

海洋微細藻類の高層化培養によるバイオディーゼル生産

生命情報工学研究センター 細胞システム解析チーム CREST

【研究成果の概要】

・異質倍数体の確認と遺伝子構造アノテーションの高精度化
 解読されたDNA配列のアセンブリを行った結果、33.0kbの Scaffold N50サイズ(297配列)が得られました。ここで推定されたゲノムサイズ(49.9Mb)から、JPCC DA0580が大規模に染色体構造を変化させている可能性が示唆されました。そこで、得られたscaffoldについてアライメントを行い、49組のscaffold対の再構築を行いました。その結果、①scaffold対での60%の遺伝子の保存と、②scaffold内の重複遺伝子の割合(1%程度)を明らかにし、さらに③高精度な物理マップを構築し、JPCC DA0580株における異質倍数性を明示化を達成しました。(図1)

・トリグリセリド代謝経路と合成酵素遺伝子群の網羅的同定
 脂肪酸関連酵素に特化したHMMプロファイルを独自に作成し、72個の合成酵素遺伝子を網羅推定しました。さらに、推定した遺伝子から代謝経路多価不飽和脂肪酸合成経路とトリグリセリド合成経路を再構築しました。その結果、JPCC DA0580ゲノムは既知の脂肪酸およびトリグリセリド合成酵素ホモログを多数保有しており、多様な飽和/不飽和/多価脂肪酸を合成可能であることを明らかにしました。(図2)

【平成24年度計画との関連】

小項目：II-1-(3)-<1> 配列情報と分子構造情報を用いた創薬支援技術開発、I-3-(1)-<2> 微生物資源や有用遺伝子の探索と機能解明

【開発技術の用途】

・バイオディーゼル生産への応用



図1: JPCC DA0580ゲノムの物理マップ

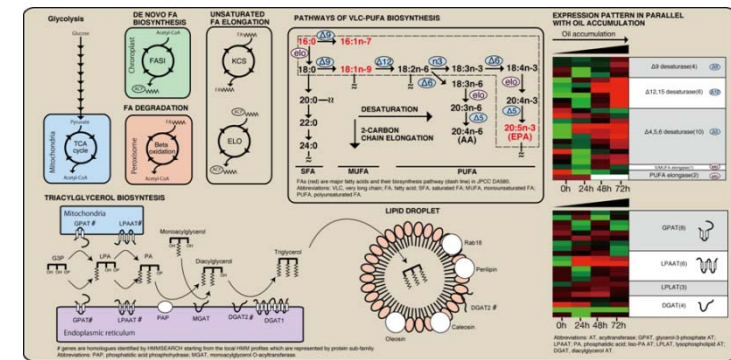


図2. 脂肪酸合成経路と関連する遺伝子群の発現変動

構造方程式モデリングなど統計手法を用いたネットワークモデルの構築と推定 生命情報工学研究センター 細胞システム解析チーム 油谷幸代

【研究成果の概要】

- ・多変量解析の1つである構造方程式モデリングに因子分析を組み合わせた解析手法を開発し、遺伝子発現データに適用することで細胞内における別要因と遺伝子の制御関係構造の推定を行いました。まず、モデル生物である出芽酵母や線虫の特徴的遺伝子発現制御モデル化によって、実験的に確認されていない細胞内におけるタンパク質の影響を推定し、論文発表を行いました。応用研究として、開発した手法を環境化学物質を曝露させたES細胞の遺伝子発現データに適用しました。その結果、毒性の違いによって、ES細胞内での遺伝子ネットワーク構造に違いが生じることが示唆されました。
- ・時系列データにおいて、複数の波形形状の類似性を表す交差相関係数と、変数間の直接的な関係性を推定する偏相関係数を組み合わせた手法を開発し、生態系で測定された酵素活性化データに適用し、生態系で行われているマクロな現象の解明を行いました。(奈良工業高等専門学校との共同研究) その結果、トリメチルアミン分解時にはコンポスト内の微生物相が大きく変化しており、それに対応して活性化された酵素が変化している事が明らかになりました。本研究成果については、国際学会にて発表を行いました。

