

# AffinDB: a freely accessible database of affinities for protein–ligand complexes from the PDB

Peter Block, Christoph A. Sotriffer, Ingo Dramburg and Gerhard Klebe\*

Department of Pharmaceutical Chemistry, Philipps-University Marburg, Marbacher Weg 6,  
D-35032 Marburg, Germany

Received June 17, 2005; Revised and Accepted October 3, 2005

## ABSTRACT

**AffinDB is a database of affinity data for structurally resolved protein–ligand complexes from the Protein Data Bank (PDB). It is freely accessible at <http://www.agklebe.de/affinity>. Affinity data are collected from the scientific literature, both from primary sources describing the original experimental work of affinity determination and from secondary references which report affinity values determined by others. AffinDB currently contains over 730 affinity entries covering more than 450 different protein–ligand complexes. Besides the affinity value, PDB summary information and additional data are provided, including the experimental conditions of the affinity measurement (if available in the corresponding reference); 2D drawing, SMILES code and molecular weight of the ligand; links to other databases, and bibliographic information. AffinDB can be queried by PDB code or by any combination of affinity range, temperature and pH value of the measurement, ligand molecular weight, and publication data (author, journal and year). Search results can be saved as tabular reports in text files. The database is supposed to be a valuable resource for researchers interested in biomolecular recognition and the development of tools for correlating structural data with affinities, as needed, for example, in structure-based drug design.**

## INTRODUCTION

Understanding the energetics of biomolecular recognition is of paramount importance for a large variety of biomedical and biotechnological disciplines. One of the most prominent examples is given by structure-based drug design where the 3D structure of a target macromolecule (most frequently a

protein) is used to identify, design or optimize small-molecule ligands, which bind tightly to the target. Obviously, such design can only be successful if the structural requirements for energetically favorable interactions and high-affinity binding are known. Much of the current knowledge has been gained from comparative analyses of different complex structures and their affinities (1,2). These analyses, however, were normally restricted to rather small sets of data, and the understanding of protein–ligand recognition is still far from being complete, as illustrated by the recurring surprises during projects of molecular design (3–6). Clearly, more data are instrumental to increase the knowledge about protein–ligand interactions and to improve not only the qualitative understanding but also the quantitative tools for estimating affinities from complex structures, such as empirical, regression-based scoring functions (7–12).

Structural data of protein–ligand complexes are available to a large and rapidly increasing extent through the Protein Data Bank (PDB) (13). This database, however, is a general resource for biomacromolecular structures that had not particularly been designed for protein–ligand complexes. Accordingly, secondary databases, such as Relibase (14,15) and PDBsum (16,17), have been developed which provide more convenient access to specific information about protein–ligand complexes (e.g. search functions for ligand structures and analysis tools for interaction patterns in Relibase). Unfortunately, neither the secondary *structural* databases nor the PDB contain any information about the binding energetics of the corresponding complex, since this information is not required to be included upon submission of structural coordinates to the PDB. However, some databases exist that collect binding data for enzymes, receptors or protein–ligand complexes in general, such as BindingDB (18) and KiBank (19), but these, in turn, are not limited to complexes with available structure and do not provide a direct link to the available 3D structure of a given complex with measured affinity.

Given the obvious need for databases that establish the missing link between structural information from the PDB

\*To whom correspondence should be addressed. Tel: +49 6421 2821313; Fax: +49 6421 2828994; Email: [klebe@staff.uni-marburg.de](mailto:klebe@staff.uni-marburg.de)  
Present address:

Ingo Dramburg, BioSolveIT, An der Ziegelei 75, 53757 Sankt Augustin, Germany

and the rather sparse and widely distributed affinity data, we started to develop AffinDB, a database of affinity values collected from the scientific literature for protein–ligand complexes of known structure. Originally intended as a simple tabular collection of affinity values related to PDB codes for in-house use only, the project has grown over time, both with respect to data content and database management, such that it has ultimately been made available to the public as a potentially valuable new resource, despite the recent appearance of other databases of similar scope, most notably PDBbind (20,21) and Binding MOAD (22). In the following, we briefly describe the database architecture and content of AffinDB, as well as the data collection procedure, give a succinct introduction to possibilities for accessing the data through the user interface, and discuss differences and similarities to other databases.

