Supplementary Information: Structure of the BAD

The optimum structure of a database will largely depend on its target users, its purpose and magnitude. The original BAD database (in 2002, 200+ cases) was created as MS-Excel file, and later versions (2004-2006, 400+ cases) were build using SQL.

To build a reliable database with relational dependence the data was transferred to an MS-Access file and divided into nine tables in line with the adsorption descriptors (see further). Each table contains the unique index for each row (unique or primary key). The relations between tables are mediated by foreign keys. Thanks to this approach the redundancy in the database is avoided, and also maintaining the coherence of the data. Moreover, extracting the relevant information form the whole database can be performed with ease because of possibility of accessing only those tables under question.

The schematic diagram depicting the architecture of BAD database i.e. collection of tables used and the relation between them, is shown in the Figure 1.
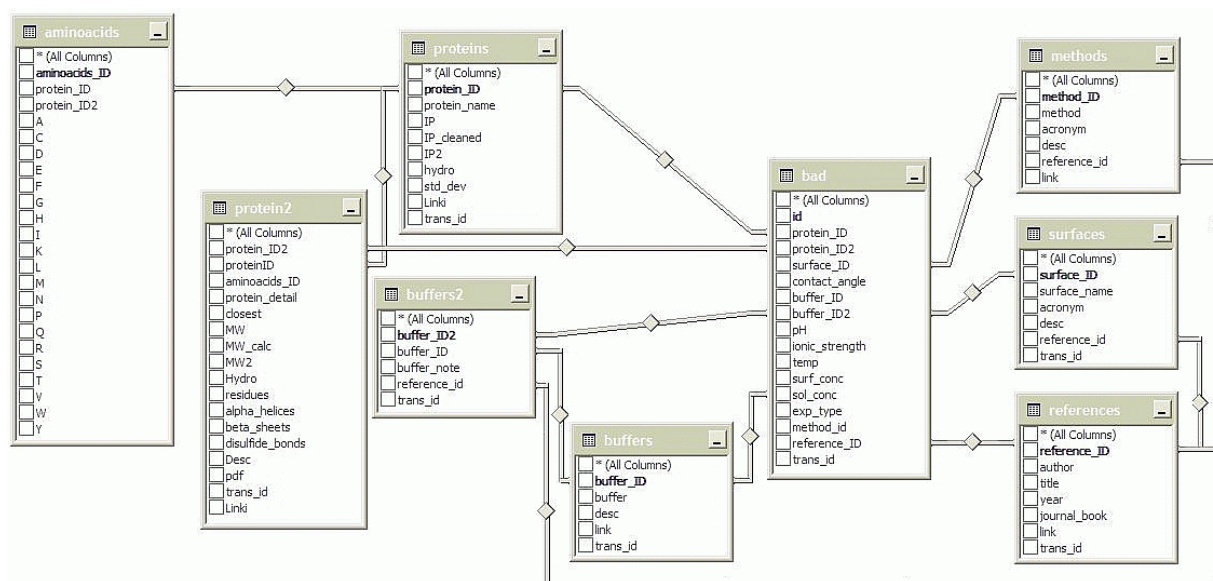
Figure 1. Schematic diagram of the BAD database architecture.

In general, the relational database is defined as a collection of relations (tables) and represented by a set of ordered lists having the same attributes that is table with data organized in rows and columns. One of the main principles of relations database model states that each row in the table has to be unique. Uniqueness for a table is guaranteed by a primary key, which is usually represented by a column that contains unique values for a particular table. In the BAD, the primary keys of all nine tables were created by adding additional column called 'ID' to avoid further ambiguities. In addition to the primary keys, we set several foreign keys, which are used to model relationships in the real world. Both keys have to share a common meaning and draw their values from the same domain. Three different types of relations can be established: "one-to–one", "one-to-many" and "many-to–many". A properly designed database needs to undergo a

normalization process to achieve the optimal structure through first to fifth normal form rules. The relational database is created, updated and administrated by the Relational DataBase Management System (RDBMS) and accessed through the Structural Query Language (SQL). This programming language for getting information from and updating a database is making use of queries. Queries take the form of command language that enables to select, update, insert, delete, fin out the location of the data and so forth.

