embnet.news

Volume 2 Nr.2 (ISSN1023-4144)

June 21st, 1995

Welcome once again to embnet.news! One of the many strengths of EMBnet is its ability to support projects in training, help and data dissemination. Such vital aspects of computational molecular biology are often difficult to manage, or to be funded by individual sites. In this issue many articles emphasise and demonstrate the collaborative role of EMBnet in these areas.

EMBnet has recently set up a virtual help desk. This has been initiated to help EMBnet as a whole. It allows the EMBnet node managers to draw on a cross-European knowledge resource for support and technical problems. Five node managers from EMBnet are involved in helping other nodes to setup systems, maintain them and troubleshoot user problems.

A recent example on the technical side has been the usage of MIME. The MIME protocol allows you to transmit letters, images and other information via electronic mail. Messages to, and swift replies from, the virtual help desk have enabled EMBnet nodes to quickly utilise this facility and communicate even more effectively. EMBnet succesfully collaborates on more than one level. The article on the extension of the GCG package is another demonstration of effective group collaboration.

Other articles in this issue of embnet.news include database development, a node focus, a node manager interview, a bioinformatics theory section, useful commands to type at your terminal and the 'node news' section. As ever, important new developments are shown.

Last but not least, the 'announcements' section details interesting meetings and events. When will YOU initiate a collaboration? Maybe at one of these meetings?

The embnet.news editorial team

Alan Bleasby (UK)
Reinhard Doelz (CH)
Robert Herzog (BE)
Andrew Lloyd (IE)
Rodrigo Lopez (NO)

BITS
BioInformatics Theory Section
STAYING ALERT ABOUT SEQUENCES IN THE DATABANKS

Philippe Alard, Belgian EMBnet node
Florian Eggenberger, Swiss EMBnet node

Molecular biologists are faced with an increasing volume of sequence data. This article presents approaches to accomodate this data from both the Belgian (BEN) and Swiss EMBnet nodes.

CSA and CSSA
BEN's new sequence arrival warning tools

BEN (the Belgian EMBnet node) offers its users the possibility of being warned of the arrival of new sequences in databanks. This relieves the user from the burden of repeated searches at regular intervals. Users interested in this service can submit queries by email at ben@vub.ac.be or by WWW at http://ben.vub.ac.be/ben/services.html
The submission is not automatic, the queries are viewed by a member of the BEN staff and are invoked manually.

CSA: Current Sequence Awareness

CSA runs on the daily updates of EMBL. It finds sequences related to a certain field of interest by looking for strings of characters in the documentation.

The syntax of a CSA query is as follows :
a string can contain any character; however, if you want to include blanks, the string must be put between quotes, e.g.

Contents

Editorial	1
BITS - Staying alert about new sequences	1
Node Focus - Biozentrum Basel, Switzerland	2
INTERviewNET - Sandor Pongor, Trieste, Italy	3
Database development - CpG Islands	4
Program development - EGCG 8.0	5
Data Transfer Technology - Passing multimedia messages	7
Tips from the computer room - Compressing and archiving	9
Node News	10
Announcements	12
A biocomputing thesaurus	16
The EMBnet nodes	17
embnet.news information	18

"protein c"; note the importance of blanks, e.g. a search for the string "synthetic" would also find a sequence with the word photosynthetic in its documentation; this can be avoided by giving the string " synthetic".

By default the search for letters is case insensitive. If you want a case sensitive search then put an underscore "_" before the letters, e.g. _C_O (which will match CO2 and CO-gene, but not coenzyme and Co.).

A query can consist of several strings, connected by the logical operators AND, OR or NOT. Logical operations can be nested at any depth by using parentheses. Note the importance of the positions of parentheses, e.g. "(virus and phage) or protease" does not give the same result as "virus and (phage or protease)".
You can choose if you want to receive the complete databank entry or only the bibliography, sequence or title.

CSSA: Current Sequence Similarity Awareness

CSSA searches for sequences that are similar to a query sequence and thus presumably derive from a common ancestral sequence and/or have a similar function. CSSA uses the program BLAST. CSSA is the natural complement of CSA. It can find sequences which are related to a query sequence but don't share any of the keywords findable by CSA. Moreover, CSSA can be very useful if you have a equence of unknown function and an initial search against the databanks yields nothing. At present, you can choose between :

- a blastn of a nucleic acid sequence against the daily updates of EMBL.
- a tblastn of a protein sequence against the daily updates of EMBL, translated in the six reading frames.
- a blastp of a protein sequence against the weekly updates of GenPept.

You must specify if you want to search for perfect matches, near perfect matches, highly related sequences or weakly related sequences.

BEN proposes to extend these services to the whole of EMBnet. Queries submitted from domains in Europe will be honoured while they remain within reasonable limits. The results will be returned by E-mail.

The database growth and search theme is now continued by the Swiss EMBnet node.

FastAlert

The alerting tools from Basel

Given the daily growth of the nucleotide sequence databases, today's scientists have to redo their sequence searches at frequent intervals in order to re-evaluate previous findings. Recently, services such as BEN's CSSA and ExPASy's SwissShop have become available to allow people to search the latest nucleotide entries.

We have implemented an automatic alerting system that allows people to search a sequence against protein or DNA database sets. The system is called FastAlert as it integrates with part of W.R.Pearson's FastA package. Access control is provided by the hostname-bound authorisation mechanism of HASSLE v5. Three actions may be initiated:

- registration
- deregistration
- obtain a copy from the most recent FASTA run

Upon registration, a query sequence is searched against the full peptide and DNA databank, including the updates. FastAlert reports contain the FastA best score list and the probability estimates for each entry as computed by the 'prdf' program, which is part of the FastA package. Unless the user has reregistered, she or he will get further reports on a weekly basis by electronic mail. Initially, a full report is generated, whereas subsequent reports contain only the entries of the current best scores list which did not appear in the previous search. As the system is not based on WWW but relies on the HASSLE protocol, the job may run on any HASSLE-compatible node whether within EMBnet or external to it.

Unlike WWW based systems our system does not require the knowledge of any URL - just click and let HASSLE's resource discovery mechanisms do the job for you. The user interface allows file input. Hence, it is not necessary to type in the sequence.
The FastAlert system runs on PC (Windows 3.x, MacOS, OS/2) and on Motif/DecWindows environments. An ASCII interface will be available soon.

Node focus

EMBnet Switzerland
Biozentrum der Universität Basel
Klingelbergstrasse 70
CH 4056 Basel, Switzerland

Reinhard Doelz, Swiss EMBnet node

General Information

The Swiss Node of EMBnet is based at the BioComputing facility of the Biozentrum in Basel. We provide sequence updates, full databases and specially formatted databases to most sites within Switzerland, including cantonal universities and the federal universities ETH and EPFL. We

also provide services to smaller sites on a UNIX cluster. All sites can access our ftp and other internet services.

Around 300 researchers at the Biozentrum of Basel University utilise the data on a VMS cluster. The tasks of the BioComputing Facility include teaching and training for the Biozentrum as well as data maintenance.

The facility is funded by Basel University and receives grants from the Schweizerische Nationalfonds as well as the Bundesamt fuer Bildung und Wissenschaften, in order to cope with EU-funded research projects.

Network access is provided via our national academic Internet provider which is called SWITCH.

The Swiss node has more than 70 GByte of disc space on 10 different units. including DEC and Silicon Graphics workstations, personal computers (Mac, and Intel-based systems). Additionally, we access the main Basel University computing centre's DEC VMS cluster and compute servers including a DEC AXP 4000 system for BLAST and FASTA searching, as well as a 4096 processor MasPar for MPsrch Smith and Waterman sequence searching.

Research projects

We were instrumental in porting the SRS system (by Thure Etzold) to UNIX and this is now one of our main database retrieval engines. Our latest activites include three main projects:

Quality control on Internet Newsgroups (Nicole Redaschi): "bionet.software.www" can be accessed via electronic mail, NEWS, and, in a graphical version, under WWW. It is moderated in order to ensure quality and relevance of postings. Articles are validated for the availability of the links, formatted, and are archived (see http://www.ch.embnet.org/bio-www).

