CSDBase: an interactive database for cold shock domain-containing proteins and the bacterial cold shock response

Michael H. W. Weber, Ingo Fricke, Niclas Doll and Mohamed A. Marahiel*

Philippus-Universität Marburg, FB Chemie, Hans-Meerwein-Straße, D-35032 Marburg, Germany

Received September 20, 2001; Accepted October 10, 2001

ABSTRACT

CSDBase (http://www.chemie.uni-marburg.de/~csdbase/) is an interactive Internet-embedded research platform providing detailed information on proteins containing the cold shock domain (CSD). It consists of two separated database cores, one dedicated to CSD protein information, and one to provide a powerful resource to relevant literature with emphasis on the bacterial cold shock response. In addition to detailed protein information and useful cross links to other web sites, CSDBase contains computer-generated CSD structure models for most CSD-containing protein sequences available at NCBI non-redundant protein database at the time of CSDBase establishment. These models were calculated on the basis of known crystal and/or NMR structures using SWISS-MODEL and can be downloaded as PDB structure coordinate files for viewing and for manipulation with other software tools. CSDBase will be regularly updated and is organized in a compact form providing user friendly interfaces to both database cores which allow for easy data retrieval.

INTRODUCTION

Proteins interacting with the biological information molecules DNA and RNA play important cellular roles in all organisms. One widespread superfamily of proteins implicated in such function(s) contains the cold shock domain (CSD). It consists of two separated database cores, one dedicated to CSD protein information, and one to provide a powerful resource to relevant literature with emphasis on the bacterial cold shock response. In addition to detailed protein information and useful cross links to other web sites, CSDBase contains computer-generated CSD structure models for most CSD-containing protein sequences available at NCBI non-redundant protein database at the time of CSDBase establishment. These models were calculated on the basis of known crystal and/or NMR structures using SWISS-MODEL and can be downloaded as PDB structure coordinate files for viewing and for manipulation with other software tools. CSDBase will be regularly updated and is organized in a compact form providing user friendly interfaces to both database cores which allow for easy data retrieval.
cold shock response (http://www.chemie.uni-marburg.de/~csdbase/).

DATABASE CONTENT, ACCESS AND USAGE

CSD protein sequences were acquired by performing a BLASTP search at NCBI non-redundant database using cold shock protein CspB from *B.subtilis* as template (http://www.ncbi.nlm.nih.gov/blast/; 25,26). Except for the number of descriptions and the number of displayed alignments which were both set to a non-limiting value, BLAST default settings were used. On the basis of experimentally determined CSD protein structures (5–8), each of the protein sequences retrieved and thereby detected to have homologies to CSD containing CspB, was subjected to comparative protein modeling using SWISS-MODEL first approach mode (http://www.expasy.ch/swissmod/SWISS-MODEL.html; 27,28). The constructed homology-models were automatically analyzed by the WHAT IF verification program (29,30) and were stored in CSDBase as multi-layered atom coordinate files (PDB format). Each layer holds the atom coordinates of a separate model where the first layer represents the model structure generated for the protein of interest while the other layers contain the coordinates of the experimental structure(s) used as template(s) for the calculations as taken from PDB (http://www.rcsb.org/pdb/). To demonstrate the quality of the generated protein models, Figure 1 shows stereo images of a typical raw model structure generated for CspB from *Bacillus caldolyticus* in comparison to its crystal structure that was experimentally determined later (r.m.s.d. = 0.87 Å). Although the theoretical structure model shown in Figure 1 fits well to the experimentally determined, the reader should be reminded, that such homology models should be regarded as raw models which often require additional inspection and further refinement.

For those sequences which allowed for generation of a protein model, all relevant information was extracted via NCBI from the original database entries and was stored together with the downloadable protein model atomic coordinate file in CSDBase. In addition to the protein sequences, accession numbers, protein names, gene names and a brief description of the proteins, the modeled regions of the protein sequences were indicated, the number of amino acids for each protein was determined and its pI and mass were calculated using the proteomic tools available at ExPASy server (http://ca.expasy.org/tools/pi_tool.html; 31). One of the interesting results obtained during this data acquisition was that, in contrast to what is sometimes believed, bacterial CSPs are not uniformly acidic but rather stretch over a broad pI range and

Figure 1. Comparison of the crystal structure of cold shock protein CspB from *B.caldolyticus* [(A) PDB accession no. 1C9O; A-chain (4)] and its computer-generated model structure counterpart as stored in CSDBase [(B) first layer of CSDBase file name CspB_baccl (this work)]. Ribbon model stereo pairs were generated with Swiss-PdbViewer version 3.7b2 (34) and images were rendered using POV-Ray version 3.1g. Superimposition (C) of CspB crystal structure (yellow) on CspB computer model (cyan) was performed using the “iterative magic fit” function of Swiss-PdbViewer (r.m.s.d. = 0.87 Å for the carbon α trace). Note that the presented computer model was generated with SWISS-MODEL (27,28) on the basis of known structures for CspB and CspA from *B.subtilis* and *E.coli*, respectively, at a time where the crystal structure of CspB from *B.caldolyticus* was not yet available.
can have a basic pI of as high as 10.28 (CspH from \textit{E.coli}). Moreover, apart from a N-terminal CSD, the genus \textit{Mycobacterium} harbors CSP sequences which contain an additional C-terminal domain of approximately 70 amino acids of unknown structure and function. Since Mycobacteria have some features in common with eukaryotes and utilize these as hosts, it might be possible that this CSD extension feature reflects further development of simple CSP proteins in the direction of higher organized Y-box proteins.

In addition to the CSP protein data and model structures described above, CSDBase provides a powerful, easy-to-use retrieval system for the relevant literature which was compiled using the PubMed interface at NCBI (http://www.ncbi.nlm.nih.gov/PMID/). CSDBase literature entries (as well as those for the protein data) can be searched by choosing from a range of pre-defined topics such as author, year of publication, journal name, etc., followed by entering a user-defined phrase (an asterisk will result in display of all entries currently available). In all cases, data output can be ordered according to different aspects. It should be noted that at this stage the content of the literature database core is restricted to bacterial CSPs and the bacterial cold shock response. We are currently working on adding the eukaryotic publications.

CSDBase can be accessed via the Internet at http://www.chemie.uni-marburg.de/~csdbase/ using Microsoft Internet Explorer version 5 or higher and is organized in a compact self-explanatory form. Note that with most other browsers, database functions remain non-operational. Further details concerning usage of CSDBase can be viewed directly by selecting the ‘Info’ submenu displayed in the title bar of every CSDBase page.

**FUTURE PERSPECTIVES**

Since the CSD-containing cold shock proteins are implicated to play a major role during the bacterial cold shock response (32,33), in an additional pilot project a third database core is currently under construction to incorporate detailed proteomic data including clickable 2D gels of cold shock stressed bacteria. Moreover, two novel CSP protein structures await release in the near future (PDB accession nos 1G6P and 1H95) which will allow us to further refine the model structures presently available in CSDBase. In this manner, it is our intention to continuously extend the capabilities of this initial version of CSDBase step-by-step to provide a powerful starting platform for researchers working on CSD containing proteins and the bacterial cold shock response.

**ACKNOWLEDGEMENTS**

We would like to thank Rolf Henrich for a quick introduction into Internet-embedded database organization. This work was supported by Sonderforschungsbereich 395 and Fonds der Chemischen Industrie.

**REFERENCES**