In the BAD the tables and the fields were selected based on examining the basic elements of the protein adsorption experiment, i.e. proteins, surfaces, buffer solutions, protein concentration in solution and on the surface, measurement method, as well as general information regarding experimental conditions. The database was divided into the following structural tables:

- *BAD* – main table containing all foreign keys and the parameters that are unique for each record;
- *Proteins* and *Proteins2* – tables containing protein descriptors and the additional information about the respective protein,
- *Amino Acids* – table containing amino acid proteins composition,
- *Buffers* and *Buffers2* – tables containing buffer descriptors and the information about buffer solution composition,
- *Surfaces* – table containing surface descriptors and the information about surfaces itself,
- *Methods* – table with measurement methods used in the experiments,
- *References* – table of references.

Technical Specification- technology
The BAD Web application was built using Visual Web Developer 2005 environment, a free Microsoft tool for creating Web sites in a variety of configurations. We also used ASP.NET and ADO.NET technology.
Active Server Pages (ASP.NET) is a new server side scripting technology for developing fully dynamic Web pages. ASP.NET uses runtime Common Language Runtime Environment (CLR) environment and functional .NET Framework that introduces the aspects of 'Cross-Language Development'. The main feature of ASP.NET environment is its high speed and simplicity in building sophisticated web services. Possibility of separation of application code (Visual Basic programming language) from the presentation layer that is graphical interface of webpage is without doubts the biggest advantage of this technology.
ADO.NET is regarded as a development of the data access technology provided by the Microsoft® *ActiveX® Data Objects* (ADO). It is a part of the base class library that is included with the Microsoft. NET Framework. ADO.NET technology provides large functionality. It is also commonly used by programmers to access and modify data stored in relational data sources like for example MS-access BAD database. Due to ADO.NET database access technology many programmatic tasks can be solved quicker and more effective.

Technical Specification- web application
The BAD Web application was created using Microsoft ASP.NET technology for building dynamic Web pages. The remote access from BAD web application to BAD relational database is possible through the standard Web browser. Visual Basic language together with the .NET framework (VB.NET) was used to write the BAD web project. The whole application is composed of several *.aspx files with the names corresponding to the adequate protein

adsorption components. Each web page contains a dynamically loaded table that presents the results of query sent to the MS-access database.

For the present moment the Web database is available in a simple lookup table form additionally with a *Search* form that enables to search the database over ranges and look for exact matches. In the near future the BAD will also have the possibility for the user to input their own data, upload the database by means of *New Data* form and attach reference publications directly to the ftp server. In order to reach this stage, an additional table called *transactions* has been already created (see database scheme Figure 1). The references to each of the table will be possible through index *trans_id*, which will store the information about the user data deposition and will track the database for uploading. At the current stage of the design the submission form enables users to insert a new data, which are sent before uploading via email server services to the administrator for the data evaluation. Routinely, the database administrator will check the database to ensure the accuracy of the information given and will decide about the update action.

The BAD database is available free at http://www.bionanoeng.com/bad/ following the link for Biomolecular Adsorption Database, or directly at http://dbweb.liv.ac.uk/bad/index.htm.

Data Organization and Variables
The BAD web based database was created using Microsoft Visual Web Developer 2005 tool and was written using Visual Basic. Currently, the database contains 767 records that can be sorted within different fields. The basic information concerning the protein adsorption is presented in the main BAD table within the following fields:

- Protein
- Surface Acronym
- Contact Angle [˚deg]
- Surface Tension [mJ/m²]
- Buffer
- pH
- Ionic Strength [M]
- Temperature [˚C]
- Surface Protein Concentration [mg/m^2]
- Solution Protein Concentration [mg/ml]
- Experiment Type
- Measurement Method
- Reference
- Notes