【平成24年度計画との関連】

小項目：II-1-(3)-<1> 配列情報と分子構造情報を用いた創薬支援技術開発

【開発技術の用途】

- ・薬剤効果の推定
- ・悪臭分解をより安全に行う技術開発の基礎

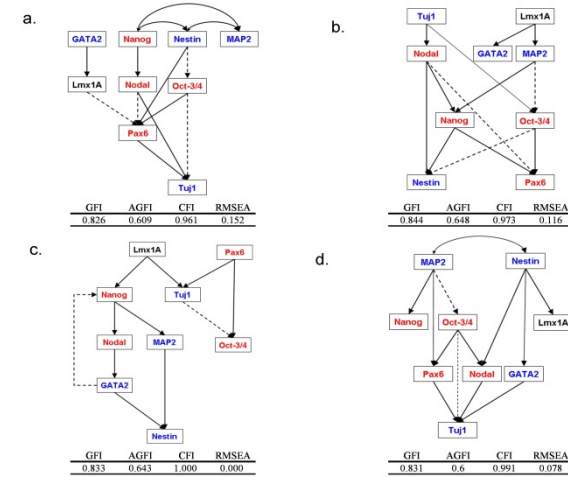


図1. ES細胞における化学毒性特異的な遺伝子ネットワークモデル

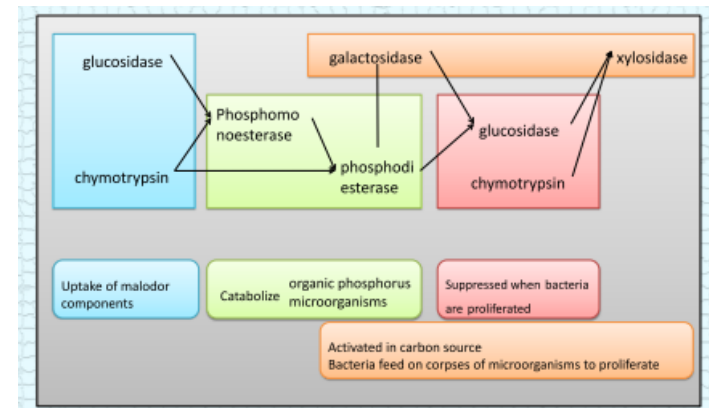


図2. トリメチルアミン分解時の酵素活性化ネットワーク

転写産物配列情報に基づく高精度遺伝子予測法の開発

生命情報工学研究センター 比較ゲノム研究班

【研究成果の概要】

・転写産物(完全長アミノ酸、cDNA、EST、RNA-Seq、ペプチド断片)配列を利用した真核生物の遺伝子構造予測法が現在までに数多く開発されています。私たちは転写産物対ゲノム配列間の様々なアラインメント法を相互に比較し、エキソン認識精度と計算速度を実データとシミュレーションを用いて検証しました。長いアミノ酸(>100aa)やcDNA(>300bp)配列を用いた実験では、私たちの開発したソフトウェア(Spaln2)が最も高精度であり速度的にも優れていることを確認しました。また、短いRNA-Seq配列(75K300bp)を用いた場合にも、試したプログラムの中で最も高精度でした。しかし、計算速度は最速のものより一桁程度遅く、改良の余地を残しています。

【平成24年度計画との関連】

小項目：II-1-(3)-<1> 配列情報と分子構造情報を用いた創薬支援技術開発、I-3-(1)-<2> 微生物資源や有用遺伝子の探索と機能解明

【開発技術の用途】

- ・ 効率的、高精度なゲノムアノテーション
- ・ 選択的スプライシングを含む、転写産物の網羅的解析(トランスクリプトーム解析)
- ・ 細胞状態(組織、疾患)特異的な転写様式の解明
- ・ 新世代DNAシーケンサーを用いたSNPs解析
- ・ 新世代DNAシーケンサーを用いたゲノムアセンブリー(参照ゲノムの有無を問わず)

