生命情報工学研究センター

Computational Biology Research Center

Life Science Database Integration Project, Workflow 2009

Toshiyuki Tashiro, Yukimitsu Yabuki, Kazuhiko Fukui, Katsuhisa Horimoto, Tamotsu Noguchi, Kiyoshi Asai

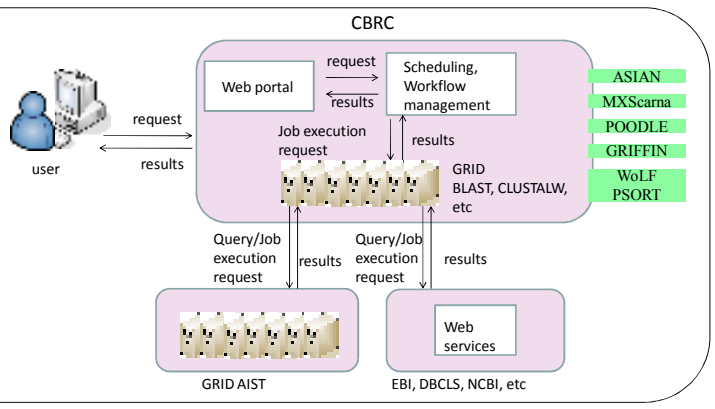
Integration of Biological Information Group, CBRC, AIST

統合DBにおけるCBRCの役割

CBRCは2007年より文部科学省が推進するライフサイエンス統合データベースプロジェクトに参画しており、データベース・解析ソフトウエアを組み合せたワークフロー技術及び分散処理技術の開発を担当している。

現在、多くのデータベース・ソフトウエアが存在するが、サイトが分散しており、これらを組み合せて使用する場合、時間と手間がかかり非効率的である。ユーザがこれら複数のデータベース・ソフトウエアを短時間に効率的に実行するためには、組合せ技術及び分散処理が不可欠である。CBRCはタンパク質立体構造に関して有益な情報を提供する3点のワークフロー及びアクティブ・ワークフローを開発し提供します。

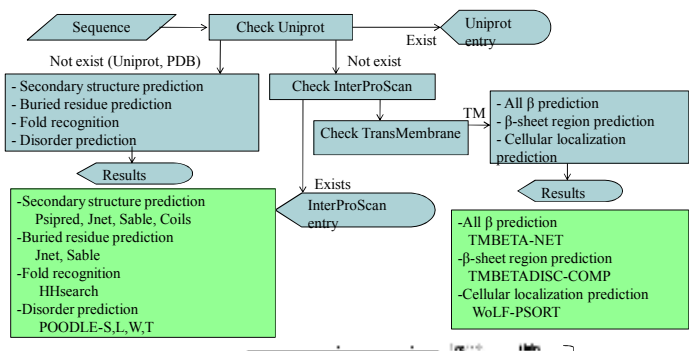
CBRC情報統合基盤サイト　http://togo.cbrc.jp



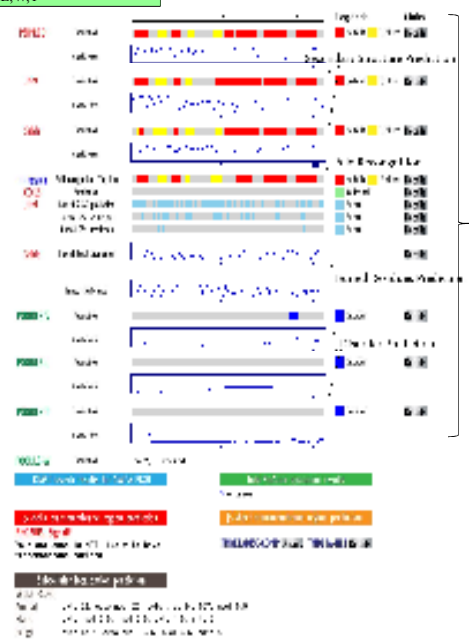
CBRC Protein Structure Workflows

Protein Annotation Workflow

本ワークフローは、立体構造未知のタンパク質に関し、構造及び機能のヒントとなる情報を幅広く実験研究者等に提供することを目的としており、各種プログラム等をGridにより効率的に分散処理し、従来と比し短時間で結果を表示する。ユーザからアミノ酸配列を受取り、二次構造予測、埋れ残基予測、フォールド認識、ディスオーダー予測、膜タンパク質オールベータ・ベータシート予測、細胞内局在予測をGridにて分散処理を行う一方、データベース検索及び疎水性予測の実行を他のサーバへ依頼し結果を取得後、全ての結果をユーザーが解析し易いよう配置し出力する。2008年12月末より一般公開。



