# Review of PHYSPROP Database (Version 1.0)

Doris E. Bloch

Information Management Division (7407), Office of Pollution Prevention and Toxics, United States Environmental Protection Agency, 401 M Street SW, Washington, DC 20460

Received October 10, 1994

The Physical Properties (PHYSPROP)[1] data base, distributed by Syracuse Research Corporation, is a valuable resource for scientists who need to obtain physical constants for a wide variety of chemicals. The file contains CAS Registry numbers, chemical names, melting points, boiling points, water solubilities, vapor pressures, dissociation constants, octanol/water partition coefficients, and Henry's law constants for about 11 000 substances. In addition, there are structure diagrams for many of the chemicals, and the structures are searchable by means of functional substructure fragments.

N.B.: The PHYSPROP is implemented in a proprietary format, as a ChemBase data base, and users are advised that to use the PHYSPROP file at all, it is mandatory that they also have the ChemBase program, available from MDL Information Systems, formerly known as Molecular Design Limited. For those who are unfamiliar with this product, a short description follows. The ChemBase data base management software is written to run on a microcomputer under MS-DOS or PC-DOS; a 386 or higher microprocessor with at least 640 KBytes of RAM is advisable. It allows the user to create data bases which can include structures, to design customized report forms or tables for display on the screen or for printing and to perform data entry of structures and/or data. It permits searching on existing data bases, such as PHYSPROP, based on either structural features such as graphic fragments, formula fragments, molecular weights, etc., or on nonstructural data, e.g., chemical name fragments. ChemBase affords chemists a system which incorporates "chemical understanding" of such features as aromaticity, while remaining both relatively easy to learn and flexible enough to enable complex searching.

The PHYSPROP package consists of two 3.5″ diskettes, plus a User's Guide which gives some description of the file and includes a full listing of the references used in the data compilation.

Loading of PHYSPROP from diskette is quite straightforward; it requires 7.4 MB of free space on the microcomputer's hard drive. The install program is very simple to use, as it decompresses just two files: the PHYSPROP data base and a data display form, also referred to simply as a form. There is no instruction in the manual on the use of ChemBase; it is assumed that the user has a good working knowledge of this product prior to using PHYSPROP and will already know how to use the data base and form. The guide contains a short introduction (one paragraph), install and start-up directions (two paragraphs), a sample form, a two page description of how the physical/chemical constants were selected, and a full bibliographic listing of the source citations; this will be adequate for experienced ChemBase users but probably will not suffice for novices.

The User's Guide was discovered to be inadequate in certain other aspects, as well. The data base field structure description is referenced but not included in the PHYSPROP User Manual. In addition, there is no information anywhere for the contents of one field, ambiguously named OH in the data base and reported in units of $cm^3$/molc s. The guide mentions inclusion of a diskette file listing the bibliographic citations on a separate floppy disk, but this was missing in the package sent for review.

The sales literature accompanying the product indicates that the PHYSPROP file contains about 7000 chemicals (records) and that PHYSPROP is "an excellent starter data base for new ChemBase, ISIS Base, or MACCS users who need a base set of chemical structures and physical properties". The User's Guide refers to "approximately 5000 chemicals; about 1000 with a complete compilation of properties". Upon opening the file, it was ascertained that over 11,000 chemicals are included. However, it is not clear how the chemicals were selected, nor is there a common thread among them. Active pharmaceuticals, pesticides, commercial, and natural products are all represented. Therefore, a "typical" customer may or may not find the set of substances of value or fit for their needs.

A PHYSPROP record is composed of information for just one chemical. In addition to the structure and chemical name fields mentioned above, the physical property information is fully referenced, with the temperature conditions, method of evaluation, e.g., experimental or estimated, and a full literature citation for each constant. The index field is the CAS Registry number, which provides for quick and easy retrieval.

The Syracuse Research Corporation has done an admirable job in pulling together data from a variety of diverse sources. The User's Guide describes the selection criteria used where multiple sources were available and also the process for determining which value was to be extracted where the sources were in conflict. The accompanying documentation lists more than 300 references which were consulted.

However, not all fields in each record have been filled in. For example, the Henry's law constant was present for only about 800 substances, and water solubility for about 880 compounds. In fact, some 4000 records lacked even a chemical structure and name. Alternatively, some records had structures and names, but no physical constants.

The contents of the chemical name field are not restricted to standard systematic nomenclature; some names are common ones, e.g., "dihydrosafrole", "thalidomide", "dichlorvos", trade names, e.g., "Aroclor", or acronyms, e.g., "LSD", and "p,p′-DDT". In any case, there is only one name for each chemical, so the user will not retrieve a match based on a systematic name or fragment if an alternative was entered in the PHYSPROP file.

In perusing the data base, several misspelled names were encountered, e.g., "Phenobarital" for "Phenobarbital", "Amitptyline" for "Amitriptyline", "Pysotigmine" for "Physo-

stigmine". In other records, abbreviated functional groups were added to or substituted for names, e.g., "O=P(OEt)-(OEt)Et" was used as a name for "Ethylphosphonic acid, diethyl ester". This would inhibit name searching. Also, in a few cases bibliographic citations found in the data base were missing from the reference listings. As Syracuse Research continues to update PHYSPROP, one hopes such errors or inconsistencies will be discovered and resolved.

The Syracuse Research Corporation also distributes other, complementary programs which calculate certain physical constants, e.g., Henry's law constants, water solubilities, as derived from SMILES structural inputs, and bibliographic data bases with references to the original sources for physical property constants. PHYSPROP is unique in that one can perform graphic structure/substructure searches to find analogs. Therefore to justify its selection by a user, the structures depicted in PHYSPROP must be very accurate; no errors were detected in reviewing a few randomly selected records.

ISISBase or MACCS users along with their ChemBase brethren might also find the PHYSPROP data of interest and they could, through creation of sdfiles, potentially convert the records to whichever format is in current usage at their site; there are only about 15 PHYSPROP records containing text associated with the structure (all relating to the composition of mixtures) which would be incompatible with the more sophisticated software and would require manual transfer.

If the user is interested either in performing structure searching for frequently referenced substances to retrieve physical constants from the literature or in identifying constants for structural analogs by means of shared substructure fragments, then PHYSPROP is a valuable addition to the automated "reference shelf". However, when the purchase is made, addition of the six month update option is recommended to take advantage of SRC's anticipated corrections of omissions and errors.

## REFERENCES AND NOTES

(1) PHYSPROP is available from Syracuse Research Corporation, Merrill Lane, Syracuse, NY 13210-4080. Phone (315)-426-3200, Fax (315)-426-3429. The cost is $750, with a $60 additional charge for a six month update.

(2) ChemBase, ISISBase, and MACCS are available from MDL Information Systems, Inc., 14600 Catalina Street, San Leandro, CA 94577. Phone (800)-635-0064.

CI940181M