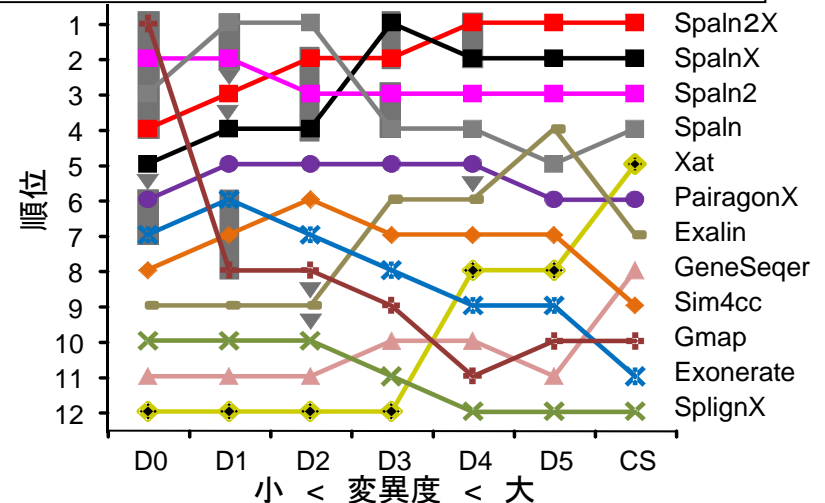


図1. アラインメント法の精度比較 (cDNA)

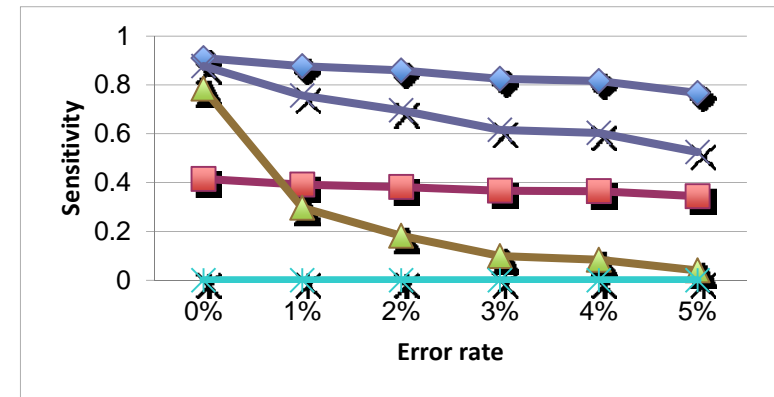


図2. RNA-Seqエキソン境界認識感度
シロイヌナズナゲノム、100bp reads

Spaln GMAP Tophat2 Blat Bowtie2

空間統計学を利用したタンパク質の機能解析手法の開発

生命情報工学研究センター 生体分子システム研究班 藤 博幸

【研究成果の概要】

・空間統計学を利用したタンパク質機能部位の予測

タンパク質の機能部位を構成するアミノ酸は保存的であり、立体構造上互いに近接していることが知られています。そこで、アミノ酸配列データの比較解析と、立体構造情報の二つから機能部位を予測するため、上記の性質を空間統計学を用いて評価することで、機能部位を同定する手法を開発しました(図1)。これによりパラログで機能分化が生じていても機能部位を正確に予測できるようになりました。

Nemoto, W. Toh, H. *BMC Struct Biol* **12**, 11 (2012)

・ケモカイン受容体ファミリーの機能分科関連サイトの同定

ケモカイン受容体はリガンド結合能とシグナリング能を持つGPCRですが、その近縁なホモログであるデコイ受容体はリガンド結合能は持つがシグナリング能は無く、ウイルス性受容体はリガンド結合がなくても構成的にシグナルを送り続けるシグナリング能力を持つことが知られています。この三者の機能差に関連するサイトをアラインメントの各サイトにおけるホモログ間でのアミノ酸組成の差に基づき検出しました(図2)。

Daiyasu, Nemoto, Toh. *Front Microbil* **3**, 264 (2012).