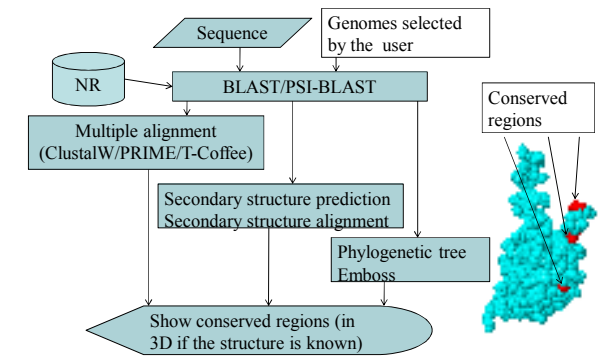
結果画面



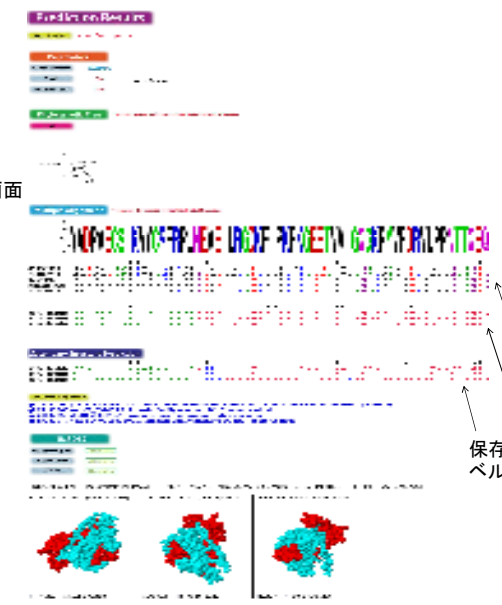
異なるプログラムの予測結果を揃えて表示し、目視認識度を向上

Comparative Protein Information Workflow

本ワークフローは、相同なタンパク質を比較することで保存部位等構造上重要な部位を表示し、実験研究者等に提供することを目的としている。ユーザからアミノ酸配列を受取り、相同タンパク質を検索、その結果からユーザーがいくつかのタンパク質を選択し、マルチプルアラインメントを実行することで保存性が高い残基を表示する。また、二次構造予測結果も同様にマルチプルアラインメントし、保存性が高い二次構造を表示する。2009年3月末より一般公開。



