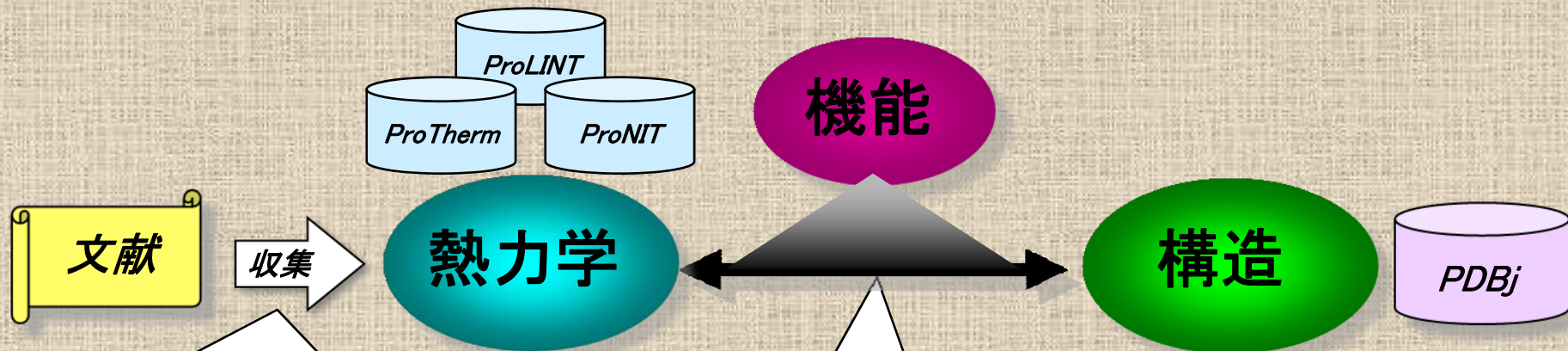
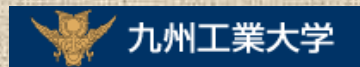


生体分子の熱力学データと構造データの統合



テキストマイニング技術

- データを含む論文の自動収集
- テキストからのデータの自動抽出

統合化技術

- 熱力学情報と構造情報のクロスレファレンス
- オントロジーの整備
- データ交換フォーマットの整備



統合データベースプロジェクト

生体分子の熱力学データと構造データの統合

ホーム プロジェクト 最新情報 メンバー コンタクト

リンク (内部)

皿井研究室

BioInfoBank

ProTherm

ProNIT

Controlled Vocabulary for
Thermodynamic Databases

リンク (外部)

文科省統合データベースプロジェクト

統合データベースセンター(LSDB)

PDBj

The National Center For
Biomedical Ontology

Protein Ontology

Biochemical Thermodynamics

IUBMB

The Gene Ontology

ダウンロードリンク

ProTherm-Data-Download

ProNIT-Data-Download

Cross-Reference of ProTherm

Cross-Reference of ProNIT

概要

九工大では、文科省の統合データベースプロジェクトの一環として、蛋白質の安定性や相互作用の網羅的な熱力学データと構造データとの統合化を進めています。

統合化として具体的には以下のことを進めています。

1. 熱力学情報と構造情報のクロスリファレンスの作成

九工大の熱力学データとPDBjの構造データを対応づけるクロスリファレンステーブルを作成します。クロスリファレンステーブルを元に、九工大が以前から構築している蛋白質熱力学データベース (ProTherm)、蛋白質・核酸相互作用データベース (ProNIT) と、PDBjとの間に、双方向にリンクを作成します。

ProTherm: 蛋白質の熱力学データベース

ProNIT: 蛋白質・核酸相互作用データベース

2. オントロジーの整備

熱力学データのオントロジーを整備します。

Controlled Vocabulary for Thermodynamic Databases

3. データ交換フォーマットの整備

XMLなど、熱力学データのデータ交換フォーマットを整備します。

ProThermのXMLデータ

ProNITのXMLデータ (暫定版)

また、ライフサイエンス統合データベースセンターと連携して、テキストマイニングによる論文の自動収集やデータの自動抽出を実施します。

最新情報

熱力学データと構造・配列データのクロスリファレンスの更新
(2009年01月08日) [詳細....](#)

ProNITデータの更新 (2009年01月08日) [詳細....](#)

熱力学データと構造・配列データのクロスリファレンスの更新
(2008年12月22日) [詳細....](#)

ProThermデータの更新 (2008年12月19日) [詳細....](#)

熱力学データと構造・配列データのクロスリファレンスの更新
(2008年12月02日) [詳細....](#)

ProNITデータの更新 (2008年12月02日) [詳細....](#)

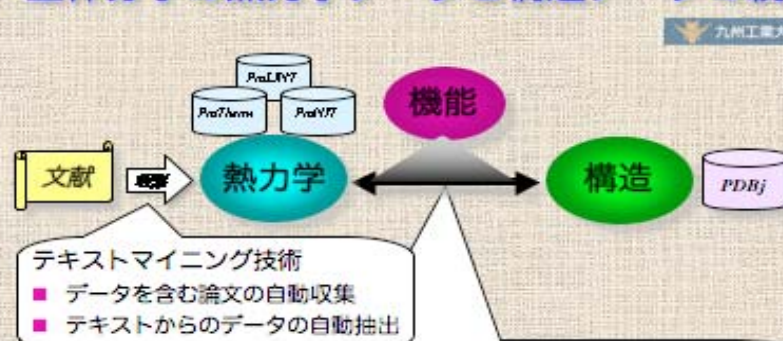
熱力学データと構造・配列データのクロスリファレンスの更新
(2008年10月29日) [詳細....](#)

ProNITのXMLデータ (暫定版)
(2008年10月20日) [詳細....](#)

ProThermデータの更新 (2008年9月8日) [詳細....](#)