Biomolecular Adsorption Database

Protein	Surface Acronym	Contact Angle [°]	Surface Tension [mJ/m²]	Buffer	pH	Ionic Strength [M]	Temperature [°C]	Surface Protein Concentration [mg/m²]	Solution Protein Concentration [mg/ml]	Experiment Type	Measurement Method	Reference	Notes
Albumin	GLS	0	0	PB	7.4	0.128	25	7.1	1	isotherm	UV-Vis Spectroscopy	[33]	>>
Albumin	MetSiO2 0.008% DDS	30.5	1.51	Hanks	7	0.15	22	2	0.02	isotherm	Ellipsometry	[31]	>>
Albumin	Cab-O-Sil M5	2.5	0.0001	PB	7	0.19	22	1.43	0.95	isotherm	UV-Vis Spectroscopy	[24]	>>
Albumin	Cab-O-Sil M5	2.5	0.0001	PB	7	0.19	22	1.31	0.79	isotherm	UV-Vis Spectroscopy	[24]	>>
Albumin	Cab-O-Sil M5	2.5	0.0001	PB	7	0.19	22	1.25	0.57	isotherm	UV-Vis Spectroscopy	[24]	>>
Albumin	Cab-O-Sil M5	2.5	0.0001	PB	7	0.19	22	1.05	0.4	isotherm	UV-Vis Spectroscopy	[24]	>>
Albumin	Cab-O-Sil M5	2.5	0.0001	PB	7	0.19	22	0.76	0.25	isotherm	UV-Vis Spectroscopy	[24]	>>
Albumin	Cab-O-Sil M5	2.5	0.0001	PB	7	0.19	22	0.46	0.13	isotherm	UV-Vis Spectroscopy	[24]	>>
Albumin	Cab-O-Sil M5	2.5	0.0001	PB	7	0.19	22	1.45	1.42	isotherm	UV-Vis Spectroscopy	[24]	>>
Albumin	Ti	82.8	24.6	PB	7.4	0.265	37	11	1	single measur.	Bicinchoninic Acid Assay	[8]	>>
Albumin	Cab-O-Sil M5	2.5	0.0001	PB	7	0.19	22	1.44	1.55	isotherm	UV-Vis Spectroscopy	[28]	>>
Albumin	GLS	0	0	PB	7.4	0.128	25	6.6	0.8	isotherm	UV-Vis Spectroscopy	[33]	>>
Albumin	GLS	0	0	PB	7.4	0.128	25	4.4	0.45	isotherm	UV-Vis Spectroscopy	[33]	>>
Albumin	GLS	0	0	PB	7.4	0.128	25	3	0.2	isotherm	UV-Vis Spectroscopy	[33]	>>
Albumin	PEUN			PB	7.4	0.128	25	6.5	0.95	isotherm	UV-Vis Spectroscopy	[33]	>>

1 2 3 4 5 6 7 8 9 10 ...

© University of Liverpool, 2008

Figure 2. Biomolecular Adsorption Database Interface – main BAD table.
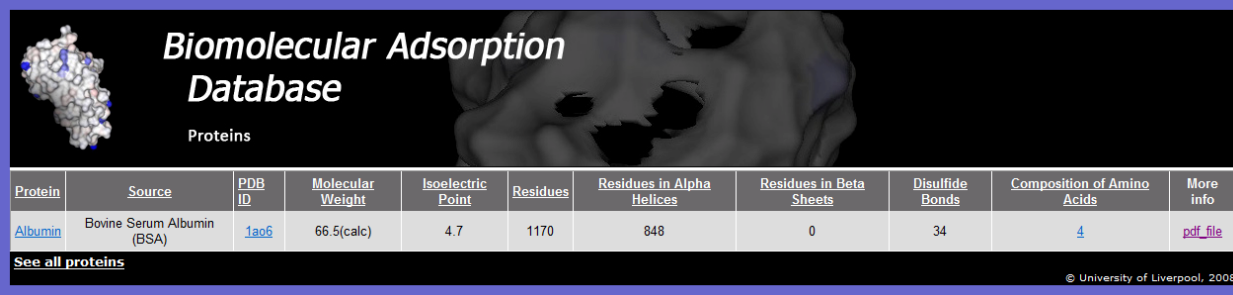
In addition to the main BAD table, there are seven detailed tables with descriptors that can be accessed either by selecting the column name (opens the whole table) or through selecting a linked entry under interest (underlined field names above) in main BAD table. Following the second option, each table opens in the new window with the selected record. In order to see all records in the table one should click link below it 'See all [field variable]', e.g., protein, contact angle, etc.

1. **Protein Table** – accessible by selecting linked protein column in the main BAD table. This gives further information regarding the protein itself, protein sequence details, short description, links to PDB file and other useful information, etc.