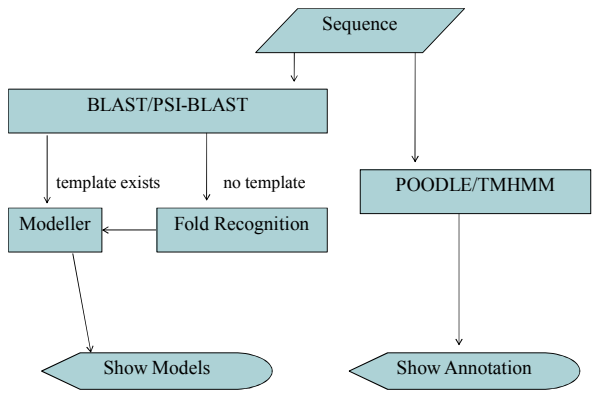
結果画面

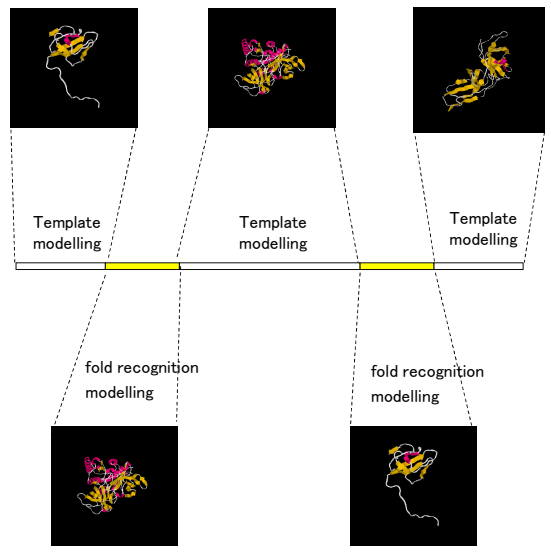


保存度のレベルは3段階

Protein Modelling Workflow

本ワークフローは、立体構造未知のタンパク質に関し、立体構造のモデリングを行う。ユーザからアミノ酸配列を受取り、BLAST/PSI-BLASTで鋳型構造を探し、モデリングプログラムを実行する。また、鋳型構造が発見できなかった領域が存在する場合は、構造認識実行後モデリングする。さらに、ディスオーダー領域や膜貫通領域などのアノテーション情報も同時に計算／検索し、モデリングの結果とともにユーザに提示する。2009年12月一般公開予定。





アクティブ・ワークフローの開発

2009年度より、ソフトウェアの利用者であるユーザが、解析用途に応じて各解析要素技術（プログラム）を組み合わせ解析構造体とし、要素技術を利用できる基盤技術環境の構築を実施しています。ここでは、自ら能動的に解析を進めるユーザを対象とし、それぞれ目的に応じて、独自のワークフローを構築し実行できる目的指向アーキテクチャーを目指したプラットフォーム型アクティブ・ワークフローの開発を行っています。

KNIME＊プラットフォームの利用

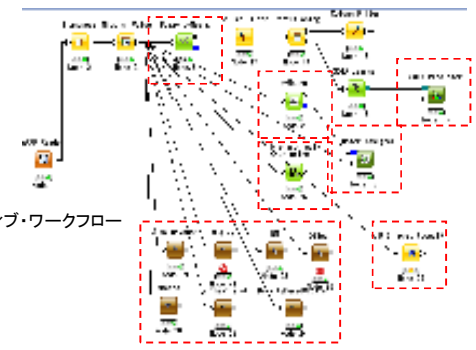
KNIMEはオープンソースで開発され、無料で利用することができるソフトウェアであり、ノードと呼ばれるプログラム群をGUI上でインタラクティブに接続し、一連のデータ解析を実現することができます。

＊ http://www.knime.org

ASIAN Active Workflow

CBRCでは遺伝子ネットワーク推定システムとしてASIAN（Automatic System for Inferring A Networkの略」）を開発しています。このシステムは、DNAマイクロアレイ等の遺伝子発現データに対し、階層型クラスタリングを適用し、クラスタ数を推定し、クラスタ間の関係性を推定するシステムです。

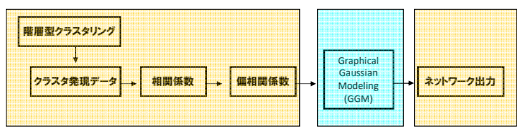
ASIANアクティブ・ワークフロー

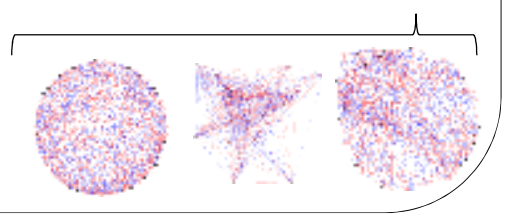


CBRC　ASIAN Server

SOAP

Local PC　　Local PC





ASIANの現状での課題に対応する、複数のデータタイプからASIANに適した形式へ変換する方法や様々なクラスタリング機能がKNIME上で既に実現されているため、ASIANの機能をKINME上のノードとして実現し、ワークフロー化することでユーザの求める機能を効率的に実現することができると考えられます。

技術を社会へ
Integration for Innovation

National Institute of Advanced Industrial Science and Technology AIST

独立行政法人
産業技術総合研究所