PROTEINS: Structure, Function, and Genetics 49:429–431 (2002)

EDITOR'S CORNER

HIVdb: A Database of the Structures of Human Immunodeficiency Virus Protease

Human immunodeficiency virus protease (HIV PR) was only discovered as encoded in the HIV genome in 1985,[1] but soon thereafter, this virus-specific enzyme was identified as a crucial target for designing drugs against acquired human immunodeficiency syndrome (AIDS).[2] With six such drugs approved to date since 1995 by the United States Food and Drug Administration and several others in current clinical trials,[3] the initial promise of rational drug design that originally seemed to be overly optimistic has been fulfilled beyond expectations. Introduction of PR inhibitors has changed the clinical outcome of AIDS and transformed an invariably fatal disease into a manageable one, although serious side effects and the development of resistance are still major unsolved problems.

Crystal structures of HIV PR were first reported in 1989[4–6] and their availability had a major role in the process of drug development (although it is clear that the structures provided only a small fragment of information necessary for drug design). All pharmaceutical companies that succeeded in creating PR inhibitor drugs and many that either discontinued the efforts or are still conducting such research have been involved in solving numerous crystal structures of HIV PR. Most of these structures were of complexes of the enzyme with inhibitors that were either potential or actual drugs, or intermediate compounds useful in drug-design efforts.[7] In addition, many academic laboratories joined the field and solved crystal and nuclear magnetic resonance structures of the complexes of HIV-1, HIV-2, and simian immunodeficiency virus (SIV) PRs with many diverse inhibitors. Because many variants of these proteins are known, attributable either to natural variation in the viral genomes or to their rapid mutation as a result of drug resistance, numerous mutant structures were also solved in pharmaceutical and academic laboratories. Although the total number of structures of these enzymes is not known, it can be estimated at many hundreds, making HIV PR the most widely studied enzyme in the history of protein crystallography.

During the last 13 years, many structures of HIV PR were published and as many as approximately 150 have been deposited at the Protein Data Bank (PDB). However, it became clear early on that many other structures, especially those solved as part of drug-development efforts and neither fully refined nor published, might ultimately be lost. This prompted the National Institute of General Medical Sciences (NIGMS) to award in 1996 an interagency agreement to the National Cancer Institute (NCI), where some of the initial structural efforts on HIV PR have taken place, to create a repository that would contain as many structures as possible, and not necessarily only the published ones. This decision led to creation of the Internet-based HIV PR Database (HIVdb).[8] After 6 years spent on creating and curating this database, the NCI effort is terminating on September 30, 2002. After that date, the National Institute of Standards and Technology (NIST), in collaboration with PDB, will take over and continue the project. This change of guard seems to be an opportune moment to remind the community about the existence of the database and the associated tools that have been created to enable its utilization.

HIVdb is an Internet-based archive of experimentally determined three-dimensional structures of HIV-1, HIV-2, and SIV PRs and their complexes with inhibitors or products of substrate cleavage. HIVdb was one of the first databases of macromolecular structures created outside of the PDB. HIVdb includes both primary structural data and the derived information for this family of three very closely related enzymes. For that reason, it serves as an example of a special subset of a general structural database that can exist on its own, as well as coexist with a larger structural database and its tools. Information regarding one particular enzyme can show in detail how the structure adapts to binding of different ligands through changes in protein–ligand interactions, and conformationally adjusts to binding under different conditions. For proteins that serve as drug-design targets, it is important to study these interactions fully and in as many complexes as possible. Careful analysis of the wild-type as well as drug-resistant mutants of HIV PR may also help in creating new drugs that would overcome the problem of resistance.[9]

The structures contained in HIVdb have either been previously deposited in the PDB, or have been obtained directly from depositors for their inclusion in HIVdb only. In the latter case, they may have been less completely refined, or even not refined at all beyond the placement of the ligand; or they may have resulted from experiments that were never fully completed and published, but nevertheless were comparable in quality to the structures deposited in PDB. Many structures unique to HIVdb came

Received 19 July 2002; Accepted 22 July 2002

from pharmaceutical companies and were used primarily as intermediate steps of drug design; under such circumstances, full refinement was not necessary. In addition, 69 structures came from the Structural Biochemistry Program of SAIC, a contractor to NCI that ceased operation 2 years ago. During its existence, the Structural Biochemistry Program performed an unusual and very successful drug-design effort outside of an established pharmaceutical company.

