

# 論文内容の要旨

論文題目:

## Representation of Signal Transduction Pathways : Toward Computational Analysis of Signal Transduction Pathways

(シグナル伝達経路の再構築：シグナル伝達経路の解析に向けて)

氏名: 福田 賢一郎

本論文では、生物のシグナル伝達機構に関する知識を形式化し知的システム上に共有するための方法論に関する研究成果を報告している。

細胞内シグナル伝達の研究は細胞外からの刺激が核内に伝達され細胞応答が引き起こされる制御のネットワークを明らかにする研究である。分子生物学の分野で近年めざましい進展を遂げており、発生における細胞の運命の決定や細胞遊走、青色光への反応など多種多様な生命現象の分子機構を説明している。しかしながら、新しい研究分野であるが故に関与する構成因子の機能分類や相互作用の分類などの知識基盤が未整備であり、その知見は自然言語および情報の流れを描いたダイアグラム図として論文中に埋没しているのが現状である。これらの知見を計算機利用可能にするために、(1) 知識表現手法、(2) 推論手法、(3) 文献からの知識の抽出手法の確立が望まれる。

博士論文の主要な章の内容は以下の通りである。

第2章では、ポストゲノム時代における生命科学研究の知識依存性について議論している。様々な体内ネットワークおよび関連研究を比較し、本研究の枠組みを提示している。

第3章では、断片的かつ記述の知識粒度が不揃いで階層的であるというシグナル伝達に関する専門知識の情報構造を明らかにし、これを自然に記述するための知識表現について論じている。シグナル伝達は a) 現象の認識; b) 構成要素の認識; c) 介在する相互作用の認識; d) 構成要素と相互作用の結線関係の認識; e) 要素・作用のより詳細な認識もしくは下流の現象の認識; というサイクルを繰り返すことで漸近的に既存知識に新規知識を追加して知見を詳細化する。この結果シグナル伝達経路の知見は抽象度の異なる知識、不完全な知識が混在した記述となる。これらの問題を解決するために複合グラフ構造に基づいた知識表現モデルを提案した。複合グラフはグラフ構造の拡張で、木  $T = (E^T, V, r)$  とグラフ  $G = (E^G, V)$  の対で定義される。各節点自身が階層的に部分グラフを内包できる構造であり、節点の包含関係は上記の木によって定義される。シグナル伝達パスウェイの各要素を複合グラフの節点と対応づけることにより、任意の粒度の知識を自由に記述することが可能になった。図1に示した例は TGF- $\beta$ /Smad および

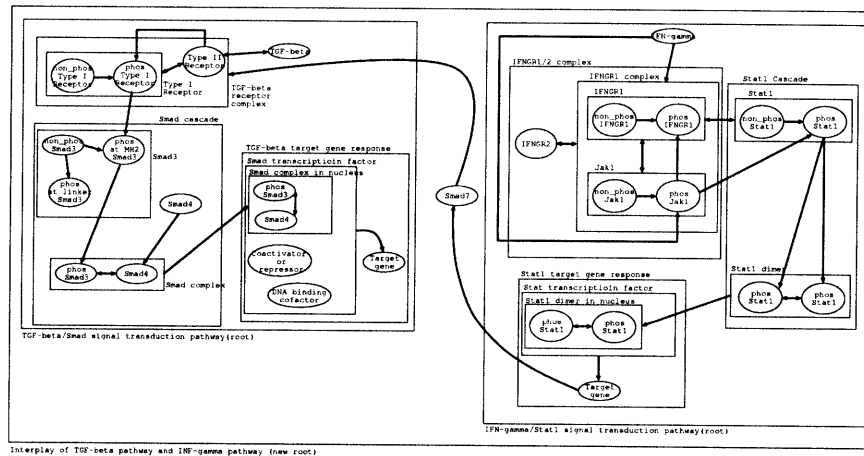


図 1: TGF-β/Smad および IFN-γ/Stat1 パスウェイの複合グラフによる表現

IFN-γ/Stat1 パスウェイの複合グラフによる表現である。

第 4 章では生物学者が行っている“生物学的に正しい推論”を事例ベースに対する仮説推論としてモデル化し、知識ベース上で実験事実と矛盾しない推論を行うシステムの設計に関して論じている。シグナル伝達知識は多段階の実験結果を傍証として生物学者が導き出した細胞メカニズムの現状における認識結果であり、あるシグナル伝達パスウェイに関する記述が生物学的に妥当であるかを判断する際には人間の判断基準を離れて客観的な形式的指標を導入することができない。このため、生物学者が正しい新規知識を獲得するプロセスを背景知識と実験データに矛盾しない観測事実を説明する仮説の獲得行為としてモデル化し、論理に基づく仮説推論の枠組みを応用した。提案手法に基づいて実際に過去の知見を組み合わせて新規の知識を獲得するシステムを実装し提案手法の有効性を実証した。

第 5 章は文献から知識を抽出する際の支援環境の設計・実装について述べている。3 章で提案した知識表現モデルが実際に生物学者が抱くシグナル伝達経路の表象を自然に表現しており、その結果、文献中の自然言語・図による記述を形式化するための入力環境を容易に構築できることを実証している。

第 6 章は生物医学文献から知識を抽出する手法について論じている。知見が文献中で自然言語によって表現されているシグナル伝達経路においては情報抽出技術の応用が期待されており、未知語の類出する同分野において適切に目的知識を抽出するための要素技術としてタンパク質名の同定方式を提案している。

第 7 章では、提案手法の応用について論じている。将来の課題としては過去事例の検索を行う方式、構成要素の一致の判定の方式および知識ベースの更新方式の確立をあげている。