286–288 *Nucleic Acids Research, 1999, Vol. 27, No. 1*

ProTherm: Thermodynamic Database for Proteins and Mutants

M. Michael Gromiha, Jianghong An, Hidetoshi Kono, Motohisa Oobatake[+], Hatsuho Uedaira and Akinori Sarai*

Tsukuba Life Science Center, The Institute of Physical and Chemical Research (RIKEN), 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

Received August 20, 1998; Revised and Accepted October 5, 1998

ABSTRACT

The first release of the Thermodynamic Database for Proteins and Mutants (ProTherm) contains more than 3300 data of several thermodynamic parameters for wild type and mutant proteins. Each entry includes numerical data for unfolding Gibbs free energy change, enthalpy change, heat capacity change, transition temperature, activity etc., which are important for understanding the mechanism of protein stability. ProTherm also includes structural information such as secondary structure and solvent accessibility of wild type residues, and experimental methods and other conditions. A WWW interface enables users to search data based on various conditions with different sorting options for outputs. Further, ProTherm is cross-linked with NCBI PUBMED literature database, Protein Mutant Database, Enzyme Code and Protein Data Bank structural database. Moreover, all the mutation sites associated with each PDB structure are automatically mapped and can be directly viewed through 3DinSight developed in our laboratory. The database is available at the URL, http://www.rtc.riken.go.jp/protherm.html

INTRODUCTION

Thermodynamic data for proteins are important for understanding the mechanism of protein stability. Nishikawa *et al.* (1) constructed a Protein Mutant Database (PMD), which covers natural and artificial mutants of proteins. Recently, Pfeil (2) collected a set of thermodynamic data on protein stability and folding. Further, mutation databases for specific objectives, such as p53 genes (3), androgen receptor genes (4), human collagen and PAX6 genes (5–7), VHL and WT1 genes (8,9) and haemophilia B, SV40 large tumor antigens (10,11) have been developed. On the other hand, there are no electronically accessible databases for thermodynamic data on protein mutant stability. In this work, we have developed a database, Thermodynamic Database for Proteins and Mutants (ProTherm), which includes several thermodynamic data (unfolding Gibbs free energy change, enthalpy change, heat capacity change, transition temperature, activity etc.), structural information (secondary structure, solvent accessibility etc.), measuring methods, experimental conditions and literature information. This database will help to understand the mechanism of protein stability. We have developed a WWW interface to facilitate searching the database and sorting outputs.

CONTENTS OF THE DATABASE

Each entry in the database is identified by a serial number and includes the following information.

Structural information. Protein name, Protein Data Bank (PDB) code (http://www.pdb.bnl.gov) for wild and mutant structures, wild and mutant residue names with residue number showing the nature of mutation, information on monomeric and oligomeric states, secondary structure, accessibility and number of transition states. The secondary structure of each mutant was assigned from the information available at PDB. The solvent accessible surface area of all the atoms and the residues were computed using the program ASC (12,13) as described in another article of ours (M.M.Gromiha, M.Oobatake, H.Kono, H.Uedaira and A.Sarai, *Protein Engng*., submitted). The three dimensional coordinates were taken from PDB (14; http://www.pdb.bnl.gov).

Thermodynamic data obtained from denaturant denaturation experiments. Unfolding Gibbs free energy change in the absence and presence of denaturant (ΔG^{H_2O}, ΔG), difference in unfolding Gibbs free energy changes in the absence and presence of denaturant ($\Delta\Delta G^{H_2O}$, $\Delta\Delta G$), temperature (T), midpoint of denaturant concentration (C_m) and slope of denaturation curve (m).

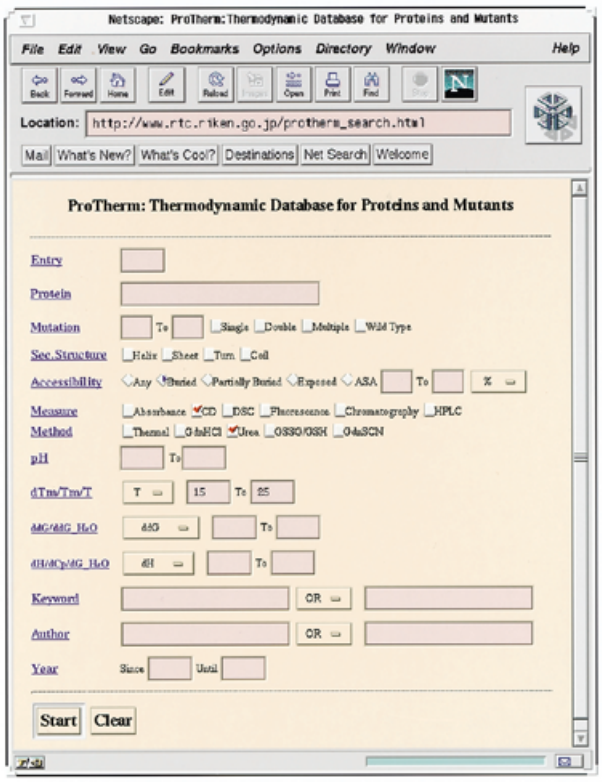
Thermodynamic data obtained from thermal denaturation experiments. Unfolding Gibbs free energy change (ΔG), difference in unfolding Gibbs free energy changes ($\Delta\Delta G$), transition temperature (T_m), transition temperature change (ΔT_m) enthalpy change (ΔH) and heat capacity change (ΔC_p).

