# コンピュータで
# ライフサイエンスする

## この際だからドライな生活も経験してみよう

東北大学情報科学研究科
中尾光之

# この際だからリアリティの外から考えてみる

- 医学・生物学にとっては何よりもリアリティが重要視される。
- ドライな手法はあくまでも支援者として、ウエットな方法で行われるリアリティの解明や追究を助ける役割を担う。
- だけど、幸運な場合には、ドライな手法によって、データの医学・生物学的理解や解釈や予想しなかった洞察がもたらされることもある。

**この際だから、ドライな生活も経験してみませんか。**
**それによってウェットな生活も充実するかもしれません。**

# 3つのDry

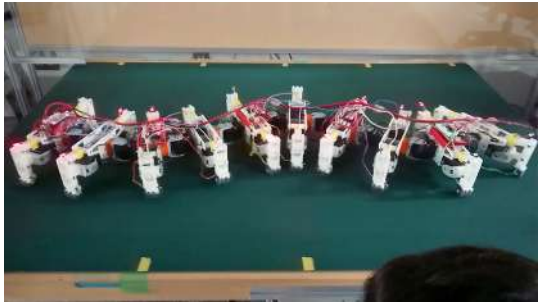
- Synthetic
- Data
- Modeling

# Synthetic Dry:生物を作っちゃえ

東北大学
電気通信研究所
石黒・加納研究室

# Synthetic Dry: 作っちゃったロボット生物たち

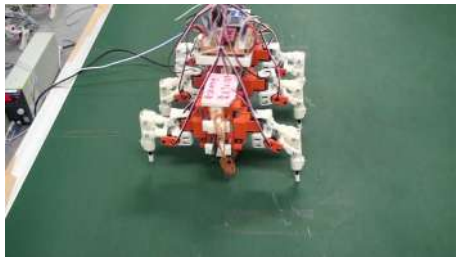
四足歩行



ムカデ

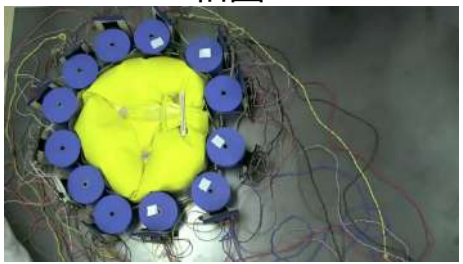
ヘビ

昆虫

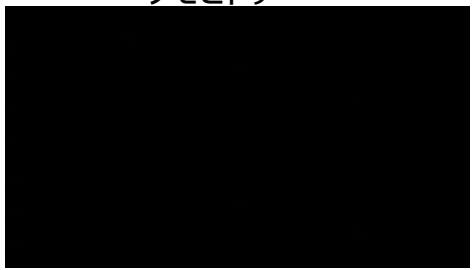


ヘビ

粘菌



クモヒトデ



東北大学
電気通信研究所
石黒・加納研究室

# Synthetic Dry: 四脚動物の歩容遷移

速度に応じて歩容（足並み or 脚運動の位相関係）が遷移する！

| 低速 | | 中速 | | 高速 |
|---|---|---|---|---|
| ウォーク（Walk） | ⇔ | トロット（Trot） | ⇔ | ギャロップ（Gallop） |

