
 解説

タンパク質と変異体の熱力学データベース (ProTherm) 入門

上平初穂, M. Michael Gromiha, 安江 虹, 皿井明倫

(受取日: 2000年10月2日, 受理日: 2000年11月1日)

Introduction to the Thermodynamic Database for Proteins and Mutants, ProTherm

Hatsuho Uedaira, M. Michael Gromiha, Jianghong An, and Akinori Sarai

(Received October 2, 2000; Accepted November 1, 2000)

Thermodynamic properties of proteins are essential for understanding the mechanism of their folding and stability. We introduce and explain the database, "ProTherm: Thermodynamic Database for Proteins and Mutants", which was developed in our laboratory and can be searched through WWW on various conditions with different options for output. The database contains more than 7000 numerical data for several important thermodynamic properties, structural information of proteins and mutants, experimental methods and conditions, functional and literature information. The database is cross-linked with PubMed, PDB, PMD, EC, and 3DinSight which is developed in our laboratory. The mutation sites and surrounding residues are automatically mapped on structures and can be directly viewed through 3DinSight. ProTherm can be accessed through the World Wide Web at: <http://www.rtc.riken.go.jp/jouhou/protherm/protherm.html>.

1. はじめに

1999年の熱測定討論会で、当時公開を始めた“WWW上で検索できる、タンパク質と変異体の熱力学データベース: ProTherm”¹⁾について発表した。最初の公開後もデータの収集数を増やし、データ項目を見直し、検索機能、リンク機能の向上に努めており、2000年1月に改訂版として第二版を発表した。²⁾ここでは、現時点でのデータ収集および改良状況を説明するとともに、タンパク質の理解のために熱力学がなぜ必要なのか、このデータベースはどんなもので、どのように利用できるか、等について以下に説明したい。

2. タンパク質の熱力学データベースはなぜ必要か

ヒトゲノムの全解析がアメリカでは完了したと報道され、今や遺伝子に基づいて、各個人の体質から、先天的な病気の傾向まで分かってしまう時代に入ろうとしている。遺伝子の情報に基づいて、時間的・空間的に制御されて創られるタンパク質は、生体中で代謝を初めとし、色々な反応の触媒、運搬、免疫など、その生物が生まれ、成長し、環境に応じて命を保ち、子孫を残し、死ぬために必要な種々の機能を担っている。分子レベルで遺伝情報を理解するにはタンパク質の分子レベルでの理解が必要である。

タンパク質が各々の機能を発揮するには、そのタンパク質に固有の立体構造が保たれていなければならない事が広く知られている。しかし、なぜ、熱力学が必要なのだろうか? アミノ酸が鎖状に長く繋がったランダムな状態から、環境に応じて熱力学的に最も安定な状態になるように折り畳まれて、各タンパク質固有の立体構造が作られる事を実験的に初めて証明したのは Anfinsen である。^{3,4)} ウシ腓臓リボヌクレアーゼを 8M 尿素の存在下で還元すると、分子内

の4本のS-S結合が切れて、活性のないランダムなポリペプチド鎖になる。Anfinsenは、この溶液から尿素を取り除きながら酸化し、S-S交換の促進剤を加えると、次第にS-S結合が元の位置に形成され、立体構造がひとりでに再形成され、それと同時に酵素活性も回復することを示した。再形成された立体構造が、最も安定な状態かどうかを決めるのは、その立体構造をもった状態のギブズエネルギーという熱力学量である。それまでに信じられていたような“人力を越えた神秘的な力”は酵素活性を回復するには必要なかったのである。

さらに、熱力学はタンパク質が、ある環境下でどのような立体構造を取り、それを保てるかどうか、どの程度安定かを定めるだけでなく、機能とも密接に関係していることも、明らかにされている。例を挙げよう。がん遺伝子産物であるc-Mybタンパク質のN端近くに、DNA結合ドメインがある。DNA結合ドメイン(アミノ酸156残基)は、各々3本のヘリックスをもった三つのリピート(R1, R2, R3)から成る。DNA結合ドメインの水溶液の熱変性の測定から、このタンパク質の熱変性過程は3状態転移であること、また、熱転移過程に中間体が存在する原因は、R1, R2, R3のうち、R2が他の二つより不安定であることが分かった。⁵⁾ R2がR1, R3より不安定な原因を明らかにするには、分子の立体構造を知る必要がある。そのため、NMRによる立体構造解析が進められた。その結果、R2には他のリピートと違って、3本のヘリックスが交差する中心部分にcavityが存在し、それがR2を不安定にしていること、⁶⁾ 更に、DNA結合ドメインがDNAに結合する際に、タンパク質の立体構造がDNAに良くフィットするためにはR2にcavityが必要であることが明らかになった。⁷⁾ これは、熱力学的性質-立体構造-結合機能の3者が密接な関係を示す1例である。