## METHODS

### Database architecture

AffinDB is based on a MySQL (4.0.24) backend machine. The web interface is written in PHP (4.3.10). The database is designed to provide supplementary information for PDB structures of protein–ligand complexes. Accordingly, AffinDB is structured by PDB code, which is the primary reference for all the data. Basic PDB meta-information about the protein is available for every PDB structure, basic ligand information is provided for ligands with more than five non-hydrogen atoms. Ligand entries of complexes are stored only once in AffinDB, i.e. in case of multiple occurrences of the same ligand in different structures, a pointer to the reference ligand molecule is used. Affinity data and related information are always associated with a specific ligand of a specific PDB structure.

### Data collection and database content

The database core is constituted by basic meta-information about all PDB structures. To obtain these data, a helper-application was generated with a Python-based Relibase toolkit and the data were retrieved from Relibase+ (14,15). A further preprocessing step served to store only ligands with more than five non-hydrogen atoms in AffinDB, using a unique and consistently created name for these molecules. The PDB meta-information provided for every entry includes the name of the protein or protein class (as given in the header information of the PDB file), the EC number (for enzymes), the protein source, the resolution of the crystal structure and the names of the authors who determined the structure. In addition, for each PDB code links to the following external databases were added: PDB (13), Relibase (14,15), MSD (23), SCOP (24) and PDBsum (16,17).

The ligand entries consist of the chemical name (as provided in the PDB file), the molecular weight and the SMILES code (25) as basic information. In addition, a 2D molecule drawing of the ligand structure is included. These drawings are generated for every unique ligand with Marvin (3.5.7) (<http://www.chemaxon.com/marvin>). Babel (1.6) (<http://www.eyesopen.com/babel>), Corina (3.1) (26), and *in-house* software was used to harmonize the format of the ligands in order to

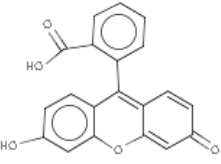
obtain best results from Marvin. This automated procedure provided correct drawings for most of the ligands. A small proportion which could either not be drawn by Marvin or gave distorted pictures had to be post-processed by hand. Titratable functional groups are always shown in their neutral state, independent of any actual protonation state.

The protein and ligand information described so far is shown by AffinDB regardless whether affinity data are already available for the PDB entry or not. The main purpose of AffinDB, however, is to provide affinity information. Affinity data are exclusively extracted from the scientific literature. Both ‘primary’ and ‘secondary’ references are taken into account. A primary reference is a paper describing the original work of the affinity measurement for the corresponding protein–ligand complex. A secondary reference, instead, is any other paper that reports an affinity value for a PDB complex; this may include publications with compilations of affinity data for the development of scoring functions or similar purposes. In a secondary paper, the affinity value for a PDB complex is often only cited, without specifying further experimental details. So far, more than 740 affinity values covering over 470 PDB complexes could be collected and stored in AffinDB (cf. Discussion).

The input of affinity data into AffinDB is implemented in the form of a wizard. After entering the desired PDB code, AffinDB provides a list of all the ligands of the PDB entry in combination with a 2D molecule drawing and a hint whether affinity information is already stored for the given ligand. After choosing the desired ligand, the user is requested to enter the affinity information. Upon submission of the data, simple checks of the data integrity are performed and the entry is flagged for review by the database curators. Only after a database curator has checked these data, they are released for public access in AffinDB.

