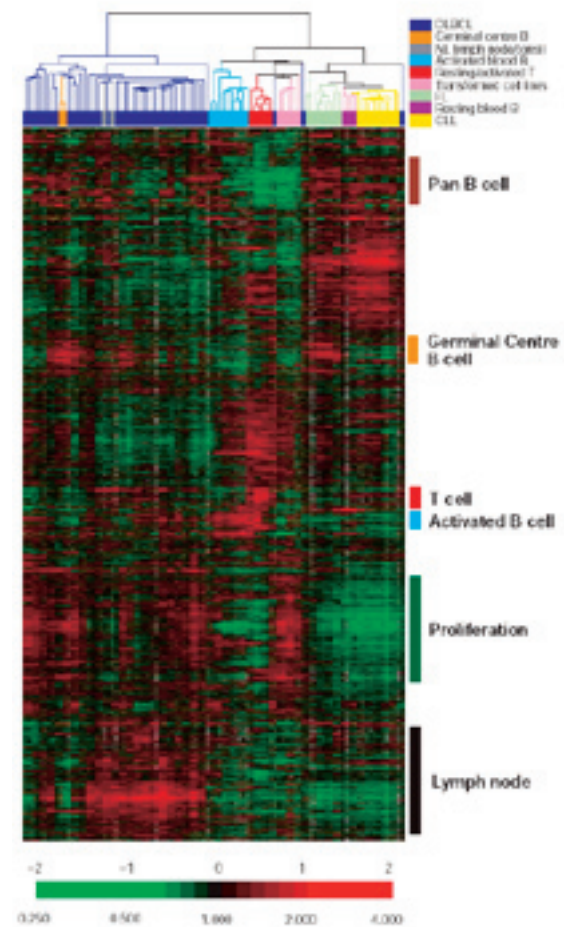

Figure 2 Hierarchical clustering. Hierarchical clustering organize samples in a tree that reflects their level of (expression) similarity. The figure shows the hierarchical clustering of 96 samples from normal and malignant lymphoma using expression profiles from almost 20 000 selected human genes. Diffuse large B-cell lymphoma (DLBCL) is known to be clinically heterogeneous in that only about 40% respond well to current treatment. The figure shows how this subtype is almost perfect clustered separately from the other samples. Furthermore, it was shown that the two sub-clusters of DLBCL (found using clustering) include patients with significantly different expected survival time indicating two distinct molecular types of DLBCL (Alizadeh et al., Nature 403, 503-511, 2000).

Common to both clustering and machine learning methods is the need to validate results in order to assess their biological or medical relevance. While early studies on the relation between gene function and gene expression only visually hinted on expression similarity within groups on functional related genes, newer studies typically test the

statistical significance of overlap between expression clusters and functional classes or even build models capable of hypothesizing novel functions for genes. Such computerized use of large amount of biological knowledge, in this case knowledge of gene function, requires structured databases and controlled vocabularies (6). Building such databases again requires reading thousands of articles in order to extract the relevant information. As a consequence, developing new tools for text-mining (automatic extraction of knowledge from text) has become one of the most important research areas in Bioinformatics (7).

Selected data analysis tools and further reading

The research community early recognized the need for standardization to facilitate data management, processing, transfer, publication and reproducibility of microarray data. This has resulted in Laboratory Information Management Systems (LIMS) sharing common data formats (MIAME: Minimum Information about a Microarray Experiment), and data warehouses such as ArrayExpress (4) and the Gene Expression Omnibus (5) publishing microarray studies. There is a wide variety of freely available systems and services applicable to gene expression data analysis on the web. Some recommended sites and directions for further reading are indicated here:

References

1. Cluster and TreeView: various clustering algorithms including a much used visualization tool for hierarchical clustering (http://rana.lbl.gov/EisenSoftware.htm).
2. J-Express: cluster and statistical analysis of gene expression data (http://www.ii.uib.no/~bjarted/jexpress/).
3. The ROSETTA system: a general toolkit for IF-THEN rule-based supervised learning (http://www.idi.ntnu.no/~aleks/rosetta). Case study using ROSETTA for functional gene expression analysis: http://www.lcb.uu.se/~hvidsten/fibroblast/.
4. ArrayExpress: public database for microarray studies by the European Bioinformatics Institute (EBI) (http://www.ebi.ac.uk/arrayexpress).
5. Gene Expression Omnibus: public database for microarray studies by the National Center for Biotechnology Information at NIH (http://www.ncbi.nlm.nih.gov/geo/).
6. Gene Ontology: a dynamic controlled vocabulary of gene and protein roles in cells (http://www.geneontology.org/).
7. PubGene: gene networks automatically extracted from literature (http://www.pubgene.uio.no/).
8. The Chipping Forecast II, freely available supplement to Nature Genetics, 32: 461 – 552, 2002 (http://www.nature.com/ng/web _ specials/)

Announcements

BEN - the Belgian EMBnet node (http://www.be.embnet.org) - has been invited to give a workshop in Bioinformatics, the 3 and 4 April 2003 at the "Université de la Mediterranée" (http://mediterranee.univ-mrs.fr/), in Marseille, France. Guy Bottu and Valérie Ledent ran the workshop in front of an attendance of 24 bioinformaticians and IT specialists.