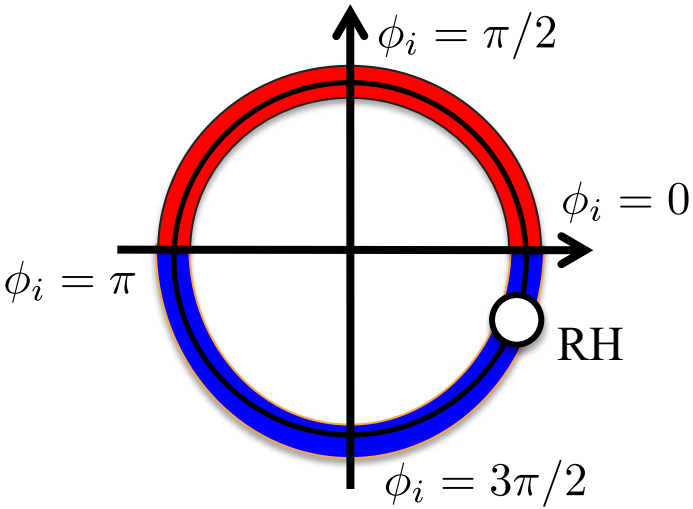
# Synthetic Dry: 四脚ロボットの歩容遷移

東北大学
電気通信研究所
石黒・加納研究室

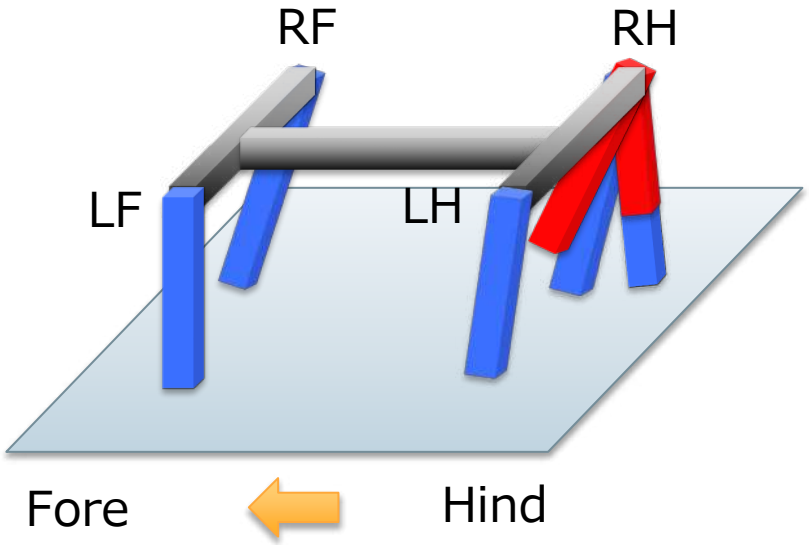
# Synthetic Dry: 四脚歩容遷移の原理について考えた

位相振動子：　$\dot{\phi}_i = \omega$

Swing phase（遊脚相）



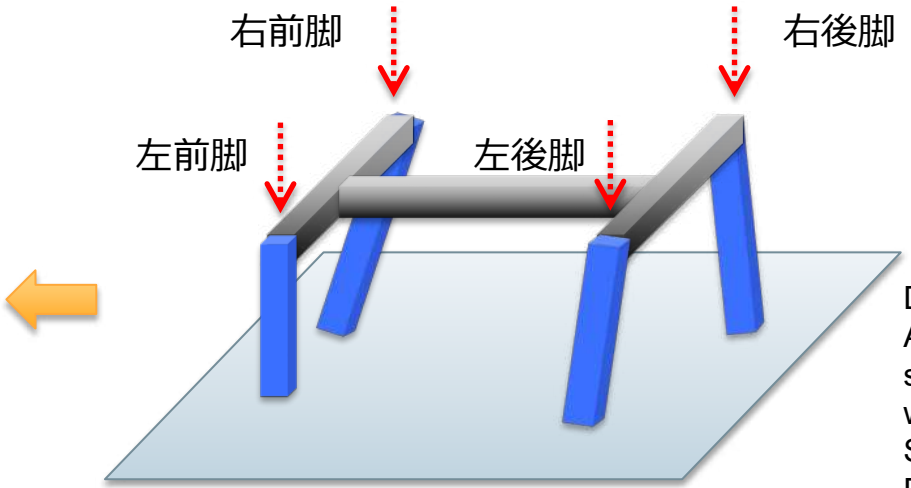

Stance phase（支持脚相）




東北大学
電気通信研究所
石黒・加納研究室