The binding affinity is thermodynamically quantified as a free energy of binding  $\Delta G_{\text{bind}}$  or as equilibrium constant (for association:  $K_a$ ; for dissociation:  $K_d$ ) for the reversible equilibrium reaction between protein P and ligand L to form the protein–ligand complex PL:  $P + L \leftrightarrow PL$ .  $\Delta G_{\text{bind}}$  and the equilibrium constants are related by the equation:  $\Delta G_{\text{bind}} = -RT \ln K_a = RT \ln K_d$ , where  $T$  is the temperature (in Kelvin) and  $R$  is the ideal gas constant ( $8.314 \text{ J mol}^{-1} \text{ K}^{-1}$ ). For enzyme inhibitors, affinities are more frequently quantified in terms of parameters derived from kinetic assays. This may either be the inhibition constant  $K_i$  (which to a first approximation may be considered as a  $K_d$  for the enzyme–inhibitor complex, thus  $\Delta G_{\text{bind}} = RT \ln K_i$ ) or the  $\text{IC}_{50}$  value, which is the inhibitor concentration leading to 50% inhibition of the enzymatic activity. In AffinDB, the affinity value is stored in the same form as published in the specified reference, i.e. without any conversion of type or unit. If available, experimental uncertainties or error margins are saved as well.

Along with the affinity value itself, also the experimental method and conditions of the affinity measurement are stored in AffinDB, if specified in the corresponding reference. This information is provided to the user through a separate ‘affinity information window’ (Figure 1), which can be opened by activating the ‘Details’ link next to the affinity entry in the main window (Figure 2). The method by which the affinity was determined is characterized by a keyword or a brief statement. Temperature and pH value at which the measurements

Affinity information for <b>FLU_PDB1FLR</b>	
<b>Chemical Name:</b> Fluorescein	
<b>Molecular Weight:</b> 331.3	
<b>Affinity value</b>	
<b>Experimental Conditions</b>	
<b>Method</b>	fluorescence quenching and polarization
<b>Temperature</b>	293
<b>pH</b>	6.8
<b>Buffer</b>	50 mM sodium phosphate
<b>Reagents/Additives</b>	n.a.
<b>Literature</b>	
<b>Author</b>	Gibson, A. L. et al.
<b>Journal/Source</b>	<a href="#">Proteins, 3, (1988), 155</a>
<b>Literature primary</b>	yes
<b>Comment</b>	
Method described by Herron in "Fluorescein Hapten: An Immunological Probe", Voss, E. W., Jr, ed Boca Raton, Florida: CRC Press, (1984), 49. - NOTE: With 40% MPD a very different value has been measured!! Additional values at different temperatures and pH 8 are reported in Fig. 3 of the publication.	

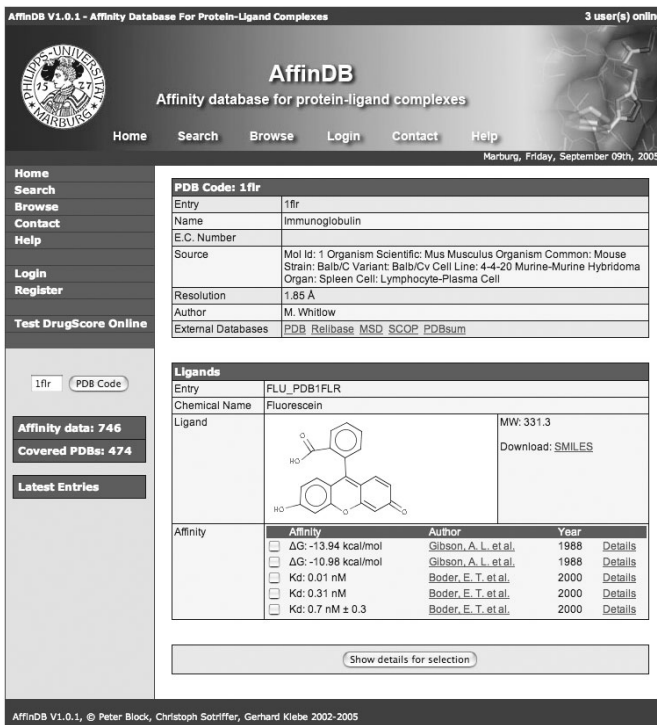
**Figure 1.** Affinity information window for one of the affinity entries for PDB complex 1FLR.

were carried out are stored separately. In addition, the buffer and any other significant reagents or additives present in the solution are reported. For the literature reference itself, the name of the first author, the title of the journal, as well as volume, year and first page of the publication are stored. The reference is linked to the corresponding PubMed entry, which provides direct access to the abstract. A flag indicates whether the reference is of primary or secondary type. Finally, comments and additional valuable information regarding the method, the reference, the structure or the affinity value itself are saved in a separate data field.

