

## 秘密計算による化合物データベースの検索技術

## 生命情報工学研究センター

### 【研究成果の概要】

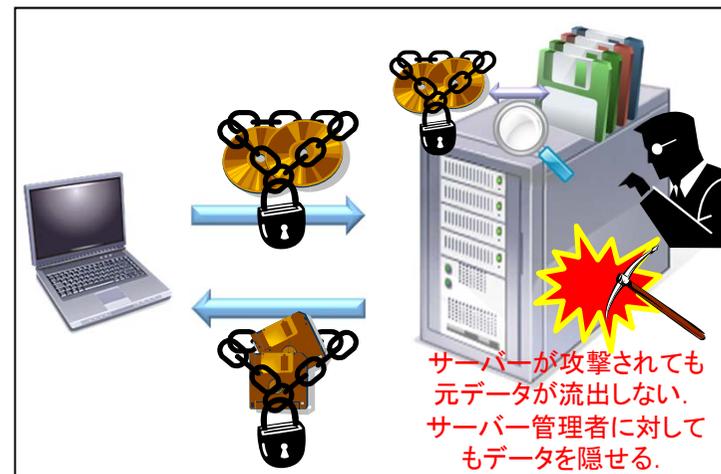
創薬などに用いられる化合物の情報は企業秘密として厳重に管理されるため、外部データベースに情報を送って類似化合物の検索を行うことが難しかった。今回開発した技術により、ユーザー側とサーバー側の双方が互いに情報を開示することなく、互いのデータを暗号化したままで比較し、検索結果だけを得ることができる。この技術では、データを加法準同型暗号により暗号化し、加算だけで検索を行えるアルゴリズムによって、大量のデータに対して実用的な速度で検索を実行できる。今回開発した技術は、企業間での安全かつ効果的な情報交換を促進し、創薬におけるオープンイノベーションに貢献すると期待される。また、本技術はゲノム等、他のデータにも応用することが可能。

### 【平成23年度計画との関連】

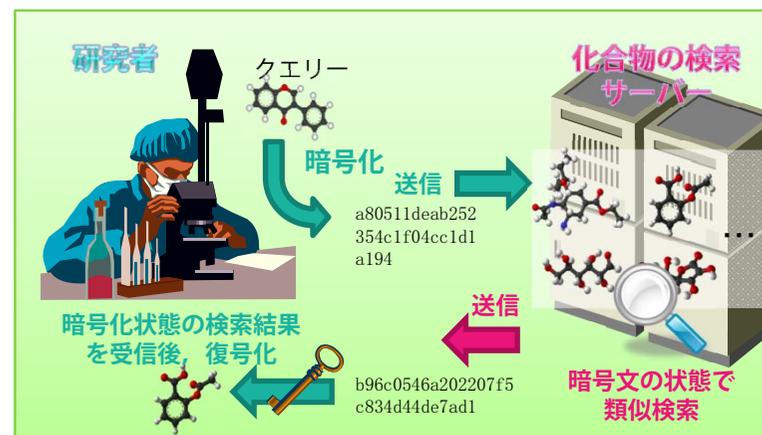
小項目:II-1-(3)-<3> バイオデータベース整備と利用技術の開発

### 【開発技術の用途】

- ・フォーカスライブラリの検索
- ・個人ゲノムによる疾患診断



全工程で暗号化されたデータを使用。情報漏えいのリスクが大幅に低下



化合物検索の概要

## 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

### 【研究成果の概要】

・プラットフォームを用いデータベースと解析ツールを組み合わせた大規模大量データ解析を可能とする技術開発

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。生命情報工学研究センターでは、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ワークフローに関する技術開発を進め、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る。(図1)

### 【平成23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  
 ・遺伝子や生体分子に関する情報の高度な利用を促進する解析サービスやワークフロー等を構築する。また、ヒトの遺伝子、RNA、タンパク質、糖鎖情報の整備及び統合を行うとともに、診断技術等の利用技術を開発する。