This workshop (http://www.esil.univ-mrs.fr/%7Ethieffry/BEN/index.html) was organised in the frame of the DESS CCI (Compétences Complémentaires en Bioinformatique; http://www.dil.univ-mrs.fr/desscci/). It focused on the Software Suites Staden and EMBOSS, the EMBOSS web interface WEMBOSS developed by BEN and the Sequence Retrieval System SRS. The focus was on the overview of the functionalities of these software packages, as well as their installation and maintenance on Linux machines.

Other courses and meetings

See our web site:

http://www.embnet.org/activities/dyn _ schedule.php

You can even use this page to announce your own courses and meetings (OBS: this service is moderated).

A guide on how to install Bioinformatics tools on MacOSX (Part I)

Erik Bongcam-Rudloff
Assistant Professor
Swedish University of
Agricultural Sciences,
The Linnaeus Centre for
Bioinformatics

MacOSX an introduction

In 1999 Apple Computer released its first version of Macintosh OS X Server. This next generation operating system was the precursor to their Macintosh OS X. This new member of the Unix family unifies the power of Unix with the ease-of-use and sophisticated applications of a consumer O/S.

Macintosh OSX is a combination of Apple's OS, NeXT Computer's OpenStep OS, and UNIX operating systems. The system has a Mach micro kernel; and a BSD layer on top of that. The kernel for MacOSX is UNIX based (Mach from Carnegie-Mellon University and 4.4BSD from the University of California at Berkley). The kernel for MacOSX is called Darwin and is an open-source development[1].

Most command-line and X11[2] programs that can be compiled under Linux[3] recompile under MacOSX without difficulty. It is often a function of the compiler (in this case gcc 3.1)[4] and the available libraries (which are also typically the same open source ones on both systems). When installing OS X, Apple puts the command-line application in a less obvious place to reduce the chance that a novice would accidentally launch it, but it is definitely there. Go to /Applications/Utilities and you will find the "Terminal", put it on the Dock (just click on it and place it on the Dock, you will make heavy use of this terminal). MacOSX is designed so that an ordinary user never needs to know anything about Unix, but it keep all the Unix power available to those that need/want it.

Elements of OS X

MacOSX[5] is made up of four basic elements when viewed in the simplest way.
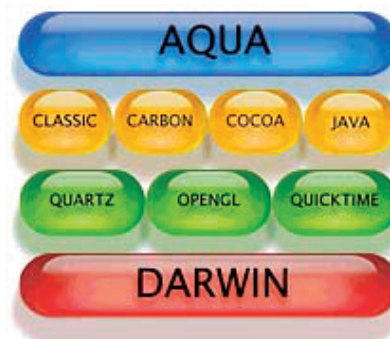
a-The kernel (Darwin)
b-The imaging layer (Quartz, OpenGL, QuickTime)
c-The application environments (Classic, Carbon, Cocoa, Java)
d-The graphical user interface, or GUI (Aqua) (We will add X11 at this level!)

The imaging, application and GUI layers sit on top of the Darwin core. Is important to keep this in mind to understand why we install the X-windows system in parallel with Apples GUI (Aqua).

MacOSX and Bioinformatics

In the first article of this series you will learn an easy way to install some of the basic and most popular Bioinformatics packages on a Macintosh running MacOSX. In the next article we will discuss how to use the programs and how to expand this collection with other tools. After this simple step by step tutorial you will have on your hands a powerful workstation to accomplish most of the common bioinformatics tasks.
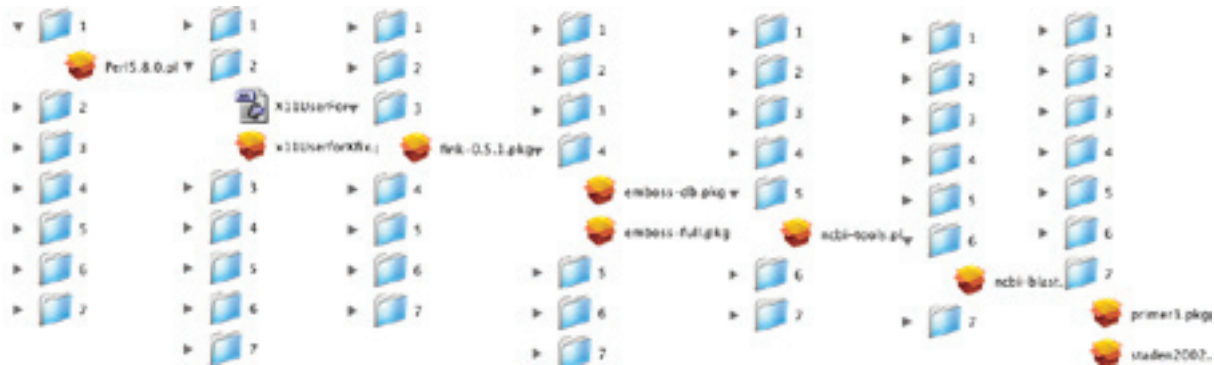
Prerequisites

A PowerPC computer running MacOSX 10.2.x. You should also install the freely available "Developers tools"[6] (Do not install Apples latest X11 (as of 020403). Use the X11 provided in this tutorial. Apples X11 have some bugs that we fixed in our release.)