【平成24年度計画との関連】

小項目：II-1-(3)-<1> 配列情報と分子構造情報を用いた創薬支援技術開発

【開発技術の用途】

- ・タンパク質の機能部位の予測
- ・ホモロジーモデリングにおける適切な鑄型の選択
- ・ホモログ間で機能差をもたらすサイトの検出

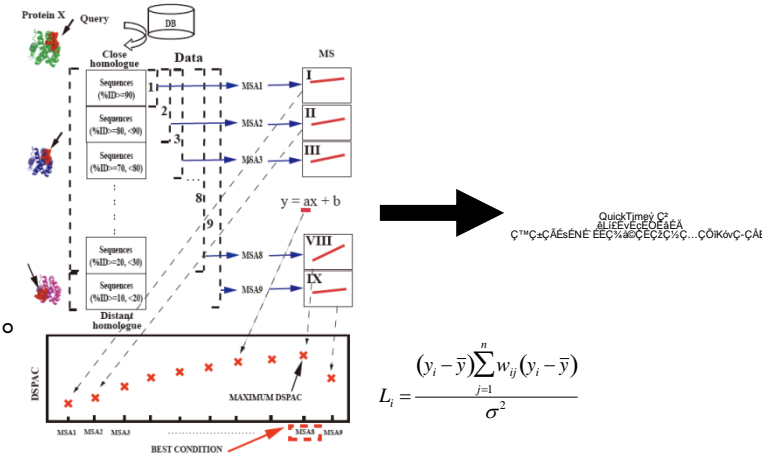


図1. 空間統計学を利用した機能部位予測

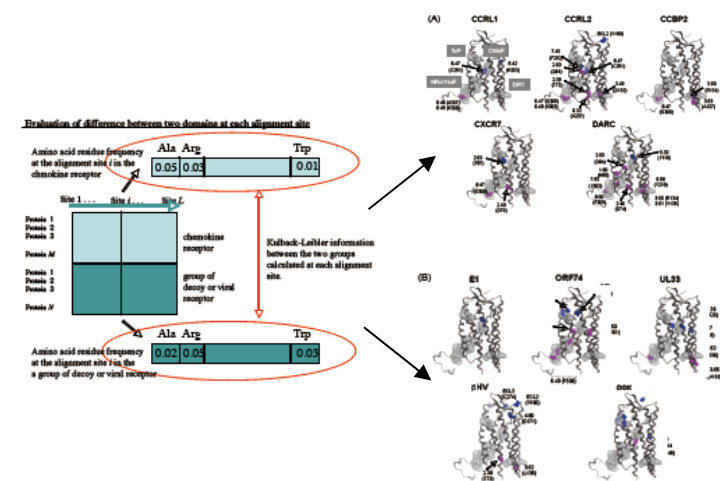


図2. ケモカイン受容体ファミリーの機能差関連サイトの予測.

染色体分配機構の解明に向けた動原体(キネトコア)構成タンパク質の相互作用解析 生命情報工学研究センター 細胞システム解析チーム 富井健太郎

【研究成果の概要】

・染色体の安定的な複製と分配は、生命現象の維持に不可欠であり、この機構に誤りが生じると、がん化など細胞に対する悪影響が生じることが知られています。染色体の複製と分配機構の解明は、複雑な細胞システムを理解する上で大変重要な基礎となります。細胞の分裂期に紡錘体が結合する染色体(のセントロメア)上に形成される、約100種類のタンパク質からなる巨大な複合体は、動原体(キネトコア)と呼ばれます。このキネトコア形成に重要な、タンパク質複合体におけるリン酸化を利用した新規の巧妙な分子間相互作用機構を共同研究者と共に明らかにしました(EMBO Journal, 2013)。