### 【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・細分化された様々なデータベースと解析ツールのシームレスな連携・統合による知的プラットフォームとして利用できる。(図2)



図1. CBRC統合DB情報基盤サイト

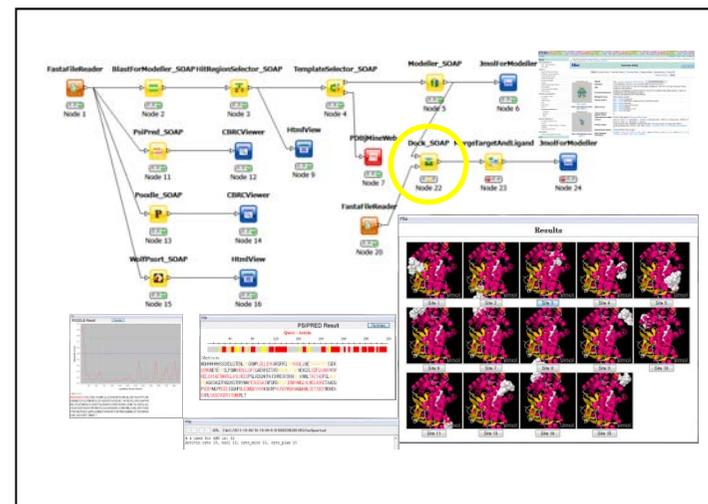


図2 プラットフォームによる解析例

## 高速近傍探索法を用いた類似リガンド結合部位の網羅的発見とそのデータベース化 生命情報工学研究センター

### 【研究成果の概要】

・タンパク質は多くの場合、他の分子(リガンド)との相互作用を介して機能を発現することが知られています。つまり、タンパク質が持つ他分子と結合する領域(リガンド結合部位)は、タンパク質の機能を特徴付ける非常に重要な場といえます。本研究では、SketchSortと呼ばれる非常に高速な近傍探索アルゴリズムを開発しました。この方法を用いることによって、通常のPCで、120万の既知及び潜在的リガンド結合部位の類似ペア(約8000万点)の列挙を約4時間で行うことが可能になりました。一般の研究者がこの類似度ネットワークを用いて研究を行えるよう、データベースPossumを立ち上げ、Web上に公開した。

### 【平成23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  
・遺伝子や生体分子に関する情報の高度な利用を促進する情報データベースやポータルサイト等を構築する。また、ヒトの遺伝子、RNA、タンパク質、糖鎖情報等の整備及び統合を行うとともに、診断技術等の利用技術を開発する。

### 【開発技術の用途】

- ・創薬における副作用探索
- ・次世代シーケンサーデータの高速アセンブリ
- ・製薬企業に蓄積されている多数の化合物の検索、クラスタリング

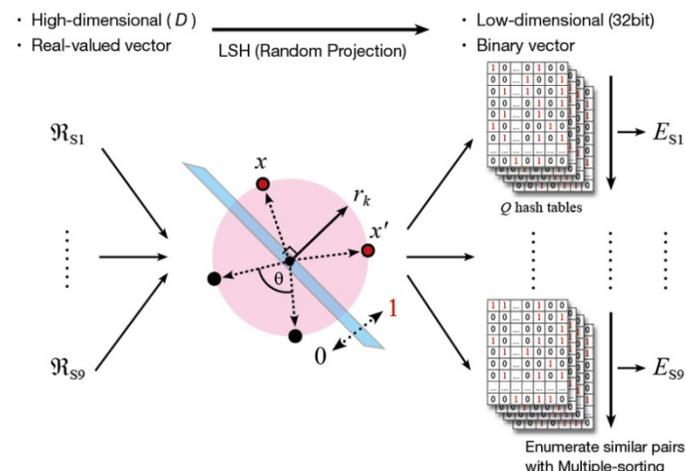


図1. SketchSortアルゴリズムの概要

Summary of your query

PDB ID	1D3Q
HET code	ADP
Chain name	Any
CATH code of query site	3.40.50.720
SCOP code of query site	c.4.1.1
EC number of query chain	1.5.99.7
Function of query protein	OXIDOREDUCTASE
Target DB	Known and Putative binding sites
Cosine similarity cut-off	0.77
Aligned length cut-off	7
Maximum number of hits	1000
Search mode	1
Assign CATH to hits	Yes
Assign SCOP to hits	Yes
Assign EC to hits	Yes

Search Result  
Number of hits for known ligand binding sites: 1000  
Number of hits for putative sites: 752

[Download results as a text file](#)