# Synthetic Dry: 四脚歩容遷移の原理はこれだ!?

$$\frac{\mathrm{d}\phi_i}{\mathrm{d}t} = \omega - \sigma N_i \cos\phi_i$$

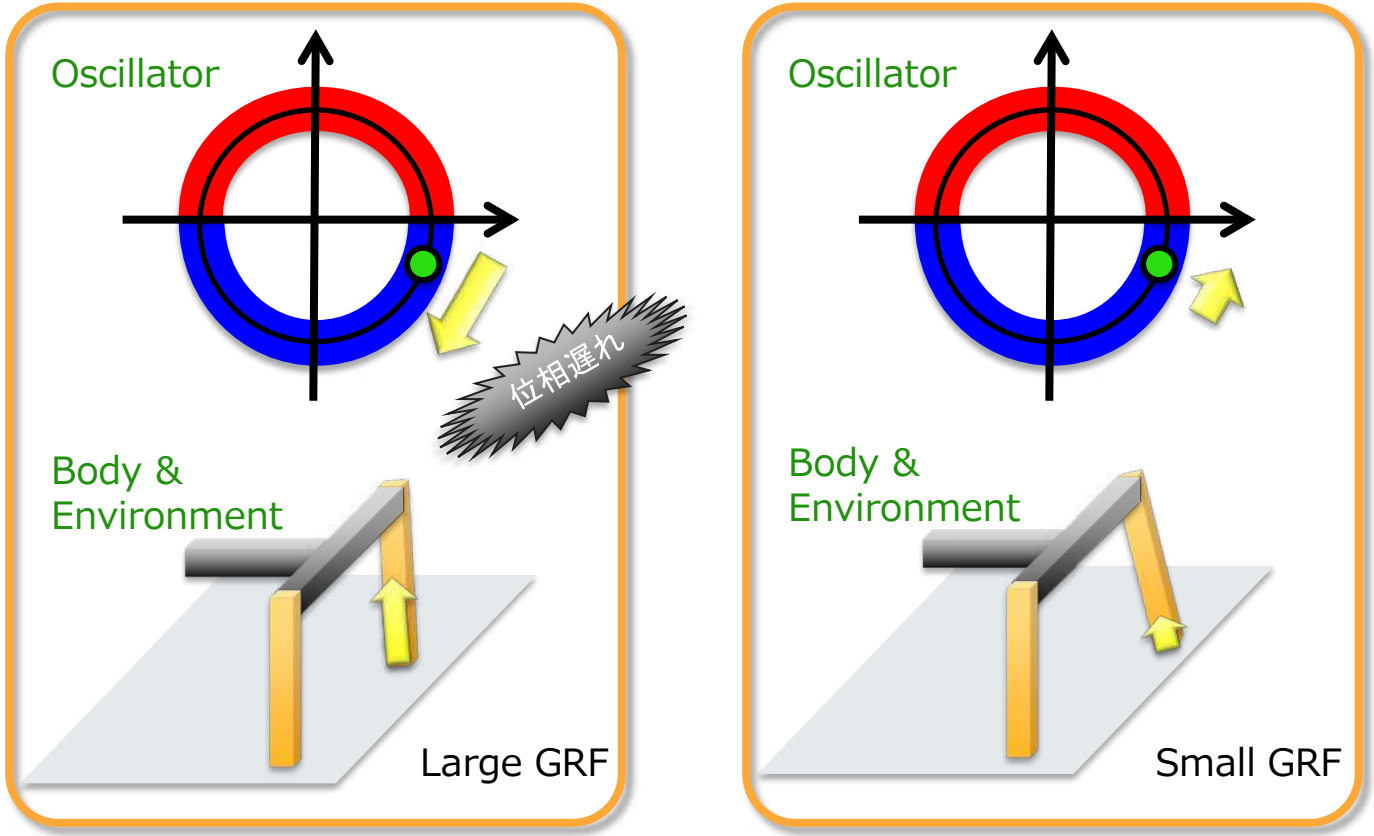



東北大学
電気通信研究所
石黒・加納研究室

D.Owaki, A.Ishiguro.
A quadruped robot exhibiting spontaneous gait transitions from walking to trotting to galloping.
Scientific Reports, 7: 277
DOI:10.1038/s41598-017-00348-9 1.