【平成24年度計画との関連】

小項目：II-1-(3)-<1> 配列情報と分子構造情報を用いた創薬支援技術開発

【開発技術の用途】

・細胞分化の効率化や制御に向けた人工的動原体の設計

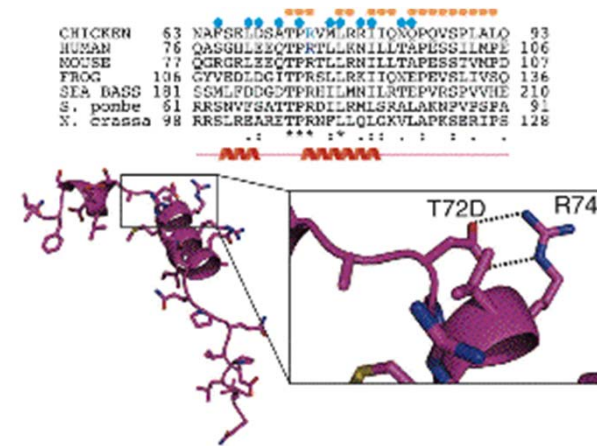


図1. 相互作用を制御するリン酸化部位のモチーフとその立体構造



図2. 相互作用機構の模式図

bisulfite処理シーケンシングデータの解析法

生命情報工学研究センター 配列解析チーム

【研究成果の概要】

哺乳類において、C塩基(シトシン)のメチル化は細胞分化に中心的な役割を担っている。bisulfite処理を施した条件でゲノム・シーケンシングを行うと、メチル化されていないC塩基はT塩基(チミン)として読まれる為、メチル化された位置が決められる。しかし、シーケンサーの誤読やゲノムの個人差などからメチル化された位置を正確に決めるのは必ずしも簡単ではない。また、全ゲノムを解析するため、計算時間を押える必要がある。先行研究ではさまざまな手法は開発され、利用されているが、感度と計算量において改良の余地は残されている。本研究では古典的な配列アラインメントを応用するだけで、競合するツールより良い性能が得られることを示した。

【平成24年度計画との関連】

小項目：II-1-(3)-<1> 配列情報と分子構造情報を用いた創薬支援技術開発

【開発技術の用途】

- ・ iPS細胞の品質管理。

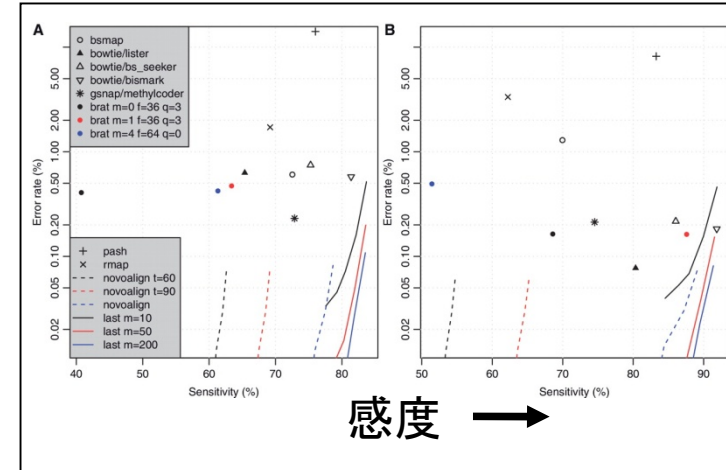


図1 本手法は競合する手法より感度が高い

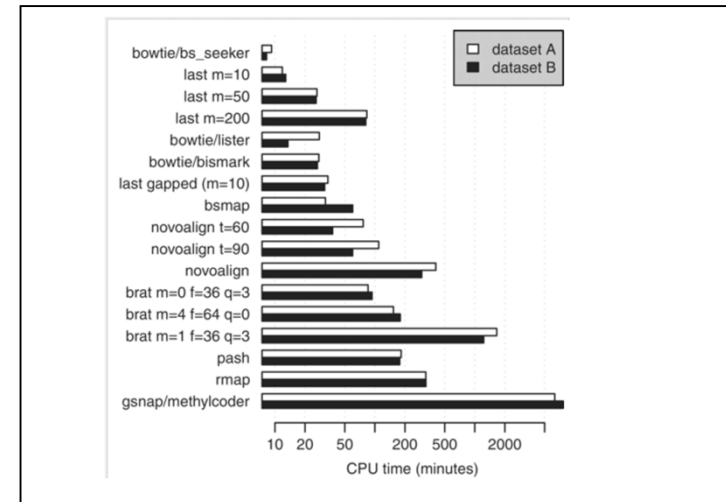


図2 本手法の計算は比較的速い

タンパク質・新規核移行シグナルの発見

生命情報工学研究センター 配列解析チーム

【研究成果の概要】

細胞質から核への移行はタンパク質の機能を制御する重要なメカニズムである。十数種類の「運び屋」タンパク質が千種類以上の「荷物」タンパク質を運ぶと考えられているが、どの運び屋がどの荷物を運ぶという仕分けの機序は十分解明されていない。本研究ではTrnという運び屋が運ぶ荷物タンパク質のアミノ酸配列を解析し、Trnともうひとつの運び屋であるImp-βが「荷札」として認識するアミノ酸配列の特徴を突き止めた。このシグナルは20残基程度の、負電荷が少なく、アルギニンが多い領域と定義できた。この研究は理研との共同研究で、配列解析は産総研側が担当し、理研側が実験を担当した。