If you fulfill these prerequisites it is now time to start installing the packages as described in this tutorial.

Part 1

Go to http://www.ebioinformatics.org
Download all the packages on macosx/software/download.htm.

2.6.0: The European Molecular Biology Open Software Suite. EMBOSS is a new, free Open Source software analysis package specially developed for the needs

Please install the packages following <u>exactly</u> the indicated steps on the next figure (it is essential to do this in the same order as some packages depend on libraries and programs installed in the previous steps!):

The following packages will be installed:

1- Perl 5.8.0[7]　(this is needed because MacOSX provides Perl 5.6)

2- X11, the X Window System that makes it possible to run X11-based applications in MacOSX. Based on the open source XFree86 project[8] . Add the fix provided, this will be changed in the future!

3- Package 3 installs Fink[9] with some libraries and programs needed for the rest of this installation. The Fink project brings the full world of Unix Open Source[10] software to Darwin and MacOSX. The people involved in the Fink project modify Unix software so that it compiles and runs on MacOSX ("port" it) and make it available for download as a coherent distribution. Fink uses Debian[11] tools like dpkg and apt-get to provide powerful binary package management. You can choose whether you want to download precompiled binary packages or build everything from source. Later on you will be able to use programs like "dselect" to install hundred of programs to your system. With our package we install Pymol[12] (among others) for demonstration purposes.

4- In this step we install EMBOSS[13] version

of the molecular biology (e.g. EMBnet[14]) user community. Together with this installation you will also install additional packages (e.g., HMMER-2.1.1[15], with Altivec-enhancements, Domainatrix[13], Phylip-3.573c[16].). Add also the second package to install the necessary databases (e.g., Rebase: the restriction enzyme database maintained by New England Biolabs. Rebase is needed for programs such as remap and restrict.).

5- Now you can add the NCBI Toolbox[17] i.e., the NCBI software tools for building bioinformatics resources. This toolbox is needed for adding the next package.

6- This package installs the most common tool in Bioinformatics: The BLAST® (Basic Local Alignment Search Tool) a set of similarity search programs. In this installation you will install the Altivec enhanced version[18] (OBS. optimised only for blastn)

7- The last step. You will install Primer3[19] and the Staden[20] package. The Primer3 program is used by the EMBOSS program eprimer. The Staden package is the version-2002.0. This is the first official release of the MacOSX version (10.2) of the package. This version requires X11 and provides a nice GUI (Graphical User Interface) to the EMBOSS package. (**OBS. The Staden package is only free for academics and you need to apply for a license**).
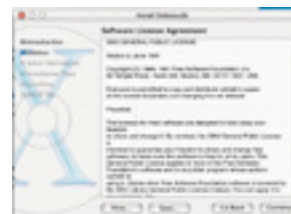
Part 2

Step 1 Click twice on the first package. The first window will ask you for an administrator password.
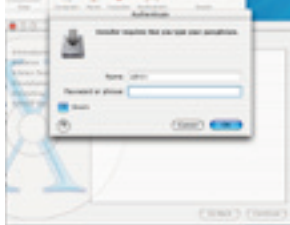
Step 2 You see a disclaimer with information about the people behind these packages. In this case Anders Nister and me. Continue if you trust us.
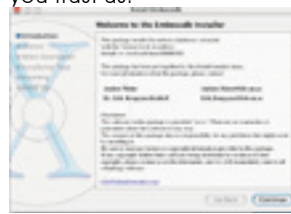
Step 3 Now you will read the license information related to each specific package

Step 4 Click agree to the license agreement (read it first!).

Step 5 Select the hard disk where you want the package to be installed (all packages install on /usr/local) Click continue.

Step 6 Click on install to perform a basic installation. The packages are installed in /usr/local.

Step 7 The installation starts, wait a few seconds.

Step 8 This is the last step of each package.
Repeat with all packages as indicated on figure 1.

Part 3

During installation you will have an automatic configuration of your environment parameters, this to tell the system that you have new programs and libraries installed. You can also adapt your new system in the way that suites your needs. I did one change for my computer (This is just an example!).

Open a new Terminal.app window and type "pico .cshrc". A text editor will pop up. Enter these lines:

```
source /sw/bin/init.csh
setenv PLPLOT_LIB /usr/local/share/emboss
set path=( /usr/local/bin /usr/local/blast/bin   /usr/local/staden/macosx-bin ${path} )
setenv DISPLAY :0.0
```

To get out of the editor, press control-O, return, control-X.
This works for me but if you are a user with some Unix knowledge you should adapt the system for your own needs!
The complete installation procedure takes less than one hour.