カロリメトリーや分光学的測定による熱変性の解析がNMRやX線立体構造解析より先に行われたのは、一つには、熱力学測定の方がサンプル量がむしろ少なく済むこと、測定・解析が、NMRやX線構造解析よりも短時間で出来るという優位性のためである。

さて、タンパク質の機能を分子レベルで理解するには、一次構造から立体構造が形成されるメカニズムの理解が欠かせない。しかし、この問題はまだ完全に解明されたとは言えない。理論によって完全に証明できない状況では、実証的に実験データを積み重ねてこの問題に迫る研究が重要でまた必要となる。すなわち、タンパク質の種々の箇所をアミノ酸置換し、立体構造、安定性および機能がどう変化するかについての研究が多数行われている。しかも、研究数は年々増える傾向にある。

それら全体を把握して理解するにはデータベース化が必

Table 1 The data included in ProTherm.

Dataset	Oct. 1998	Aug. 2000
Total	3317	7114 (214%)*
Accessible surface area (ASA)**		
Buried	1317	2269 (172%)
Partially buried	918	1246 (136%)
Exposed	826	1219 (148%)
Mutation		
Wild type	429	2460 (573%)
Single	2434	3941 (162%)
Double	353	538 (152%)
Multiple	101	175 (173%)
Secondary structure		
Helix	1241	1900 (153%)
Strand	789	1264 (160%)
Turn	242	475 (196%)
Coil	846	1179 (139%)
Measurement		
Circular dichroism	1709	2870 (168%)
Fluorescence	1052	1743 (166%)
Calorimetry	489	1990 (407%)
Method		
Thermal	1628	3718 (228%)
GdnHCl	1094	2038 (186%)
Urea	573	1314 (229%)
References		
Proteins	61	492 (807%)
Articles	245	711 (290%)

*) The percentage was calculated with respect to Oct. 1998.

***) Buried: ASA < 20 %; partially buried: 20 % < ASA < 50 %; exposed: ASA > 50 %.

要との認識が高まり、著書も発行されている。^{8,9)} しかし、データベースとしては、多数の熱力学データが収集されているのみではなく、熱力学の必要な(知りたい)項目をユーザーが選んでWWW上で検索でき、立体構造データベースや変異体データベースとリンクして、総合的に考察できるような機能が必要である。ProThermはこのような機能を備えるよう設計・構築されたデータベースである。以下にこのデータベースの内容(3節)、検索方法(4節)、と他のデータベースとのリンク法(5節)について解説する。

3. 蛋白質と変異体の熱力学データベース ProThermの収集データ内容

このデータベースの入力件数は2000年8月現在約7100件、すなわち公開時のデータ数の200%以上に達し(Table 1参照)、次のようなタンパク質の構造や熱力学情報が収集されている。

3.1 タンパク質関連データベースのコード

熱力学データを載せているタンパク質について構造、活性、等の関連データベースに記載がある時には、これをデータに入れている。関連データベースとしては、

PDB (Protein Data Bank, タンパク質データバンク, 野生型のみでなく, 変異体についてもデータベースに記載があれば, そのコード番号も記述),

PMD (Protein Mutant Database, タンパク質の変異体とその活性に関する情報),

EC (Enzyme code, 酵素活性に関する情報),

PIR (Protein Information Resource, アミノ酸配列情報),

SWISS-PROT (The Swiss Institute of Bioinformatics と The European Bioinformatics Institute の共同で作られたアミノ酸配列情報)。(リンクするデータベースについては5節参照)

3.2 タンパク質と変異体の立体構造情報

タンパク質名とそのPDBコード(既述), 天然状態における集合数, 野生および変異アミノ酸名(1変異体, 多変異体等), 変異位置: アミノ酸配列上の変異の位置, 変異箇所二次構造とASA(%)

3.3 測定情報と熱力学データ

測定法: カロリメトリー, 分光学(法(CD, 蛍光, 可視・紫外吸収)等

測定条件: pH, 緩衝剤名と濃度, 添加塩等, 測定温度(T)

