

## 生体分子の熱力学データと構造データの統合

(蛋白質およびその相互作用に関する熱力学データについて、対応する構造データと統合する)

国立大学法人九州工業大学

### 1. 委託事業の9月末時点の判断基準になる目標 (以前にお出しいただいたもの)

- ・蛋白質熱力学データベースおよび蛋白質・核酸相互作用データベースのそれぞれ新規 700 件程度の熱力学データと構造データのクロスレファレンスを作成する。蛋白質・蛋白質相互作用データは格納用データベースシステムを完成させる。
  - ・蛋白質・核酸相互作用データを XML 化し公開する。
  - ・統合 DB センターと連携して熱力学データの文献収集とデータ抽出の自動化技術の開発に必要な解析やツールの評価を行う。
- 9 月末までに実用的な自動収集が行えるようにすることを目標とする。

### 2. 9 月末時点の達成状況

- ・蛋白質熱力学および蛋白質・核酸相互作用のそれぞれ新規分データ約 700 件と、対応する構造データのクロスレファレンスを作成した。蛋白質・蛋白質相互作用データについては、格納用のデータベースのプロトタイプを作成し試験を行っている。
- ・上記の蛋白質と核酸の相互作用の熱力学データの更新分について XML フォーマットに変換した。
- ・熱力学データを含む文献を自動的に収集する目的で TogoDoc をカスタマイズし、収集の効率が大幅に改善しほぼ実用の域に達した。

### 3. 上記達成状況を踏まえたプロジェクト終了までの目標

- ・蛋白質熱力学データおよび蛋白質・核酸相互作用データの収集を継続し、構造データとのクロスレファレンスを作成する。蛋白質・蛋白質相互作用データは格納用データベースシステムを完成させる。
- ・これらのデータは XML 化し公開する。
- ・熱力学文献自動収集用にカスタマイズした TogoDoc は実用段階に達したので、今後実地の作業で使用し、こまかいチューニングを行いさらに使いやすいようにする。

### 4. 成果の概要

蛋白質熱力学データベースおよび蛋白質・核酸相互作用データベースは、それぞれ 25,000 件、10,700 件のデータを蓄積し、対応する構造データとのクロスレファレンスを作成し統合化した。類似のデータベースが存在しないため、この分野では重要なリソースとして世界中の研究者に利用されている。1 月～7 月のアクセス (ページリクエスト数) は約 46,000 回となっている。これらのデータベースはすでに 200 件以上の論文に引用され、データベースを用いて解析を行った研究論文も数多く出版されている。

データベース構築で最も時間のかかるのが熱力学データを含む文献の収集とデータの抽出であるが、TogoDoc をカスタマイズすることにより、文献収集の効率を大幅に改善することができた (DBCLS との共同研究)。