

**Xi Chen\***

**Yuhmei Lin**

**Michael K. Gilson**

Center for Advanced  
Research in Biotechnology,  
University of Maryland  
Biotechnology Institute,  
9600 Gudelsky Drive,  
Rockville, MD 20850

---

## The Binding Database: Overview and User's Guide

**Abstract:** *The large and growing body of experimental data on molecular binding is of enormous value in biology, pharmacology, and chemistry. Applications include the assignment of function to biomolecules, drug discovery, molecular modeling, and nanotechnology. However, binding data are difficult to find and access because they are available almost exclusively through scientific journals. BindingDB, a public, web-accessible database of measured binding affinities, is designed to address this problem. BindingDB collects data for natural and modified biomolecules and for synthetic compounds, and provides detailed experimental information. Currently, measurements by isothermal titration calorimetry are fully supported; measurements by enzyme inhibition will soon be included as well. The web site allows data to be searched by a range of criteria, including binding thermodynamics, sequence homology, and chemical structure, substructure, and similarity. Experimentalists are encouraged to publicize their data by entering it into BindingDB via the online forms. Such data can be updated or revised by the depositor, if necessary, and will remain publicly accessible. User involvement and feedback are welcomed.* © 2002 Wiley Periodicals, Inc. Biopoly (Nucleic Acid Sci) 61: 127–141, 2002; DOI 10.1002/bip.10076

**Keywords:** *Binding Database; molecular binding; biomolecules; drug discovery; molecular modeling; nanotechnology; binding affinities*

---

### INTRODUCTION

Molecular recognition—specific noncovalent association—is fundamental to biology in functions such as defense, metabolism, sensing, and differentiation. Specific binding also plays a special role in medicine because most drugs act by binding specific molecular targets and thus altering their activities. As a consequence, molecular recognition is a significant focus of research in the biosciences. In addition, the aim of

mimicking and ultimately understanding the physical basis of biomolecular binding has helped motivate the development of synthetic *chemical* systems that also display binding and recognition. Such systems are of interest not only as biomimetics but also as a basis for self-assembling structures and nanodevices.

It is thus not surprising that there is a large and growing body of experimental data on noncovalent binding. Qualitative binding data, such as those from yeast two-hybrid experiments, are valuable and tend to be

---

Correspondence to: Michael Gilson; email: gilson@umbi.umd.edu

\* Present address: Vertex Pharmaceuticals, Inc., 130 Waverly Street, Cambridge, MA 02139

Biopolymers (Nucleic Acid Sciences), Vol. 61, 127–141 (2002)

© 2002 Wiley Periodicals, Inc.

easier to obtain than quantitative data. However, as biology matures, quantitative data become increasingly important as a basis for modeling and prediction. Quantitative data are also critical for a range of molecular design applications, such as drug design, and for elucidating the physical basis of molecular recognition and improving the predictivity of models for binding.

Despite their value, experimental data on binding—both qualitative and quantitative—have historically been difficult to retrieve because they have been accessible only in scientific journals, which offer limited search capabilities even when indexed in online databases. For example, there has been no method of determining what proteins or small molecules are known to bind a target protein and its homologs. A workshop held in 1997 therefore considered the development of a community database that would gather binding data and make it accessible via the World Wide Web. Participants concluded that such a database could be of substantial value, and envisioned a number of scenarios for its use, for example:

*Simulation of Regulation Networks:* Binding data from the database are used to assemble a quantitative model of a signaling pathway involved in the control of the cell cycle.

*Discovery of Lead Compounds:* A scientist identifies a protein involved in a disease process, then searches the database for drug-like compounds already known to bind its homologues. These compounds become candidates for lead compounds directed against the new target.

*Prediction of Side Effects:* Two promising drug candidates are identified. Chemical similarity searches in the binding database reveal that a compound similar to one of the candidates binds a metabolically important protein. R&D is therefore focused on the other candidate.

*Evaluation and Reanalysis of Data:* A researcher measures a binding affinity that deviates from published results for the same system. Reanalysis of raw data from the prior study in the binding database allow the discrepancy to be traced to an arithmetic error in the data analysis.

*Development of Models for Structure-Based Drug Design:* A research group downloads 500 binding measurements for proteins of known structure and uses these data to parameterize a computer model for binding that is then used in drug-design software.

*Elucidation of the Mechanisms of Biological Effectors:* A small, organic molecule is found to be active in a growth-inhibition screen. A search of the binding database uncovers a very similar

compound that binds a protein known to be involved in cell-cycle regulation. This result yields a testable hypothesis for the mechanism of the original effector.

*Nanotechnology:* It is hypothesized that structured gels composed of self-assembling units could be used to separate and purify organic solutes from mixtures. The binding database is used to find low molecular weight “snaps and fasteners” that drive assembly of the materials to be tested.

Workshop participants also proposed guidelines for the development of a binding database. In particular, the database should include binding affinities for biopolymers such as proteins, nucleic acids, and carbohydrates, along with small organic molecules such as drug candidates and synthetic host and guest molecules. In addition, sophisticated interpretation and evaluation of data should be supported by inclusion of experimental details, perhaps beyond the level of detail acceptable for publication in a scientific journal. The database should be kept publicly accessible via the World Wide Web, and its interface should allow data to be retrieved according to a range of search criteria, including reactant name, range of binding affinities, reactant chemical identity and chemical substructure, and reactant sequence similarity. Finally, the web interface should permit experimentalists to submit new data for deposition into the database.