熱変性: 転移温度(T_m), 転移エンタルピー(ΔH), 転移のギブズエネルギー変化(ΔG), 変異体と野生型の ΔG の差($\Delta\Delta G$), 転移に伴う熱容量変化(ΔC_p)

変性剤変性: 変性剤名, 測定温度(T), 変性中点の変性剤濃度(C_m), ΔG の変性剤濃度依存性(m), ΔG の変性剤濃度ゼロへの外挿値(ΔG^{H_2O}), 変異体と野生型の ΔG^{H_2O} の差($\Delta\Delta G^{H_2O}$)

熱変性と変性剤変性に共通な変性情報: 転移の状態数, 変性の可逆性に関する情報

酵素活性: 変異体の相対活性(%), ミカエリス定数(K_m), 酵素活性(ミカエリス・メンテン式の反応最大速度/酵素濃度, k_{cat}), 解離定数(K_d)

3.4 文献情報

原論文: 著者, 掲載誌, 発表年, 巻, 頁, キーワード, Remarks

これらのデータの入力例として, ヒトLysozyme変異体の熱変性の場合の例をFig.1に示す。(この図を画面上で見ると, 後述のFig.5の検索結果画面の左端の入力番号をクリックする。)

アドレス <http://www.rtc.riken.go.jp/jouhou/protherm/protherm.html> の移動

```

NO. 2135
PROTEIN Human Lysozyme
EC NUMBER EC:3.2.1.17
PDB NO. 1L2Y
PDB URL http://www.rcsb.org/pdb/
PDB URL 47 F
NO. MOLECULES 1
SEQUENCE C
ASA 84.0
STATE 1
dS_H2O
dG_H2O
dH
dH 0.1
Tm 65.4
dFm 0.2
dH 119.4
m
dCp 1.560
pH 2.8
BUFFER NAME glycine hydrochloride
BUFFER CONC 50 mM
ION NAME 1 -
ION CONC.1 -
REFERENCE 200
METHOD Thermal
REVERSIBILITY yes
ACTIVITY
ACTIVITY_Bm
ACTIVITY_Rcat
ACTIVITY_Ed
KEY WORDS human; lysozyme; protein; stability; susceptibility;
site-directed mutagenesis; protein structure
REFERENCE
AUTHOR Henning TS, Petersen H, Soehnle B, Matuchniak M,
Eberhart M.
REMARKS
    
```

Fig.1 An example of the input data of ProTherm. Entry No. 2135 for human lysozyme is shown.

4. ProThermの利用法一: 検索と結果の表示

このデータベースを実際の検索に利用するためには, データベースにアクセスして, 必要な検索項目を入力し, 検索を実行する。その手順について次に説明する。

4.1 ProThermへアクセスとデータベースの統計情報

次のアドレスを指定すると,

<http://www.rtc.riken.go.jp/jouhou/protherm/protherm.html>

Fig.2に示すような初期画面が表示される。Fig.2の初期画面の中央付近に, Search, Database statistics, TutorialおよびWhat's Newの項目がある。

Searchをクリックすると, 次節に述べるように検索が開始される。Database statisticsはこのデータベースに現在収録されているデータについて, タンパク質名一覧, 著者一覧, 文献一覧等の情報が示される他, タンパク質の各種の分類毎のデータ数が得られる (Table 1参照)。What's Newには新情報やタンパク質立体構造表示ソフト (RasMol) のインストール法などが載っている。初めてのユーザーのためにTutorialの表示があり, これをクリックすると, 次の三つの例について, 検索項目の選び方や表示方法の選択とその結果がどのように示されるかを知ることができる。

例1. 検索条件 変異の位置と数: ヘリックス上の1変異体; 熱変性に依る変性; pH 7

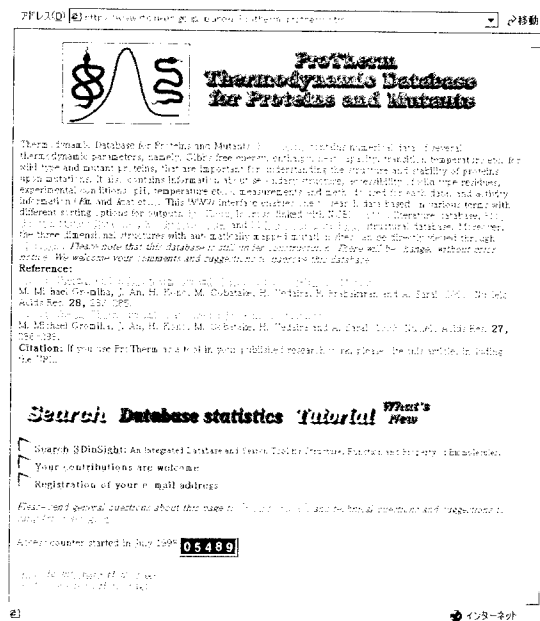


Fig.2 Home page of ProTherm access site: <http://www.rtc.riken.go.jp/jouhou/protherm/protherm.html>.