この数式に従って，各脚を動かすだけで，四脚動物と同じような振る舞いが発現する！

# Synthetic Dry: 四脚歩容遷移の原理はこれだ!?




支持脚であり続けようとする！

遊脚になることができる！

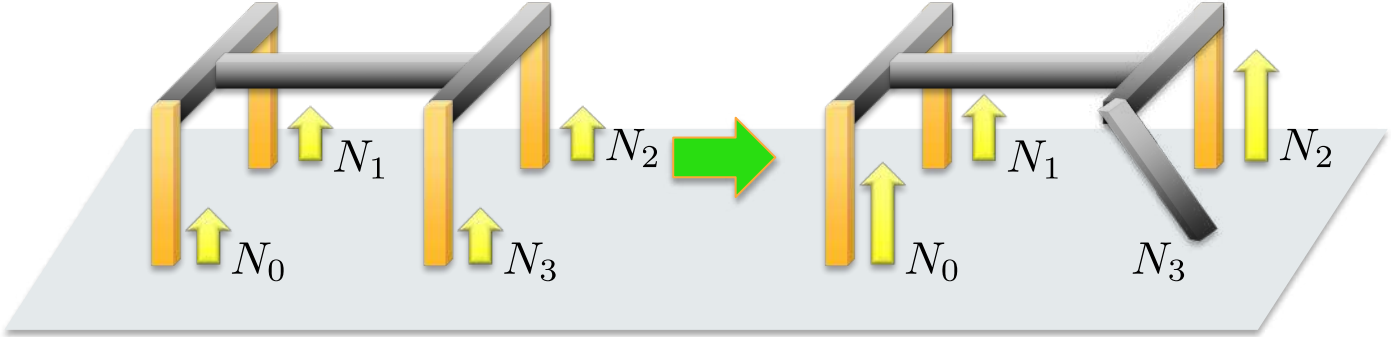
東北大学
電気通信研究所
石黒・加納研究室

# Synthetic Dry: 床反力と脚状態の相互作用が鍵！

$$\dot{\phi}_i = \omega - \sigma N_i \cos \phi_i$$

$N_i$ ： $i$ 番目の脚にかかる床反力

ある脚が遊脚になると…




床反力情報には脚間の位相差がどのようになっているかに関するリッチな情報が埋め込まれている！　これを脚間協調に徹底的に活用すればよい！

東北大学
電気通信研究所
石黒・加納研究室

# 3つのDry

- Synthetic
- **Data**
- Modeling

# Data Dry:データベース・ツールを使ってみよう

## データベース・ツールをこれから学ぼうとする時に役立つサイト

生命科学系DB・ツール使い倒し系チャンネル：「統合TV」
https://togotv.dbcls.jp/
「統合TV」は、生命科学分野の有用なデータベースやツールの使い方を動画で紹介するウェブサイト

# Data Dry:データベース・ツールを使ってみよう

**標準的な流れ**

(1) 生データの特徴を理解する
(2) 生データの補正をする
(3) 本解析を行う

**ツールを選ぶ時には**

(1) 解析手法の中身を理解する
(2) 解析結果の特徴を理解する

## 例：マイクロアレイを用いた遺伝子のクラスタリング

**前処理**

(a-1) マイクロアレイ生データの特性の理解（可視化）
(a-2) データのフィルタリング、補正
(a-3) 補正の確認（可視化）

**本解析**

(b-1) サンプル重複の可視化と処理
(b-2) 遺伝子間距離の選択
（ユークリッド距離、ピアソン距離、etc）
(b-3) 可視化方法の選択
・1D: 階層的クラスタリング
（クラスタリング法：最短距離法、最長距離法、etc）
・2D: MDS
・2D: tSNE（近傍数）
・2D: UMAP

