

## NIST Facilitates the Development of Medicines for HIV/AIDS

*A collaborative effort between NIST, the National Institutes of Health (NIH), and Rutgers University is addressing the needs of the pharmaceutical industry through the development of the HIV Structural Database (HIVSDB). The goal of the data resource is to receive, annotate, archive and distribute standardized 3-D structural data for AIDS related molecules and their complexes, with special emphasis on drug design and development.*  
(<http://xpdb.nist.gov/hivsdb/hivsdb.html>).

T.N. Bhat (Div. 831)

As the war against AIDS enters its third decade, pharmaceutical companies and organizations are continuing to research new drugs and treatments to thwart this persistent and deadly infection. Each year millions of dollars of both public and private investments target AIDS research. However, pharmaceutical companies are faced with rising pressure to reduce (or at least control) costs as well as time to market. There is a growing need for advanced informatics tools and standardized structural protein data entry methods that enable pharmaceutical researchers to streamline the drug discovery process.

NIST, in partnership with NIH's National Cancer Institute, and the National Institute of Allergy and Infectious Disease, and Rutgers University, developed the NIST HIV Standard Reference Database, SRD 102. The database contains a collection and classification of data from thousands of structures of HIV protease molecules and their inhibitors. The substances are arranged according to their chemical structure. Three-dimensional animations allow for the manipulation and viewing of the molecules from all directions. The database presently contains over 1,500 thousand structures, with another 1,000 structures to be added in the updated version to be released shortly. The NIST database, believed to include all known active drugs for HIV protease, provides a centrally located resource

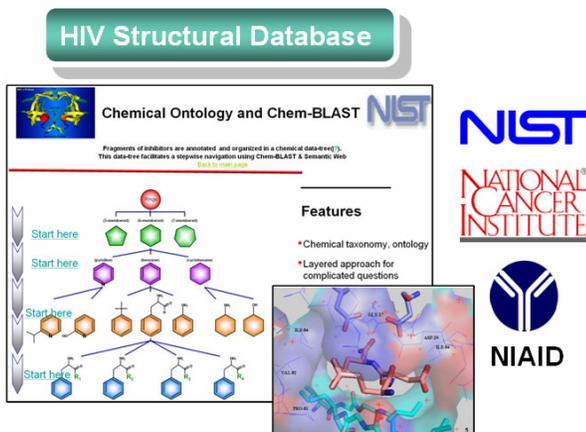
where all involved in AIDS research can exchange structural information for mutual benefit.

Novel new tools and techniques to annotate and distribute structural data with special emphasis on drug-design have also been developed. One of these methods is called Chem-BLAT (Chemical Block Layered Alignment of Structure Technique). This method uses a Semantic Web concept that aims for a more intelligent online experience. Web search engines that use this method are written to be more intuitive and accurate in processing data and finding results for end users. These search engines are "data aware". The "data aware" search engine guides the user via queries that are indexed according to the standard concepts of navigation. It guides users to formulate complex queries on chemical *fragments*. The users can specify chemical *fragments* using drop-down lists, text strings, or molecular sketches shown as hyperlinks. These queries are specified in several layers – in each layer the user can further refine the query element (*probe*) to reduce unwanted hits. This is designed to link the expertise of the structural biologist with that of the researcher in drug design and development. See example at:

<http://esw.w3.org/topic/HCLS/ChemicalTaxonomiesUseCase>

Since the public release of the new HIV SRD by NIST in July 2004, its multi-million hits make it one of NIST's most accessed databases.

**Testimonial:** "HIV protease is not only the most studied enzyme from a structural point of view, but it is of critical importance in the development of anti-AIDS drugs. Therefore, it was important to have all of the information on HIV protease curated and available on one site. ... Indeed, this database has become an important resource for the research community."  
James C. Cassatt, Ph.D., NIH, National Institute of General Medical Sciences, AIDS Coordinator.



**Dissemination:** The NIST Standard Reference Database 102 is listed on several database websites, including the most prominent international biological databases in the biosciences, such as the National Library of Medicine, the Protein Databank, and SWISSPROT) and technical libraries. It is widely used by the scientific community and more broadly disseminated through numerous database websites and AIDS educational services.