

# システム生物学・創薬インフォマティクスにおける統計科学のモデリング技術

吉田 亮 モデリング研究系・データ同化研究開発センター 准教授

膨大な生命情報の蓄積は、ゲノム解析や創薬研究における化合物データ解析など、バイオサイエンスの研究開発の在り方に大きな変化をもたらしています。昨今の超高速シーケンサの普及により、大量のゲノムシーケンスが世界中の研究拠点において驚異的なペースで蓄積されており、ビッグデータ・ソリューションとしての統計科学への期待は益々高まっています。また医薬品開発の現場では、化合物アッセイデータにもとづく候補化合物の薬理活性・毒性プロファイルの予測や、デノボ化合物の分子設計など、統計科学の解析技術は様々な局面において役立っています。ここでは、バイオサイエンスに対する統計科学の取り組みとして、(1) ベイズ・モデリング技術を使った生体分子ネットワークの構造設計と制御、(2) カーネル逆像推定にもとづくデノボ化合物の分子設計法、を紹介いたします。

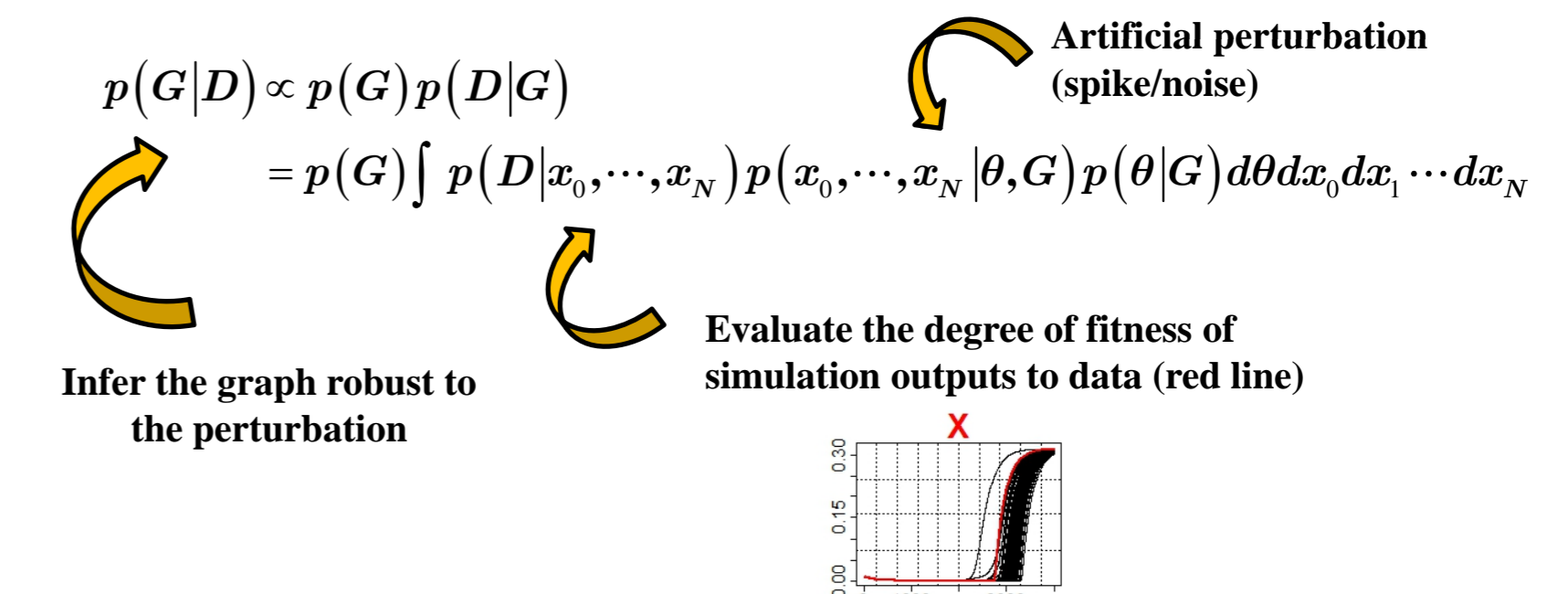
## 生体分子ネットワークの逆問題：ロバストなネットワークを設計する

遺伝子発現・シグナル伝達・代謝反応など、細胞プログラムの本質的な機能は生体分子の相互作用ネットワークに書き込まれています。本研究室では、生体分子ネットワークの動作・設計原理の解明に役立つバイオモデリング技術とデータ解析手法の開発を行っています。ここでは一例として、細胞のロバスト性に関する研究を紹介します。

生物は様々な外的変化に柔軟に対応しながら、生命システム全体の安定性や恒常性を維持しています。例えば、外界温度や生化学パラメータの変化に対して、個々の素過程は高い揺らぎを示しますが、生物の発生や分化のプログラムは驚くほど厳密な規則性を保ちながら正確に遂行されます。近年の分子生物学の研究は、生命システムのロバスト性を維持するメカニズムとして、(1) 反応経路の冗長性; (2) 生化学物質の冗長性; (3) ネットワークの構造的安定性; (4) モジュール性などを明らかにしてきました。