【平成24年度計画との関連】

小項目：II-1-(3)-<1> 配列情報と分子構造情報を用いた創薬支援技術開発

【開発技術の用途】

- ・ 癌など、細胞制御異常疾病に対する創薬の標的探索と作用機序解明

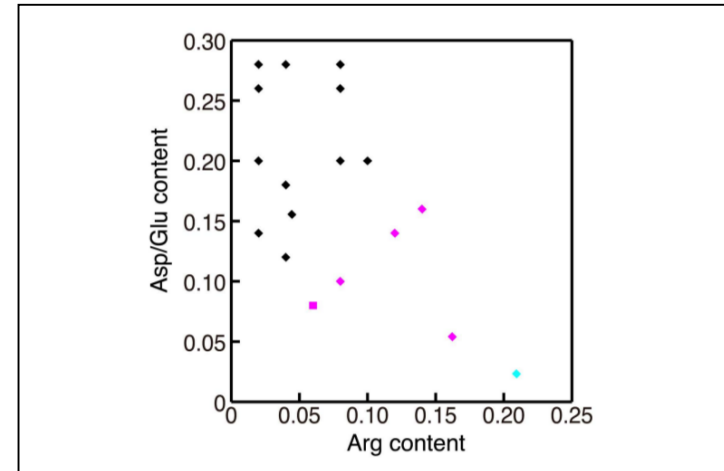


図1 新規シグナル(色)はアルギニンが多く、負電荷が少ない。

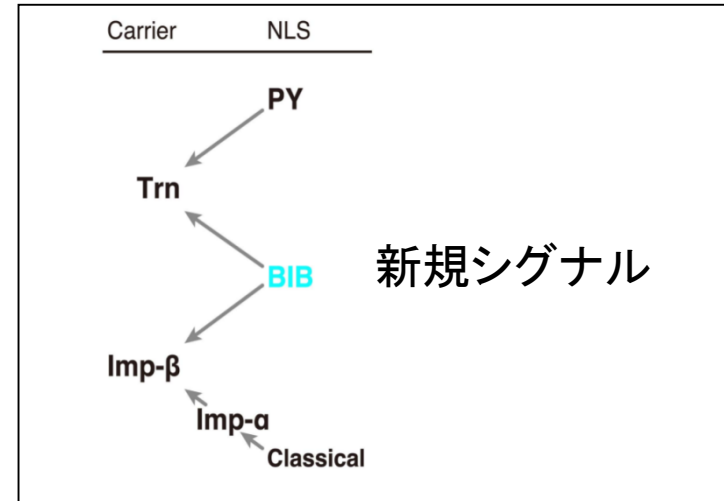


図2 新規シグナルはTrnにもImp-βにも認識される。

ライフサイエンス情報統合基盤技術開発

生命情報工学研究センター 情報基盤統合チーム 福井一彦

【研究成果の概要 Outline of your achievement】

ライフサイエンス研究分野では、その細分化や専門化に伴い、様々なデータベース、解析ツール、予測ソフトウェアが開発され膨大な数に達しており、これらを有効かつ効率よく利用するには様々な課題を克服する必要があります。情報基盤統合チームでは、これら分散するソフトウェア解析やデータベース検索等を、一つ一つユーザが実行するのではなく、一連の処理の流れを定義することにより、効率的に短時間に実行するためのプラットフォームを利用したワークフローの技術開発を進め、統合化を目指した実用的な情報基盤開発に取り組んでいます。またデータベースの標準化(RDF化)に伴い、開発した高度な解析ツール群を広く利用可能とするために、セマンティック技術に対応したフレームワークを利用して、解析ツールにRDF入出力機能を追加し、連携型データと組み合わせた大規模解析を可能とする解析サービスやツール群のオントロジー開発を行っています。