- 例2. 検索条件 変異の位置: buried領域; 尿素変性; 温度15~25℃; CD測定
出力条件 測定温度と残基数の順に出力
- 例3. 検索条件 変異体: Ala (溶媒露出度0~50%) から任意の残基への変異; 論文の発表時期: 1990~1997年
出力条件 年代順; 溶媒露出度(%)が高い順; 変異残基のアルファベット順

さて、各ユーザの欲しい情報を得るには、Fig.2の画面でSearchをクリックする。するとFig.3のような検索画面となる。

4.2 検索

Fig.3の画面で、検索したい項目にVをつけるか、または検索範囲を指定する。

選べる検索項目は、Fig.3に示されるように上から順に、① 入力番号、② タンパク質名、③ 変異特性: 置換前後のアミノ酸名と変異数 (1置換体, 2置換体, 多置換体, 野生型)、④ 変異箇所の二次構造上の位置: ヘリックス, β シート, ターン, コイル、⑤ 変異箇所の溶媒露出度: 全て (無指定), buried, partially buried, exposed; ASAの範囲指定 (%またはA)、⑥ 測定法: 吸光度, CD, DSC, 蛍光クロマトグラフィー, HPLC, その他、⑦ 変性方法: 熱, GdnHCl, 尿素, GSSG/GSH, GdnSCN、⑧ pH: 範囲を指定、⑨ 測定温度 (T), 熱転移温度 (T_m), 変異による T_m の差 (ΔT_m) の何れかを選び、範囲を指定、⑩ 変異による変性のギブズエネルギーの差: 熱変性 ($\Delta\Delta G$) か、変性剤

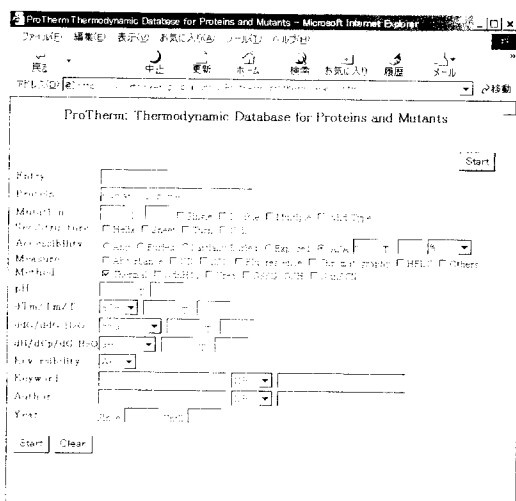


Fig.3 Search interface of ProTherm. In this example, protein name, ASA >60%, thermal denaturation are selected from the menu.

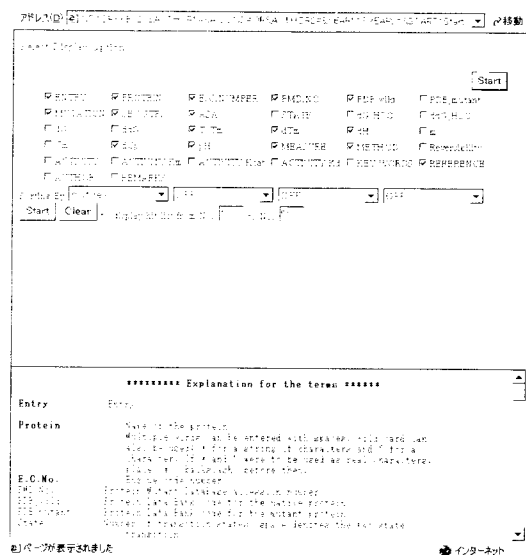


Fig.4 Display and sorting options of ProTherm. In this example, entry number, protein name, EC number, PMD number, PDB code for wild type, mutation site, secondary structure at the site, ASA (%), T/T_m , ΔT_m , ΔH , ΔC_p , pH, measure, method and reference are selected for the output. The explanations of the terms which are used in ProTherm are described in the lower part.