Now you are ready to test your new "Bioinformatics Workstation".
The first thing you do is to start the X11 environment, do this with the icon X11 in your Applications folder. Then open a new Terminal.app window and type: "pymol"
You should see this when testing the demo files:
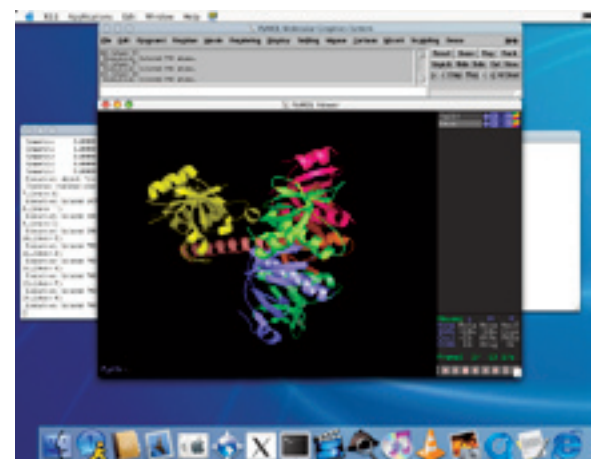
If you want to test the Staden package:
Open a new Terminal.app window and type: "spin"

This starts spin, one of the Staden package modules. in Demo mode (Apply for a license!) but using the demo files and some testing you should be able to see this using plotorf on the EMBOSS menu:
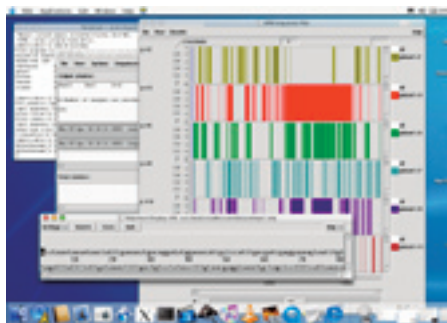
The EMBOSS package can also be used directly on the command line (on the Terminal.app). Test a DNA sequence (in fasta format) using the EMBOSS program plotorf.
You should see something similar to this if X11 is working correctly.
We installed the png[21] (Portable Network Graphics) libraries with Fink and compiled EMBOSS with libpng so you are also able to produce png images.

I also strongly recommend to buy a three button mouse for your Mac. You will need it for most Bioinformatics applications under X11. MacOSX has support for it, the three button mice work without any configuration changes, just plug in!

The next article will discuss:
a- how to install GUI- and web-interfaces to the command line programs,
b- how to install other packages and
c- how to build your own databases to use the blast programs

Acknowledgements:

References

1- http://developer.apple.com/darwin/
2- http://www.xdarwin.org/
3- http://www.linux.org/
4- http://gcc.gnu.org/
5- http://www.apple.com/macosx/
6- http://developer.apple.com/tools/macosxtools.html
7- http://www.perl.org/
8- http://www.xfree86.org/
9- http://fink.sourceforge.net/
10- http://www.opensource.org/
11- http://www.debian.org/
12- http://pymol.sourceforge.net/
13- http://www.emboss.org/
14- http://www.embnet.org
15- http://hmmer.wustl.edu/
16- http://evolution.genetics.washington.edu/phylip.html
17- http://www.ncbi.nlm.nih.gov/Ftp/index.html
18- http://www.apple.com/downloads/macosx/math _ science/applegenentechblast.html
19- http://www-genome.wi.mit.edu/genome _ software/other/primer3.html
20- http://www.mrc-lmb.cam.ac.uk/pubseq/
21- http://www.libpng.org/pub/png/

Protein Spotlight

Protein Spotlight (ISSN 1424-4721) is a periodical electronic review from the SWISS-PROT group of the Swiss Institute of Bioinformatics (SIB). It is published on a monthly basis and consists of articles focused on particular proteins of interest. Each issue is available, free of charge, in HTML or PDF format.
http://www.expasy.org/spotlight

We provide the EMBnet community with a printed version of issues 32 and 33.
Please let us know if you like this inclusion.

THE THINGS WE FORGET
By Vivienne Baillie Gerritsen

Why remember when it is so important to forget? That sounds promising in a day and age when techniques to enhance one's memory are as popular as a cup of tea. It is important not to forget your mother's birthday, your husband's cell phone number, the date you have to give in your income tax sheet but it is very desirable to forget all the items you float past in a supermarket, the names you scan through in a phone book, the sign posts you sweep past on your way home. We would like to think that we are the masters of what we choose to remember and what we choose to forget. No doubt, to a certain degree we are. But some initial sifting is done on the molecular level and for our own benefit. The understanding of the biochemistry involved in the process of learning, memory…..and forgetfulness is still very much in its infancy, but we do know of one protein, protein phosphatase 1, which certainly seems to be at the heart of this primordial sift.

Figure 1   Now, how did it go?

In humans, the necessity to forget is best portrayed by those who suffer from what has been called the 'Savant Syndrome'. This disorder was first described in the late 1900s by the English clinician Dr J. Langdon Down (1828-1896) – after whom by the way the Down Syndrome is named - who was then the Medical Superintendent of the Royal Earlswood Asylum for Idiots. In fact, the disorder was initially termed 'Idiot Savant' but with time the word 'idiot' was discarded because of its contemporary pejorative connotation. The Savant Syndrome was brought to public awareness thanks to Dustin Hoffman's performance in 'Rain Man' in the 1980s, where Hoffman played the part of an autistic savant based on the life of the 'real'

Rain Man, Kim Peek[1], born in 1951 and still alive today. While reading a book, Kim's parents would run his finger along each sentence and before the age of two, Kim was already memorizing every book that was read to him. In fact, he has read and can recite 7600 books - amongst many other forms of expertise.