情報提供：
東北大学情報科学研究科
生命情報システム科学分野
大林　武先生

# Data Dry: タンパク質の立体構造予測

タンパク質立体構造は実験で得るのが難しいし、そもそも水溶性でないなら解けない場合もある。立体構造が判れば創薬研究をはじめとする疾病治療に応用できる。

**立体構造**

I-TASSER

https://zhanglab.ccmb.med.umich.edu/I-TASSER/

HHPred

https://toolkit.tuebingen.mpg.de/tools/hhpred

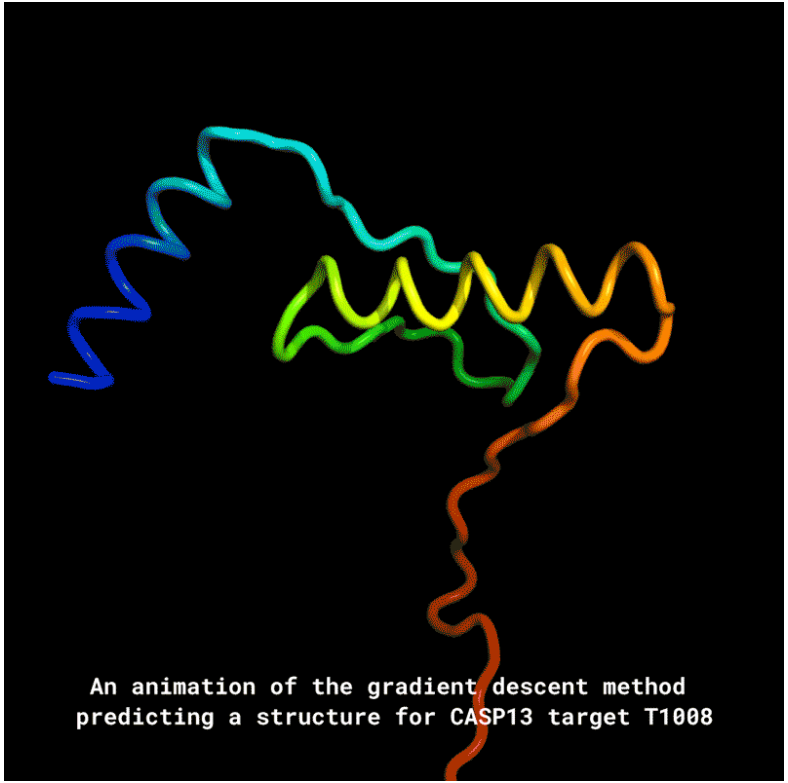
**二次構造予測**

PsiPred

http://bioinf.cs.ucl.ac.uk/psipred/

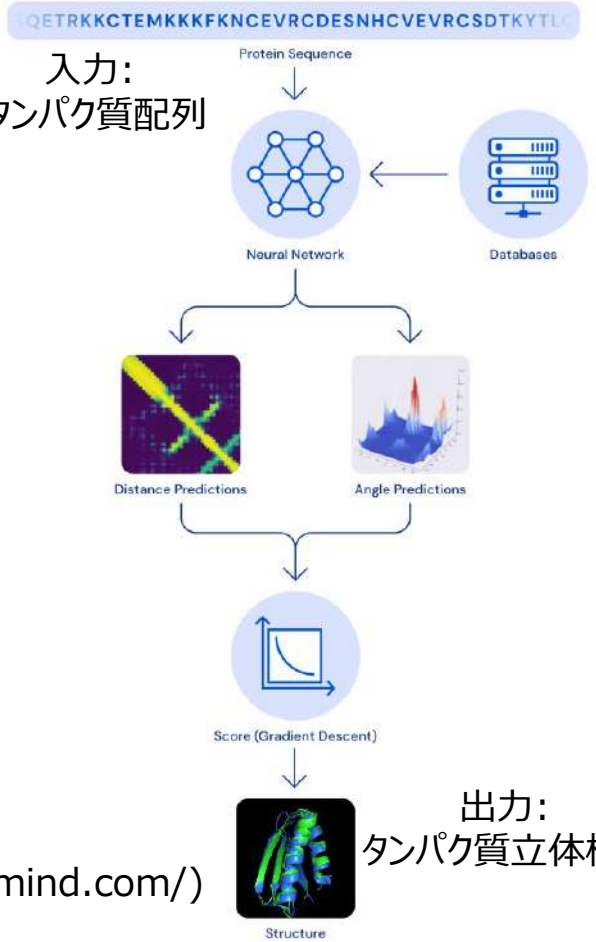
# Data Dry: 機械学習によるタンパク質立体構造予測

## Google DeepMind による AlphaFoldの実現




(https://deepmind.com/)