変性 ($\Delta\Delta G^{H_2O}$) かを選び、範囲を指定、⑪ 変性エンタルピー変化 (ΔH), 熱容量変化 (ΔC_p), ギブズエネルギー変化 (ΔG , または ΔG^{H_2O}) の範囲を指定、⑫ 可逆性、⑬ キ

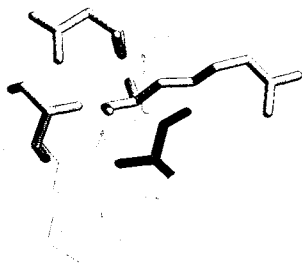


Fig.8 Display of protein mutant and surrounding residues by RasMol. In this example, the surrounding residues of Ala 47 in human lysozyme are shown. The central residue, surrounding residues within 4 Å and surrounding residues between 4 and 8 Å are shown in black, thin gray and dark gray, respectively.

変異アミノ酸位置をクリック→変異アミノ酸から0~4 Å および0~8 Åにあるアミノ酸の一覧が次のように表示される。

PDB 1LZ1 TARGET 47 ALA

0~8 Å 45 TYR 46 ASN 49 ASP 50 ARG 51 SER 48 GLY

0~4 Å 45 TYR 46 ASN 48 GLY

この画面のRASMOLをクリック→Fig.8のように変異位置付近の立体構造が図示される。この図で変異アミノ酸残基が黒色に、0~4 Å内および4~8 Åにあるアミノ酸がそれぞれ薄い灰色と濃い灰色で表示される。(原図はカラー)

PDBコードをクリック→3DinSight (当研究室で開発した構造、機能と物性を統合したデータベース^{10,11)} が起動し、タンパク質全体の立体構造が画面に表示される。(図略)

右端の文献欄をクリック→PubMedが起動し原論文の題や要旨が見られる。(Fig.9)

6. データベースのタンパク質立体構造の理解、設計等への寄与と今後の拡充予定

熱力学データベースを種々の観点から検索し、統計処理を行うことによって多数の実験データに基づいて帰納的にタンパク質の立体構造を議論することが可能になる。これらの議論が、タンパク質の理解に寄与するのみでなく、理論計算や、タンパク質の設計に役立つことが期待される。すでにこのような全般的性質の抽出に関する研究が進められている。¹²⁾

熱力学データベースの内容と操作性、機能をさらに充実させるため、絶えずプログラム等に改良を加えている他に、改善を予定中の項目は、

(1) データの内容に以下の項目を加える：タンパク質の分子量、アミノ酸残基数、由来、タンパク質の測定の際の濃度、ファントホッフエンタルピー (ΔH^{FH} , カロリメトリックエンタルピーと区別)、酵素活性の熱処理温度依存性による熱安定性のデータ。

(2) 検索項目にタンパク質の由来、分子量、アミノ酸残基数、を加える。

(3) リンクするデータベースに疾病に関連した変異体のデータベースを加える。

(4) データベースの利用価値は、含まれるデータ数が多いほど、また、質が高いほど高い。データの信頼性の判断は、学術雑誌に掲載されたデータであることを一つの基準にし、変性の可逆性の記述があるかどうかを収集項目に加えた。また、ProThermに収録したデータの著者にE-mailを送り、データの確認とコメントを依頼する予定。継続的にデータ数を増加させ、データベースを管理するには、人力、費用等の長期間にわたる確保も今後の課題の一つと思われる。ご意見、コメントを sarai@rtc.riken.go.jp までお

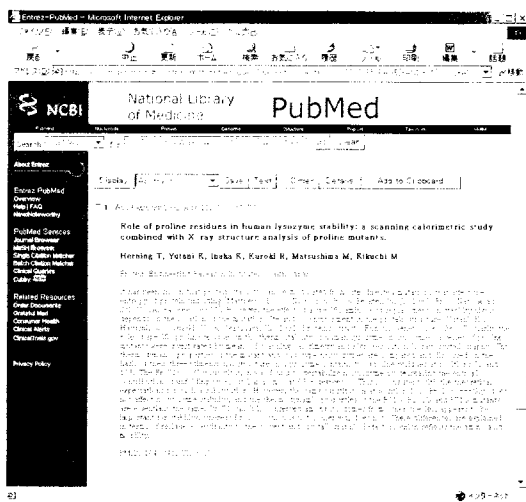


Fig.9 PubMed literature database linked from ProTherm result page (Fig.5), showing the corresponding reference with title and abstract.