No. of hits	PDB ID	HET code	HET_chain	Cos-value	Aligned length	RMSD(Ca)	Function	CATH code	SCOP code	EC number
No.1	1D94	ADP	A	0.996	29	0.13	ELECTRON TRANSPORT	3.40.50.720	c.4.1.1	1.5.99.7
No.2	1D94	ADP	B	0.996	29	0.13	ELECTRON TRANSPORT	3.40.50.720	c.4.1.1	1.5.99.7
No.3	1E1L	FAD	A	0.813	29	1.32	OXIDOREDUCTASE	3.40.50.720	c.4.1.1	1.18.1.2
No.4	2IHD	ADP	A	0.994	29	0.11	OXIDOREDUCTASE	3.40.50.720	c.4.1.1	1.5.99.7
No.5	2IHD	ADP	B	0.992	29	0.13	OXIDOREDUCTASE	3.40.50.720	c.4.1.1	1.5.99.7
....										
No.1001	3K09	PRB	X	0.91	27	2.03	OXIDOREDUCTASE	0.0.0.0	0.0.0.0	1.8.1.14
No.1002	3DVC	PRB	X	0.891	26	3.18	OXIDOREDUCTASE	3.40.50.720	0.0.0.0	1.1.1.1
No.1003	1D5I	PRB	X	0.889	26	2.99	OXIDOREDUCTASE	3.40.50.720	c.2.1.2	1.1.1.100
No.1004	2Q45	PRB	X	0.825	26	3.18	OXIDOREDUCTASE	3.40.50.720	c.2.1.2	1.1.1.1

図2. Webサーバ

## ヒト細胞・細胞分化データベースCELLPEDIA 生命情報工学研究センター

### 【研究成果の概要】

平成21年度に公開したヒト細胞の網羅的分類体系に基づく細胞情報データベースを著名な学術論文の出版社であるOxford JournalのDatabase誌に報告した。ヒト分化細胞2,260種および幹細胞66種の独自分類バックボーン上へ遺伝子発現、細胞画像、論文等を統合して検索できるようにした細胞情報データベースシステム。

### 【平成23年度計画との関連】

小項目：1-(3)-② システム生物学的解析を用いた創薬基盤技術の開発

- ・ 転写制御、シグナル伝達、代謝に代表される、細胞内のネットワーク、パスウェイ等の推定やシミュレーションにより、創薬に必要な化合物の設計と合成、標的分子を推定する技術を開発する。

### 【開発技術の用途】

- ・ 異なる細胞間での比較遺伝子発現解析を行うことで細胞分化に有効な転写因子等の探索が可能である。
- ・ 今後、増加すると考えられる人工細胞の検査のため、リファレンス細胞として情報を提供することが期待される。
- ・ 細胞進化の解析支援が可能。

The screenshot displays the CELLPEDIA database interface. At the top, there's a navigation bar with 'Top | Manual | Data submission' and 'Release Notes'. The main content area includes a search bar with 'Human Cell Taxonomy' selected. Below the search bar, there are several panels: 'Examples of Human Differentiated Cell' showing images of 'Capitulum trophi' and 'Mouth Fibroblast (Cell ID:10020)'; 'Gene Expression Neighbors of Human Cell' showing a tree view of gene expression data; 'Gene Expression' showing a heatmap and a table of gene expression data; and 'Gene Ontology' showing a list of GO terms. Overlaid on the screenshot are six callout boxes with labels (a) through (f) pointing to specific features.

- (a) 細胞画像401
- (b) 細胞の形態
- (c) 関連文献422
- (d) オントロジー
- (e) 分化する細胞へのリンク934
- (f) 遺伝子発現データ970

ヒト細胞を大学、研究所、UCSF、Stanford大、企業などから画像、遺伝子発現、論文を登録して公開。Hatano et al. Database, Vol. 2011, bar046, 2011.

## タンパク質間相互作用制御化合物のインシリコスクリーニング 生命情報工学研究センター

### 【研究成果の概要】

・ NEDO「化合物等を活用した生物システム制御基盤技術開発」PJにおいて、タンパク質タンパク質ドッキング計算から相互作用部位を予測し、リガンドドッキング計算や分子動力学計算を用いて化合物をスクリーニングする戦略を基本に、様々なタンパク質間相互作用標的と制御化合物の探索を実施した。複合体タンパク質の結晶構造を用いた統計解析によれば、タンパク質-タンパク質間相互作用表面は、800 Å<sup>2</sup>を1区間とした複数のパッチ構造で形成されていることが知られており、PJで実施した複合体タンパク質標的をこの基本構造や天然物スクリーニング結果等を参考に、図のような分類を新規に提案した。(図1)