The present paper describes BindingDB, the database that was constructed to meet these guidelines (Figure 1). Previous publications have focused on the technical implementation of BindingDB<sup>1,2</sup>; this one takes the user's perspective. It begins with a brief discussion of the data and then provides a user manual for the scientist wishing to find existing data in BindingDB or to deposit new data. Additional sections discuss levels of usage, database policies, relationships to other molecular interaction databases, and aims for further development of this resource.

## DATA STORED IN BINDINGDB

BindingDB currently contains information on about 400 binding reactions involving a total of about 90 biopolymers and 150 other types of molecule. These data all were obtained by isothermal titration calorimetry (ITC) and are fit by the reaction model  $A + B \leftrightarrow AB$ . Enzyme inhibition data can be stored at this time, and it should soon be possible to search for and enter these data online. Enzyme-inhibition data can now be accessed as well. Additional measurement techniques

and binding models will be incorporated in the future. It is worth noting that each experimental technique requires special attention because BindingDB includes a detailed description of methods and results.

The information stored about each measurement can be grouped into four general categories: Experimental Results and Details, Reactants, Reaction Solution, and References. These categories are now briefly discussed; further details are available in prior publications<sup>2</sup> and online at <http://www.bindingdb.org/bind/info.html>.

*Experimental Results and Details:* This central category contains the thermodynamic results—free energy, enthalpy, and entropy of binding, in the case of ITC—along with the reaction model, key experimental conditions such as pH and temperature, and experimental details such as instrumentation and data fitting method.

*Reactants:* Reactants in BindingDB may be polymers, such as proteins; small organic molecules; or complexes of polymers and/or nonpolymers. Each reactant is identified by one or more names and polymers are also described by sequence while small molecules, such as drug candidates, are described by MDL Molfiles (<http://www.mdli.com/downloads/literature/ctfile.pdf>) and/or SMILES strings.<sup>3</sup> These formats provide a full description of each molecule in terms of elements and bonds. A complex—for example, an enzyme–coenzyme complex that binds an inhibitor—is described in terms of its constituents and their stoichiometry. Reactants in BindingDB can be linked to descriptions of the same molecules in other databases, such as the Protein Data Bank.<sup>4,5</sup>

*Reaction Solution:* The solution in which a reaction is studied is described in terms of the base solvent—e.g., 100% water, or water/ethanol at mole fractions of 50%—along with the pH buffer and other solutes such as salts or reducing agents.

*References:* This category stores publication information related to each BindingDB data entry, and includes PubMed reference IDs when applicable.

## SEARCHING THE DATABASE

BindingDB provides a range of search options, as described below, but all search results are reported in a uniform format. The results are displayed initially as an overview in the form of a summary table (Figure 2)

in which each row corresponds to one measurement and summarizes the names of the reactants, the experimental conditions, and the binding data. Further details of each measurement can be obtained by clicking the “More” link at the end of the corresponding row. This brings up a detail page (Figure 3) that displays thermodynamic results in greater detail, along with journal references, a summary of the reaction solution, and chemical descriptions of the reactants. Small organic molecules for which chemical structures are available are displayed graphically and sequences of proteins, DNA, and other polymers are provided as text. Links on this page lead the user to further details, including the detailed composition of the reaction solution, the type of instrument used in the study, and other databases with further information on publications and reactants.

The simplest search method is to enter a keyword or the name of a reactant or an author in the search form on the BindingDB home page, <http://www.bindingdb.org> (Figure 1). More specific search methods, accessible via the links on the left-hand side of the home page, are now described.

### Searching for Names, Words, and Numbers

Several text-based searches are possible. First, the **Reactant Name** searches allow the user to search for all measurements involving one or two specified reactants. These searches check all stored synonyms for each reactant, and partial matches are included in order to generate a broad search. Nonetheless, searches by reactant name can miss data because some synonyms may not be included with the data and because new synonyms can arise in the literature after data have been entered. As discussed below, it is therefore preferable to search by sequence in the case of proteins and nucleic acids and by chemical structure for other organic molecules. It is also possible to search for binding measurements with citations by a selected **Author**. Here, unlike a reactant name search, only exact matches are retrieved. Finally, each data entry in BindingDB is associated with descriptive **Keywords** or phrases selected by the person depositing the data. Examples might include “antibody,” “kinase,” or “heat-shock protein.” The keyword search yields all data in entries with matching or partially matching key words. For example, a search for “protein” would return entries with the key phrase “heat-shock protein.”

BindingDB allows users to find binding measurements with thermodynamic changes—**Free Energy**, **Enthalpy**, or **Entropy**—in specified ranges. For ex-

The screenshot shows the homepage of the Binding Database (BindingDB). At the top, there is a navigation bar with links: Home, Search, Deposit, SiteMap, About us, Email us, and Info. Below this is a search bar with a "Go" button and a "Help" link. The left sidebar is titled "Search by" and lists various search criteria: Reactant Name, Two Reactant Names,  $\Delta G^\circ$ ,  $\Delta H^\circ$ ,  $-\Delta S^\circ$ , Author, Substructure, BLAST, Keyword, and Advanced. Below this is an "Enter Data" section with links for Start New Entry, Continue Deposition, Revise Existing Entry, Send Comments, and Site Evaluation. The main content area contains a paragraph describing the database's purpose, a list of policies for data deposition, and a "Project Update" section with a list of events from 1997 to 2001. The footer includes a copyright notice: "Last updated February 6, 2001 ©2000 BindingDB. All rights reserved."