Though this type of disorder is fascinating, it can be a bit of a burden on the normal psyche. It no doubt depicts less an overwhelming faculty to memorize things than an awesome inability to forget things. The mechanisms, underlying the process of learning and memory, were introduced by the American psychologist Donald Hebb (1904-1985) who developed the theory which states that synapses are the key elements of memory. The term 'synaptic plasticity' soon bounced into the playing field. It is believed that memory is the result of a synaptic change in size, shape and sensitivity brought on by the synthesis of various proteins. And this change is the work of complex biochemical pathways - one of which protein phosphatase 1 (PP1) belongs to.

What exactly does PP1 do? PP1 hinders the synthesis of proteins involved in synaptic plasticity, hence in the process of memory. And how does it do this? As it name suggests, it is a phosphatase and, as all phosphatases, it deprives another molecule of its phosphate thereby modifying its function. In this event, PP1 binds to the gene transcription factor CREB (cyclic AMP-dependent response element binding) and dephosphorylates it. CREB is subsequently inactivated and the transcription of a number of 'memory' genes checked.

When though does PP1 pick its moment to inhibit CREB? Obviously, PP1 should not always be active otherwise we would remember nothing at all. The German psychologist Hermann Ebbinghaus (1850-1909) was the first person to study memory scientifically and showed very elegantly –

---

[1]
http://www.wisconsinmedicalsociety.org/savant/kimpeek.cfm

what we know instinctively – that one memorizes information which is given in little spurts (distributed) much better than information given in one large gush (massed). This may remind some (or not, precisely) of wearisome, unending, informative lectures and brain-cramming sessions before exams, which in the long run have not left much of an impression on our memories.

When PP1 is inhibited in mice, there is no difference between distributed learning and massed learning. This means that the gene transcription factor CREB is left to get on with its job unimpeded whatever the situation. It is thought that in the event of massed learning, the synthesis of PP1 increases, thereby checking CREB function and putting a constraint on what we learn and remember. In the event of distributed learning, PP1 synthesis does not have the chance to increase so sharply and its effect on CREB is weaker. What is more, the longer the interval between two sessions, the better the imprint of the memory, probably because the level of PP1 decreases and CREB is given time to settle and function as it should. So, in effect, our memorizing faculties depend on a certain PP1 threshold.

More to the point, when PP1 is inhibited *after* the memory process, the memory is long-lasting, which means that in its natural state, PP1 has a direct effect on forgetfulness.

The findings are interesting. Memory decline as we get older is probably due, not to the wear and tear of our brain as we could easily imagine or, for that matter, to a lack of space, but to an increasing amount of PP1. It shouldn't be too difficult to conjure up some memory enhancing therapy then. But it is. PP1 is 'just' a phosphatase. Phosphorylation and dephosphorylation are among the most common biochemical processes and are used as on/off switches in many biological instances. Inhibiting PP1 could have drastic effects – its clients are not only the gene transcription factor CREB. However, the knowledge of PP1 function does shed light on the very complex path of learning, memory and forgetfulness and could pave the way to subtle therapies in the event of Alzheimer's disease for instance, brain traumas and indeed the withering of our memories due to the simple passing of time.

Cross-references to Swiss-Prot

P08129: human, *Mus musculus* (mouse), *Rattus norvegicus* (rat) and *Oryctolagus cuniculus* (rabbit) protein phosphatase PP1-alpha
P37140: human protein phosphatase PP1-beta
P36873: human protein phosphatase PP1-gamma
P37139: *Mus musculus* (mouse) and *Rattus norvegicus* (rat) protein phosphatase PP1-gamma

References

1. Silva A.J., Josselyn S.A.
   The molecules of forgetfulness
   Nature 418:929-930(2002)
   PMID: 12198533

2. Sweatt J.D.
   Memory mechanisms: the yin and yang of protein phosphorylation
   Curr. Biol. 11:R391-R394(2001)
   PMID: 11378403

3. Gross M.
   Mysteries of the memory molecules
   http://www.chemsoc.org/chembytes/ezine/2002/gross_dec02.htm

MOODY WALLPAPER
By Vivienne Baillie Gerritsen

How would you like to have wallpaper that changes color according to the seasons? Imagine a warm ochre hue for the cold winter months and a fresh yellowy tint for the hot summer months. Fantasizing? No, not really. Researchers in Madagascar recently discovered a very peculiar protein from the epidermis of a rare species of chameleon, *Chamaeleo differensis*, which is in fact a pigment and takes on different hues depending on various environmental criteria. The protein was baptized, in an outburst of scientific originality, chameleonin.

Chameleons are found mainly in Africa, Europe and Asia but over 50% of the world's chameleon species live in Madagascar. The first description of a chameleon dates back to the late 18th century. It was made by Jeanne Baret, the traveling companion of the renowned botanist Philibert Commerson (1728-1773), whom she accompanied in his numerous travels, amongst which a visit to the island of Madagascar. She remarked the peculiarity of the independently revolving eyes, the curled prehensile tail, the swiftness of the chameleon's tongue and, naturally, its capacity to change color. In her diaries, Jeanne Baret also mentions the use of dried chameleon flesh in folk medicine. Still today in Northern Africa, women who fear that their men have strayed a little add parts of chameleon to their food believing that this restores their fidelity.