### 【平成23年度計画との関連】

小項目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

・創薬タンパク質-タンパク質相互作用を標的としたバーチャルスクリーニングでは、複合体形成の作用機序別に効率よく実験と計算機を連携することでスクリーニング効率を向上する。

### 【開発技術の用途】

・タンパク質間相互作用の様式に応じてインシリコスクリーニングの戦術を変えることで、効果的な化合物探索が実現できます。  
・タンパク質間相互作用様式毎のフォーカスライブラリーなどの設計が期待される。(図2)

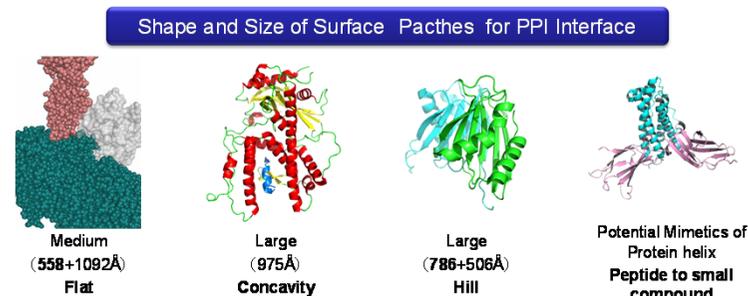


図1. インシリコスクリーニングにむけたタンパク質間相互作用の分類

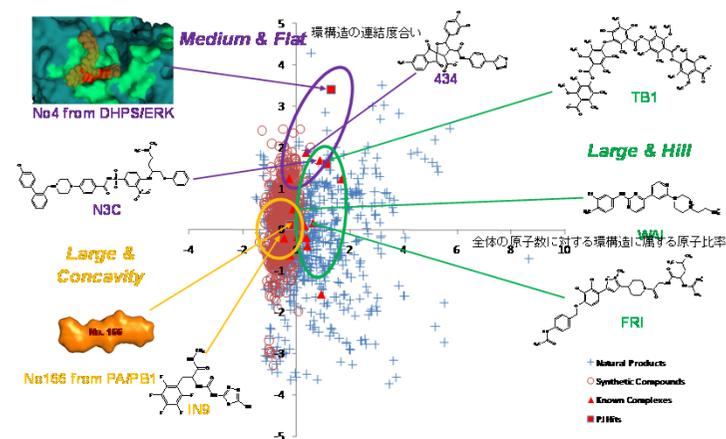


図2. 合成化合物と天然物による化合物空間の解析と、タンパク質間相互作用様式ごとの阻害剤分布の対応付け



## ネットワーク解析技術に基づく高精度・高効率創薬基盤技術開発 生命情報工学研究センター

### 【研究成果の概要】

平成22年度に改良した制御ネットワーク候補の絞り込み技術をiPS細胞、ES細胞、始原生殖細胞等の多能性幹細胞や前立腺がん、肝がん、糖尿病等の疾患細胞に適用し、少数要因候補制御因子を同定し、共同研究者により実験検証が開始された。さらに新規に、予後データ等のアウトカムデータから分子機能を推定する技術の開発と網羅的タンパク質量計測システムの開発とを開始した。

### 【平成23年度計画との関連】

小項目：II-1-(3)-<2>システム生物学的解析を用いた創薬基盤技術の開発

- ・ 要因候補遺伝子の絞り込みを10分子程度に限定する。
- ・ アウトカム重視の解析法により、信頼性の高い要因候補を同定可能にする
- ・ 網羅的タンパク質量計測システムの開発を完成し、プロテオミクスデータについて要因候補タンパク質の同定を行う。

### 【開発技術の用途】

- ・ 疾患等の要因遺伝子の実験検証を促進し、創薬開発の効率化に貢献する
- ・ 直接的機能を担うタンパク質候補やアウトカム指向の分子機能同定により、高精度な創薬開発に貢献する。