のデータを例に取ると、

ECコードをクリック→ECが起動し、活性情報が得られる。(Fig.6)

PMDをクリック→PMDが起動し、変異体関連情報が得られる。(Fig.7)

願いしたい。なお、検索画面や検索方法等に随時変更があり得るので、ご了解いただきたい。

文 献

- 1) M. M. Gromiha, J. An, H. Kono, M. Oobatake, H. Uedaira, and A. Sarai, *Nucleic Acids Res.* **27**, 286-288 (1999).
- 2) M. M. Gromiha, J. An, H. Kono, M. Oobatake, H. Uedaira, P. Prabakaran, and A. Sarai, *Nucleic Acids Res.* **28**, 283-285 (2000).
- 3) C. B. Anfinsen, *Science* **181**, 223-230 (1973).
- 4) 例えば、解説：上平初穂，現代化学 No.268, 24-29 (1993).
- 5) A. Sarai, H. Uedaira, H. Morii, T. Yasukawa, Y. Nishimura, K. Ogata, and S. Ishii, *Biochemistry* **32**, 7759-7764 (1993).
- 6) K. Ogata, S. Morikawa, H. Nakamura, H. Hojo, S. Yoshimura, R. Zhang, S. Aimoto, Y. Ametani, Z. Hirata, A. Sarai, S. Ishii, and Y. Nishimura, *Nature Struct. Biol.* **2**, 309-320 (1995).
- 7) K. Ogata, C. Kanei-Ishii, M. Sasaki, H. Hatanaka, H. Nagadoi, M. Enari, H. Nakamura, Y. Nishimura, S. Ishii, and A. Sarai, *Nature Struct. Biol.* **3**, 178-187 (1996).
- 8) H.-J. Hinz, "Thermodynamic data for biochemistry and biology", Springer-Verlag, NY (1986).
- 9) W. Pfeil, "Protein stability and folding: a collection of thermodynamic data", Springer-Verlag, NY (1998).
- 10) J. An, T. Nakama Kubota, and A. Sarai, *Bioinformatics* **14**, 188-195 (1998).
- 11) T. Nakama, J. An, Y. Kubota, and A. Sarai, *Bioimages* **5**, 59-64 (1997).
- 12) M. M. Gromiha, M. Oobatake, H. Kono, H. Uedaira, and A. Sarai, *Protein Eng.* **12**, 549-555 (1999).

要 旨

タンパク質の立体構造は熱力学的に最も安定な状態という条件で決められる。そのため、タンパク質の変異体と野生型の立体構造の安定性を実測、比較する多数の研究がなされている。我々はこれらの熱力学的測定値を収集し、立体構造および機能データベースとリンクさせ、WWW上で検索可能な熱力学データベース (ProTherm) を構築した。これを一般公開し、質と量の向上に努めている (アドレス: <http://www.rtc.riken.go.jp/jouhou/protherm/protherm.html>)。本論文では、このデータベースの内容と機能について説明し、利用法について解説する。



上平初穂 Hatsuho Uedaira
理化学研究所 筑波研究所, The Institute of Physical and Chemical Research, RIKEN Tsukuba Institute, TEL. 0298-36-9082, FAX. 0298-36-9080, e-mail: uedaira@rtc.riken.go.jp
研究テーマ：タンパク質立体構造の熱力学
趣味：コーラス



M. Michael Gromiha
理化学研究所 筑波研究所, The Institute of Physical and Chemical Research, RIKEN Tsukuba Institute, TEL. 0298-36-9082, FAX. 0298-36-9080, e-mail: gromiha@rtc.riken.go.jp
研究テーマ：Development of protein thermodynamic database, Structure prediction and stability of proteins, Protein-DNA interactions
趣味：reading



安江虹 Jianghong An
理化学研究所 筑波研究所, The Institute of Physical and Chemical Research, RIKEN Tsukuba Institute, TEL. 0298-36-9082, FAX. 0298-36-9080, e-mail: ajh@rtc.riken.go.jp
研究テーマ：バイオインフォーマティクス
趣味：写真



皿井明倫 Akinori Sarai
理化学研究所 筑波研究所, The Institute of Physical and Chemical Research, RIKEN Tsukuba Institute, TEL. 0298-36-9082, FAX. 0298-36-9080, e-mail: sarai@rtc.riken.go.jp
研究テーマ：タンパク質・核酸の構造安定性・相互作用
趣味：アウトドアスポーツ, 旅行