HASSLE v5 (F.Eggenberger) has been completely redesigned to feature a structured application programmer's interface (API) and to comply with the scientist's preferences for generic graphical user interfaces on the Mac and PC. The layered design has been implemented in ANSI C using the NCBI toolbox and the IBM class library user interface (ICLUI) for the graphical user interface (GUI) on PC, Mac, and Motif/DecWindows platforms. Nicole Redaschi applied 'software inspection' (i.e., read written code and compare it with the design specification), paired with extensive documentation. HASSLE v5 runs on most UNIX flavours, various flavours of IP implementations on VMS, Windows 3.x, MacOS, and OS/2. Beta testers are welcome.

Sequence updating in heterogenous clusters (R.Doelz): A next-generation management system to track, format and validate data on various streams of information updates has been created. The HASSLE-based central information processor automatically keeps track of incoming information, validates it and makes status reports on the various databases available on http://www.ch.embnet.org/development/status/HASSLE.STATUS. The system uses HASSLE for failsafe operation, resulting in correct operation even if the main computer running the system is unavailable due to operational reasons.

Further reading: http://www.ch.embnet.org/

INTERviewNET

In this issue Alan Bleasby talks to Sandor Pongor from ICGEB Trieste.

AJB: Hello Sandor, perhaps we can start with a short description of what goes on at your site.

SP: ICGEB, as an international organisation, has 50 so called signatory countries, in Asia, Central+South America, East+Central Europe etc. Our 700 users log in from these countries, usually from molecular biology institutes affiliated to ICGEB. Otherwise, we work very much like other nodes. As of March we have a new systems manager, Maria Verina, who took over Valeria Bevilaqua's place.

AJB: You have produced the SBASE database. Can you tell us something about this?

SP: I was always fascinated why interesting domain homologies are so easily overlooked. The idea of the SBASE domain library came about because we figured that most biologists will never dig into sequence databases deeper than simple database searching. So we decided to make a sequence collection of functional domains which can be searched by the popular programs and gives an instant answer. Other than this simple trick, we add information to the sequences, we cluster them into similarity groups, cross-reference them to all the major databases - for this we have been using www since about a year. Recently we started to add minireviews of the domain types. By the way, the name SBASE comes from Segment Database, the initial name of the project. It is not the Sandor database.

AJB: There are various methods for looking at motifs; for example consensus sequences [heaven help us], patterns and matrixes. How does SBASE fit into this scheme?

SP: Motifs are generalized representations that one has to build and test first. Generalizations on the other hand are always biased against the rare examples. In contrast, SBASE is not a motif collection but a sequence database, and it contains both the typical and the atypical sequences. In this way one can notice rare homologies including those for which

there are no known motifs yet. Simply put, motif collections are good for known and well-characterized domain-types. SBASE on the other hand may also help to find novel motif-candidates in new sequences.

AJB: What software and databases would you recommend to people looking for motifs?

SP: Naturally, I use SBASE with or without reason... More seriously, when looking for known motifs, I mostly use PROSITE and PRINTS because of the thorough literature reviews. To find a new motif in a new sequence, I first use a BLAST search on SBASE. This is now possible through the e-mail server, which also gives a graphic output of the homology regions. For this purpose we use the HSPCRUNCH program of Eric Sonhammer and Richard Durbin, which is a very useful program for handling BLAST outputs. Generally speaking, I recommend everyone to use the e-mail or www servers for motif searches - simply because this guarantees that the database is updated and the software is functional.

AJB: Being an EMBnet node which provides many training courses perhaps you can tell us something about them and your view of their importance?

SP: We give one major computer course each year and several smaller primer courses for new users. The major course is usually in the summer and usually takes 10 days. This year we will concentrate on database searching, with emphasis on e-mail and www servers. This will be a practical course for 40 students coming from 13 countries.

AJB: Has there been a benefit in your courses in being able to tap into external EMBnet speakers?

SP: For the third year now our course is run by Jack Leunissen and David Judge. This year we will also have Thure Etzold, Rob Harper and David Landsman from NCBI, in addition to Amos Bairoch and Martin Bishop who have been coming here for several years. EMBnetters provide a unique source of expertise for us.

AJB: The next EMBnet meeting is in Madrid. Have you any plans?

SP: Beer!

The SBASE database is available via anonymous ftp from the ICGEB server located at icgeb.trieste.it in the pub/SBASE directory. Information on SBASE via WWW is available at http://base.icgeb.trieste.it/sbase/"

Database Development

CpG Islands

Rodrigo Lopez, Norwegian EMBnet node

Introduction

This article describes the materials and methods employed to produce and maintain the CpG island database (Larsen et al., 1992) which is provided by the Biotechnology Centre of Oslo and the Norwegian EMBnet node.

Definitions

CpG islands are short and dispersed regions of unmethylated DNA with a high frequency of CpG dinucleotides relative to the bulk genome. Although the cytosines in the CpG dinucleotides in mammalian DNA are highly methylated, a small fraction of the genome contains very stable methylated CpGs in all tissues.
CpG islands are commonly defined as regions of DNA of at least 200 bp in length and that have a G+C content above 50% and a ratio of observed vs. expected CpGs close to or above 0.6. This criterion was originally established by Gardiner-Garden and Fromer (1987).
In the haploid genome it is estimated that about 30,000 such islands exist (Bird, 1987). In the bulk of the genome outside these islands, CpG is depleted to about 20% of the expected value, and CpGs are partly or completely methylated. CpG suppression is thought to be caused by mutation following deamination of methylated cytosine to thymine.
CpG islands are found to be associated with 5' ends of housekeeping genes and of some tissue-specific genes. They have open chromatin and it has been postulated that they may be sites of interaction between transcription factors and promoters.

Materials and Methods

The CpG island database deals at present with human genes appearing in major releases of the EMBL nucleotide sequence database but it is hoped that in the future it will include islands from other mammalian species. The present release is the third one and is derived from EMBL release 40.
Several programs have been developed for the initial screening of the database. That is, the process of scanning the entire human section of EMBL for gene entries that are complete. Genes that are spread through several entries are assembled providing that appropriate information is annotated in the feature table. The parser used for this task is a program developed using the GCG (Devereux et al., 1984) procedure library. Once the primary index of database entries is completed, another parser is used to

retrieve primary references to the sequences. The entry index and the references are then used to establish that no duplicates exist between 'new' and 'old' entries and if they do, the new entry is used.
The next step in the process is to analyse the sequences for CpG islands. This is done by a program called CpGPlot, which has a built-in parser that screens entries for CpG island criteria. The length must be at least 200 bp. G+C % must be above 50% and the Observed/ Expected CpG ratio has to be close to or above 0.6. During the first two releases of the database the analysis and detection is done by scanning the graphical output of CpGPlot. The program was originally developed at EMBL as part of GCGEMBL program package (see the EGCG article) under the supervision and patience of Peter Rice. The original program has undergone several modifications. The latest being the ability to write to a file any islands which satisfy the criteria.
The potential islands are further screened to identify, for example, well known CpG rich sequences such as the repetitive elements ALU and LINE. Entries are not automatically discarded as it has been observed that repetitive elements co-exist with real CpG islands in a number of entries.
Following CpG detection, and according to the annotation in the EMBL entries, potential promoters are identified as well as the position of an island with respect to the entire gene sequence. This forms the basis for classifying the islands as upstream (or including the transcription start), islands covering one or several transcription units or islands covering the entire gene. SpI sites are searched for all islands and annotated relative to an island sequence.
The next step involves identifying the expression pattern associated with CpG islands and finally, assembling all the information into a flat-file formatted database that uses a description line concordant with the EMBL database.