【平成24年度計画との関連】

小項目：II-1-(3)-<3> バイオデータベース整備と利用技術の開発

【開発技術の用途】

- ・ライフサイエンス研究分野における専門化・細分化されたデータベースや解析ツールの迅速な変化や新規情報技術へ柔軟に対応可能なワークフロー解析として利用できます。
- ・セマンティック技術の利用によりデータベースと解析ツールを組み合わせたシームレスな大規模大量データ解析を可能とする技術開発へと繋がります。



図1. CBRC統合DB情報基盤サイト

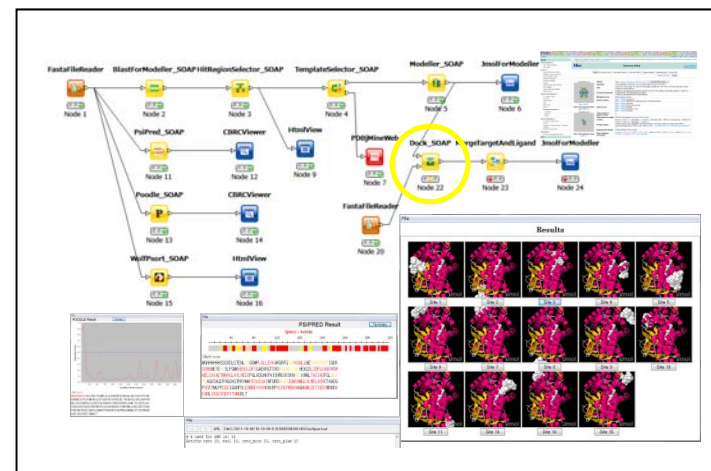


図2. プラットフォームによる解析例

楕円曲線符号を用いた高速な秘密検索アルゴリズム

生命情報工学研究センター 機械学習研究班 津田宏治

【研究成果の概要】

前年度より開発中の秘密計算を用いた化学化合物データベースの検索技術をさらに発展させるため、後期戦略予算「データベースの秘匿検索技術の開発」を獲得し、以下の成果を得た。楕円曲線暗号を用いることにより、暗号演算部分の実装を10倍程度高速化した。新規暗号を実装したことで、既存特許に抵触する危険性を回避することができた。数百万件の化学分子データベースであるPubChemに適用を行い、実用的な検索システムを作り上げた。

【平成24年度計画との関連】

小項目：II-1-(3)-<3> バイオデータベース整備と利用技術の開発

・バイオデータベースの利用に関する暗号理論を用いた情報保護技術を開発・発展させ、実用的なシステムとすることで、オープンイノベーションの促進を目指す。

【開発技術の用途】

- ・秘密計算技術の実用化による生物学的データベース検索におけるセキュリティの改善
- ・本技術の応用により、安全なゲノム診断を行うことも可能



図1：データベースの秘密検索の背景

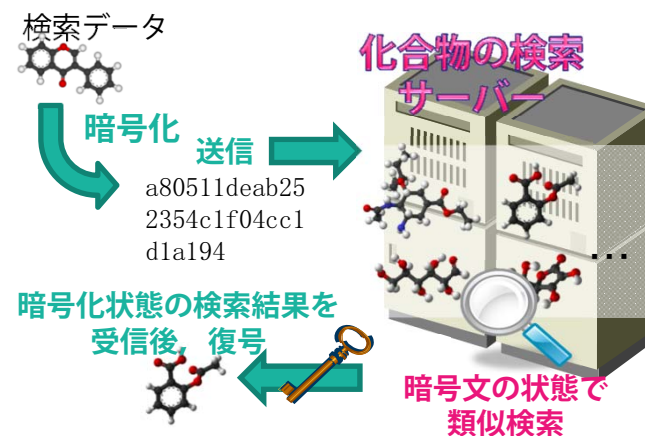


図2：暗号技術を用いた秘密検索