

From Genotype to Phenotype: Navigating Nature's Information Landscape at NCBI

David Lipman `lipman@void.nlm.nih.gov`

National Center for Biotechnology Information (NCBI),
National Library of Medicine, National Institutes of Health,
Bethesda, MD 20894, USA.

Abstract

For several years the National Center for Biotechnology Information (NCBI) has provided integrated access to linked DNA, protein, and bibliographic data through the online information retrieval system, Entrez. This information space has now been expanded to include genomic data (physical maps, genetic maps, and sequence alignments) and three dimensional structure data. The "Genomes" division presents genome level views of a large number of complete chromosomes, from organelle, through virus and phage, to completely sequenced chromosomes from yeast or bacteria, to integrated genetic and physical maps and contig'ed sequence islands from eukaryotes such as human and drosophila. Following the Entrez tradition, the chromosome views are tightly linked to DNA and protein sequence records, MEDLINE citations, and the new three dimensional structure division.

The structure information in Entrez is from a new NCBI database called MMDB (Molecular Modeling Database), derived from the Brookhaven Protein DataBank 3-dimensional structures (currently over 3,000 biomolecules). MMDB is a database of ASN.1-formatted records, not PDB formatted records. MMDB is capable of archiving conventional structure data as well as future descriptions of biomolecules, such as those generated by electron microscopy (surface models). In addition to the typical text queries and Entrez links from sequence and bibliographic records, structural "neighbors" are also computed using a new algorithm for 3-dimensional structure comparison. Structure data from Entrez may be viewed in 3D, with real-time rotation, using the public domain graphics programs RasMol or Kinemage, and soon with an NCBI-designed 3D viewer which can take advantage of some of the unique features of MMDB.