Conclusions

The average length of widespread expression CpG islands is about 240 bp according to the detection criteria. But there is no typical size. Most are between 200 and 1400 bp with a majority of islands being 200 - 400 bp. However, the average size of islands associated with limited expression is slightly smaller.
Some genes are associated with more than one CpG island. Few widely expressed genes are associated with more than one island, whereas a large percentage of genes of limited expression have more than one island.
Almost all islands cover one or more exons. This implies that CpG islands may be used to identify transcripts since they tend to cover exons. Furthermore, CpG islands cover the whole or part of the promoter regions containing canonical TATA and CAAT boxes and include GC-rich promoters lacking a typical TATA box. Such promoters have been considered typical of housekeeping genes.
All analysed genes with widespread expression have an island covering the transcription start, and a few of these genes have an additional island 3' of the transcription start. Less than half of tissue specific expressed genes, and genes with limited expression, are associated with CpG islands and cover one or more exons.
The CpGPlot program is available as part of the EGCG package. This package may be obtained from the Sanger Centre or from the Norwegian EMBnet node via anonymous ftp :

- ftp.sanger.ac.uk/pub/pmr/egcg8/
- ftp.no.embnet.org/pub/egcg/egcg8

The CpG Isle 3.0 database may be obtained from:

- ftp.ebi.ac.uk/pub/databases/cpgisle and
- ftp.no.embnet.org/cpgisle

References

Bird, A.P. (1987). CpG islands as gene markers in the vertebrate nucleus. Trends Genet. 3: 342-347
Devereux, J. Haeberli, P. and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic Acid Res. 12: 387-395
Gardiner-Garden, M. and Frommer, M. (1987). CpG islands in vertebrate genomes. J. Mol. Biol. 196: 261-282
Larsen, F. Gundersen, R. Lopez, R. and Prydz, H. (1992) CpG islands as gene markers in the human genome Genomics 13: 1095-1107

Program development

EGCG 8.0

Peter Rice , Sanger Centre, Hinxton, UK
Rodrigo Lopez, Norwegian EMBnet node
Reinhard Doelz, Swiss EMBnet node
Jack Leunissen, Netherlands EMBnet node

Historical Background

EGCG is an abbreviation of "Extended GCG". The package began life at EMBL in Heidelberg, Germany as a small collection of programs to support EMBL's research activities, in particular the development of automated DNA sequencing. The programs that resulted were published [1] and released on the new EMBL Network File Server [2] under the name "GCGEMBL".

In 1989, Rodrigo Lopez came to EMBL from the Norwegian EMBnet node for a sabbatical visit, and developed several new programs which were included in the next release "GCGEMBL 6.0", along with additional collections "GCGDBASE" for database maintenance and "GCGQUICK" for enhanced Quick Searching programs [3].

When GCG 7.0 was released, these collections were merged and renamed as "EGCG". At the time, the "E" stood for "European" but we had already intended it to be changed to "Extended" once a non-European contribution appeared. EGCG 7.0 was distributed on the EMBL Network File Server, but was also included on the new EMBL ftp site, and was submitted to GCG for distribution as "unsupported" software to all their licensed VMS users.

When GCG released a Unix version, and as EMBnet nodes began to migrate to Unix systems, the most popular EGCG programs were ported to Unix by Reinhard Doelz (Swiss EMBnet node), Jack Leunissen (Netherlands EMBnet node) and Rodrigo Lopez in time to be distributed with GCG 7.2 as "unsupported".

Jaakko Hattula, a summer student from Tampere University of Technology in Finland, completed the addition of a command line interface to all GCG programs, and contributed to the development of some new ideas for EGCG graphics.

Current Status

When the earliest test versions of GCG 8.0 appeared, it was clear that conversion would be a non-trivial task. We undertook a major rewrite of some sections of the EGCG code, including the development of our own procedure libraries EGENLIB and EAPPLIB, which has made our code much easier to maintain. Additional advantages include the potential to run any program in batch with minor code changes, and to provide many extensions to the graphics options.

Unlike GCG, the Unix and VMS versions of EGCG share common source code and documentation files. New applications are developed mainly on Unix and tested on VMS with very few problems.

Users of the EMBL Quick service will be happy to know that we have continued to support this in EGCG 8.0. Several users at the Sanger Centre have found Quick Searching to be extremely useful, and we will continue to maintain and extend these programs.

We do not, however, have any plans to support the GCG WPI user interface. Although the EGCG programs will work with WPI, we have no time to spend on writing the configuration files.

EGCG is, like GCG, written in both Fortran and C. We will continue to maintain support for both languages.

Distribution

EGCG is no longer distributed with GCG. We prefer to distribute the package directly to users through FTP, allowing us to more rapidly provide new programs and updates, and to keep control of our own software. The EGCG programs are distributed freely to academic users thourghout the world. The only requirement is that users must have a valid GCG licence to build the programs.

We have a new E-mail address for support. All questions and comments about EGCG can now be sent to egcg@embnet.org .

Program Examples

There are over 70 programs in EGCG 8.0. Old favourites include PRETTYPLOT which produces boxed sequence alignments; CPGPLOT which identifies CpG islands; SIGCLEAVE which identifies signal peptide cleavage sites; EQUICKSEARCH which runs QuickSearch faster and with less system resources than GCG's version (and QUICKMATCH which removes false hits); GELSTATUS and GELPICTURE which report on fragment assembly projects.

Unix users will find several new programs that were only in the EGCG 7 VMS release. These include MAPSELECT which creates local enzyme data files; PEPCOIL which identifies coiled-coil regions and leucine zippers; NEWFEATURES which analyses and edits the EMBL feature table; TPROFILESEARCH which compares protein profiles to DNA databases; and PEPALLWINDOW which plots hydrophobicity for aligned sequences.

New programs in release 8.0 include POLAND which simulates transition curves for DNA and RNA; PRETTYBOX which produces shaded sequence alignments; and STSSEARCH which searches a DNA database for matches to primer pairs.

The EMBnet Connection

Throughout the development of the EGCG package, there have been excellent collaborations with the EMBnet community. Academic software such as EGCG, developed for a limited number of users, cannot under normal circumstances be maintained to a sufficient standard for reliable public use. Only the large number of EMBnet users, and the efforts of the EMBnet support personnel, have made it possible to produce a robust package. In return, we receive many suggestions for further developments which we can provide and test locally and then easily incorporate into the next release.

Acknowledgements

We are very grateful to (in alphabetical order) Rein Aasland, Wilhelm Ansorge, Peer Bork, Thure Etzold, Toby Gibson, Tom Kristensen, David Mathog, Franc Pattus, Kate Rice, Christian Schwager, Peter Sibbald, Julie Thompson, Hartmut

Voss, Gert Vriend and Rick Westerman for their many contributions and critical comments as users of the EGCG Programs.

We are also deeply indebted to the staff of GCG Inc. who provided rapid and helpful answers to our many questions during the development of the programs. Many thanks to Irv Edelman, Maggie Smith, Donald Katz, Michael Hogan, Joseph King, Mary Schulz and especially John Devereux.

Into the Future

Since EGCG was first released with GCG, it has become very popular at a number of sites around the world. We now have a backlog of program submissions to include in EGCG which we plan to include as soon as the C interface to the procedure library is in place. Contributing authors in the backlog include William Pearson and Rainer Fuchs. Additional contributions are always welcome.

We also have to work on improvements to our procedure library. After consulting with the users, we intend to extend the scope of EGCG beyond the GCG package. We are interested especially in supporting a wider range of data formats, providing an improved user interface and providing documentation more suited to the needs of novice users. Past experience leads us to strongly prefer European solutions in these areas. At present we are looking at SRS [4], HASSLE [5], Compugen's Bioccelerator [6], www2gcg from the Belgian EMBnet node, the UK CCP11 project and a number of other options.

Availability

Current major version: 8.0 beta
URL: ftp://ftp.sanger.ac.uk/pub/pmr/egcg8
E-mail contact: egcg@embnet.org
pmr@sanger.ac.uk

References

[1] Edwards A. et al. Automated DNA Sequencing of the Human HPRT Locus. Genomics 6:593-608 (1990).
[2] Stoehr P.J. and Omond R.A. The EMBL Network File Server. Nucleic Acids Res. 17:6763-6764 (1989).
[3] Devereux J. A Rapid Method for Identifying Sequences in Large Nucleotide Sequence Databases.
Ph.D. Thesis, 1988 (University Microfilms Inc., Ann Arbor, Michigan, USA.)
[4] Etzold T. The Sequence Retrieval System (SRS) on the WWW. EMBnet News 1(2):5-6 (1994)
[5] Doelz R, Eggenberger F. and Wadley, R. Biocomputing on a Server Network. EMBnet News 1(2):6-8 (1994)
[6] Esterman L. Bioccelerator: A Currently Available Solution for Fast Profile and Smith-Waterman Searches. EMBnet News 2(1): 5-6 (1994)

Data transfer technology

Passing Multimedia Messages

The EMBnet Virtual Help Desk Team
Alan Bleasby, Weiyun Chen, Reinhard Doelz, Leon Esterman, Peter Gad, Jack Leunissen, and Rodrigo Lopez

Electronic communication is improving rapidly. Electronic Mail, once intended to send text messages only, has become a very versatile tool to pass several articles in the same message, and have text, audio, or graphics included. World-Wide Web browsers like Netscape and Mosaic utilise these capabilities too. This allows the launch of local 'viewers'.

