



## BioWAP, mobile Internet service for bioinformatics

Pentti Riikonen<sup>1, 2</sup>, Jorma Boberg<sup>2, 3</sup>, Tapio Salakoski<sup>2, 3</sup> and Mauno Vihinen<sup>1, 4,\*</sup>

<sup>1</sup>Institute of Medical Technology, FIN-33014 University of Tampere, Finland,

<sup>2</sup>Department of Computer Science, University of Turku, FIN-20520 Turku, Finland,

<sup>3</sup>Turku Centre for Computer Science TUCS, FIN-20520 Turku, Finland and <sup>4</sup>Tampere University Hospital, FIN-33520 Tampere, Finland

Received on February 27, 2001; revised on May 21, 2001; accepted on May 30, 2001

### ABSTRACT

**Summary:** We have developed a new Internet service, which provides mobile access to bioinformatics databases and software tools. The BioWAP service facilitates access to basic bioinformatics databases and analysis tools from everywhere without a PC or a laptop computer. Both open source bioinformatics program suites and Internet services, which are not designed for mobile Internet access, were utilized in the BioWAP service.

**Availability:** The BioWAP service starting page can be browsed with any WAP terminal from <http://bioinf.uta.fi/wml/welcome.wml>.

**Contact:** Mauno.Vihinen@uta.fi

Experimental approaches generate a vast amount of biological information, which has to be organized, stored and analyzed. The majority of the common bioinformatics analyses can be performed on the Internet. The Internet access is not always available outside of the research institutes. Here, we describe to our knowledge the first system that allows access to all the central biological databases as well as to the basic analysis tools from anywhere by using a mobile communication terminal such as a cellular phone.

The integrated BioWAP package provides scientists an easy access to biological knowledge on the Internet. Through the BioWAP service it is possible to search information from all the major nucleotide and protein sequence databases, study structural information and address data on immunodeficiency mutation databases. The available databases are listed in Table 1. The BioWAP service provides some 20 different analysis tools.

Currently, the Wireless Application Protocol (WAP) is a global standard for the presentation and wireless delivery of information to mobile terminals.

Due to the limitations of the portable devices, WAP services cannot use e.g. HTML-based pages. WAP services

**Table 1.** Databases accessible with the BioWAP

Sequence databases	EMBL, SWISS-PROT, TrEMBL
Structure	PDB
Immunodeficiency mutation databases	BLMbase, BTKbase, CYBAbase, CYBBbase, CD3Ebase, CD3Gbase, CD40Lbase, JAK3base, NCF1base, NFC2base, RAG1base, RAG2base, SH2D1Abase, ZAP70base
Enzyme	Enzyme commission nomenclature
Protein domain	PROSITE

have to be developed separately for the mobile platform. BioWAP is a part of the mBIOINFO service suite that provides access to all the major bioinformatics data. The user interfaces of different WAP terminals vary a lot. The BioWAP service can be utilized also with the most common WAP phones with small displays.

Analysis and search services of BioWAP are divided to seven main categories as follows: (1) amino acids, codons; (2) DNA/RNA sequences; (3) protein sequences; (4) PDB structures; (5) Enzyme Commission nomenclature; (6) mutation databases; and (7) SRS search.

Amino acids can be queried by one letter or three letter abbreviations or by full names to obtain abbreviations. Because it is difficult to remember the genetic code, the service is designed so that either codons or amino acids can be used to query from translation information. Of the translation tables the most common tables are included for normal, alternative yeast, bacterial, vertebrate mitochondrial, and yeast mitochondrial translation.

The BioWAP service facilitates searching and analyzing DNA/RNA or protein sequences. Sequence identification or accession number is required to retrieve sequences from local databases. All local database searches are based on *bioperl* sequence retrieving functions. When the user has

\*To whom correspondence should be addressed.

typed the identification of the sequence, he can then select one or more of the following analysis services: general properties, references, single cutting restriction enzymes, enzyme recognition sites, patterns, and alignment for nucleotide sequences and general properties, patterns and alignment for amino acid sequences. Some of these analysis services require additional input from the user.

Here, the following EMBOSS routines were utilized (<http://www.uk.embnet.org/Software/EMBOSS/>). The program *restrict* is used in single cutting restriction enzyme analysis, program *fuzznuc* for pattern analysis and program *needle* in alignment calculations. The *needle* uses comparison table EDNAMAT in the nucleotide sequence alignment and EBLOSUM62 in protein alignment calculations. Program *pepstats* calculates protein General properties and program *fuzzpro* is used in Prosite (Hofmann *et al.*, 1999) pattern analysis.