Protein descriptors:

- Protein – protein name
- Source- information about protein source
- PDB ID – linked Protein Data Bank Identifier
- Molecular Weight [Daltons] - abbreviation calc means that the molecular weight was calculated form the amino acids composition given PDB file by means of Matlab software. Otherwise molecular weight was taken from the literature
- IsoElectric Point- abbreviation cleaned means IsoElectric Point of the cleaned structure

- Protein Surface Tension [mJ/m²] – calculated protein surface tension
- Residues – number of residues of the protein
- Residues in Alpha Helices – number of residues in the alpha helices structures
- Residues in Beta Sheets – number of residues in the beta sheets structures
- Disulfide Bonds- number of disulfide bonds of the protein
- Composition of Amino Acids - index that leads to Amino Acids Composition table
- More info – PDF file containing details about selected protein

Biomolecular Adsorption Database
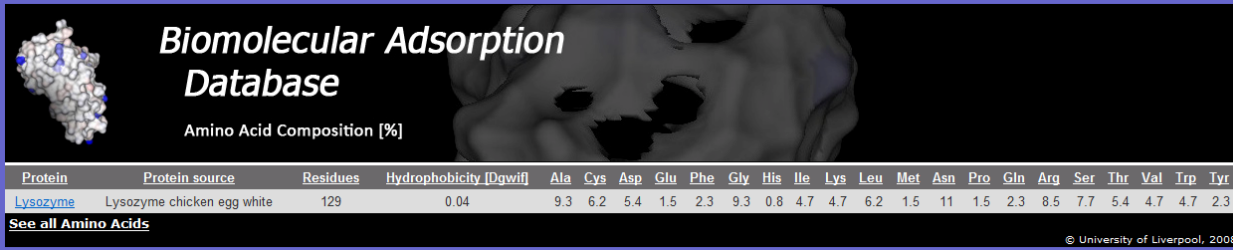
Proteins

Protein	Source	PDB ID	Molecular Weight	Isoelectric Point	Residues	Residues in Alpha Helices	Residues in Beta Sheets	Disulfide Bonds	Composition of Amino Acids	More info
Albumin	Bovine Serum Albumin (BSA)	1ao6	66.5(calc)	4.7	1170	848	0	34	4	pdf_file

See all proteins

© University of Liverpool, 2008

Figure 3. Biomolecular Adsorption Database Interface – Proteins table with one protein record.

2.   **Amino Acids Composition Table** – accessible by selecting linked amino acid index in the protein table. This gives information about percentage of particular amino acid in the protein sequence and protein hydrophobicity [Dgwif].

Biomolecular Adsorption Database

Amino Acid Composition [%]

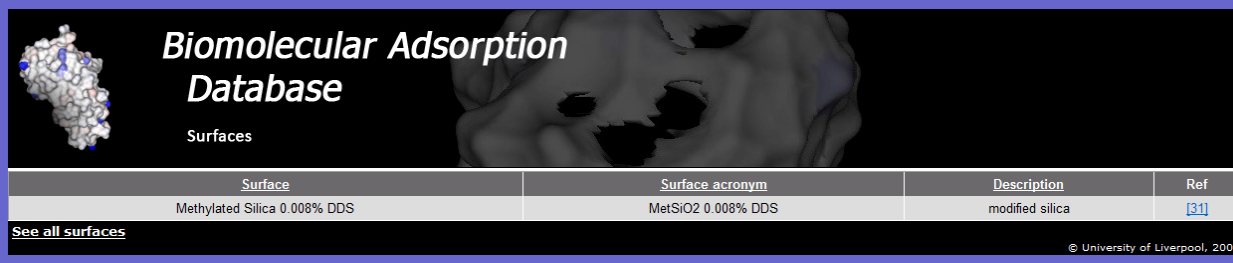| Protein	Protein source	Residues	Hydrophobicity [Dgwif]	Ala	Cys	Asp	Glu	Phe	Gly	His	Ile	Lys	Leu	Met	Asn	Pro	Gln	Arg	Ser	Thr	Val	Trp	Tyr
Lysozyme	Lysozyme chicken egg white	129	0.04	9.3	6.2	5.4	1.5	2.3	9.3	0.8	4.7	4.7	6.2	1.5	11	1.5	2.3	8.5	7.7	5.4	4.7	4.7	2.3

See all Amino Acids

© University of Liverpool, 2008

Figure 4. Biomolecular Adsorption Database Interface – Amino Acids Composition table with Lysozyme record.