Experimental methods and conditions. pH, measurements and method.

*To whom correspondence should be addressed. Tel: +81 298 36 9082; Fax: +81 298 36 9080; Email: sarai@rtc.riken.go.jp

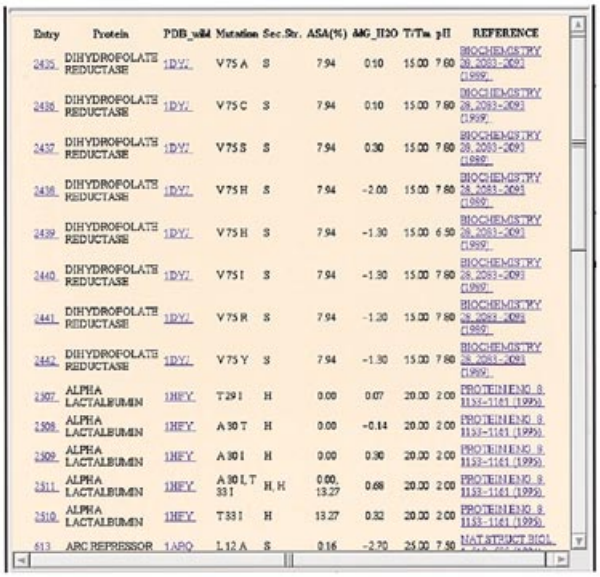
[+]Present address: Faculty of Science and Technology, Meijo University, 1-501 Shiogamaguchi, Tenpaku-ku, Nagoya 468-8502, Japan

a



b

c



Functional information. Enzyme activity, binding constants etc.

Literature information. Keywords, reference, authors and remarks.

DATABASE STATISTICS

At present ProTherm contains 3317 entries from 61 different proteins (of which three dimensional structures are known for 57 proteins) with 2434 single and 353 double mutations. In terms of secondary structures, 1241 mutations are in helical segments, 789 in strand, 242 in turn and 846 in coil region. According to solvent accessibility, 1317 mutations are at buried, 918 mutations are at partially buried and 826 are at exposed regions. A majority of data are obtained by circular dichroism, CD (1709), followed by fluorescence, Fl (1052) and differential scanning calorimetry, DSC (489). Further, 1628 data are obtained with thermal denaturation, and 1094 and 573 data from GdnHCl and Urea denaturation, respectively.

ACCESS TO ProTherm

ProTherm can be directly accessed through the World Wide Web at http://www.rtc.riken.go.jp/protherm.html . Various search options are available in the interface, as shown in Figure 1, and briefly explained below.

(i) Retrieving data for a particular protein.

(ii) Specifying the type of mutation as single, double, multiple or wild type. Further, specifying any wild type residue and/or mutant residue is possible.

(iii) Specifying secondary structures having mutations, helix (H), strand (S), turn (T) and coil (C).

(iv) Searching data based on solvent accessibility and solvent accessible surface area (ASA; in % or $Å^2$) of relevant residue. The mutations are classified into buried (ASA < 20%), partially buried (20% ≤ ASA ≤ 50%) and exposed (ASA > 50%).

(v) Extracting data for a particular measurement (CD, DSC, Fl etc.) and a specific method (Thermal, GdnHCl, Urea etc.).

(vi) Limiting data for a particular range of T, T_m, ΔT_m, ΔG, $\Delta\Delta G$, ΔG^{H_2O}, $\Delta\Delta G^{H_2O}$, ΔH, ΔC_p and pH.

(vii) Extracting data with authors, publication year and key words.

(viii) Specifying output format by selecting various output items and by sorting with publication year, wild type residue, mutant residue, residue number, secondary structure, accessibility, pH, T, T_m, ΔT_m, ΔH, ΔG, $\Delta\Delta G$, m, C_m, ΔG^{H_2O} and $\Delta\Delta G^{H_2O}$.