### Database access

The database is freely accessible at <http://www.agklebe.de/affinity>. Data can be retrieved via the PDB code, by defining specific search queries using the affinity search form, or simply by browsing.

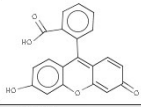
Upon specifying a PDB code in the data entry field on the left navigation bar of the main window (cf. Figure 2), the summary information for the PDB entry is shown. If an affinity value for the ligand is available in the database, it is displayed below the ligand structure, along with the first author and the year of the publication which reports this value. If additional affinity values are available from other references, these are displayed as well, each in a separate line (cf. Figure 2). Further details can be requested for each affinity entry. Searching for a specific PDB entry with the affinity search form (accessible through the 'Search' link in the left navigation bar) yields only a result if an affinity value is already associated with the corresponding PDB entry.



**PDB Code: 1flr**

Entry	1flr
Name	Immunoglobulin
E.C. Number	
Source	Mol Id: 1 Organism Scientific: Mus Musculus Organism Common: Mouse Strain: Balb/C Variant: Balb/Cv Cell Line: 4-4-20 Murine-Murine Hybridoma Organ: Spleen Cell: Lymphocyte-Plasma Cell
Resolution	1.85 Å
Author	M. Whitlow
External Databases	PDB Relibase MSD SCOP PDBsum

**Ligands**

Entry	FLU_PDB1FLR
Chemical Name	Fluorescein
Ligand	
MW: 331.3	Download: <a href="#">SMILES</a>

Affinity	Author	Year	
<input type="checkbox"/> $\Delta G$ : -13.94 kcal/mol	Gibson, A. L. et al.	1988	<a href="#">Details</a>
<input type="checkbox"/> $\Delta G$ : -10.98 kcal/mol	Gibson, A. L. et al.	1988	<a href="#">Details</a>
<input type="checkbox"/> Kd: 0.01 nM	Boder, E. T. et al.	2000	<a href="#">Details</a>
<input type="checkbox"/> Kd: 0.31 nM	Boder, E. T. et al.	2000	<a href="#">Details</a>
<input type="checkbox"/> Kd: 0.7 nM $\pm$ 0.3	Boder, E. T. et al.	2000	<a href="#">Details</a>

Show details for selection

**Figure 2.** Main window showing a PDB entry with affinity data in AffinDB. PDB complex 1FLR is used as an example. Five different affinity values measured in different studies and under different conditions are available for this complex. The left navigation bar provides fast access to all functionalities of AffinDB.

In the affinity search form, a variety of queries for affinity data and related information can be defined. It is possible to search for affinities of a certain range of magnitude and for measurements carried out at a specific temperature and/or pH range. Affinities for certain enzyme classes or PDB codes may be retrieved, as well as affinities for ligands of a certain molecular weight range. Also the affinity values published by a certain author or within a specified time frame can be requested, and the retrieved affinity values may be limited to those obtained from primary literature sources.

AffinDB generates tabular reports for displaying affinity search results and for browsing through the database (Figure 3). The format of the tables consists of six columns providing the drawing of the ligand structure; the PDB code (linked to the summary information for the PDB entry), the affinity value in the originally reported form as well as converted to the negative base-10 logarithm [i.e. as  $pK_i$ ,  $pK_d$ ,  $pIC_{50}$  (relative to the standard concentration of 1 mol/l)]; the pH value of the measurement, the first author of the publication (with a link to the PubMed entry); and the year of the publication. Tables reporting search results can be saved as 'csv' file, which is an ASCII file with semicolon-separated columns and one affinity entry per line.