**FIGURE 1** The entrance page of BindingDB. The left pane lists the different search options and deposition entry points. The center frame provides an “Express Search” function and a brief introduction to the database. The right pane provides information on the project.

ample, one can search for all binding reactions with a change in standard free energy  $\Delta G^\circ$  more negative than  $-20$  kJ/mole. Units of kcal/mole may also be selected and will be carried through to the display of results.

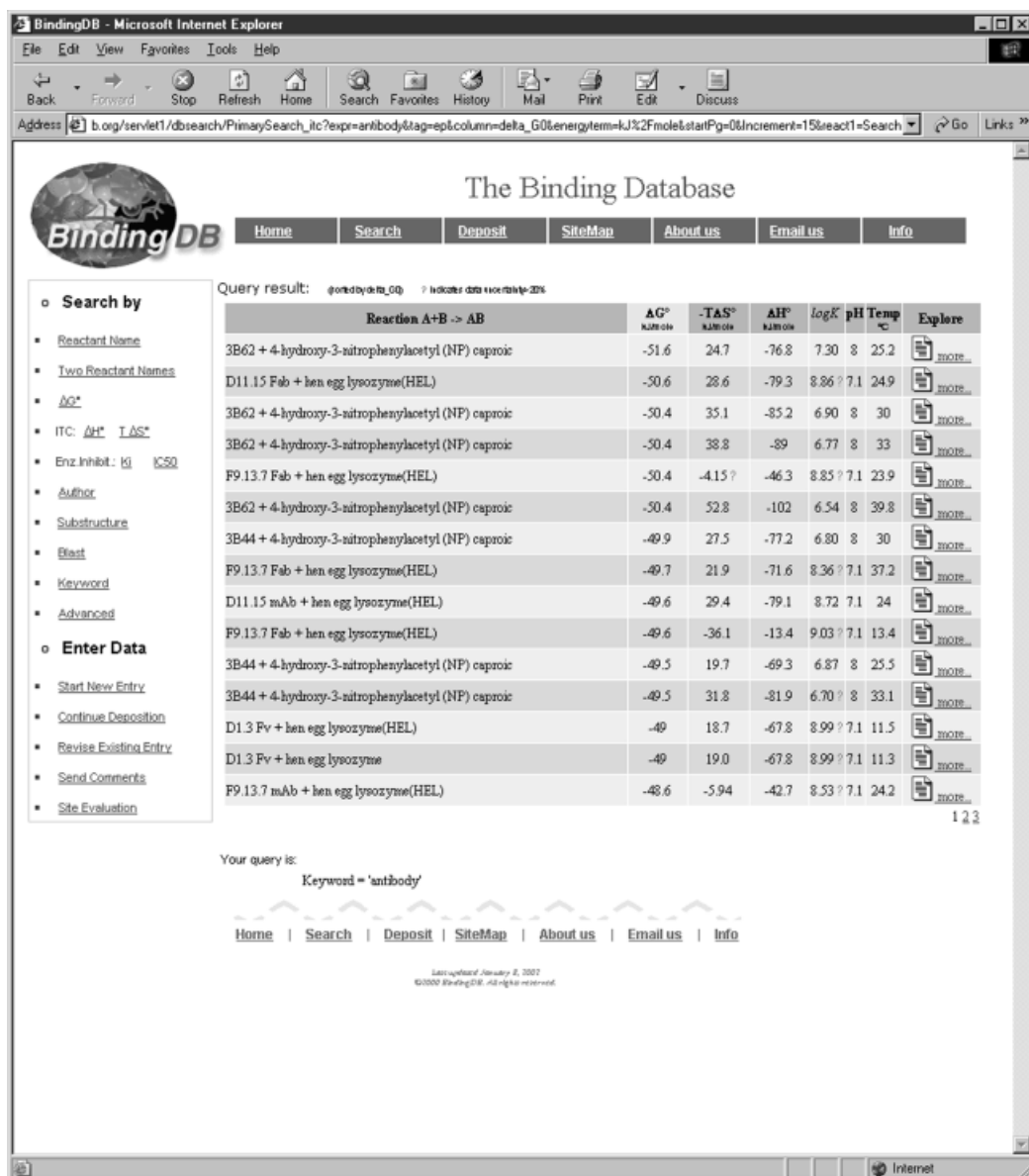
Finally, the **Advanced Search** method allows numerical and text-based search criteria to be combined (Figure 4); the results of such a search will satisfy all the criteria simultaneously (logical AND). For example, a user interested in the tightest binding reactions involving cyclodextrins might search with reactant name “cyclodextrin” and  $\Delta G^\circ < -15$  kJ/mole. Note

that it is not necessary to fill out all the text boxes in the advanced search form; empty boxes are simply ignored.

In all of these searches, the user may request that results be presented in either **kcal/mole** or **kJ/mole**. The results are sorted by  $\Delta G^\circ$  by default, but it is also possible to have them sorted by  $\Delta H^\circ$ ,  $-\Delta S^\circ$ ,  $\log K$ , pH, or temperature.