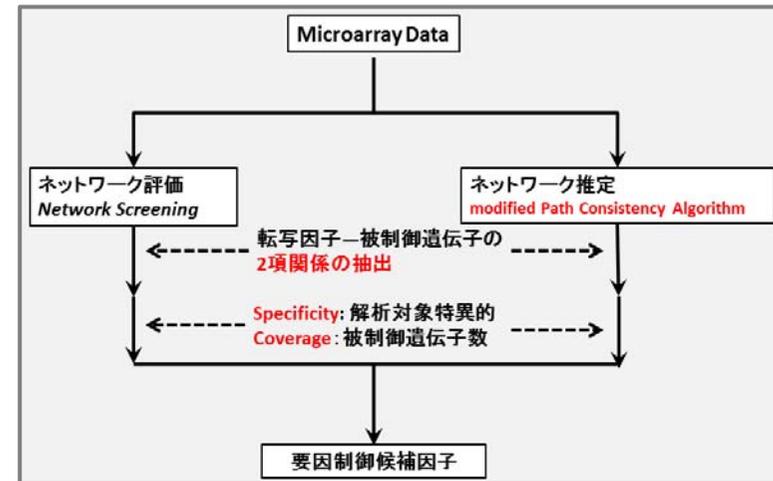


図1 ネットワーク解析による少数要因候補制御因子の絞り込み

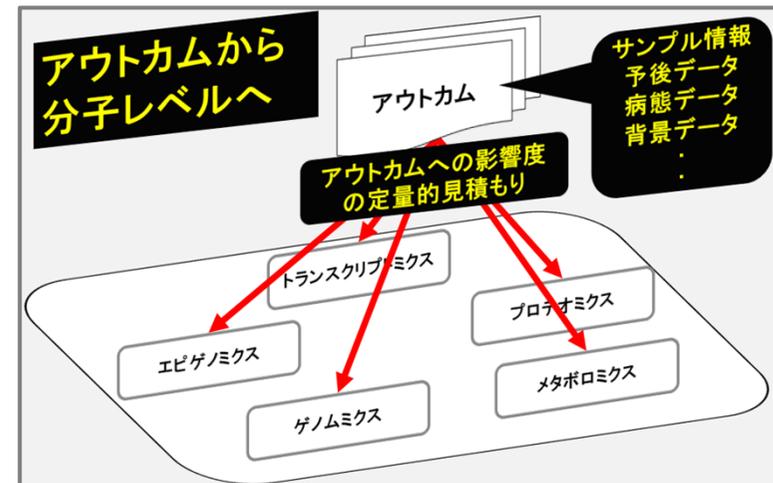


図2：アウトカムデータ指向の分子機能推定技術の開発

## タンパク質・核移行シグナル予測法の開発

## 生命情報工学研究センター

### 【研究成果の概要】

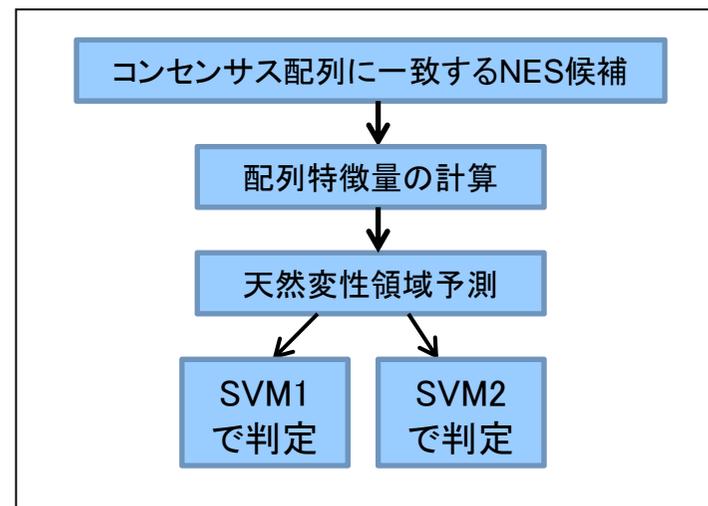
癌遺伝子 (P53など) とウイルス・タンパク質 (HIV-1 Revなど) の制御に重要である配列要素、核外移行シグナル (NES) の予測法 (NESsential) を開発した。NESと天然変性領域の関係をモデルに取り入れ、予測精度を高めた。更に、性能比較を行い、NESsentialが新規NES候補のスクリーニングに有効である、世界初かつ唯一の手法であることを示した。