## DISCUSSION

AffinDB has been designed to provide fast and easy access to affinity data. The popular MySQL backend was chosen as database machine, since MySQL offers a speed-optimized SQL engine. Using the scripting language PHP, special care

AffinDB V1.0.1 - Affinity Database For Protein-Ligand Complexes 3 user(s) online

**AffinDB**  
Affinity database for protein-ligand complexes



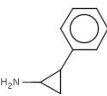

Home Search Browse Login Contact Help

Marburg, Friday, September 09th, 2005

Home Search Browse Login Register Test DrugScore Online

PDB Code

Affinity data: 746  
Covered PDBs: 474  
Latest Entries

Found affinity information (7 entries)						
Ligand	MW	PDB	Affinity	pH	Author	Year
 MW: 121.18	11n1	Ki: 11.00 mM pKi: 1.96	8.0	Kurinov I.V. et al.	1994	<a href="#">Details</a>
 MW: 150.25	11n1	Ki: 20.00 mM pKi: 1.70	8.0	Kurinov I.V. et al.	1994	<a href="#">Details</a>
 MW: 134.2	11n1	Ki: 13.30 mM pKi: 1.88	8.0	Kurinov I.V. et al.	1994	<a href="#">Details</a>
 MW: 114.21	11ng	Ki: 1.17 mM pKi: 2.93	8.0	Kurinov I.V. et al.	1994	<a href="#">Details</a>

**Figure 3.** Tabular report, showing part of the search results produced by a query for affinity data published by a specific author ('Kurinov') in a primary reference.

was taken to generate a clearly structured layout which enables fast and easy navigation. Since all the data can be accessed and retrieved directly via the web browser, the user does not have to install any special software to work with AffinDB.

Using current PC hardware (CPU: Athlon XP 2400+), AffinDB executes search queries in <0.1 s, fairly independent of the complexity of the query. The representation of the tabular report including the 2D molecule drawings needs between 0.1 and 5 s, depending on the number of hits (5 s if all entries are retrieved). The representation of a PDB entry takes up to 0.2 s, depending on the number of affinity data available for that entry. These values reflect only a server-side benchmarking. Obviously, the real speed also depends on the client-side hardware, the Internet connection and the browser.

Data collection for AffinDB is a very time-consuming process which can hardly be automated since scientifically educated readers are required to critically extract the relevant data from the scientific literature. In contrast to other databases (cf. below), we decided to include all affinity data found during literature research for a given PDB complex. Multiple affinity entries may, thus, be available for certain structures. These may reflect measurements with different methods or under different experimental conditions (e.g. PDB 1FLR; cf. Figure 1), or it may be due to additional reports from secondary references, which allows the user to trace back in which context the corresponding complex and its affinity have already been used. Only purely redundant data are not included (e.g. if the value is reported in the same paper as  $K_d$  and  $\Delta G$  derived thereof).

The current coverage of more than 470 PDB structures derives from a priority selection made upon constructing the database. The initial basis was formed by compilations of affinity values from secondary references concerning empirical scoring functions. Owing to discrepancies among some of the values and to obtain more detailed information, primary references were also retrieved for part of this initial set. Subsequently, the database was augmented by seeking affinity data for PDB complexes of different datasets, such as a docking test set of validated structures (27) or datasets for certain target classes (e.g. carbonic anhydrases and trypsin-like serine proteases). Furthermore, published data from our own laboratory were also directly included.

AffinDB is a valuable resource for anyone interested in correlating structural data with binding energetics and complements other databases of similar subject, specifically the Ligand-Protein Database LPDB (28), the Protein-Ligand Database PLD (29), PDBbind (20,21), and—published shortly after first submission of this paper—Binding MOAD (22). LPDB is a compilation of 262 PDB complexes with affinity data. Since it also provides scoring values, docked ligand poses ('decoys') and ligand files setup for docking, LPDB is primarily intended to serve as a dataset for testing and developing docking and scoring methods. It does neither provide details nor references for the affinity values. The same is true for PLD, which contains 485 complexes and experimental binding energies for 344 of them. PLD can be searched by using a variety of single search criteria, but no combined search queries are possible. Similar to AffinDB it is freely accessible to anybody over the Internet, whereas PDBbind and Binding MOAD require a registration before granting academic users a free login account. The latter two databases offer by far the largest amount of affinity values, both covering well beyond 1700 complexes in their latest updates. Details about the affinity measurement and experimental conditions, however, are not included, which is an information provided by AffinDB for data retrieved from primary references. In summary, although there is certainly some overlap among the structure-affinity databases recently arisen from independent efforts, there are clear differences in focus, design and content, rendering each database on its own and in mutual combination an indispensable tool for the scientific community as long as affinities are not reported by the PDB and/or no common repository for biomolecular affinity data exists.

