

Recent Developments in ProTherm: Thermodynamic Database for Proteins and Mutants

K. Abdulla Bava¹

bava@rtc.riken.go.jp

Michael M. Gromiha²

michael-gromiha@aist.go.jp

Hatsuho Uedaira¹

uedaira@rtc.riken.go.jp

Koji Kitajima¹

kkitajim@rtc.riken.go.jp

Akinori Sarai¹

sarai@rtc.riken.go.jp

¹ RIKEN Tsukuba Institute, 3-1-1 Koyadai, Tsukuba, Ibaraki 305-0074, Japan

² Computational Biology Research Center, AIST, 2-41-6 Koto-ku, Tokyo 135-0064 Japan

Keywords: protein stability, protein mutants, database, free energy change, solvent accessibility

1 Introduction

Due to the rapid progress in genome analysis, complete genome sequences of many organisms become available. Now proteins are the next target for intensive study, as evidenced by the emerging fields such as structural genomics and proteome analysis. In order to elucidate molecular function of proteins, we need to have knowledge of their amino acid sequence and structure. The sequence and structural information, however, is not enough to infer the protein function, because protein is a microscopic entity and its behavior obeys the law of thermodynamics. Thus, the information of thermodynamic quantities of proteins and their interactions is as fundamental as sequence and structural information. Although the databases for sequence and structure are well established and available, databases for thermodynamic quantities on protein stability and interactions are few. Thermodynamic data for proteins are essential not only for understanding the mechanism of protein folding and stability, but also for designing stable mutants. The compilation of thermodynamic data along with the sequence, structural, functional and literature information would be a valuable resource for such studies [2, 4]. Thus, we have designed an electronically accessible database, ProTherm [3], including several thermodynamic parameters, along with sequence and structural information, experimental methods and conditions, and literature information. Further, we have provided a WWW interface to facilitate searching the database, sorting and visualizing the results. Currently ProTherm contains about 12,000 entries.

2 Content and Organization of ProTherm

ProTherm contains protein information such as protein name, source; mutational information such as amino-acid mutation(s), mutation position, secondary structure and accessible surface area; experimental conditions; thermodynamic data such as unfolding Gibbs free energy, enthalpy and heat capacity, transition temperature, free energy of unfolding by denaturant denaturation, slope of denaturation curve and midpoint of denaturant concentration; literature information, keywords and remarks.

ProTherm is implemented into 3DinSight, a relational database system for structure, function and property of biomolecules [1]. This facilitates more efficient search and retrieval of data by flexible

queries, and enables users to gain insight into the relationship among structure, thermodynamics and function of proteins. The data in ProTherm can be retrieved through a form-based interface with several search, display and sorting options, and 3D structures are visualized with automatically mapped mutation sites and their surrounding amino acids.

3 Database Statistics

We update Protherm frequently. At present, it contains 11764 entries from 555 proteins with 5091 wild type, 5601 single and 768 double mutations. In terms of secondary structures, 2424 mutations are in helical segments, 1704 in strand, 765 in turn and 1777 in coil region. Based on solvent accessibility, 3241 mutations are at buried, 1651 mutations are at partially buried and 1540 are at exposed regions. A majority of data are obtained by circular dichroism (4295), followed by scanning calorimetry (3859) and fluorescence (2399). Further, 6579 data are obtained with thermal denaturation, and 2884 and 2160 data from GdnHCl and urea denaturation, respectively.

4 New Features

The current version has the following new features: We have included the information about the source of each protein and assigned unique Protein Data Bank (PDB) code with a link to all homologous PDB entries. A table showing the frequency of amino acid substitutions for the data available in ProTherm is provided. Further, the free energy change and transition temperature obtained with thermal denaturation experiments have been separated for analysis. The data for both the calorimetric and van't Hoff enthalpy changes have been included for calorimetric experiments. Molecular weight and amino-acid length for each protein are also added.

5 Availability of ProTherm

ProTherm is freely available at <http://www.rtc.riken.go.jp/jouhou/protherm/protherm.html>.

Suggestions and other materials for inclusion in the database are welcome and should be sent to protherm@rtc.riken.go.jp.

References

- [1] An, J., Nakama, T., Kubota, Y., and Sarai, A., 3DinSight: An integrated relational database and search tool for structure, function and property of biomolecules, *Bioinformatics*, 14:188–195, 1998.
- [2] Gromiha, M.M., Oobatake, M., Kono, H., Uedaira, H., and Sarai, A., Role of structural and sequence information for predicting the stability changes: Comparison between buried and partially buried mutations, *Prot. Eng.*, 12:549–555, 1999.
- [3] Gromiha, M.M., Uedaira, H., An, J., Selvaraj, S., Prabakaran, P., and Sarai, A., ProTherm, thermodynamic database for proteins and mutants: Developments in version 3.0, *Nucleic Acids Res.*, 31:301–302, 2002.
- [4] Mendes, J., Guerois, R., and Serrano, L., Energy estimation in protein design, *Curr. Opin. Struct. Biol.*, 12(4):441, 2002.