3.   **Surfaces Table** - accessible by selecting linked surface acronym in the main BAD table.
Surface descriptors:
- Surface – surface full name
- Surface Acronym – surface abbreviation
- Description
- Ref – linked reference id that leads to publication/s containing detailed information about surface under interest

Biomolecular Adsorption Database

Surfaces

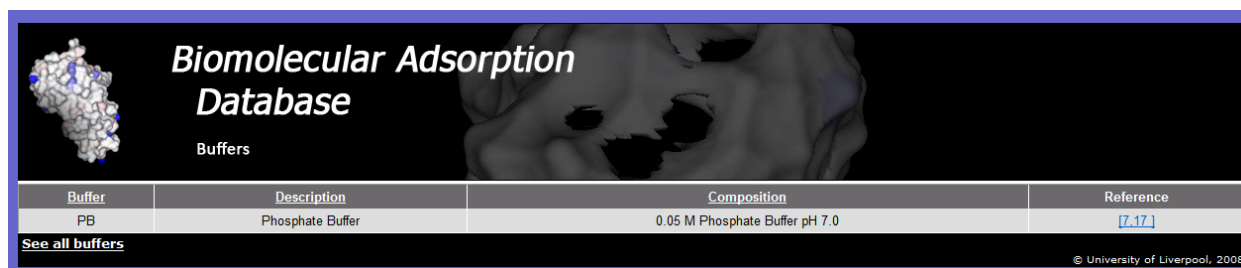| Surface	Surface acronym	Description	Ref
Methylated Silica 0.008% DDS	MetSiO2 0.008% DDS	modified silica	[31]

See all surfaces

© University of Liverpool, 2008

Figure 5. Biomolecular Adsorption Database Interface – Surfaces table with one record of PEUN surface acronym.

4. **Buffers Table** - accessible by selecting linked buffer name in the main BAD table.
Buffer descriptors:
- Buffer
- Description – full buffer name
- Composition – composition of buffer solution
- Ref – linked reference id that leads to publication/s containing detailed information about buffer solution under interest
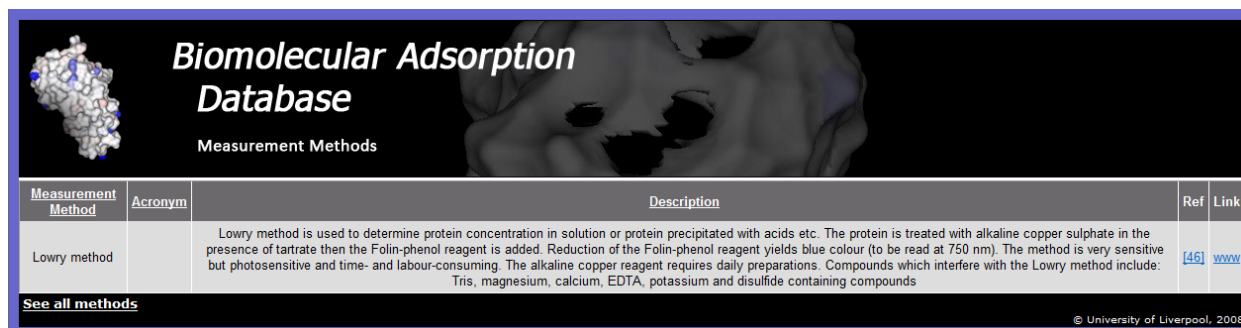

Figure 6. Biomolecular Adsorption Database Interface – Buffers table with one record of PB buffer.

5. **Methods Table** - accessible by selecting linked measurement method in the main BAD table.
Methods table fields:
- Measurement Method
- Acronym
- Description
- Ref – linked reference id that leads to publication/s containing detailed information about measurement method under interest
- Link – link to external source

Figure 7. Biomolecular Adsorption Database Interface – Methods table with one record for Lowry method.

6. **References Table** - accessible by selecting linked reference id in the main BAD table.
References table fields:
- ID – reference identification

