

1000 Databases for the Bioinformatician

by W. Touw, E. Krieger, G. Vriend
CMBI, Radboudumc, Nijmegen (NL)

Very (too) many projects in bioinformatics are directed towards protein secondary structure prediction, and we have provided -for longer than the existence of the Internet - free access to the [DSSP](#) software and databases needed for these projects.

The [CMBI](#) protein structures facilities further include [HSSP](#), [WHAT IF](#) and its derived [WHAT CHECK](#) and [PDBREPORT](#), the [PDBFINDER](#), several improved [PDB](#) variants ([PDB_REDO](#), [BDB](#)), and a few more. Some of these facilities are also available in the [eBioKit](#) (Figure 1), either through [MRS](#) or via [YASARA](#). [WHAT IF](#) can now calculate a wide variety of features for the whole (useful subset of the) [PDB](#).

Together with the lists of sequence unique chains in the [PDB](#) ([PDB_SELECT](#)), these data could potentially spur a flurry of prediction software activities. Currently, 20 datasets are available at swift.cmbi.ru.nl/gv/lists/, but the potential for new datasets is unlimited.

These first 20 datasets fall in five main categories: 1) elementary geometric aspects, such as bond, torsion angles and surface areas; 2) amino acid contact prediction- this got a big boost recently, but research has focused on the reduction of false-positive prediction, rather than the equally important definition of what is a contact.

The second group of sets therefore deals in many ways with amino acid contacts in proteins; 3) symmetry contacts, contacts with ions, and salt-bridges; 4) sets of augmented PDB files in which, for example, symmetry calculations have been worked out; 5) 'other' datasets.

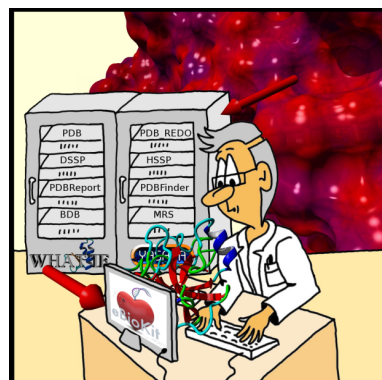


Figure 1. Andreas working with the [eBioKit](#).

Users can download individual files or entire datasets. New datasets will be made on request, provided that [WHAT IF](#) can produce the requested data with existing options. We encourage people to ask for novel datasets, because this can only stimulate the important field of protein structure bioinformatics.

A series of PDB-related databanks for everyday needs. Touw WG, Baakman C, Black J, Te Beek TA, Krieger E, Joosten RP, Vriend G. Nucleic Acids Res. 2015 Jan 28;43 (Database issue):D364-368. doi: 10.1093/nar/gku1028. PMID: 25352545



Broadening the bioinformatics infrastructure to unicellular, animal and plant science

by E. Bongcam-Rudloff and A. Gisel

AllBio was the product of responding to a former KBBE call entitled 'Supporting the development of Bioinformatics Infrastructures for the effective exploitation of genomic data: Beyond health applications'. A group of 'old' EMBRACE members - most of them members of [EMBnet](#) - accepted the challenge, and formulated a proposal, applying the experience that this group had acquired and developed in past projects.

[The project – AllBio –](#) was awarded to a consortium of [10 partners](#) from 8 countries, giving them the opportunity to increase bioinformatics awareness and to spot the bioinformatics needs and bottlenecks in non-human-health fields, such as animal, microbiology and plant science. The project was based on so-called '[test-cases](#)', in which the aforementioned biological communities formulated data-analysis problems they faced but for which they did not know the right approaches or have the right tools to solve them.

AllBio collected more than 60 test-cases from across Europe, and selected 15 that represented [generally encountered bioinformatics problems](#). The AllBio partners, together with specifically selected bioinformatics specialists, organised several events between bioinformaticians and biologists to discuss and try to solve these problems. The most successful events were problem-specific hack-a-thons based on a series of community-building



workshops, where several teams physically sat together for several days and, on the spot, produced successful software solutions. Some of these new software packages have been published and made available for other end-users.

As a Coordination Support Action, AllBio was very successful in a) demonstrating how, in a very cost-efficient but productive way, to solve such data analysis problems, bringing together specialists from diverse disciplines, and b) delivering some outstanding solutions by applying these strategies. For more information visit www.allbioinformatics.eu



EMBnet.digest

EMBnet.Spotlight is a quarterly release of InFocus sections published in EMBnet.digest (www.embnet.org/embnet-digest), EMBnet's monthly publication that provides a round-up of news from the community. The InFocus section features member activities, projects, initiatives, etc., especially from new members, that may be of interest both to the network and to EMBnet's associated communities, societies and projects.