入力:
タンパク質配列




出力:
タンパク質立体構造

# Data Dry: DeepMindによる立体構造予測

- 多重配列アラインメントに由来する特徴量の作成している。これには配列のプロファイル情報や、残基間の共進化情報が含まれる。これらを入力として、各残基の二面角や各残基間の距離のヒストグラムをDNN（Deep Neural Network）を用いて予測する。
- AlphaFoldは、各残基の二面角および残基間の距離に対するタンパク質固有のポテンシャルに対して、連続最適化問題を解く準ニュートン法の一種であるL-BFGS法を用いるという非常にシンプルなアプローチで構造予測を行った。

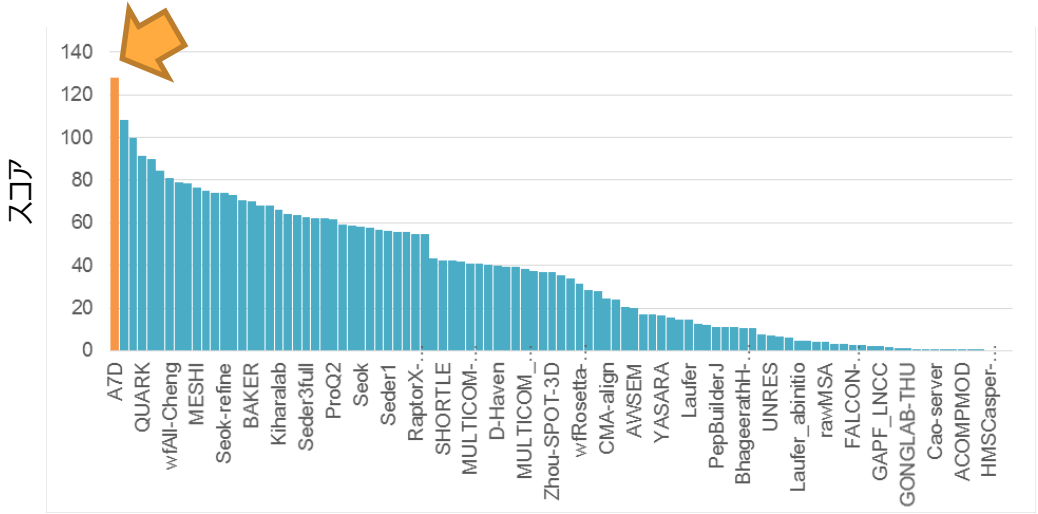
東北大学大学院情報科学研究科　中村司さんの解説 「実験医学」2019年7月号
https://www.yodosha.co.jp/jikkenigaku/nhpd/9784758125215/c1.html

よほど詳しくないと手に負えないようですがAlphaFoldのソースコードがGitHubにあるようです。
https://github.com/deepmind/deepmind-research/tree/master/alphafold_casp13

# Data Dry: AlphaFoldの衝撃

CASP（Critical Assessment of Techniques for Protein Structure Prediction）

1994年より隔年開催されてきた伝統的なタンパク質立体構造予測のコンテスト。タンパク質立体構造予測はタンパク質構造研究者の最終目的のひとつであり、生命情報学分野の著名な研究者が技術を競い合ってきた。




近年最も強かった2位のチームのスコア（108）を約20%も上回るスコア（128）を記録。20年にも渡る立体構造予測の技術の蓄積を超えた衝撃的な出来事。

情報提供：
東北大学情報科学研究科
データ科学国際共同大学院担当
山田和範先生

# Data Dry:バイオデータのテンソル分解

- テンソル分解のアイデアは古くからある。

  F.L. Hitchcock. The expression of a tensor or polyadic as a sum of products. J Math Phys 6:164-189, 1927.