- Author/s
- Title
- Year
- Source
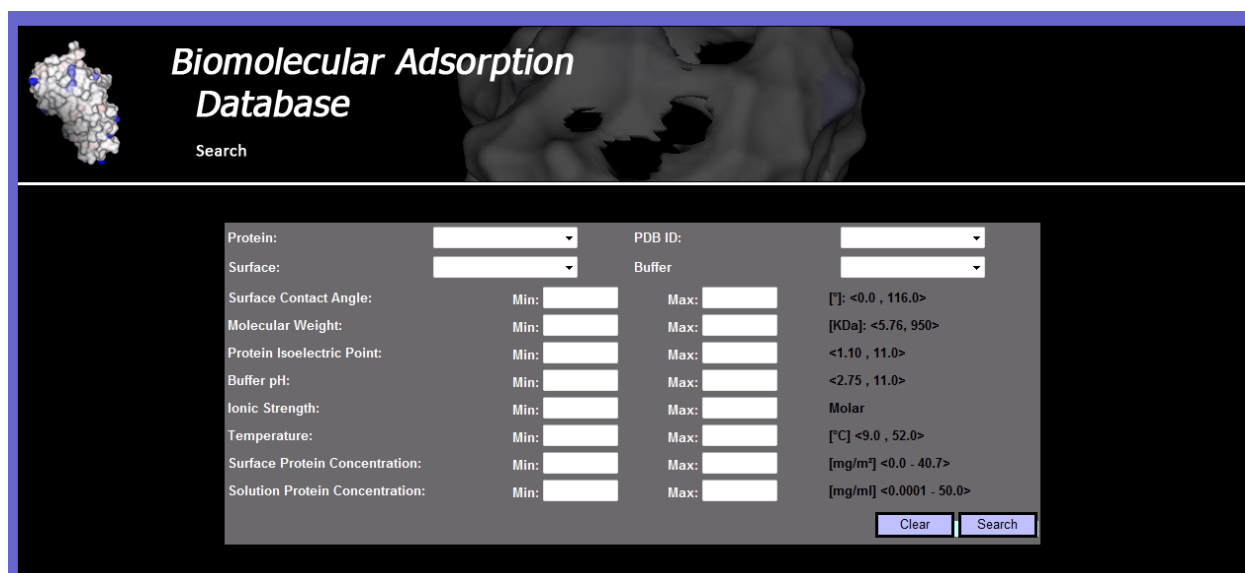- Link

Figure 8. Biomolecular Adsorption Database Interface – References table with three records.

Search and sorting capabilities

The data sorting capability can be stated as an important BAD database feature. In each Data Grids (tables) the column headings contain links that, when clicked, cause the table to use that column as the sort key. The columns, both that contain numeric and nonnumeric entries, can be sorted in ascending or descending order. If the entry contains a null value, then it will appear as a first, in front of all other not null entries. Note that this feature is not applied for all of the columns in the main BAD table, i.e., selection of protein column opens protein table but does not sort proteins neither in ascending nor in descending order.

Except of looking up and sorting possibilities of the data stored in BAD, the information can also be searchable between specified ranges. The user can retrieve amount of protein adsorbed and other relevant data by querying the database using the search form "SEARCH BAD" (Figure 9). The user can specify the name of protein, surface; buffer or PDB id by directly typing term into text box or choosing it form the drop-down lists provided. When a drop-down list is activated it displays a list of values contained in the database form which user may select one. To avoid mistakes while typing, it is recommended to use drop-down lists if they are provided. For the experimental parameters (Contact Angle, pH, IsoElectric Point, Ionic Strength, Temperature, Surface Protein Concentration, Solution Protein Concentration, Molecular Weight) two text boxes are provided in order to specify the minimum and maximum value for the search. The search can be also performed using combination of parameters. Additionally, the search form gives information about the units and ranges for each parameter according to the data already contained in the BAD database. Since the information given in the BAD database is displayed by means of several tables, the result of the search is not always given immediately i.e. shown in the main table under search form. In order, for example, to view reference details the user has to enter references table through reference id in the main BAD table or to view molecular weight the user has to enter proteins table through protein name. Inside the tables of search results some of the measuring techniques are designated by acronyms. Similarly, surface names and buffer used are viewed by acronyms in order to save display space. There are two buttons provided below the search form: Search and Clear. Selecting search button the search results according to specified by user values are displayed. Clear button clears the search form. Figure 9 presents the layout of the 'Search BAD' form.

Figure 9. Biomolecular Adsorption Database 'Search BAD' form.

Additional Materials

The BAD database contains additional materials that deliver detailed information about protein adsorption. The PDB id column in the protein table provides direct links to the corresponding Protein Databank Page for a chosen protein, from which structural, geometry and sequence details can be viewed and downloaded. The last column in protein table specifies a link to the Sigma PDF file containing product details for a chosen protein name. For each reference (References table) link to Science Direct page is provided, from which article abstract or full article and further references can be obtained. In each of particular BAD table there is a special column dedicated to contain link to additional materials, although some information has not been delivered yet.