Why is it that chameleons change color in the first place? It is the popular belief that they change for camouflage purposes and that they can produce any color on the market. This is not the case. The change of color in chameleons – which can occur in only 20 seconds - is an answer to the surrounding temperature, light and mood. An exasperated chameleon does not portray the same colors as a composed one; a warm chameleon does not look like a cool one; a sexually aroused chameleon bears no resemblance to one that shows no lust. And all this is due to a number of pigment cells, or chromophores, which lie under the surface of the skin. Hormonal, light or temperature differences affect the distribution of the pigments and the light

which is reflected off the epidermis offers a panel of colors, which range from browns to greens to blues, reds, yellows, black and white.
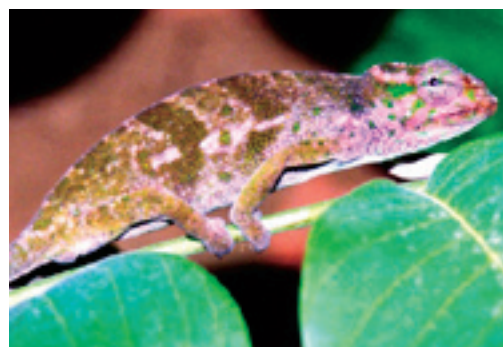

Figure 1   *Chamaleo differensis* above 18º Celsius.

Chameleonin, however, does not function in the same way at all. Unlike the other pigments, it is not found in the chromophores under the skin but on the skin's surface itself and seemingly only in *Chamaeleo differensis*. Chameleonin is not distributed regularly on the chameleon's skin but appears in patches along the animal's spine. In fact, it is thanks to this odd distribution that the pigment was discovered in the first place. Talisker and his research team were studying the distribution of color on *Chamaeleo differensis'* body under different conditions. To speed things up, part of the team were carrying out their research in the main lab while the others had moved down to the basement. In time, they noticed that despite carrying out identical tests on the chameleons, those in the basement sported a row of very small bright orange spots down their spine. In a true rational procedure, the chameleons from the basement were moved upstairs and vice-versa. And, lo and behold, the ex-lab chameleons also developed the spinal row of bright orange spots.

The next question was what was it that caused this orange eruption? After much investigation, it became apparent that it was the difference of temperature between the lab and the basement that was responsible for the color change. A closer inspection of the chameleons' skin actually revealed the same row of spots at any given temperature. However, they remain

invisible unless the temperature is quite low, because they adopt the natural yellowish/beige hue of the chameleon. Chameleonin is a globular protein with an end which protrudes like an antenna. Below 18ºC the protein's globular entity is deformed; the 'antenna' undergoes structural stress and its shape is modified accordingly. It is this change of shape which results in a change of color along the chameleon's spine, from the yellowish-brown hue to a bright orange. This phenomenon is similar to that which exists in lobsters, whose shell turns from blue to red when boiled*.

It is not clear why *Chameleo differensis* portrays this colorful display in the first place. Since all the individuals studied were in fact males and females don't seem to sprout orange patches, it is thought that it may be part of the mating game. *Chameleo differensis* does indeed mate only at night when temperatures are relatively low and the bright orange spots are clearly visible in the dark. One surprising development is the very recent use of chameleonin in interior decoration. Wallpaper, smeared with the protein trapped in a solution, is now available on the market and all the rage in Paris, London and New York. Above 18ºC, the wallpaper glows yellow; below 18º, a bright orange appears. The results are so encouraging that they will be stretched to the world of fashion design, materials and makeup. The next step, naturally, is to modify chameleonin's antenna to produce a wide range of colors. However, as Talisker remarks, 'a sitting-room is rarely below 18ºC and if it is, no-one would be sitting there, what's more there are perhaps more pressing issues to address than the color of your wallpaper and evening dress.'

* Read Spotlight issue 26.

Cross-references to Swiss-Prot

P99834: *Chameleo differensis* chameleonine

References

1. Talisker J., Oban T., Scapa W.
   Chameleonin: a chromomodulated protein from *Chameleo differensis*
   Helv. J. Zool. 271:45-32(2003).

2. Dimaki M., Valakos E.D., Legakis A.
   Variation in body temperatures of the African Chameleon *Chamaeleo africanus* Laurenti, 1768
   and the Common Chameleon *Chamaeleo chamaeleon* (Linnaeus, 1758)
   Belg. J. Zool. 130:87-91(2000).

3. Raxworthy C.J.
   A truly bizarre lizard
   http://www.pbs.org/edens/madagascar/creature3.htm

National Nodes

Argentina
Oscar Grau
IBBM, Facultad de Cs. Exactas, Universidad Nacional de La Plata
Email: grau@biol.unlp.edu.ar
Tel: +54-221-4259223 Fax: +54-221-4259223
http://www.ar.embnet.org

Australia
Sonia Cattley
RMC Gunn Building B19, University of Sydney,NSW, 2006
Email: scattley@angis.org.au
Tel: +61-2-9531 2948
http://www.au.embnet.org

Austria
Martin Grabner
Vienna Bio Center, University of Vienna
Email: martin.grabner@univie.ac.at
Tel: +43-1-4277/14141
http://www.at.embnet.org

Belgium
Robert Herzog, Marc Colet
BEN ULB Campus Plaine CP 257
Email:rherzog@ulb.ac.bemcolet@ulb.ac.be
Tel: +32 2 6505146 Fax: +32 2 6505124
http://www.be.embnet.org