The obvious need for a standard to describe multimedia or multipart messages has been met with the MIME definitions. MIME is an abbreviation for 'Multipurpose Internet Mail Extensions' and covered by the Internet standards RFC 1521 and RFC 1522. MIME and its associated software is freely available and offers a way to interchange text in languages with different character sets, and multi-media email among many different computer systems that use Internet mail standards.

If you are bored with plain text e-mail messages, thanks to MIME you now can create and read e-mail messages containing features like:

- character sets other than ASCII, and otherwise enriched text
- images, movies, sounds
- PostScript, Brookhaven structures, other special file formats

However, as mentioned above, MIME types are useful in other areas. The use of the World-Wide Web has become more exciting with the ability to launch dedicated programs, 'Viewers', which will treat messages as input for specific actions. In biology, the use of protein structure coordinates in MIME-encoded pages has been used successfully (see below).

In order to learn more about MIME, you might wish to look at the WWW page at http://www.oac.uci.edu/indiv/ehood/MIME/MIME.html for details. There is comprehensive documentation. An FAQ (Frequently Asked Question Sheet) is available in the newsgroup comp.mail.mime, or you might reference the document at http://www.lib.ox.ac.uk/internet/news/faq/archive/mail.mime-faq.mime0.html

MIME on electronic Mail

MIME type messages will be displayed without notice if you use a program for mail reading which can handle these

messages automatically.
The most recent mail environments (including ones like EMAIL/PMDF or PINE) support MIME text without problems.

Troubleshooting MIME on electronic Mail

There are two potential scenarios which make life difficult:

1) If you receive a mail message by a program which is MIME-unaware, (such as the system-standard 'mail' on UNIX or VMS systems), a header will be displayed telling you that only the first part will be decoded and readable, whereas the other parts usually represent a very unreadable area even if this part was encoded from normal text (so-called BASE64 encoded).

2) Similar to 1), you might receive a text which is unreadable despite the fact that your mail reader knows about MIME. Unfortunately, some mail forwarders on the international networks rewrite the headers in order to reduce 'mailer babble' and get the reply-to: field right. MIME messages add special fields into the header in order to display multipart messages. These fields are, if not explicitly considered, thrown away by the mailer and even MIME aware readers will fail to interpret the message.

Solution:

A minimalist MIME-reading program, munpack, is available from this URL:
ftp://ftp.andrew.cmu.edu/pub/mpack/
which is basically the MIME equivalent of uudecode/binhex. Source as well as binaries for DOS, Macintosh, OS/2 and other small computer operating systems are available.

MIME on the World-Wide Web

Browsers like Netscape and Mosaic can be configured to use so-called 'helper applications' which will allow viweing of images or other data outside of the WWW browser program. The most beneficial MIME types in this area are structural data ('coordinates') of biological macromolecules, currently proteins. Roger Sayle has written the program RASMOL, which is freely available, and can be configured to be used as a 'viewer' of data, such as a specially encoded Brookhaven protein structure file. You might refer to the documentation of the course 'Principles of Protein Structure', which is organised by Birkbeck College in collaboration with the Virtual School of Natural Sciences (VSNS) of the Globewide Network Academy (GNA). See the following URL for details:

Protein Structure course :
http://www.cryst.bbk.ac.uk/PPS/index.html
RASMOL software : ftp://ftp.dcs.ed.ac.uk/pub/rasmol
Mosaic configuration : http://www.ncsa.uiuc.edu/SDG/Software/Mosaic/Docs/mailcap.html

Further reading :

Tools test:
PC Mag. recently published an article about MIME and tested various MIME compatible mail agents. The following URLs provides some of the stuff-info PC Mag. worked on.
•http://www.ziff.com:8001/~pcmag/1408/pcm00054.htm

ECSMail mail/news client
This is a Windows-based mail and USENET news client, now available in most PC labs. Try
•http://www.wlv.ac.uk/pub/wincomms/news/ecsmail.html

PINE for VAX/VMS
A UK group has investigated software (preferably free) for handling MIME format mail on VMS in their own environment. This document is likely to be incomplete in many respects, and should be regarded only as describing work-in-progress at this stage.
•http://d1.ph.gla.ac.uk/~flavell/vms-pine.html

pine, pc-pine:

Pine files and documentation are available via FTP or WWW:
•ftp://ftp.cac.washington.edu/pine
•http://www.cac.washington.edu/pine

Eudora (MAC, PC/WINDOWS)

•http://www.qualcomm.com/quest/EbyQCIntro.html

Z-Mail (Various platforms)

•http://www.z-code.com/Z-Code/zcode.html

Tips from the computer room

Compressing, uncompressing and archiving files

Andrew Lloyd, Irish EMBnet node

This note offers help with two distantly related problems. 1) How to make the best use of your quota on a multi-user system. 2) How to handle certain files, utilities and programs that you have obtained from another computer (for example by ftp).

All computer systems have a finite amount of storage space and many system managers will set up each user account with a quota or allocation, limiting the number and/or size of the user's files. By preventing users from indefinitely accumulating files (as from BLAST searches) which are largely redundant or no longer required, these quotas are desirable. But by preventing users from having all necessary information from their many projects readily available, quotas are a necessary evil. One way to optimise your quota is to selectively compress certain files or indeed directories full of files. This can reduce the size of files by as much as 75% and allow you to have more data "to hand" without recourse to back-up tapes. There are two common programs for compressing and uncompressing files:

compress

```
% compress filename
```

will compress 'filename' to a (smaller) file called 'filename.Z'. To restore the original file do:

```
% uncompress filename.Z
```

(`% compress -d filename.Z` has the same effect)

gzip

The utility gzip generally achieves better compression ratios than compress:

```
% gzip filename
```

will compress 'filename' to a (much smaller) file called 'filename.gz'.
To restore this file do:

```
% gzip -d filename.gz
```
or
```
% gunzip filename.gz
```

Not surprisingly, this more efficient program takes longer to run than compress, but on most systems, unless the files are very big, it should not make a lot of difference. gzip can also compress a directory full of files, 'recursively' (compressing all the files in all directories below that point in the directory tree). Thus:

```
% gzip -r directoryname
```

will compress all the files in a directory called directoryname while

```
% gzip -rd directoryname
```

will decompress it.
Finally the option -v with both gzip and compress will printout the amount of compression achieved:

```
% gzip -v filename
% compress -v filename
```

All this should give you clues as to why files that are obtainable over the internet, by - say - anonymous ftp, often have a .gz or .Z suffix: they take up less room on the ftp site computer and also they will take a shorter time to transfer. When you have brought such files to your
own computer do:

```
% uncompress typical-ftp-file.Z
```
or
```
% gunzip typical-ftp-file.gz
```

tar

Often on ftp sites you will find files with .tar.Z or .tar.gz extensions. Such files have been 'archived' before compression; they typically contain a number of different files and/or directories which must be
'unpacked' before use. For example, having brought the file ftp-archive-tar.Z to your computer, you must first uncompress it:

```
% uncompress ftp-archive-tar.Z
```

then unpack it:

```
% tar -xsf ftp-archive-tar
```

This is likely to create a number of different directories and files.

You can create your own archives by:

```
% tar -cf myarchive.tar mydirectory
```

which will archive the contents of 'mydirectory' into the file 'myarchive.tar'. Thus tar can be useful in tidying up your files by reducing their number and can also be used as a prelude to compression.

Node.news

The Norwegian EMBnet node

The Norwegian node is continuously striving to improve the quality of its service. To improve the access to the ever growing sequence databases the main node CPU has been upgraded to a 200Mhz R4400 which represents a boost of approximately 50% over the original 100 Mhz R4000.