The data on structures contained in HIVdb can be divided into three categories. In the first category are data already available from the PDB, fully checked, and annotated. The second category contains data comparable in quality with those from PDB but not deposited to the general database. The third category contains data deposited to HIVdb without full validation, some of them not completely refined. However, data from all categories are very extensively checked before their inclusion in HIVdb and their provenance is explicitly marked. At the time of transfer from NCI to NIST, the NCI database contained 276 structures, 148 taken from PDB, and 128 that were unique entries in HIVdb. Of the latter, almost 90 structures fit into the third category as outlined above. In addition, another 18 structures were awaiting release in PDB and were scheduled to become part of HIVdb as soon as that action would take place. For identification purposes, the structures were annotated with a three-part code, hivXXXcom: the first three letters, *hiv*, are uniform code for all database entries; *XXX* is the number of the structure from a particular depositor in HIVdb; and *com* is an abbreviation of the origin—whether company or academic laboratory. We also keep the original PDB labels for those entries that are present in both databases.

The final archive version of HIVdb that resulted from the work at NCI can be found at http://mcl1.ncifcrf.gov/hivdb. That archive version of the database contains 282 structures processed by the time of the changeover to NIST and will not be updated any further. The new HIVdb portal is located at NIST at http://srdata.nist.gov/hivdb and will be updated and developed as the number of solved and released structures of this family of enzymes increases.

The original site launched in 1996 at NCI–Frederick is a collection of flat files and coordinates, not unlike the version of PDB that was available on-line at that time. That approach was chosen because of the lack of standard formats for both the official and unofficial entries. A user can access the particular data for each entry through a table providing an interface between general summary and details, pictures, experimental results, structure diagrams, and other related information. This version of the database utilizes some special tools that were specifically developed or implemented, for example, conversion of PDB-formatted files into other formats compatible with molecular modeling packages and a simple search of argument strings. Other tools include a program for multiple structure fitting, difference distance matrix analysis, and a program computing geometrical parameters for selected structures. A form for facilitating submission of new structures is also part of the original HIVdb. The database site also serves as an interface portal for reaching other HIV-related information, for example, sequence databases of HIV PR, or databases listing the development of resistance caused by treatment with HIV inhibitors. One of the most useful pages is the HIV PR database gallery, showing schematically major structural features of the enzyme, its binding to inhibitors, secondary structure elements, and the placement of residues that were mutated under pressure of various inhibitors.

Many comments from the users as well as new requirements for data structure and organization led to extensive changes in the database during the period of its existence at NCI. At the end, it became clear that the original form of the database required major modifications and that in the future HIVdb should exist in much closer cooperation with the PDB. Although HIVdb offers more information on the PR structures than can be found in the PDB, the new database offers many specialized tools that did not exist in the former. The presence of such tools helps in better understanding the structural phenomena related to HIV PR. For such reasons, a new HIV PR database portal was developed and was recently launched at NIST at http://srdata.nist.gov/hivdb in collaboration with the Biotechnology Division of NIST. Both versions of HIVdb are now accessible simultaneously.

The NIST version of the database uses an ORACLE database system and the user can handle the information as in any other database utilizing a modern database system. The most remarkable advantage of the new format of HIVdb is the availability of data in xml format and the possibility to create a structured query defined by the user. The new version of the database is extensively linked to other resources dealing with HIV PR or its inhibitors, especially those dealing with resistance. Important links are those to the Los Alamos HIV PR resistance database, http://hivweb.lanl.gov/content/index; Stanford HIV PR sequence database, http://hivdb.stanford.edu; and to the National Institute of Allergy and Infectious Disease site of anti-HIV compounds, http://www.niaid.nih.gov/daids/dtpdb/intro.htm.