Brasil
Gonçalo Guimaraes Pereira
Laboratório de Genômica e Expressão - IB
UNICAMP-CP 6109
13083-970 Campinas-SP, BRASIL
Tel: 0055-19-37886237/6238
Fax: 0055-19-37886235
Email: goncalo@unicamp.br
http://www.lge.ibi.unicamp.br

Canada
Laura Brown
Canadian Bioinformatics Resource, National Research Council Canada, Institute for Marine Biosciences,
Email: manager@cbr.nrc.ca
Tel: +1-902-426 7310 Fax: +1-902-426 9413
http://www.ca.embnet.org

Chile
Dr. Ricardo Baeza-Yates
Dept. of Computer Science, Santiago,
Email: rbaeza@dcc.uchile.cl
Tel: N/A
http://www.embnet.cl

China
Jingchu Luo
Room 303, Exchange Centre, Peking University
Email: luojc@cbi.pku.edu.cn
Tel: +86-10-6275 9001
http://www.cbi.pku.edu.cn

Colombia
Emiliano Barreto Hernández
Instituto de Biotecnología
Universidad Nacional de Colombia
Edificio Manuel Ancizar
Bogota - Colombia
Tel: +571 3165027 Fax: +571 3165415
Email : ebarreto@ibun.unal.edu.co
http://bioinf.ibun.unal.edu.co

Cuba
Ricardo Bringas
Centro de Ingeniería Genética y Biotecnología, La Habana,
Email: bringas@cigb.edu.cu
Tel: +53 7 218200
http://www.cu.embnet.org

Denmark
Hans Ullitz-Moeller
BioBase, University of Aarhus
Email: hum@biobase.dk
Tel: +45-86-13 9788
http://www.dk.embnet.org

Finland
Kimmo Mattila
CSC, Espoo
Email: Kimmo.Mattila@csc.fi
Tel: +358 9 457 2708
http://www.fi.embnet.org

France
PLAZA Jean-Marc
INFOBIOGEN, Evry
Email: plaza@infobiogen.fr
Tel: +33 1 60 87 37 11 Fax: +33 1 60 87 37 96
http://www.fr.embnet.org

Germany
Sandor Suhai
EMBnet node at the German Cancer
Research Center
Department of Molecular Biophysics (H0200)
Email: genome@dkfz.de
Tel: +49-6221-422 342 Fax: +49-6221-422 333
http://www.de.embnet.org

Greece
Babis Savakis
Institute of Molecular Biology and Biotech-
nology
Heraklion, Crete
Email: savakis@nefeli.imbb.forth.gr
Tel: +30-81-391 114 Fax: +30-81-391 104
http://www.imbb.forth.gr

Hungary
Endre Barta
Agricultural Biotechnology Center
Szent-Gyorgyi A. ut 4. Godollo,
Email: barta@abc.hu
Tel: +36 30-2101795
http://www.hu.embnet.org

India
H.A.Nagarajaram
Laboratory of Computational Biology &
Bioinformatics
facility, Centre for DNA Fingerprinting and
Diagnostics
(CDFD), Hyderabad
Email: han@www.cdfd.org.in
Tel: +91 40 7155607 / 7151344 ext:1206
Fax : +9140 7155479
http://www.in.embnet.org

Israel
Eitan Rubin
INN (Israeli National Node) Weizmann
Institute of Science
Department of Biological Services, Biologi-
cal Computing Unit, Rehovot
Email: Eitan.Rubin@weizmann.ac.il
Tel: +972- 8-934 3456
http://www.il.embnet.org

Italy
Cecilia Saccone
CNR - Institute of Biomedical Technologies
Bioinformatics and Genomic Group
Via Amendola 168/5 - 70126 Bari (Italy)
Email: saccone@area.ba.cnr.it
Tel. +39-80-5482100 - Fax. +39-80-5482607
http://www.it.embnet.org

Mexico
Cesar Bonavides
Nodo Nacional EMBnet, Centro de Investig-
ación sobre
Fijación de Nitrógeno, Cuernavaca, Morelos
Email: embnetmx@cifn.unam.mx
Tel: +52 (7) 3 132063
http://embnet.cifn.unam.mx

The Netherlands
Jack A.M. Leunissen
Dept. of Genome Informatics
Wageningen UR, Dreijenlaan 3
6703 HA Wageningen, NL
Email: Jack.Leunissen@wur.nl
Tel: +31 317 484074
http://www.nl.embnet.org

Norway
Rune Groven
The Norwegian EMBnet Node
The Biotechnology Centre of Oslo
Email: admin@embnet.uio.no
Tel: +47 22 84 0535
http://www.no.embnet.org

Poland
Piotr Zielenkiwicz
Institute of Biochemistry and Biophysics
Polish Academy of Sciences Warszawa
Email: piotr@pl.embnet.org
Tel: +48-22 86584703
http://www.pl.embnet.org

Portugal

Pedro Fernandes
Instituto Gulbenkian de Ciencia
Unidade de Bioinformatica
2781-901 OEIRAS
Email: pfern@igc.gulbenkian.pt
Tel: +351 214407912 Fax: +351 2144079070
http://www.pt.embnet.org