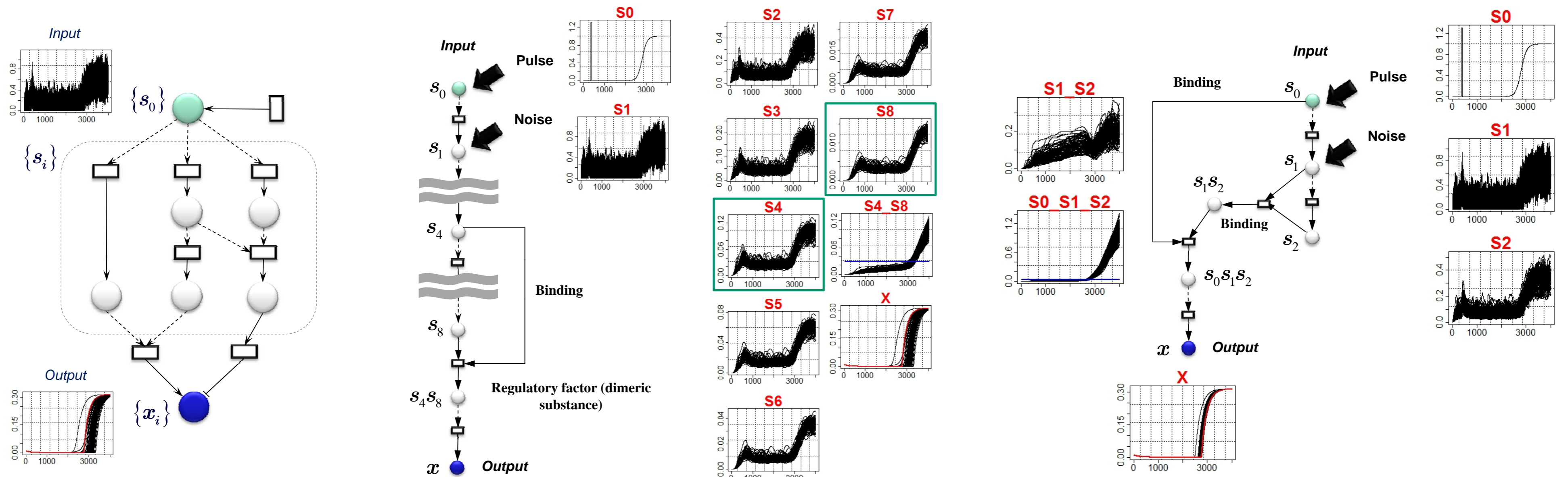
本研究室では、このようなロバストなネットワーク・システムを自動設計するためのモデリング手法の開発を行っています。技術基盤は階層ベイズのモデル設計原理と確率的なグラフ探索アルゴリズムが礎となっています。直観的に説明すると、次のようになります：システムに任意の摂動を与えたもとの、ベイズ的なモデル設計規準に従ってモデル(ネットワーク構造)を改変していくことで、摂動の影響を緩和しつつ観測されたデータを安定的に再現できるロバストなネットワーク・システムを得ることが可能ということです。このような技術は、ロバスト性の維持に関わるネットワーク・モチーフの発見や代謝経路の工学設計に応用することができます。

ベイズ統計学のモデル設計原理とロバスト性の関係：システムの摂動をモデル化した上で、モデルの事後確率が高くなるようにネットワーク・システムの構造Gを改変していけば、摂動の影響を取り除きながらデータのパターンDを安定的に再現できるモデルを自動構築することができる。



例：入力シグナルにパルス+ノイズを与えたもとの、出力Xの反応遅延と位相の揺らぎを同時に除去できるネットワーク構造を設計する。(赤：観測データ)

推定されたネットワーク構造：左のモデルの主要構成因子は、直列に連結された8個の中間変数によるシグナル伝達経路である。これに加え、二つの中間変数が複合体を形成し、これが出力Xを最終的に制御している。出力Xのパターンを見ると、摂動の影響をある程度吸収することができるが、Xの発火タイミングは元のシグナルに比べて遅延傾向にある。これに対して右のモデルは、摂動に対してロバストであると同時にシグナルのタイミングを正しくXに伝達している。



## 創薬支援システムの開発：薬を設計する ※山下博史さん(総研大)、伊庭幸人准教授(統数研)との共同研究

本研究室では、カーネル法を用いたドラッグデザインの新しい手法を開発しています。ここで、次の二つの研究課題について紹介します：(1) 標的分子に対する候補化合物の薬理活性や毒性プロファイルを予測するための判別モデルの開発;(2) 薬理活性や毒性プロファイルの向上を目的としたデノボ化合物の設計

化合物の化学構造を縮約表現するためのツールとして、新しいグラフカーネルの設計に取り組んでいます。化学構造カーネルの設計は、個々の化合物をラベル付き無向グラフで表現することから始まります。グラフカーネルの本質は、任意の二つの化合物に共通する部分構造を数え上げ、化合物の類似性を定義することです。このような研究は既に1990年代後半から始まっており、現在までに数パターンの代表的な手法が確立されています。従来法は、類似性の評価において、完全に一致する部分構造のみを反映しているため、数原子のミスマッチがあるような部分構造を「類似度」として考慮しません。そこで我々は、部分構造の完全一致性という制約を緩和することが可能な、より柔軟な化学構造カーネルを開発しました。我々はこれを原子環境カーネルと呼んでいます。

医薬品開発におけるもう一つの課題は、薬理活性や毒性プロファイルが向上するように化合物を改変することです。既存化合物のアッセイデータから判別モデルを構成する過程で、活性サンプルに特異的な幾つかの化学構造を知ることができます。ここでは、この過程で得られた複数の化合物を「混ぜる」ことで、新しい化合物を設計する方法を紹介します。

アイデアは非常に単純です。まずは、混合対象の化合物の特徴ベクトルを凸結合した上で、再生核ヒルベルト空間上における新規化合物と混合化合物のノルムをポテンシャル関数にセットし、ギブス分布を構成します。このギブス分布からラベル付き無向グラフをサンプリングすることで、再生核ヒルベルト空間における混合化合物の内点付近にどのような化合物が分布しているかを知ることができます。我々はフラグメント・マルコフ連鎖モンテカルロ法という手法を使ってこのような化合物グラフのランダム・サンプリングを実現しました。

フラグメントMCMCシミュレーションによるデノボ化合物生成