Many studies dealing specifically with HIV PR that have been performed in the last decade benefited from the availability of multiple structures of this enzyme as provided by HIVdb. In addition, HIV PR became a favorite target for general studies evaluating structural similarities and differences in multiple structures with the aim of increasing the understanding of enzyme function, structure, and mechanism, mainly because of the availability of a large number of structures. In one of the most comprehensive studies utilizing HIVdb,[10] HIV PR was selected as a model system for an analysis of the interactions of an enzyme with its inhibitors. That study analyzed in detail protein flexibility and emphasized correlations between protein sequence, structure, and the influence of inhibitor binding on the dynamic properties of the enzymes. Other recent publications that referred to HIVdb discussed the future of structural databases and their specialized subsets from a more general point of view[11] and reported new structures of drug-resistant mutants of HIV PR.[12] Currently, a movement is underway to create an even larger database that would contain all available structural data

relating to HIV and its structural components. If that project succeeds, HIVdb will be an important part of any such future repository that could better help to explain the phenomenon of HIV infection from a structure–function point of view. In the meantime, it is hoped that HIVdb will continue its usefulness for the scientists working in the field of designing drugs against AIDS, or investigating general features of structure–function relationships in proteins.

ACKNOWLEDGMENTS

We acknowledge the interest and support of Marvin Cassman, former director of NIGMS, and James Cassatt, Director of the Division of Cell Biology and Biophysics, NIGMS, who provided the resources and encouragement for this project. HIVdb would not have been possible without their help.

Jiri Vondrasek
Macromolecular Crystallography Laboratory,
National Cancer Institute at Frederick,
Frederick, MD 21702; and
Institute of Organic Chemistry and Biochemistry,
Czech Academy of Sciences,
Prague 6, Czech Republic
Alexander Wlodawer
Macromolecular Crystallography Laboratory,
National Cancer Institute at Frederick,
Frederick, MD 21702

REFERENCES

1. Ratner L, Haseltine W, Patarca R, et al. Complete nucleotide sequence of the AIDS virus, HTLV-III. Nature 1985;313:277–284.
2. Katoh I, Yasunaga T, Ikawa Y, Yoshinaka Y. Inhibition of retroviral protease activity by an aspartyl proteinase inhibitor. Nature 1987;329:654–656.
3. Wlodawer A. Rational approach to AIDS drug design through structural biology. Annu Rev Med 2002;53:595–614.
4. Navia MA, Fitzgerald PM, McKeever BM, et al. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. Nature 1989;337:615–620.
5. Wlodawer A, Miller M, Jaskólski M, et al. Conserved folding in retroviral proteases: crystal structure of a synthetic HIV-1 protease. Science 1989;245:616–621.
6. Miller M, Schneider J, Sathyanarayana BK, et al. Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 Å resolution. Science 1989;246:1149–1152.
7. Wlodawer A, Erickson JW. Structure-based inhibitors of HIV-1 protease. Annu Rev Biochem 1993;62:543–585.
8. Vondrasek J, van Buskirk CP, Wlodawer A. Database of three-dimensional structures of HIV proteinases. Nat Struct Biol 1997; 4:8.
9. Erickson JW, Burt SK. Structural mechanisms of HIV drug resistance. Annu Rev Pharmacol Toxicol 1996;36:545–571.
10. Zoete V, Michielin O, Karplus M. Relation between sequence and structure of HIV-1 protease inhibitor complexes: a model system for the analysis of protein flexibility. J Mol Biol 2002;315:21–52.
11. Berman HM. The past and future of structure databases. Curr Opin Biotechnol 1999;10:76–80.
12. Mahalingam B, Louis JM, Reed CC, et al. Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. Eur J Biochem 1999;263:238–245.