### Searching by Sequence and Chemistry

Although searches by reactant name can be useful, they may miss some data because the database may



**FIGURE 2** Example of a Summary page listing the first 15 results from a search for binding reactions involving the keyword “antibody.” The “more” links under “Explore” leads to details of each reaction. (See Figure 3.) Question marks in the table indicate results with reported uncertainties >20% of the reported value.

not list all synonyms. In addition, one may be interested in reactants that are similar but not identical to a given biopolymer or compound. BindingDB therefore supports searches by **sequence homology** for proteins and nucleic acids, and by **chemical identity**, **substructure**, and **similarity** for other organic compounds.

Binding measurements that involve a protein or nucleic acid similar or identical to a user-supplied sequence can be found by pasting the sequence of interest into the text window in the BLAST<sup>6</sup> search

page and choosing either a DNA or protein search (Figure 5). The results are returned initially as a standard BLAST report with hits ranked by significance ( $E$  value). Clicking on the name of a sequence in the report then retrieves all binding reactions involving this molecule. Links are also provided in the BLAST report to entries in other databases, such as the Protein Data Bank.

BindingDB enables three types of chemical searches for organic compounds other than proteins and nucleic acids: search for a specific compound;



The screenshot shows the BindingDB website interface in Microsoft Internet Explorer. The main content area is titled "Reaction Details" and contains the following information:

- Reactants:** Cell Reactant: F9.13.7 Fab; Syringe Reactant: hen egg lysozyme(HEL)
- Measurement:** Meas. Tech.: ITC; Entry Date: 12/11/00
- Thermodynamic Data:**
  - $\Delta G^\circ$ :  $-49.7 \pm 1.3$  (kcal/mol)
  - $\text{Log}_{10}K_a$ :  $8.36 \pm 8.07$
  - $\Delta H^\circ$ :  $-71.6 \pm n/a$  (kcal/mol)
  - Corrected for  $\Delta H_{\text{cal}}^\circ$ : n/a
  - Corrected for  $\Delta H_{\text{sol}}^\circ$ : yes
  - Protons Released: n/a
  - $\Delta C_p$ :  $-2.72 \pm 0.16$  (kcal/mol-K)
  - $\Delta S^\circ$ :  $-0.0706 \pm 0.009$  (kcal/mol-K)
- Other Parameters:** pH:  $7.1 \pm n/a$ ; Temperature:  $37.2 \pm 0.1$  (°C); Ionic Strength: n/a; Pressure: n/a; S.D. of the Fit: n/a; Stoich. Param.: n/a; Stoich. Param. Fixed: yes
- Solution:** Solute: 150 mM NaCl; Solvents: H2O; pH Buffer: 10 mM NaH2PO4/Na2HPO4
- Citations:** Schwarz, Frederick, Tello, Diana, Goldsbum, Fernando, Maruzza, Roy, Poljak, Roberto. Thermodynamics of antigen-antibody binding using specific anti-lysozyme antibodies. *Eur J Biochem* 228: 388-394 (1995) [Abstract][PubMed]
- More Info:** ITC RUN data, Solution Info, Instrument Info.
- F9.13.7 Fab:**
  - Source: n/a; Purity: n/a
  - Prep. Method: Pepsin-cleaved antigen binding fragment of F9.13.7 mAb
  - Synonyms: F9.13.7 Fab |
  - Other Links: IEBI
  - Type: Protein Complex
  - Mol. Mass: n/a; Comments: null
  - Components: This complex has following 6 components
  - Synonyms: F9.13.7 Fab chain H |
  - Other Links: PDB
  - Type: Antibody
  - Topology: Linear
  - Mol. Mass: 23602 Dalton
  - Organism: Mus Musculus
  - Description: n/a
  - Residue Count: 221
  - Sequence: QQLCQPGAE LKQDGA LKELKRAADYTFI FVWENFVQD PQGLFWIDE EEPIDYFHYIKERFRAL LKRSSTAYE QELSLSD RAVFPAALY YGHSFDVLD YKQGHVIV SRAKIPPV YSLAGSAQ DREIVTGLCL VEDVYEP VV WVSRSGLS GQVITPAVLQ IELVLEIV FVPSRPEI TVICVQKPA SFRVDEGIVP
  - Synonyms: F9.13.7 Fab chain L |
  - Other Links: PDB
  - Type: Antibody
  - Topology: Linear
  - Mol. Mass: 23794 Dalton
  - Organism: Mus Musculus
  - Description: Antibody to hen egg lysozyme
  - Residue Count: 214
  - Sequence: DQMLQIHL LKLRSDA LDKASQEH FVLPWVQDEP DQVLELVY TALEKSVH SFRGQSID YLEIDNLEQE DDAFYQQ QYTLFVYDQ GTELEKAD AAFVIVDFP SREQLKQA PAVLEKFFY FREDVAVKID QERQKQ VLWVWIDQQRD #YVRESHLI LKEDVYEM FVYCAERT STGVAQSS KAE
- hen egg lysozyme(HEL):**
  - Source: Sigma Chemical Co.
  - Purity: n/a
  - Prep. Method: n/a
  - Synonyms: hen egg lysozyme(HEL) |
  - Other Links: CAS

**FIGURE 3** Top of a Reaction Details page for an ITC Binding reaction, showing information on the reactants, thermodynamic results, reaction solution, citations, etc. Further information about ITC runs, Solution, and Instrument can be accessed via the hyperlinks.