AffinDB encourages users to contribute data and submit references to papers with affinity data for PDB complexes. After registering for upload, an input form can be accessed which facilitates the submission of all relevant data in a clear format. Data submitted by users do not directly enter the database, but must first undergo revision by the database curators. This should ensure high fidelity of the affinity data collected from literature and reported by AffinDB.

## ACKNOWLEDGEMENTS

The help of the following project co-workers and students of pharmacy in collecting and processing literature data is gratefully acknowledged: Kathrin Austrup, Petra Cordes, Julia Engel, Silvia Funke, Steffen Hoeffft, Stefanie Meseth, Birte Roessing, Sebastian Vollmer. The database has been awarded with the first prize of the web award 2005 of the Molecular

Graphics and Modelling Society, German Section (May 2005, Erlangen, Germany). Funding to pay the Open Access publication charges for this article was provided by the Scoring Function Consortium.

*Conflict of interest statement.* None declared.

## REFERENCES

- Klebe, G. and Bohm, H.J. (1996) What can we learn from molecular recognition in protein–ligand complexes for the design of new drugs? *Angew Chem. Int. Ed. Engl.*, **35**, 2588–2614.
- Babine, R. and Bender, S. (1997) Molecular recognition of protein–ligand complexes: applications to drug design. *Chem. Rev.*, **97**, 1359–1472.
- Mueller, M.M., Sperl, S., Sturzebecher, J., Bode, W. and Moroder, L. (2002) (*R*)-3-Amidinophenylalanine-derived inhibitors of factor Xa with a novel active-site binding mode. *Biol. Chem.*, **383**, 1185–1191.
- Lange, U.E.W., Baucke, D., Hornberger, W., Mack, H., Seitz, W. and Hoffken, H.W. (2003) D-Phe-Pro-Arg type thrombin inhibitors: unexpected selectivity by modification of the P1 moiety. *Bioorg. Med. Chem. Lett.*, **13**, 2029–2033.
- Brenk, R., Naerum, L., Gradler, U., Gerber, H., Garcia, G.A., Reuter, K., Stubbs, M.T. and Klebe, G. (2003) Virtual screening for submicromolar leads of tRNA-guanine transglycosylase based on a new unexpected binding mode detected by crystal structure analysis. *J. Med. Chem.*, **46**, 1133–1143.
- Specker, E., Bottcher, J., Lilie, H., Heine, A., Schoop, A., Muller, G., Griebenow, N. and Klebe, G. (2005) An old target revisited: two new privileged skeletons and an unexpected binding mode for HIV-protease inhibitors. *Angew Chem. Int. Ed. Engl.*, **44**, 3140–3144.
- Wang, R., Liu, L., Lai, L. and Tang, Y. (1988) SCORE: a new empirical method for estimating the binding affinity of a protein–ligand complex. *J. Mol. Model.*, **4**, 379–394.
- Wang, R., Lai, L. and Wang, S. (2002) Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput. Aided. Mol. Des.*, **16**, 11–26.
- Bohm, H.J. (1994) The development of a simple empirical scoring function to estimate the binding constant for a protein–ligand complex of known three-dimensional structure. *J. Comput. Aided. Mol. Des.*, **8**, 243–256.
- Bohm, H.J. (1998) Prediction of binding constants of protein ligands: a fast method for the prioritization of hits obtained from *de novo* design or 3D database search programs. *J. Comput. Aided. Mol. Des.*, **12**, 309–323.
- Head, R.D., Smythe, M.L., Oprea, T.I., Waller, C.L., Green, S.M. and Marshall, G.R. (1996) VALIDATE: a new method for the receptor-based prediction of binding affinities of novel ligands. *J. Am. Chem. Soc.*, **118**, 3959–3969.