熱力学データのポキャプラー
(2008年8月15日) [詳細....](#)

生体分子の熱力学データと構造データの統合



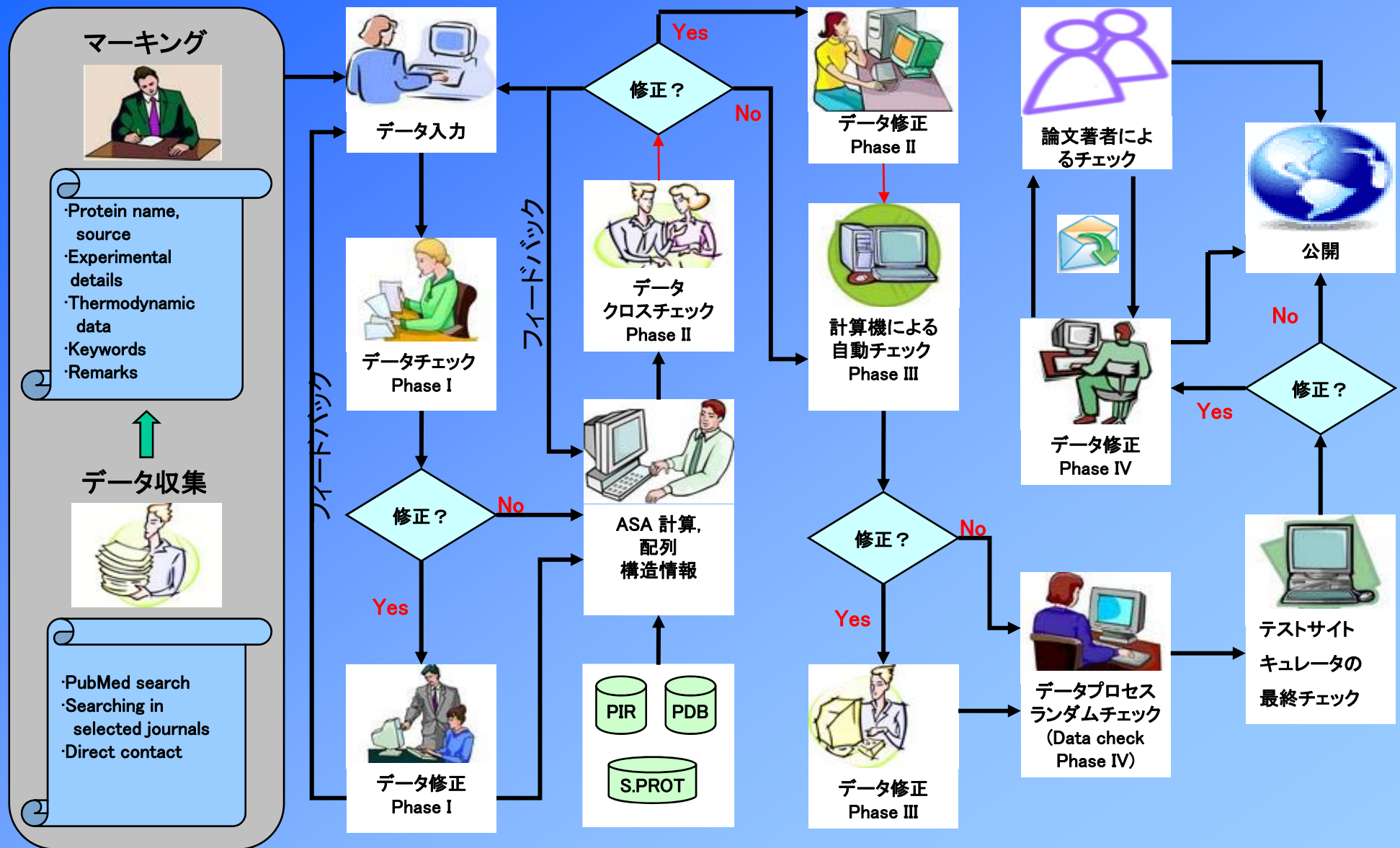
20年度実績の要約

- 蛋白質と変異体の熱力学データと構造データを対応させるクロスレファレンスを約1,100件作成
- 蛋白質と核酸の相互作用の熱力学データを対応させるクロスレファレンスを約700件作成
- 蛋白質・蛋白質相互作用データベースのスキーマの設計を行った
- 蛋白質・核酸相互作用熱力学データについてXMLフォーマットのプロトタイプを作成
- 熱力学データのControlled Vocabularyを作成
- テキストマイニングにより文献検索と文献からのデータ抽出を自動化する方法を統合DBセンターと連携して開発中
- 統合データベースプロジェクト用のページを作成

20年度のデータ入力について

- 本プロジェクトの元となるデータベースの入力部分は、過去9年にわたって科研費研究成果公開促進費の支援で行ってきたが、20年度の申請が通らなかった。
- 蛋白質と変異体の熱力学データベースについては当初目標の1,000件を達成できたが、蛋白質・核酸相互作用データベースについては当初目標の1,300件を達成できなかった。
- 21年度の研究成果公開促進費の申請も通らなかった。
- データ入力の必要経費については別紙参照。

文献からのデータ収集とデータベース構築プロセス



21年度業務計画

- 蛋白質と変異体の熱力学データと構造データを対応させるクロスレファレンスを約700件作成する
- 蛋白質と核酸の相互作用の熱力学データを対応させるクロスレファレンスを約700件作成する
- 蛋白質・蛋白質相互作用は格納用データベースを作成する
- 熱力学データについてXMLフォーマットを公開する
- 熱力学データのオントロジーの整備を進める
- テキストマイニングにより文献検索と文献からのデータ抽出を自動化する方法を統合DBセンターと連携して進める