Detailed tutorials describing the usage of the present ProTherm are available at the home page. As an example, the necessary items to be filled or selected to search data for mutations in buried regions by urea denaturation and CD measurement at

Figure 1. An example of searching conditions, display and sorting options, and results of ProTherm. (**a**) Main menu for the search options of ProTherm. In this example, items, buried (accessibility), CD (measure), urea (method) are selected from the menu and T ($\Delta T_m/T_m/T$) is specified by filling the boxes for the values from 15 to 25°C. (**b**) Display and sorting options of ProTherm. In this example, entry, protein, PDB wild, mutation, secondary structure, ASA, $\Delta\Delta G^{H_2O}$, T, pH and reference are selected for the output. T and residue number are chosen for sorting the results in the order of priority. (**c**) A part of the results obtained from ProTherm.

temperatures between 15 and 25°C are shown in Figure 1a. In Figure 1b, we show the items to be selected for the output and sorting options. In the sorting procedure, the first item has the topmost priority. In this figure, entry, protein, PDB wild, mutation, secondary structure, ASA, $\Delta\Delta G^{H_2O}$, *T*, pH and reference are selected for the output. The selected outputs are sorted with temperature as the first priority and residue number as the second priority. The final results obtained from the search conditions (Fig. 1a) and sorting options of necessary items (Fig. 1b) are shown in Figure 1c.

LINKS TO OTHER DATABASES

Each entry in ProTherm is linked to Enzyme Code, EC (http://www.expasy.ch/sprot/enzyme.html) and PMD (1; ftp://ftp.nig.ac.jp/pub/db/mutant/), through which one can obtain functional information. It is also linked to the wild type and mutant three dimensional structures of proteins in PDB (http://www.pdb.bnl.gov), and all the mutation sites associated with each PDB structure are automatically mapped and viewed through 3DinSight (15; http://www.rtc.riken.go.jp/3DinSight.html), developed in our laboratory. The references for all data are connected to PUBMED literature database of NCBI (http://www.ncbi.nlm.nih.gov/Entrez/medline.html).

AVAILABILITY AND CITAITON OF ProTherm

The database is freely accessible at http://www.rtc.riken.go.jp/protherm.html . If you use ProTherm as a tool in your published research work, please cite this article, including the URL, http://www.rtc.riken.go.jp/protherm.html . Suggestions and other materials for inclusion in the database are welcome and should be sent to Akinori Sarai (Email: sarai@rtc.riken.go.jp).

ACKNOWLEDGEMENT

M.M.G. thanks RIKEN for providing financial support to develop the database.

REFERENCES

1 Nishikawa,K., Ishino,S., Takenaka,H., Norioka,N., Hirai,T., Yao,T. and Seto,Y. (1993) *Protein Engng.*, **7**, 733.
2 Pfeil,W. (1998) *Protein Stability and Folding: A Collection of Thermodynamic Data*. Springer, NY.
3 Beroud,C. and Soussi,T. (1998) *Nucleic Acids Res.*, **26**, 200–204.
4 Gottlieb,B., Lehvaslaiho,H., Beitel,L.K., Lumbroso,R., Pinsky,L. and Trifiro,M. (1998) *Nucleic Acids Res.*, **26**, 234–238.
5 Dalgleish,R. (1998) *Nucleic Acids Res.*, **26**, 253–255.
6 Brown,A., McKie,M., van Heyningen,V. and Prosser,J. (1998) *Nucleic Acids Res.*, **26**, 259–264.
7 Cooper,D.N., Ball,E.V. and Krawczak,M. (1998) *Nucleic Acids Res.*, **26**, 285–287.
8 Beroud,C., Joly,D., Gallou,C., Staroz,F., Orfanelli,M.T. and Junien,C. (1998) *Nucleic Acids Res.*, **26**, 256–258.
9 Jeanpierre,C., Beroud,C., Niaudet,P. and Junien,C. (1998) *Nucleic Acids Res.*, **26**, 271–274.
10 Giannelli,F., Green,P.M., Sommer,S.S., Poon,M.-C., Ludwig,M., Schwaab,R., Reitsma,P.H., Goossens,M., Yoshioka,A., Figueiredo,M.S. and Brownlee,G.G. (1998) *Nucleic Acids Res.*, **26**, 265–268.
11 Robinson,C.G. and Pipas,J.M. (1998) *Nucleic Acids Res.*, **26**, 295–296.
12 Eisenhaber,F. and Argos,P. (1993) *J. Comp. Chem.*, **14**, 1272–1280.
13 Eisenhaber,F., Lijnzaad,P., Argos,P., Sander,C. and Scharf,M. (1995) *J. Comp. Chem.*, **16**, 273–284.
14 Bernstein,F.C., Koetzle,T.F., Williams,G.J., Meyer,E.E.Jr., Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) *J. Mol. Biol.*, **112**, 535–542.
15 An,J., Nakama,T., Kubota,Y. and Sarai,A. (1998) *Bioinformatics*, **14**, 188–195.