The WWW server http://www.no.embnet.org/ now hosts many new manuals. The html versions of GCG 8.0 Program Manual, Sections of EGCG, the Phylip phylogeny package, XtalView and WhatIf are all on-line to help users gain insight on what these programs are and how to use them. For Phylip and XtalView documents a special file handler was written. All manuals are accessible from the main page. Trembl - Translated EMBL - has been added to the main sequence libraries on the node. HSSP has been added to the Structural databases.

The Netherlands EMBnet node

SRS vs. 4.0 ASCII interface.

A quick e-mail poll on the desirability of an ASCII type userinterface for SRS4.0 in March 1995, resulted in an overwhelmingly positive response. It was expected that a potentially fast solution to
produce such an interface could be realised using the tcl/Tk toolkit. This has prompted the development and coding of an SRS ASCII client in tcl/Tk by Gijs Schaftenaar at the CAOS/CAMM Center, in close cooperation with Thure Etzold at EMBL.
A paper on this development will be published in the next issue of this Newsletter. Now we will limit ourselves to mentioning that, partially funded by EMBnet, the application of tcl/Tk in the able hands of Gijs, within 3 months has resulted in b-testing and implementation being underway at several EMBnet nodes. It is expected that the interface will be running soon at all EMBnet nodes with an SRS implementation.
A nice example of how accessibility of expertise in EMBnet can be used.

HGMP-RC node news (United Kingdom)

HGMP-RC Computing Services has a new addition to the team. Frank Visser has recently joined us from CAOS/CAMM in the Netherlands, where he was working with David Featherstone providing a service to the Genome project there.

Frank's primary job is to document and maintain the linkage packages and to give support and advice to people who wish to use them. He will also be helping to maintain our collection of genome databases.

We now have the following set of linkage programs:

Linkage: We currently use the 'fastlink' version of the original Lathrop and Lalouel Linkage package. These programs are for maximum likelihood estimation of recombination rates, calculation of lod score tables, and analysis of genetic risks.

Crimap: The main purpose of Crimap is to allow rapid, largely automated construction of multilocus linkage maps (and facilitate the attendant tasks of assessing support relative to alternative locus orders, generating LOD tables, and detecting errors). Although originally designed to handle codominant loci (eg. RFLPs) scored on pedigrees "without missing individuals", such as CEPH or nuclear families, it can now (with some caveats) be used on general pedigrees, and some disease
loci.

Map: This program is used to evaluate map distance, the likelihood of alternative locus orders and improve a trial map. The program analyses multiple pairwise lods under an interference parameter and, optionally, a given error frequency.

Mapmaker: Mapmaker is an interactive computer package for constructing genetic linkage maps, and for mapping genes underlying complex traits using those linkage maps.

Homoz: Homoz is a software package which allows very rapid multipoint mapping of disease genes in nuclear pedigrees, including homozygosity mapping. It includes a newly developed algorithm which makes the computation of multipoint LOD scores in small pedigrees extremely fast.

Pedpack: Pedpack is used for managing and analysing pedigrees and genetic traits, and for drawing pedigrees of arbitrary complexity.

APM: APM uses the Affected Pedigree Member (APM) method to analyze pedigrees for linkage between a diseased trait and a number of marker loci.

Simlink: a Linkage Simulation program, Simlink is used to estimate the probability, or power, of detecting linkage given family history information on a set of identified pedigrees.

Fastmap: Fastmap provides an approximate multipoint lod score for a disease against a number of markers from supplied two point lod scores. This is far faster than doing a full multipoint analysis.

COMDS: analysis of segregation, linkage and association under an oligogenic model.

Dolink and QDB: A linkage database and data output utility, it essentially consists of form-based facilities to edit data and facilities to generate reports.

Handlink: Handlink tabulates lodscores for any given set of recombination fractions and sums them over a number of families.

GAS: GAS (Genetic Analysis System)is designed to perform various types of fast accurate analyses of genetic data.

Plus a set of small utility programs distributed by Jurg Ott, such as Liped.

Frank is currently working on moving these programs onto the two IBM 590's we recently purchased to run linkage programs on.

BEN, the Belgian EMBnet Node

In June 1995, BEN has rounded the cape of 120 contractual users (to which must be added users working on someone elses account and the local users of the Departement of Molcecular Biology of the University of Brussels). BEN has stopped charging CPU usage.

Hardware/Operating systems

BEN has moved to a DEC Alpha 5000/3000 with operating system OSF/1 3.0. The old Sun Sparcstation remains available because of some not portable software.

Databases

BEN maintains a nonredundant nucleic acid databank consisting of EMBL + GenBank (GenBank entries excluded on base of common accession number) and a nonredundant protein databank consisting of Swiss Protein Bank + PIR (PIR entries excluded on base of cross-references in Swiss Protein Bank entries). Both are accessible through GCG, SRS, Gopher (WAIS-indexed) and BLAST. For Gopher and BLAST, searches are possible in the complete databank, only in the new sequences or in a subsection of EMBL/GenBank. The obsolete Vecbase has been replaced by Vector-ig from the NCBI repository.

SacchDB (Saccharomyces cerevisiae) and c21DB (human chromosome 21) have been added to BEN's collection of genomic databases.
A collection of codon usage tables in GCG format is available, including a translation of the CUTG database.
BEN is cooperating with the Consultory (Eksaarde) in making the Malaria Antigen Database (MAD) publically accessible (at http://ben.vub.ac.be/malaria/mad.html). MAD is presently being integrated under SRS.

Software

The software packages Staden, Pearson-FASTA and PHYLIP have been installed. The program Gibbs, that makes multiple gapless local sequence alignments, has been installed and rendered accessible through a GCG-like shell.

Network

BEN has a WWW server http://ben.vub.ac.be with an interface to SRS and the experimental WWW2GCG interface to the GCG software (developed by Dr. Marc Colet, accessible with password only). The WWW2GCG can be obtained, send E-mail to mcolet@dbmdec5.ulb.ac.be.

Teaching

BEN organizes each month a two day course about sequence analysis software and Internet tools, preceded by an optional introductory day for the computer illiterate. From March 1993 until today, the training sessions of BEN were attendend by a total of 312 people."

Special services

Current Sequence Similarity Awareness (CSSA): Users can submit sequences to BEN and be warned about the arrival in the databanks of similar new sequences. CSSA works with a BLAST against a databank generated on-the-fly with the dayly updates of EMBL or the weekly updates of GenPept.

INCBI, the Irish EMBnet node

INCBI has now been running under government funding for a year. The number of users appears to be reaching an asymptote at 350.

All databases and software are supported on a DEC AXP 3000 600 with 96MB memory and 12GB of storage. While we are running an SRS server (http://acer.gen.tcd.ie/srs/srsc), most of our users still use ACNUC to interrogate GenBank and PIR. This is available in an X-windows version.

Our WWW home page (http://acer.gen.tcd.ie/) is installed running NCSA httpd.

Homology searches are carried out locally using a menu driven command line assembler written by Laurent Duret (Laboratoire de Biometrie, Lyon, France), but also through GCG and remotely via MSU.

Network Entrez is available (but very slow after noon!)

Databases available include: GenBank, EMBL, PIR, Swissprot, Prosite, Owl, Hovergen.

Swiss EMBnet node (BioComputing Basel)

Development on the SRS package

Within an EU-funded project, the Swiss EMBnet node collaborates with Thure Etzold in Pat Argos' Group at EMBL and an Italian partner on the integration of SRS in electronic networks. Version 4_06 of SRS (by Thure Etzold) has been installed under CERN httpd 3.0. A tutorial for the installation of the CERN server and SRSWWW on UNIX platforms was posted to the newsgroups "bionet.software.srs" and "bionet.software.www".

Smith and Waterman Searching

Massively parallel searching goes into local production phase with the successful integration of the MasPar computer into the distributed advanced management system at BioComputing Basel. An interface was created to run the MPsrch software (Coulson et al.; Intelligenetics, Inc.) via the HASSLE protocol. Local user interfaces are both GCG and PC-based graphical user interfaces. Searchable databases include protein and DNA databases. Updates on the MasPar are on weekly or even daily basis, depending on the source.

