Web alert

Eukaryotic gene regulation

Chemistry & Biology 2000, **6**:R26

Transcription is the first step in the pathway that yields functional protein or RNA molecules as specified by the information encoded in cellular DNA. This month's Web alert focuses on web resources related to the transcription of protein-encoding genes in eukarya. Resources in this area collate sequence, structure, biochemical and mutagenesis information on transcription factor proteins and DNA sites and regions involved in transcription.

The protein-encoding genes in eukaryotic cells are transcribed by RNA polymerase II in concert with an array of basal transcription factors. The assembly of these basal factors into the complex responsible for transcription initiation is nucleated by the binding of the transcription factor TFIID to a promoter, a region of DNA that contains all the information as to when a gene should be transcribed. Steve Buratowski's Basal Transcription Factor Information web site (at Harvard Medical School) provides a good introduction. If you are interested in the DNA sequences of promoters for particular genes then check out the Eukaryotic Promoter Database (EPD). Here you will find a collection of sequences of promoters for particular genes and information on their experimentally determined transcription start sites.

Regulatory transcription factors are anchored (directly or indirectly) to specific DNA sequences distal to the core promoter, often within regions of DNA known as enhancers or silencers, which up or down regulate transcription respectively. For information on these extended regulatory regions of eukaryotic genes, as well as transcription factor binding sites and expression patterns for specific genes, look to the Transcriptional Regulatory Region Database (TRRD).

The TRANSFAC database is an excellent starting point for information on regulatory transcription factors and the DNA sites they bind. In the database you can search the Factors table for a transcription factor, and TRANSFAC will provide a wealth of information including amino acid sequence, a domain map, and a synopsis of other data such as the effects of mutations on DNA binding.

Several resources offer collated information about particular classes of transcription factors. The Homeodomain Resource at the NIH contains sequence, structure and cytogenetic information on homeodomain transcription factors. Mutagenesis data are included for homeodomain proteins known to be involved in human genetic diseases and disorders.

The Nuclear Receptor Resource at Georgetown University offers much information on hormone receptors, from mutational data to where to find antibodies to a particular receptor. Also of interest is the Histone Sequence Database, which includes information on the post-translational modifications of histones, such as acetylation, known to be important in transcriptional regulation.

Experimentally determined structures of protein–DNA complexes involved in transcription can be found at the Rutgers nucleic acid database, and the coordinates in .pdb format can be downloaded for viewing in your favorite viewer. Two popular programs for viewing pdb files are RasMol, and Swiss-Pdb Viewer, both of which can be downloaded free.

The growth in information about transcription on the web is, of course, driven by new experimental data. One technological advance that has spawned a wealth of new data in this field is gene array technology, which has made it possible to determine the effects of a single molecular species on the transcription of every gene in a particular genome simultaneously. In this vein, the Genome-Wide Expression Page, which is part of Richard Young's web site at MIT, documents the effects of removing single yeast transcription factors from cells on the transcription of other yeast genes.

Because of its flexible, linkable nature, the web is an excellent place to find information on the complex networks that regulate life. Indeed, resources related to transcription are now not only linked to each other but also to resources on interconnected processes such as signal transduction. It is easy to imagine a time when the simple databases of individual processes in a cell, such as translation, replication, recombination and transcription, will be fully integrated and will provide a cyber map that mirrors the cellular blueprint.

Jason W Chin, Department of Chemistry, Yale University, New Haven, CT 06511;
jason.chin@yale.edu

Web sites and URLs

Basal Transcription Factor Information	http://tfiib.med.harvard.edu/transcription/basaltx.html
Eukaryotic Promoter Database	http://www.epd.isb-sib.ch/
Transcriptional Regulatory Regions Database	http://www.bionet.nsc.ru/trrd/
TRANSFAC	http://transfac.gbf.de/
Homeodomain Resource	http://genome.nhgri.nih.gov/homeodomain/
The Nuclear Receptor Resource	http://nrr.georgetown.edu/nrr/nrr.html
Histone Sequence Database	http://genome.nhgri.nih.gov/histones/
Nucleic Acid Database	http://ndbserver.rutgers.edu/NDB/
RasMol	http://www.umass.edu/microbio/rasmol/
Swiss-Pdb Viewer	http://www.expasy.ch/spdbv/
Genome-Wide Expression Page	http://web.wi.mit.edu/young/expression/