search for compounds that include a user-specific fragment or substructure; and search for compounds similar to one provided by the user, where similarity is defined in terms of molecular “fingerprints”<sup>7</sup> (Daylight Chemical Information Systems Inc.: <http://www.daylight.com/dayhtml/doc/theory/theory.finger.html>;

ChemAxon, Inc.: <http://www.jchem.com/doc/admin/GenerFP.html>). The user’s query molecule or fragment can be entered with a graphical chemical draw program at the website (Figure 6). Alternatively, the user can use his or her preferred chemical draw program to generate a standard Molfile representing the

**FIGURE 4** Advanced search page. The names of one or both reactants can be specified, and reactions with standard free energy and enthalpy of binding can be selected; the interface allows units to be given in kJ/mol or kcal/mol. These criteria can be combined with publication criteria such as author names and publication dates. Fields left blank in this form do not affect the search.

molecule and then cut and past the file into a text window at the web site. The structure imported in this way can also be edited prior to query, if so desired. The chemical draw program allows entry of a number of rather detailed criteria (<http://www.bindingdb.org/bind/chemsearch/marvin/features.html>). For example, the user can allow or disallow a list of elements at a given atomic site in a molecule. Once the structure of interest has been specified, the search is performed with one of three options:

- Exact returns only reactants that *match* the drawn molecule
- Substructure finds molecules *containing* the drawn fragment.
- Similarity returns molecules *similar* to the drawn structure to within the user-specified threshold.

The search returns a page of compounds that meet the criterion, in a three-dimensional (3D) graphical format (Figure 7); clicking on the “Search” button for each compound retrieves all the binding reactions in the BindingDB that involve it. The applets that display the compounds also can provide a complete Molfile. This can be useful to users wishing to save or



**FIGURE 5** BLAST search page. The protein or DNA sequence to be used for the search is pasted or typed into the text box and the Search button is clicked.

analyze the full specification of a compound in BindingDB.

## ENTERING BINDING DATA

Scientists are invited to increase the visibility and usefulness of their data by entering it into the database. Data are considered appropriate if the detailed methods used to generate them are published in a refereed journal. This policy allows acceptance of binding data that are not published but that represent a direct continuation of a published study. The most straightforward deposition method is to use the online data deposition forms, as now discussed; the direct

deposition of data in the form of XML files is discussed below.

## Using the Online Data Deposition Forms

The deposition of new data begins with the **Start New Entry** button at the main page of the BindingDB web site. During the deposition process, the user is free to leave the site and later continue the entry process by using the **Continue Deposition** link on the home page, which makes the data already entered available for completion. Even after a data entry is completed and published to the web, it can be revised by the depositor by following the **Revise Existing Entry** link on the BindingDB home page. This generates a