New configuration of hardware introduced

Disk purchases of this year have been 24 GByte so far in order to cope with the extended volume of data. This enabled us to avoid network clashes caused by extensive use of NFS. Performance of the DEC/AXP 4000 compute server was enhanced by adding local disk space as well as switching to the OSF/1 operating system. The compute server is accessed via HASSLE and does not currently need to allow interactive work.

Announcements

EMBL WWW Nucleotide Sequence Submission

The EMBL Nucleotide Sequence Database at the European Bioinformatics Institute (EBI) announces a World Wide Web (WWW) submission tool which provides a fast and convenient way to submit nucleotide sequence data and associated biological information to the database.

The system is compatible with every known WWW browser (e.g. Mosaic, Netscape, Macweb, Lynx) on all the common platforms (i.e. X-windows, PC's and Macintosh).

You may access the WWW submission tool through the EBI WWW server at:

http://www.ebi.ac.uk/

under the heading "EMBL Nucleotide Sequence Database", or use the direct URL:

http://www.ebi.ac.uk/subs/emblsubs.html

When using the system for the first time, you are kindly requested to read the general help document provided through the first page of the submission tool.

After connecting to the EBI WWW server, a sequence of forms will enable you to prepare your data submission as efficiently and easily as possible. During any stage of submission, you may find a context sensitive help by selecting the help button provided.

EBI staff will process your submission within 24-48 hours of receipt and then by e-mail communicate to you the database accession number(s) assigned to your data.

Database entries created by the new WWW submission tool and submitted to the EMBL Nucleotide Sequence Database at the EBI will be exchanged and shared among the International Collaboration of Nucleotide Sequence Databases (DDBJ, EMBL, GenBank) and it is therefore only necessary to submit to one database.

For further information, please contact:

datalib@ebi.ac.uk
Benny Shomer
External Biological Liasion Officer,
EMBL outstation - The EBI,
Hinxton Hall, Hinxton,Cambridge CB10 1RQ, UK

Tel: +44-223-494437
Fax: +44-223-494468
Email: bshomer@EBI.ac.uk
http://www.ebi.ac.uk/ebi_docs/staff/benny.html

Bioinformatics: Computer Methods in molecular Biology (Practical Course)

Co-sponsored by EMBnet

12-21 July, 1995 ICGEB, Trieste, Italy

TOPICS:
Computer Communications, Networking, File Transfer
Electronic Mail, Bulletin Boards
Molecular Biology Databases
Sequence Homology Searches, Alignments
Sequence Patterns, Distant Protein Homologies
Genome Projects

ORGANIZER: SANDOR PONGOR
(pongor@icgeb.trieste.it)

Registrations to this course were closed on May 31.

FURTHER INFO: Ms. Diana Viti,ICGEB, Padriciano 99, Trieste 34012, ITALY

Telephone: +39-40-3757333
Fax: +39-40-226555
Telex: 460396 ICGEBT I
E-mail: viti@icgeb.trieste.it

International Conference on Molecular Structural Biology

Organised by the Austrian Chemical Society

Vienna (Austria), September 17-20, 1995

TOPICS:
1) The Impact of Molecular Biology on Structural Biology
2) Biomolecular Structure Determination
   a) X-Ray Diffraction
   b) NMR Spectroscopy
3) Dynamics and Function of Biomolecules
4) Computational Methods
5) Protein Engineering and Design

The scientific programme is designed to cover a broad range of disciplines in molecular structural biology. Topics will be presented in the form of plenary lectures, posters, discussions and exhibitions.

Sunday 17.9 Registration,Welcome Cocktail,Honorary Lecture by H.Neurath

Monday 18.9
1. The impact of Molecular Biology on Structural Biology
C.Cantor (USA) "Application of Streptavidin and Genetically Engineered Variants"
T.Blundell (GB) "Protein Super Families: Genome Analyses and Drug Discovery"

2. Biomolecular Structure Determination
- X-Ray Crystallography
R.Huber (D) "Proteolytic Enzymes and Their Natural Inhibitors - New Functions and New Structures and No End in Sight"
M.Walkinshaw (CH) "Structures and Biological Implications of Series of Cyclophilin-Cyclosporin Crystal Complexes"
I.Schlichting (D) "Ligand Binding in Heme Proteins Studied by Kinetic X-Ray Crystallography"

Tuesday 19.9
- NMR Spectroscopy
O.Jardetzky (USA) "Flexibility and Function in a DNA Binding Protein - the Trp-Repressor of E.Coli"
P.Roesch (D) "The Structure of Iron-Sulfur Proteins in Solution"
C.Dobson (GB) "The Structural Basis of Protein Folding"

3. Dynamics and Function of Biomolecules
J.Lakowicz (USA) "Fluorescence Spectroscopic Studies of the Structure and Dynamics of Biomolecules"
W.Kuehlbrandt (D) "Structure Determination of Membrane
Proteins by High-Resolution Electron Microscopy"
H.Frauenfelder (USA) "Dynamics and Function of Biomolecules"

Wednesday 20.9
4. Computational Methods
M.Karplus (USA) "Simulations of Protein Folding From the Native to the Denatured State and Back Again"
J.Skolnick (USA) "A Hierarchial Approach to the Prediction of Protein Structures"
H.-J.Boehm (D) "New Computational Tools for Structure Based Drug Design"

5. Protein Engineering and Design
A.Fersht (GB) "Pathways of Protein Folding"
L.Presta (USA) "Protein Engineering of Immunoglobulins and Immunoglobulin-Like Domains"
M.Mutter (CH) "De Novo Protein Design"

CALL FOR POSTERS

All those intending to participate in the conference are invited to submit posters presenting original work. Abstracts prepared according to the instructions given below should arrive NOT LATER THAN JUNE 30th,1995.

The authors will be informed about the provisional acceptance in July, 1995. The presentation of the posters will be finally approved and thus the contribution included in the Book of Abstracts when at least one of the registers before August 1st, 1995. Contributors of outstanding posters will be chosen by the scientific committee to give a 20 minute lecture.
Head of Scientific Committee: P.Schuster

EXHIBITON

An exhibition of instruments, accessories, software, literature and other items is planned. Companies interested in displaying their products are kindly requested to contact the conference secretariat.

INDUSTRIAL SATELLITE MEETING

A half day industrial symposium is planned. Those companies who are interested in presenting up-to-date results obtained on their products in the form of 20 minute lectures are kindly requested to contact the conference secretariat.

Registration Fees (before August 1):

Regular Participant	4000 ATS
GOeCH Member	3500 ATS
Student	2000 ATS

The registration fee includes the Book of Abstracts, the welcome cocktail and buffet, the invitation to the Viennese Rathaus, a ticket for the piano concert, coffee breaks and tram tickets.

If you are interested in a folder of the second announcement including a registration form and the social program, please contact:

A.Kungl
Gesellschaft Oesterreichischer Chemiker
AG Biophysikalische Chemie
Nibelungengasse 11
A-1010 Wien, Austria
Tel.: (43) 1 587249
FAX.: (43) 1 587966
e-mail: msb95@helix.mdy.univie.ac.at

Basel Forum on Biosafety

October 17, 1995
Safety of Transgenic Crops. Environmental and Agricultural Considerations

ZLF (Zentrum fuer Lehre und Forschung),
Kleiner Hoersaal, Hebelstrasse 20, CH-4055 Basel, Switzerland

SCOPE

To expand public access and awareness on the progress in the safety assessment of genetically modified crops, the Basel Forum on Biosafety 1995 will survey current efforts in safety research and regulation of transgenic crop varieties.

Environmental issues such as gene transfer invasiveness, and target and non-target effects, will be discussed on the basis of actual case studies.

The audience will be offered an overview on the approaches aimed at identifying and managing adverse impacts of genetically modified plants on the ecosystem. The current state of knowledge and the trends in safety research will be reviewed.

Organized and supported by: Agency for Biosafety Research and Assessment of Technology Impacts of the Swiss Priority Programme Biotechnology (BATS), Basel Swiss National Science Foundation, Berne, Federal Office of Environment, Forests and Landsape, Berne, Switzerland.

PROGRAM

Safety of Transgenic Crops. Environmental and Agricultural Considerations.