### 【平成23年度計画との関連】

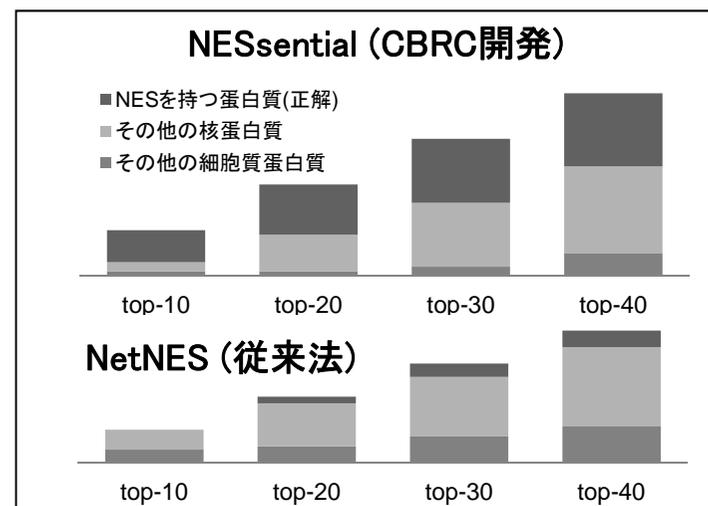
小項目: 1-(3)-① 配列情報と分子構造情報を用いた創薬支援技術開発。H23年度計画には、NES予測には明確に触れていないが、新たに発展したテーマ。

### 【開発技術の用途】

- ・ゲノム配列から新規NES候補の発見
- ・新型ウイルスにおける変異の影響予測
- ・癌腫瘍における癌遺伝子変異の影響予測
- ・NES変異が関与している薬剤耐性獲得の解析



NESsential、NES予測の流れ図



正解率は従来法を大幅に上回る

## 次世代創薬標的探索のためのインターフェイス予測技術の開発 生命情報工学研究センター

### 【研究成果の概要】

Gタンパク質共役型受容体(GPCR)は市販薬物の3割が標的とする、創薬における重要なタンパク質です。近年、GPCR同士が複合体を形成して機能することが明らかとなり、また、この複合体形成とパーキンソン病など様々な疾患との関連性が指摘されています。複合体を形成する際のインターフェイスを高精度で同定し、既存のGPCR標的薬とは異なる作用機序に基づく、創薬支援を目指している。

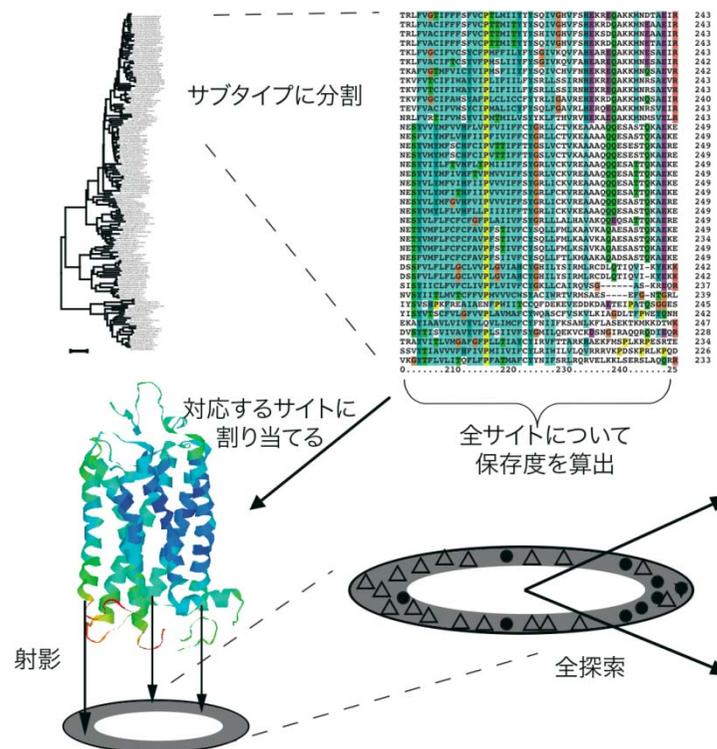
### 【平成23年度計画との関連】

小項目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

・アミノ酸配列情報と立体構造情報を利用し、GPCRが複合体を形成する際のインターフェイスを予測するサービスとDBを構築している。

### 【開発技術の用途】

- ・医薬品・実験試薬が標的とすべき領域の同定
- ・GPCR多量体形成が関与する疾患メカニズムの理解



<http://grip.cbrc.jp/>