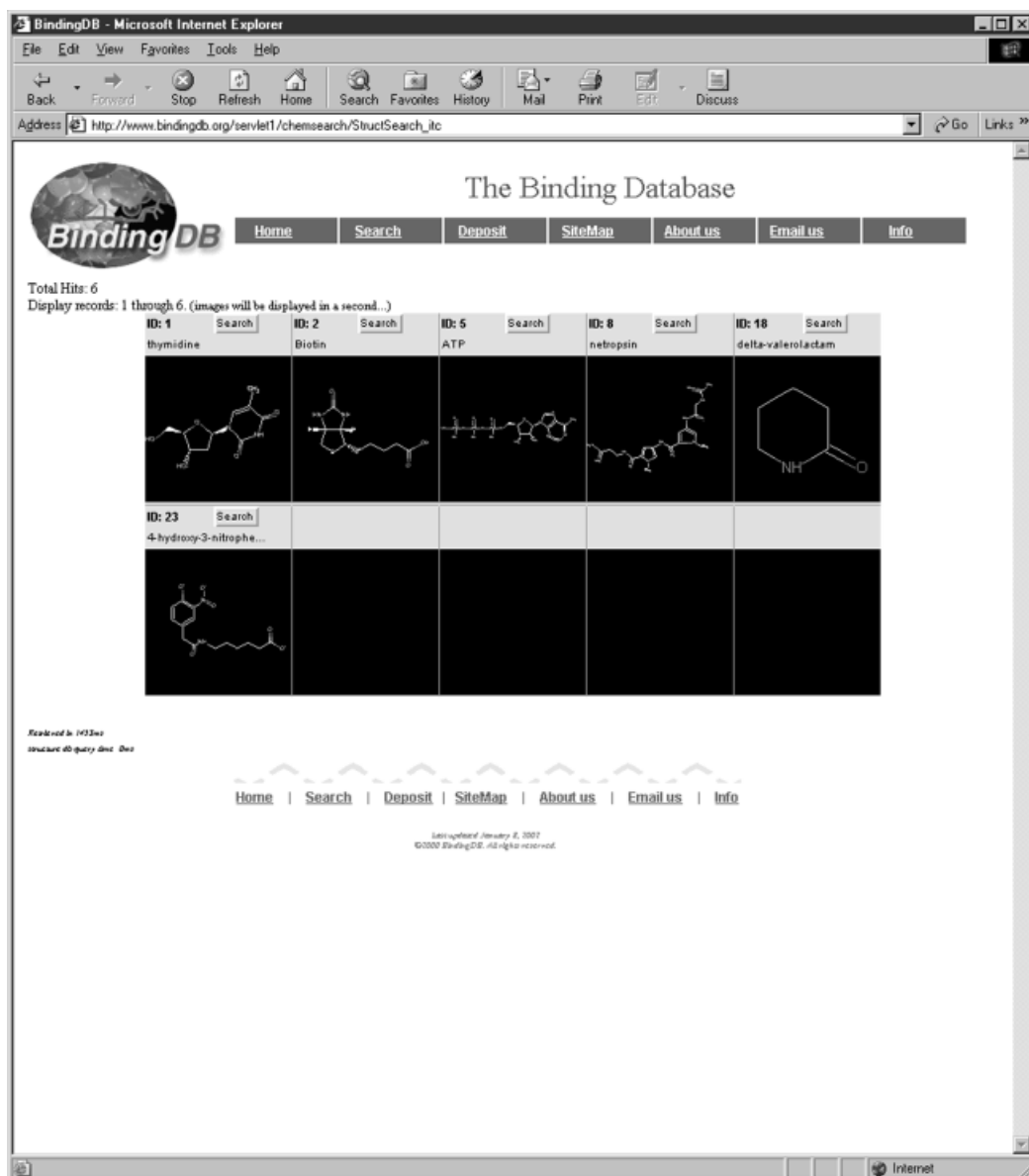


**FIGURE 6** Chemical (sub)structure search page. The target molecule can be drawn in the applet provided at the web page; clicking the “More” link brings up the periodic table and additional options. For example, the query shown specifies that neither P or S is acceptable at the leftmost position, and either C or N in the middle position. The radio buttons on the left allow search by substructure, chemical similarity, or exact chemical match.

duplicate of the existing entry and allows the depositor to edit the data via the normal deposition forms and to provide a description of the reason for the revision. The resulting data entry will be published to the web and the original entry will be saved but marked as obsolete to prevent it from being found in routine searches.

To start a new entry, the user chooses a measurement technique and a reaction model, as shown in Figure 8. Currently BindingDB accepts data gener-

ated with isothermal titration calorimetry matching the reaction model  $A + B \leftrightarrow AB$  and it should soon be possible to accept measurements by enzyme inhibition. The user’s email address is also requested as an identifier. When the **Begin** button is clicked, a unique deposition ID is generated that consists of the email address and a random 5-digit number; this ID is emailed to the user because it is needed to continue or revise the entry. The Begin button also brings up the main data deposition forms.



**FIGURE 7** Results page for a substructure search. The view of each molecule can be adjusted by double-clicking its image. Binding reactions involving each of the 6 molecules found in the search can be accessed via the “Search” button.

The left-hand side of the data deposition forms (Figure 9) contains the list of data categories and the right-hand side initially displays instructions. Clicking a category hyperlink on the left brings up the corresponding data form on the right, and it is recommended that the categories be filled out in the order presented because categories lower in the list can depend on data already entered. For example, one cannot enter **Results** before entering the appropriate **Reactants**. However, the user is free to go back and revise data in a previous category if needed. Multiple items are allowed in most data categories. For exam-

ple, Citations allows entry of up to 5 papers. Note that it is not necessary to enter the citations in numerical order, and that it is not a problem to enter citations 1, 2 and 3, and then to delete citation 1. Although each data form contains a number of data items to be entered, only those marked with a blue check are mandatory to complete the data entry.

Extensive help is available for the depositor. The **Instructions** link on the left brings up the detailed instructions shown initially, and each data item in the forms has associated **Example** and **Help** information. Each data category is now briefly described.

**FIGURE 8** Starting page for a new data entry. Currently only Isothermal Titration Calorimetry data can be entered, but this will be extended to enzyme inhibition and other methods.

*Entry* is a study that contains one or more binding measurements, and the Entry data category contains general information about an entry. It also allows the user to choose immediate release of the data or an indefinite “hold”; for example, a hold until final acceptance of an associated publication.

*Entrant* contains information about the person entering the data.

*Citations* contains information about scientific publications related to an entry. A paper can be cited because it contains the actual binding data entered here, or because it contains other related information, such as the sequence of a protein or

the method used to measure the binding affinities. Currently up to 5 citations are allowed. As noted above, it is possible to include unpublished data based upon a published method, so long as the *method* used to generate them is published in a refereed journal. In such cases, BindingDB can be used as the journal name and the depositor may provide an abstract describing the study.

*Reactant List* contains the list of all the reactants studied in this entry. The full list of reactants is entered here, and is then automatically listed in the **Results** forms below. *Reactants in Syringe* are those injected into the calorimetry vessel and

The screenshot shows a web browser window with the URL `http://www.bindingdb.org/jsp/forms/checkLogin.jsp?depid=d028.carb.nist.gov-60416&SUBMIT=CONTINUE+SESSION`. The page title is "The Binding Database". The main content area is titled "Enter Data in Category Entry" and contains a form for "Isothermal Titration Calorimetry". The form has several fields, some with blue checkmarks indicating they are mandatory. The fields and their values are:

Release Status <input checked="" type="checkbox"/>	e.g. Release without delay ( <a href="#">help</a> )	<input checked="" type="radio"/> Release without delay <input type="radio"/> Hold until notice
Entry Title <input checked="" type="checkbox"/>	<a href="#">Example (help)</a>	Role of Salt Bridge Formation in Antigen-Antibody Interaction
Keywords	e.g. antibody, drug design, DNA, HEL... ( <a href="#">help</a> )	hen egg lysozyme, entropy, HyHEL10, antigen-antibody binding
Journal <input checked="" type="checkbox"/>	e.g. J Biol Chem ( <a href="#">help</a> )	J. Biol. Chem.
Journal volume	e.g. 12 ( <a href="#">help</a> )	271
First page <input checked="" type="checkbox"/>	e.g. 122 ( <a href="#">help</a> )	32612
Year <input checked="" type="checkbox"/>	e.g. 2000 ( <a href="#">help</a> )	1996
PubMed ID	e.g. 10747923 ( <a href="#">help</a> )	8955089
Your Name	e.g. Smith, John ( <a href="#">help</a> )	

The sidebar on the left contains the following categories and their sub-items:

- Reactants in Syringe
  - Small Molecules 1 2 3 4 5 6 7 8
  - Proteins & Polynucleotides 1 2 3 4 5
  - Complexes 1 2 3 4
- Reactants in Cell/Vessel
  - Small Molecules 1 2 3 4 5 6 7 8
  - Proteins & Polynucleotides 1 2 3 4 5
  - Complexes 1 2 3 4
- Solutions 1 2 3 4 5
- Data Fitting Methods 1 2 3
- Instruments 1 2 3
- Results 1 2 3 4 5 6 7 8 9 10

At the bottom of the sidebar, there are links for "Check & Submit", "Re-Login", "Instructions", and "Your Session ID: d028.carb.nist.gov-60416".

**FIGURE 9** Sample data deposition page. Clicking a data category on the left brings up the corresponding form on the right. Mandatory data items are marked with a blue check. When the data have been entered into the form, the Save button must be clicked.

the Reactants in *Cell/Vessel* are those initially in the vessel. A reactant can be a *small molecule*, a *protein or polynucleotide*, or a *complex*. A small molecule is a small, possibly drug-like molecule normally identified via a Molfile or a SMILES string. *Proteins & Polynucleotides* comprise proteins, peptides, and polynucleotides, for which a sequence should be entered. *Complexes* here are reactants best described as complexes of multiple components, such as a protein-protein dimer. To enter a complex, first enter its parts as Small Molecules or Proteins &

Polynucleotides, then use the Complexes category to put these parts together in the correct stoichiometry.

*Solutions* contains a description of each solution used in the data entry, including information on the solvent, solutes, and pH buffer, if any. The user also assigns each solution an arbitrary name that will be used later when entering the Results. *Data Fitting Methods* describes the method and software used to extract affinities and other results from the experimental data.

*Instruments* allows the instrument(s) used to obtain

The data is ready for submission. Please review data below and click the "Submit" button. You can also return to the entry pages to modify the data. After submission, the data can still be revised at this web-site, if necessary.

Category	Started	Complete items by ID	Incomplete items by ID
Entry	Yes	✓	
SyringeReactant Small Molecules	Yes	1,	
SyringeReactant Proteins & Polynucleotides	No		
SyringeReactant Complexes	No		
CellReactant Small Molecules	Yes	1,	
CellReactant Proteins & Polynucleotides	No		
CellReactant Complexes	No		
Solution	Yes	1,	
Data Fit Method	Yes	1,	
Instrument	Yes	1,	
Results	Yes	1,	

Submit

**FIGURE 10** Top of the "Check & Submit" page. This page checks the data saved so far and reports on whether each data category has been completed. If the Entry meets the minimal submission criteria, a "Submit" button will be shown.

the data to be described. At present, only calorimeters are included, but other instruments will be included as additional measurement techniques are added to BindingDB.

*Results* contains the measured affinities and related experimental data, such as enthalpy and entropy changes, along with key conditions such as pH and temperature. Data entered previously under Reactants, Solutions, Data Fitting Methods, and Instruments are each made available in pull-down menus for convenience and to ensure the formation of correct relational links. Thus, the

user can now report results for any previously entered syringe reactant with any reaction cell reactant, with any data fitting method, etc. Because experiments are often repeated with only minor changes in conditions, a "Replicate Result" option is provided to copy an existing Result record into a new one, to be modified and saved.

Data saved while filling out these forms remain available if the user stops and resumes later. However, the session will probably time out in order to conserve

resources for other users. If this happens, the user can select **Re-Login** to resume the interrupted session without the need to reenter an email address or session ID.

Clicking the **Check & Submit** button at any time during the entry process causes the software to check the completeness of the entry (Figure 10), and missing items are flagged in the resulting report. Once the entry is complete, it can be submitted to the main database. Data submitted during this entry process are not entered into the main database itself until it has been reviewed by a member of the BindingDB staff.

### Depositing Data via Flat File

Binding DB also allows data to be contributed in the form of an XML file. This mechanism allows third-party software, such as that used to control a microcalorimeter, to interact directly with BindingDB. It may also be useful for uploading large quantities of data from other sources, and the XML format should also be useful as a means of exporting data from BindingDB. The Data Type Definition (DTD) for ITC and enzyme inhibition data in BindingDB can be accessed at <http://www.bindingdb.org/bind/info.html>.

### PROVIDING FEEDBACK

In order to improve the usability of the BindingDB, we have solicited and continue to be strongly interested in feedback from users and prospective users. The site provides three mechanisms for this.