- Eldridge, M.D., Murray, C.W., Auton, T.R., Paolini, G.V. and Mee, R.P. (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to estimate the binding affinity of ligands in receptor complexes. *J. Comput. Aided. Mol. Des.*, **11**, 425–445.
- Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N. and Bourne, P.E. (2000) The Protein Data Bank. *Nucleic Acids Res.*, **28**, 235–242.
- Gunther, J., Bergner, A., Hendlich, M. and Klebe, G. (2003) Utilising structural knowledge in drug design strategies: applications using Relibase. *J. Mol. Biol.*, **326**, 621–636.
- Hendlich, M., Bergner, A., Gunther, J. and Klebe, G. (2003) Relibase: design and development of a database for comprehensive analysis of protein–ligand interactions. *J. Mol. Biol.*, **326**, 607–620.
- Laskowski, R.A., Hutchinson, E.G., Michie, A.D., Wallace, A.C., Jones, M.L. and Thornton, J.M. (1997) PDBsum: a Web-based database of summaries and analyses of all PDB structures. *Trends Biochem. Sci.*, **22**, 488–490.
- Laskowski, R.A., Chistyakov, V.V. and Thornton, J.M. (2005) PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids. *Nucleic Acids Res.*, **33**, D266–D268.
- Chen, X., Liu, M. and Gilson, M.K. (2001) BindingDB: a web-accessible molecular recognition database. *Comb. Chem. High Throughput Screen.*, **4**, 719–725.
- Zhang, J., Aizawa, M., Amari, S., Iwasawa, Y., Nakano, T. and Nakata, K. (2004) Development of KiBank, a database supporting structure-based drug design. *Comput. Biol. Chem.*, **28**, 401–407.
- Wang, R., Fang, X., Lu, Y., Yang, C.Y. and Wang, S. (2005) The PDBbind Database: Methodologies and Updates. *J. Med. Chem.*, **48**, 4111–4119.
- Wang, R., Fang, X., Lu, Y. and Wang, S. (2004) The PDBbind database: collection of binding affinities for protein–ligand complexes with known three-dimensional structures. *J. Med. Chem.*, **47**, 2977–2980.
- Hu, L., Benson, M.L., Smith, R.D., Lerner, M.G. and Carlson, H.A. (2005) Binding MOAD (mother of all databases). *Proteins*, **60**, 333–340.
- Boutselakis, H., Dimitropoulos, D., Fillon, J., Golovin, A., Henrick, K., Hussain, A., Ionides, J., John, M., Keller, P.A., Krissinel, E. *et al.* (2003) E-MSD: the European Bioinformatics Institute Macromolecular Structure Database. *Nucleic Acids Res.*, **31**, 458–462.
- Murzin, A.G., Brenner, S.E., Hubbard, T. and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.*, **247**, 536–540.
- Weininger, D. (1988) SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *J. Chem. Inf. Comput. Sci.*, **28**, 31–36.
- Gasteiger, J., Rudolph, C. and Sadowski, J. (1990) Automatic generation of 3D-atomic coordinates for organic molecules. *Tetrahedron Comput. Meth.*, **3**, 537–547.
- Nissink, J.W.M., Murray, C., Hartshorn, M., Verdonk, M.L., Cole, J.C. and Taylor, R. (2002) A new test set for validating predictions of protein–ligand interaction. *Proteins*, **49**, 457–471.
- Roche, O., Kiyama, R. and Brooks, C.L.III (2001) Ligand-protein database: linking protein–ligand complex structures to binding data. *J. Med. Chem.*, **44**, 3592–3598.
- Puvanendrapillai, D. and Mitchell, J.B.O. (2003) L/D protein ligand database (PLD): additional understanding of the nature and specificity of protein–ligand complexes. *Bioinformatics*, **19**, 1856–1857.