BioWAP can extract essential data from PDB files including references, resolution and the locations of secondary structural elements if given in the annotation of the entry. PDB search services included are general properties, references, and secondary structures.

Enzyme Commission nomenclature numbers and descriptions are hard to remember. Therefore, EC codes can be searched either with keywords or EC numbers.

Immunodeficiency-causing mutation data can be searched from the IDbases (Väliäho *et al.*, 2000; Vihinen *et al.*, 2001) (Table 1). In addition to the actual mutation description on three levels (genome, RNA, and protein) the databases also contain patient-specific clinical information. We have implemented two types of keyword queries in mutation search, namely Patient Identification Number (PIN) (Vihinen *et al.*, 1999) and location-based search. The search results include the number of families having mutations matching to search keyword and references to these mutations.

We have developed an easy to use mobile interface to SRS (Etzold and Argos, 1993) making it possible to search for sequences from all sequence databases available at SRS. The system gives a list of entries found and their accession numbers, which can be used for further analyses. The result of an SRS query is temporarily stored in the cache memory of the system. Further analysis can be done with methods discussed above. The primary connection site in our system is the European Bioinformatics Institute service in UK (<http://srs6.ebi.ac.uk>). If no connection can be established, two other sites are tried at the Center for Scientific Computing in Finland (<http://srs.csc.fi:8002>) and at the BC Genome Sequence Center in Canada (<http://srs6.bcgsc.bc.ca>).

A typical, complete WAP service has at least three elements: the WAP terminal, the WAP gateway, and the www server. The WAP terminal is usually a mobile phone or a handheld computer equipped with WAP browser software.

The WAP gateway is a link between the Internet and a mobile phone network. The services that the user wants to access are fetched and compiled by the WAP gateway and consequently sent to the WAP terminal. WAP services are client-server applications similar to www services with the exception of the WAP gateway. The www server is the core of BioWAP services, because it handles users requests, starts appropriate analysis software, and transmits results via the WAP gateway.

Any WAP gateway that allows access to our Internet site (<http://www.bioinf.uta.fi/wml/welcome.wml>) can be used with the BioWAP system. Although the service is run in Finland, the need for making expensive long-distance calls can be avoided by accessing the service via local gateways. We maintain two free WAP gateways for those clients who cannot use local gateways. The list of gateways in each country can be found from <http://www.wapdrive.com>. Instructions on how to configure the WAP terminal to use one of our gateways are available from <http://bioinf.uta.fi/biowap/gateway.html>. The BioWAP service has been successfully tested with several devices manufactured by Ericsson, Nokia, Palm and Siemens.

The present mobile terminals are constrained in regards to computing power, the size of memory and display. However, only the programs requiring large input or output cannot be performed by the BioWAP. New developments in mobile communication, especially the increased speed of GPRS-GSM and 3G systems (UMTS), will add to the attraction of mobile Internet and mobile bioinformatics. WAP is going to remain as an application platform in the future. We will continue to add new features to the BioWAP.

## REFERENCES

- Etzold,T. and Argos,P. (1993) SRS—an indexing and retrieval tool for flat file data libraries. *Comput. Appl. Biosci.*, **9**, 49–57.
- Hofmann,K., Bucher,P., Falquet,L. and Bairoch,A. (1999) The PROSITE database, its status in 1999. *Nucleic Acids Res.*, **27**, 215–219.
- Vihinen,M., Arredondo-Vega,F.X., Casanova,J.L., Etzioni,A., Giliani,S., Hammarström,L., Hershfield,M.S., Heyworth,P.G., Hsu,A.P., Lähdesmäki,A., Lappalainen,I., Notarangelo,L.D., Puck,J.M., Reith,W., Roos,D., Schumacher,R.F., Schwarz,K., Vezzoni,P., Villa,A., Väliäho,J. and Smith,C.I. (2001) Primary immunodeficiency mutation databases. *Adv. Genet.*, **43**, 103–188.
- Vihinen,M., Lehvälaiho,H. and Cotton,R.D. (1999) Immunodeficiency mutation databases. In Ochs,H.D., Smith,C.I.E. and Puck,M. (eds), *Primary Immunodeficiency Diseases. A Molecular and Genetic Approach*. Oxford University Press, New York, pp. 443–447.
- Väliäho,J., Riikonen,P. and Vihinen,M. (2000) Novel immunodeficiency data servers. *Immunol. Rev.*, **178**, 177–185.