1. **User Survey.** A comprehensive, anonymous user-survey can be accessed via the BindingDB home page at <http://www.bindingdb.org/survey/survey.html>. The questionnaire evaluates the following aspects of the BindingDB: overall impressions, data and content, search and retrieval, display of results, user-friendliness, and online help and feedback. Most of the questions ask the user to provide a score on a scale from 1 to 5, and a number of questions also allow for free-text answers and suggestions from the user.
2. **"Email Us" Links.** Users who prefer not to fill out the comprehensive evaluation may offer comments and suggestions via the email links displayed on most pages.
3. **Error Reports.** The Reaction Details page provides an error report link for use when a user believes that he or she has found an error in the data. The email message generated by this link

identifies the suspect page to facilitate evaluation of the data by BindingDB staff.

### USAGE STATISTICS

Interest in BindingDB is strong although its data holdings are still modest, at several hundred binding reactions. Thus, from November 2000 to June 2001, the web site attracted approximately 50,000 hits from about 2000 different hosts from a range of academic institutions and pharmaceutical and biotech companies worldwide.

### FUTURE DIRECTIONS

The BindingDB has successfully met the design goals specified in its 1997 workshop ([http://www.bindingdb.org/bind/update/workshop\\_rep1b.pdf](http://www.bindingdb.org/bind/update/workshop_rep1b.pdf)). Plans for further development are now discussed.

### Expanding the Data Collection and Scope

Filling BindingDB with useful data is a central challenge, which will be met by a multipronged strategy. First, the BindingDB staff will continue to deposit data drawn from the literature. We aim to develop software that can facilitate this task by at least partly automating the process of data extraction from articles. It may be appropriate to focus part of this effort on measurements related to key disease areas. From the standpoint of physical chemistry, it would valuable to include the binding affinities associated with proteins and nucleic acids of known structure and hence represented in the PDB. Second, it is expected that contributions from experimentalists via the online forms will become increasingly significant. We are currently soliciting contributions and reviewing the deposition forms with an eye to further simplification in order to minimize the effort required to use them. Finally, we look forward to collecting existing caches of binding data from other sources and making them publicly available via BindingDB.

Not only the number but also the scope of the data included in BindingDB will increase. In particular, measurements by additional techniques, such as enzyme inhibition and BiaCore, will be included. Given sufficient interest from the scientific community, the data dictionary will also be expanded to handle raw instrument data. This would permit users to reanalyze and evaluate the fundamental data themselves.



## Enhancing the User Interface

New data analysis and data mining tools will become important as the holdings of BindingDB expand, as will new ways to access the data. Some of the new capabilities will be implemented within BindingDB, while others will emerge as BindingDB is integrated more closely with related databases.

*Tools for Structure-Affinity Correlation:* Methods of visualizing 3D macromolecular structures and interpreting them in the context of measured affinities will be included, as will methods for finding proteins based upon 3D structural similarity. The latter are expected to use accepted online applications, such as CE (<http://cl.sdsc.edu/ce.html>), VAST (<http://www.ncbi.nlm.nih.gov/Structure/VAST/vastsearch.html>), or DALI (<http://www.ebi.ac.uk/dali/>).

*Tools for Data Visualization and Analysis:* Methods of exploring the data will be provided. For example, an application could be developed to show the overall distribution of binding affinities within the database, or some segment of it, and then allow the user to zoom in to individual entries for further analysis. Visualization may also help to derive data correlations. Thus, BindingDB could provide a dynamically generated graph that plots measured binding free energy vs chemical similarity of small molecules that bind to HIV protease. Such correlations may be useful for discovering commonalities among active molecules that would then be useful for molecular design. Similarly, it will be possible to precompute and make available quantitative structure-activity relationships for series of data within BindingDB. It may also be of interest to provide a server-side ligand-protein docking and scoring program to help users generate pre-

dictions in cases where binding data is not available.

*Integration with Other Databases:* BindingDB is intrinsically an integrating database, since it provides data on the interactions of molecules that are frequently represented in other databases. Links with other databases will enable users to, for example, quickly determine what other proteins a protein found in SwissProt is known to bind or to learn whether a ligand in the BindingDB affects the growth of cancer cells via the NCI compound database.

*Supporting Data Downloads:* BindingDB has already received requests to download multiple binding data. An FTP server will permit such downloads in XML format, as discussed above. It may be helpful to allow users to select a number of entries, and then download them as a batch file.

## REFERENCES

1. Chen, X.; Liu, M.; Gilson, M. *Combinat Chem High Throughput Screening* 2001, in press.
2. Chen, X.; Lin, Y.; Liu, M.; Gilson, M. *Bioinformatics* 2001, accepted.
3. Weininger, D. *J Chem Inf Comp Sci* 1988, 28, 31–36.
4. Sussman, J. L.; Lin, D.; Jiang, J.; Manning, N. O.; Prilusky, J.; Ritter, O.; Abola, E. E. *Acta Crystallogr D Biol Crystallogr* 1998, 54, 1078–1084.
5. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. *The Protein Data Bank. Nucleic Acids Res* 2000, 28(1), 235–242.
6. Altschul, S. F.; Madden, T. L.; Schaffer, A. A.; Zhang, J.; Zhang, Z.; Miller, W.; Lipman, D. J. *Nucleic Acids Res* 1997, 25, 3389–3402.
7. Flower, D. R. *J Chem Inf Comput Sci* 1998, 38, 379–386.