Russia

Sergei Spirin
Biocomputing Group, Belozersky Institute
Moscow
Email: sas@genebee.msu.su
Tel: +7-095-9328825
http://www.genebee.msu.su

Slovakia

Lubos Klucar
Institute of Molecular Biology SAS Bratislava
Email: klucar@embnet.sk
Tel: +421 7 5941 2284
http://www.sk.embnet.org

South Africa

Ruediger Braeuning
SANBI, University of the Western Cape,
Bellville
Email: ruediger@sanbi.ac.za
Tel: +27 (0)21 9593645
http://www.za.embnet.org

Spain

José M. Carazo, José R. Valverde
EMBnet/CNB, Centro Nacional de Biotec-
nología, Madrid
Email: carazo@es.embnet.org,
jrvalverde@es.embnet.org
Tel: +34 915 854 505 Fax: +34 915 854 506
http://www.es.embnet.org

Sweden

Nils-Einar Eriksson, Erik Bongcam-Rudloff
Uppsala Biomedical Centre, Computing
Department, Uppsala, Sweden
Email: nils-einar.eriksson@bmc.uu.se
erik.bongcam@bmc.uu.se
Tel: +46-(0)18-4714017, +46-(0)18-4714525
http://www.embnet.se

Switzerland

Laurent Falquet
Swiss Institute of Bioinformatics, CH-1066
Epalinges/Lausanne
Email: Laurent.Falquet@isb-sib.ch
Tel: +41 (21) 692 5954 Fax: +41 (21) 692 5945
http://www.ch.embnet.org

United Kingdom

Alan Bleasby
UK MRC HGMP Resource Centre, Hinxton,
Cambridge
Email: ableasby@embnet.org
Tel: +44 (0) 1223 494535
http://www.uk.embnet.org

Specialist Nodes

EBI

Rodrigo López
EBI Embl Outstation, Wellcome trust Genome
Campus
Hinxton Hall, Hinxton, Cambridge, United
Kingdom
Email: rls@ebi.ac.uk
Phone: +44 (0)1223 494423
http://www.ebi.ac.uk

ETI

P.O. Box 94766
NL-1090 GT Amsterdam, The Netherlands
Email: wouter@eti.uva.nl
Phone: +31-20-5257239
Fax: +31-20-5257238
http://www.eti.uva.nl

EU

Dr Bernard Mulligan
DG Research - European Commission
Brussels BELGIUM
Email: bernard.mulligan@cec.eu.int

Hoffmann-La Roche

Martin Ebeling
F. Hoffmann-La Roche Ltd, Basel, Switzerland
Email: martin.ebeling@roche.com
Phone: +41-61-68-71398
http://www.roche.com

ICGEB

Sándor Pongor
International Centre for Genetic Engineering
and Biotechnology
AREA Science Park, Trieste, ITALY
Email: pongor@icgeb.trieste.it
Phone: +39 040 3757300
http://www.icgeb.trieste.it

LION Bioscience

Peter Rice
LION Bioscience AG, Heidelberg, Germany
Email: Peter.Rice@uk.lionbioscience.com
Phone: +44 1223 224700
http://www.lionbioscience.com

MIPS

H. Werner Mewes
Email: mewes@mips.embnet.org
Phone: +49-89-8578 2656
Fax: +49-89-8578 2655
http://www.mips.biochem.mpg.de

Pharmacia - Biovitrum

Timothy Wood
Biovitrum, Stockholm, Sweden
Email: timothy.wood@eu.pnu.com
Phone: +46 (8) 695 9134
http://www.pharmacia.se

Sanger Institute

Michelle Clamp
Welcome Trust Sanger Institute, Wellcome
Trust Genome
Campus, Hinxton Hall, Hinxton CB10 1SD,
Cambridge, United Kingdom
Email: michele@sanger.ac.uk
Phone: +44-1223-494 967
Fax: +44-1223-494 919
http://www.sanger.ac.uk

UMBER

Terri Attwood
School of Biological Sciences, The University
of Manchester
Oxford Road, Manchester M13 9PT, UK
Email: attwood@bioinf.man.ac.uk
Phone: +44 (0)61 275 5766
Fax: +44 (0) 61 275 5082
http://www.bioinf.man.ac.uk/dbbrowser

TECH-MGR

Email: tech-mgr@embnet.org
The team gives support to EMBnet nodes
and helps them with maintenance and
troubleshooting.
The team is formed of experienced system
administrators and programmers who ensure
the availability of local services for all EMBnet
users.

EMBnet.news

ISSN 1023-4144

Dear reader,

If you have any comments or suggestions
regarding this newsletter we would be very
glad to hear from you. If you have a tip you
feel we can print then please let us know.
Please send your contributions to one of the
editors. You may also submit material by
Internet E-mail.

Past issues of embnet.news are available as
PostScript or PDF files (ISSN 1023-4144). You
can get them by anonymous ftp from:
the EMBnet organisation Web site
http://www.embnet.org/download/embnet-
news
the Belgian EMBnet node
ftp://ftp.be.embnet.org/pub/embnet.news
the UK EMBnet node
ftp://ftp.uk.embnet.org/pub/embnet.news
the EBI EMBnet node
ftp://ftp.ebi.ac.uk/pub/embnet.news

Submission deadline for next issues:

June 30, 2003
October 31, 2003