09.30 - 09.45 Welcome, O. Kaeppeli, BATS, Basel, CH
09.45 -10.30 From small-scale to large-scale release of transgenic crops. Are there unsolved issues regarding the impacts on the ecosystem? P. Dale, John Innes Center, Norwich, UK
10.30 - 11.00 Break
11.00 - 11.45 Herbicide tolerant crops. Impacts on agriculture and environment. E.Rasche, ArgEvo GmbH, Frankfurt, D
11.45 - 12.30 Virus resistant crops. New viral pathogens from transgenic plants? P. Malnoe, Federal Research Station of Agronomy, Changins, CH
12.30 - 12.40 Discussion
12.40 - 14.00 Lunch
14.00 -14.45 Horizontal gene transfer from transgenic plants into plant associated bacteria and soil bacteria. K. Smalla, Biologische Bundesanstalt Braunschweig, D
14.45 - 15.15 Coffee
15.15 - 16.00 Insect tolerant maize. Interactions with the ecosystem. E. Jarchow, Ciba Ltd., Basel, CH
16.00 - 17.00 Final Discussion,F. Pythoud, Federal Office of Environment, Forests and Landsape,Berne, CH

BATS

http://www.ch.embnet.org/BATS/
Dr. Elisabeth Schulte
Clarastrasse 13
CH-4058 Basel, Switzerland
phone ++41 61 690 93 10
fax ++41 61 690 93 15 +33 (76) 88-9594
+33 (76) 88-5494 (fax)

Pacific Symposium on Biocomputing Kohala Coast, Island of Hawaii - January 3-6, 1996

Call For Papers, Abstracts and Demonstrations
http://cgl.ucsf.edu/psb/psb.html

The first Pacific Symposium on Biocompting (PSB), will be held January 3-6, 1996 in Hawaii. PSB will bring together top researchers from the North America, the Asian Pacific nations, Europe and around the world to exchange research results and address open issues in all aspects of computational biology. Replacing and extending the last three years of Biotechnology Computing Tracks at the Hawaiian International Conference on System Sciences, PSB will provide a forum for the presentation of work in databases, algorithms, interfaces, visualization, modelling and other computational methods, as applied to biological problems, with emphasis on applications in data-rich areas of molecular biology. In addition, PSB intends to attract a more balanced combination of computer scientists and

biologists by reducing some of the barriers to the attendance of biologists reported by participants of more specialized events.

To provide focus for the very broad area of biological computing, PSB is organized into a series of tracks, minitracks and workshops. The tracks identify the focus research topics for the meeting; minitracks are opportunities for presentation of work in emerging research areas; and workshops provide forums for discussions of significant infrastructure and other non-research issues. In addition, PSB offers two invited keynote talks, an general paper track, poster and demonstration sessions, and plenty of opportunity for informal public discussion.

All paper submissions will be rigorously peer-reviewed, and accepted papers will be published in an archival proceedings volume. Paper publication is required for oral presentation. Researchers wishing to present their research without official publication are encouraged to submit an abstract for the poster or demonstration sessions.

Important dates:

Expression of interest to session chair:	immediate
Paper submissions due:	July 27, 1995
Notification of paper acceptance:	Sept. 11, 1995
Final paper deadline	October 2, 1995
Early registration deadline	October 2, 1995

Each track, workshop and minitrack has a chair who is responsible for organizing that session. Please contact the specific session chair relevant to your interests for further information.

Tracks:

* The Evolution of Biomolecular Structures and the Structure of Biomolecular Evolution.
chairs: Richard A. Goldstein & Russ B. Altman
contact: richardg@Chem.LSA.UMich.Edu
+1 (313) 763-8013 (tel) +1 (313) 747-4865 (fax)

* Interactive Molecular Visualization
chairs: Michael Teschner & Chris Henn
contact: micha@basel.sgi.com
+41 (61) 641-0903 (tel) +41 (61) 641-1201 (fax)

* Stochastic Models, Formal Systems and Algorithmic Discovery for Genome Informatics
chairs: Tom Head, Takashi Yokomori, Katsumi Nitta & Kiyoshi Asai
contact: tom@math.binghamton.edu
+1 (607) 777-2278 or leave message at +1 (607) 777-2147
+1 (607) 777-2450 (fax)

* Discovering, Learning, Analyzing and Predicting Protein Structure
chairs: Richard H. Lathrop & A. Keith Dunker
contact: dunker@jaguar.csc.wsu.edu
+1 (509) 335-2430 (tel) +1 (509) 335-9688 (fax)

Minitracks:

* Population Modelling
chairs: John Conery and Ross Kiester
contact: conery@cs.uoregon.edu
+1 (503) 346-3973 (tel) +1 (503) 346-5373 (fax)

* Hybrid Quantum and Classical Mechanical Methods for Studying Biopolymers in Solution
chairs: Martin Field & Jiali Gao
contact: mjfield@ibs.fr
+33 (76) 88-9594 (tel) +33 (76) 88-5494 (fax)

* Models of Control Systems in Biology
chair: Seth Michelson
contact: Seth.Michelson@syntex.com
+1 (415) 354-7142 (tel) +1 (415) 354-7554 (fax)

* Computational Studies on the Design of Protease Inhibitors
chair: C.N. Hodge
contact: hodgecn@lldmpc.dnet.dupont.com
+1 (302) 695-3698 (tel) +1 (302) 695-2813 (fax)

PSB Workshops provide opportunities for discussion about issues that are significant to the computational biology community. They will feature presentations of current research as well as forums for other kinds of communication. Please contact the chairs for more information about presentation requirements.

* Internet Tools for Computational Biology
chair: Reinhard Doelz
contact: doelz@ubaclu.unibas.ch
+41 (61) 267-2078 (tel) +41 (61) 267-2247 (fax)
http://www.ch.embnet.org/psb/

* Biocomputing Education: Challenges and Opportunities
chairs: Susan J. Johns, Steven M. Thompson, & A. Keith Dunker
contact: thompson@jaguar.csc.wsu.edu
+1 (509) 335-0533 or -3179 (tel) +1 (509) 335-9688 (fax)

For questions, or for information about submissions of papers, posters or demonstrations that do not fit into any of the above sessions, contact the conference co-chairs:

Dr. Teri Klein
Computer Graphics Laboratory
University of California, San Francisco
San Francisco, CA 94143-0446 USA
phone: +1 (415) 476-0663, fax: +1 (415) 502-1755
email: klein@cgl.ucsf.edu

or
Dr. Lawrence Hunter
Lister Hill Center, National Library of Medicine
Bldg. 38A, MS-54, 8600 Rockville Pike
Bethesda, MD 20894 USA
phone: +1 (301) 496-9300, fax: +1 (301) 496-0673
email: hunter@nlm.nih.gov

Participation fee: $375. Hotel accommodation: $130 +tax. For registration and site information, contact the PSB conference coordinator:
Sharon Surles
PSB Conference Coordinator
IS, Inc., Suite 203, 5330 Carroll Canyon Rd
San Diego, CA 92121 USA
phone: +1 (619) 658-9782, fax: +1 (619) 658-9463
psb@intsim.com

A biocomputing thesaurus

API	Application Programmer Interface
CCP11	Computer Collaborative Project
EBI	European BioInformatics Institute (EMBL Outstation)
EGCG	Extensions to the GCG package (See article by Rice et al.)
eMail	Electronic Mail
EMBL	European Molecular Biology Laboratory
EMBnet	European Molecular Biology network
EPFL	Ecole Politechnique Federale de Lausanne, Switzerland
ETH	Eidgenoessische Technische Hochschule Zuerich, Switzerland
EU	European Union
Entrez	Database retrieval program created by the NCBI
FAQ	Frequently Asked Questions
FASTA	Sequence database searching program created by W.Pearson
GByte	GigaByte , 1 000 000 000 bytes
GCG	Genetics Computer Group Inc., commercial vendor of the GCG package
GNA	Globewide Network Academy, and initiative for teleteaching
GUI	Graphical User Interface
HASSLE	Network protocol created by the Swiss node of EMBnet
HGMP	Human Genome Mapping Project, based at Cambridge/United Kingdom
HSSP	Database of homology-derived structures of proteins (Sander et al.)
ICGEB	International Center for Genetic Engineering and Biotechnology, Trieste
ICLUI	IBM Class Library User Interface
IG	Intelligenetics Inc.; commercial vendor of the Intelligenetics Suite
INCBI	Irish National Centre for BioInformatics
IP	Internet Protocol, connecting 3-5 millions of computers
MasPar	Name of a hardware vendor for computers of the same name
MIME	Multipurpose Internet Mail Extensions, conventions for sending eMail
MPsrch	Sequence searching application on the MasPar computer (Coulson et al.)
MSU	Mail Submission Utility, allows to send data by eMail conveniently
NCBI	National Center for Biotechnology Information, Washington, US
NCSA	National Center for Supercomputing Applications, US
NFS	Network File System, a program system to share files on local networks
OS	Operating System, the program which keeps computers alive
OSF/1	Tradename for Digital Equipment's flavour of UNIX operating systems
PC	'Personal Computer', can be anything from low-end to very powerful
PHYLIP	Program Package created by J.Felsenstein for Phylogenicity
PIR	Protein Identifiaction Resource International, a protein database vendor
PMDF	Program package to deal with electronic mail
PSB	Abbreviation for 'Pacific Symposium on Biocomputing'
RASMOL	Program package by R.Sayle to display protein structures
RFC	Request for Comments, documents describing Internet Protocol conventions
SRS	Program for biological database browsing created by T.Etzold(EMBL)
SRSWWW	Version of the Sequence Retrieval System running in the WWW environment
tcl/Tk	Programing environment for the creation of graphical user interfaces
Trembl	'Translated EMBL', a SRS-based compilation of the EMBL DNA data library
URL	Universal resource Locator, the unique name of a document in the WWW
USENET	world-wide messaging system to transport electronic bulletin boards
VMS	Digital Equipment's proprietary operating system
VSNS	Virtual School of Natural Sciences, a teleteaching initiative
WAIS	full-text database searching system, mostly used in WWW environments
WPI	Wisconsin Package Interface, graphical user interface to the GCG package
WWW	World-Wide Web, an information system to browse the international networks
WWW2GCG	The Belgian EMBnet Node's system to use the GCG package via WWW

The EMBnet Nodes

[AT] EMBNet (martin.grabner@cc.univie.ac.at)
VIENNA BIOCENTRE
University of Vienna, Vienna, Austria

[BE] BEN (rherzog@ulb.ac.be)
Brussels Free Universities,
Rhode-St-Genese, Belgium

[CH] Biocomputing Basel (info@ch.embnet.org)
Biozentrum der Universitaet,
Basel, Switzerland

[CH] ROCHE (doran@embl-heidelberg.de)
Hoffmann-La Roche,
Basel, Switzerland

[CH] SWISSPROT (bairoch@cmu.unige.ch)
Med. Biochem. Dept. CMU, University of Geneva
Geneva, Switzerland

[DE] EMBL (datalib@EMBL-Heidelberg.de)
European MolecularBiologyLaboratory,
Heidelberg, Germany

[DE] GENIUS
(dok419@genius.embnet.dkfz-heidelberg.de)
DKFZ,
Heidelberg, Germany

[DE] MIPS (mewes@mips.embnet.org)
Max-Plank-Institut fur Biochemie,
Martinsried, Germany

[DK] BIOBASE (hum@biobase.aau.dk)
BioBase,
Aarhus, Denmark

[ES] CNB (carazo@samba.cnb.uam.es)
Centro national de Biotechnologia
CSIC, Madrid, Spain

[FI] CSC (Heikki.Lehvaslaiho@csc.fi)
Centre for Scientific Computing,
Espoo, Finland

[FR] CEPH (claude@genethon.fr)
GENETHON,
Evry, France

[FR] INFOBIOGEN (dessen@infobiogen.fr)
INSERM,
Villejuif, France

[GR] EMBnet Node (savakis@myia.imbb.forth.gr)
Institute of Molecular Biology and Biotechnology,
Heraklion, Greece

[HU] EMBnet (remenyi@abc.hu)
Agricultural Biotechnology Centre,
Godollo,Hungary

[IL] INN (isestern@weizmann.weizmann.ac.il)
Weizmann Institute of Science,
Rehovoth, Israel

[IT] CNR (marcella@area.ba.cnr.it)
Consiglio Nationale delle Ricerche,
Bari, Italy

[IT] ICGEB (pongor@genes.icgeb.trieste.it)
International Centre for Genetic Engineering,
Trieste, Italy

[IE] INCBI (atlloyd@acer.gen.tcd.ie)
Irish National Centre for BioInformatics,
Dublin, Ireland

[NL] CAOS (jackl@caos.caos.kun.nl)
Katholieke Universiteit,
Nijmegen, Netherlands

[NO] BiO (rodrigol@biotek.uio.no)
Biotechnology Centre of Oslo,
Oslo, Norway

[PL] IBB (piotr@ibbrain.ibb.waw.pl)
Institute of Biochemistry and Biophysics,
Polish Academy of Sciences, Warsaw, Poland

[PR] EMBnet (pfern@gulbenkian.pt)
Instituto Gulbenkian de Ciencia,
Oeiras, Portugal

[SE] EMBnet.se (gad@perrier.embnet.se)
Computing Department, Biomedical Centre,
Uppsala, Sweden

[UK] HGMP (mbishop@hgmp.mrc.ac.uk)
Human Genome Mapping Project Resource Centre,
Hinxton, Cambridge, United Kingdom

[UK] SEQNET (bleasby@daresbury.ac.uk)
DRAL Daresbury Laboratory,
Daresbury, United Kingdom

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 in the Tips from the computer room section, please let us know. Submissions for the BITS section are most welcome, but please remember that we cannot extend space beyond two pages per article. Please send your contributions to one of the editors. You may also submit material by Internet E-mail to:

emb-pub@dl.ac.uk

If you had difficulty getting hold of this newsletter, please let us know. We would be only too happy to add your name to our mailing list. This newsletter is also available on-line using any WWW client via the following URLs:

The Online version(ISSN 1023-4152) :

- ***http://www.dl.ac.uk/embnet.news/vol2_2/contents.html***
- ***http://ben.vub.ac.be/embnet.news/vol2_2/contents.html***
- ***http://www.ch.embnet.org/embnet.news/vol2_2/contents.html***
- ***http://biomaster.uio.no/embnet.news/vol2_2/contents.html***

A Postscript version (ISSN 1023-4144) is also available. You can get it by anonymous ftp from:

- ***s-ind2.dl.ac.uk in the directory pub/embnet.news/***
- ***dbmdec5.ulb.ac.be in the directory pub/embnet.news/***
- ***ftp.no.embnet.org in the directory pub/embnet.news/***

Back issues.

Online:

- ***http://www.ch.embnet.org/embnet.news/info.html***

Postscript by ftp:

- ***s-ind2.dl.ac.uk in the directory pub/embnet.news/***
- ***dbmdec5.ulb.ac.be in the directory pub/embnet.news/***
- ***ftp.no.embnet.org in the directory pub/embnet.news/***

Publisher:

EMBnet Administration Office.
c/o J.Franklin,
ASFRA BV,
Voorhaven 33,
1135 BLEDAM.
The Netherlands

Editorial Board:

Alan Bleasby, DRAL Daresbury, UK
(bleasby@daresbury.ac.uk)
FAX +44-1-925-603100
Tel +44-1-925-603351

Reinhard Doelz, Basel University, CH
(doelz@comp.bioz.unibas.ch)
FAX +41-61-2672078
Tel +41-61-2672247

Robert Herzog, BEN, Free University Bruxelles, BE
(rherzog@ulb.ac.be)
FAX +32-2-6509767
Tel +32-2-6509762

Andrew Lloyd, INCBI, Trinity College Dublin, IE
(atlloyd@acer.gen.tcd.ie)
FAX +353-1-679-8558
Tel +353-1-608-1969

Rodrigo Lopez, BiO, University of Oslo, NO
(rodrigol@biotek.uio.no)
FAX +47-2-2694130
Tel +47-2-2958756

embnet.news
Vol.2, No.2, 1995
21st June 1995
ISSN 1023-4144