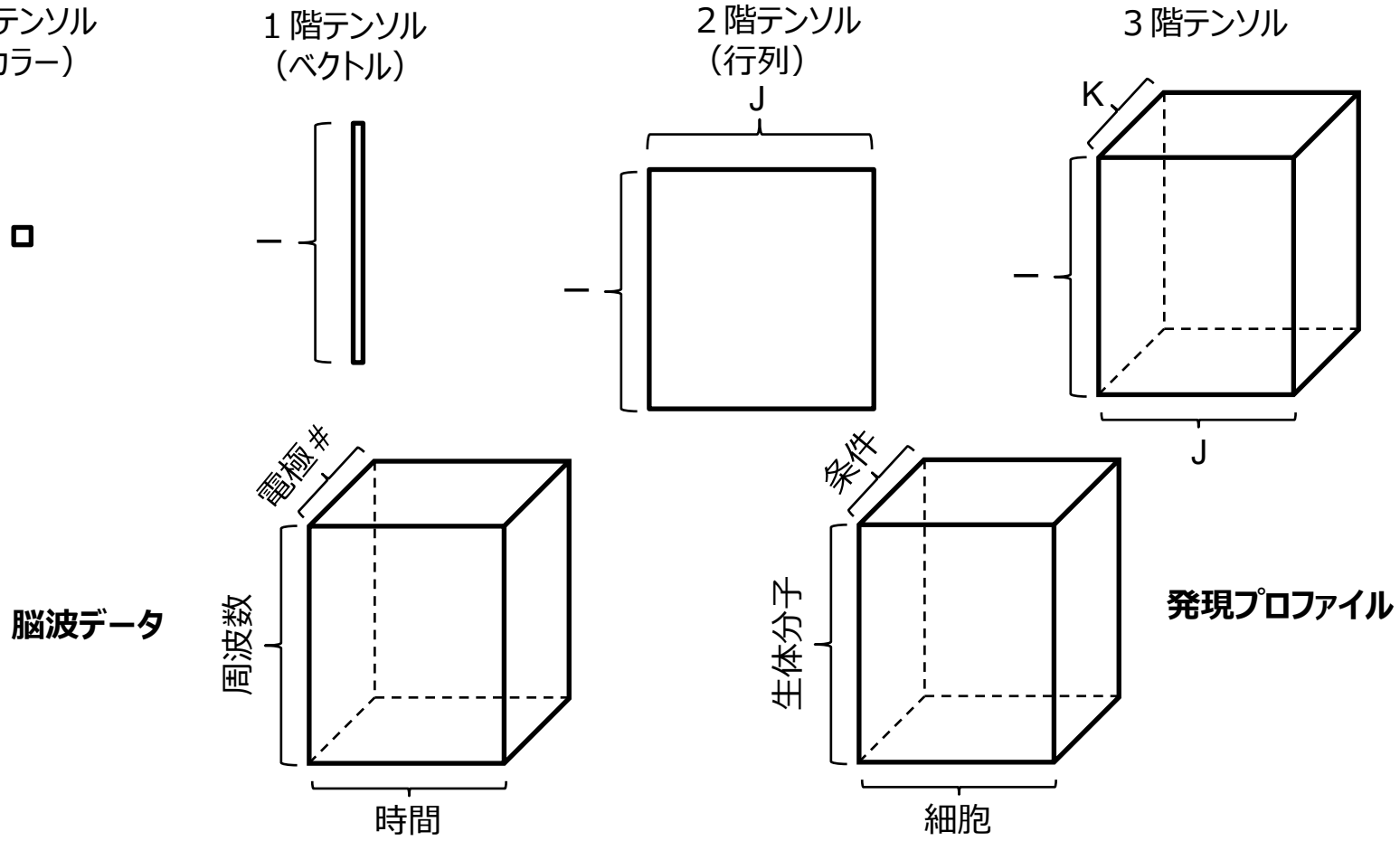
- 近年では、データマイニングや機械学習における、データの低ランク近似（縮約）やノイズ除去の観点から研究されている。

  A.Cichocki, R.Zdunek, S. Amari. Nonnegative matrix and tensor factorization. Lecture Notes, IEEE Signal Processing Magazine, 142-145, Jan 2008.
  D.Hong, T.G. Kolda, J.A. Duersch. Generalized canonical polyadic tensor decomposition. SIAM Review 62: 133-163, 2020.

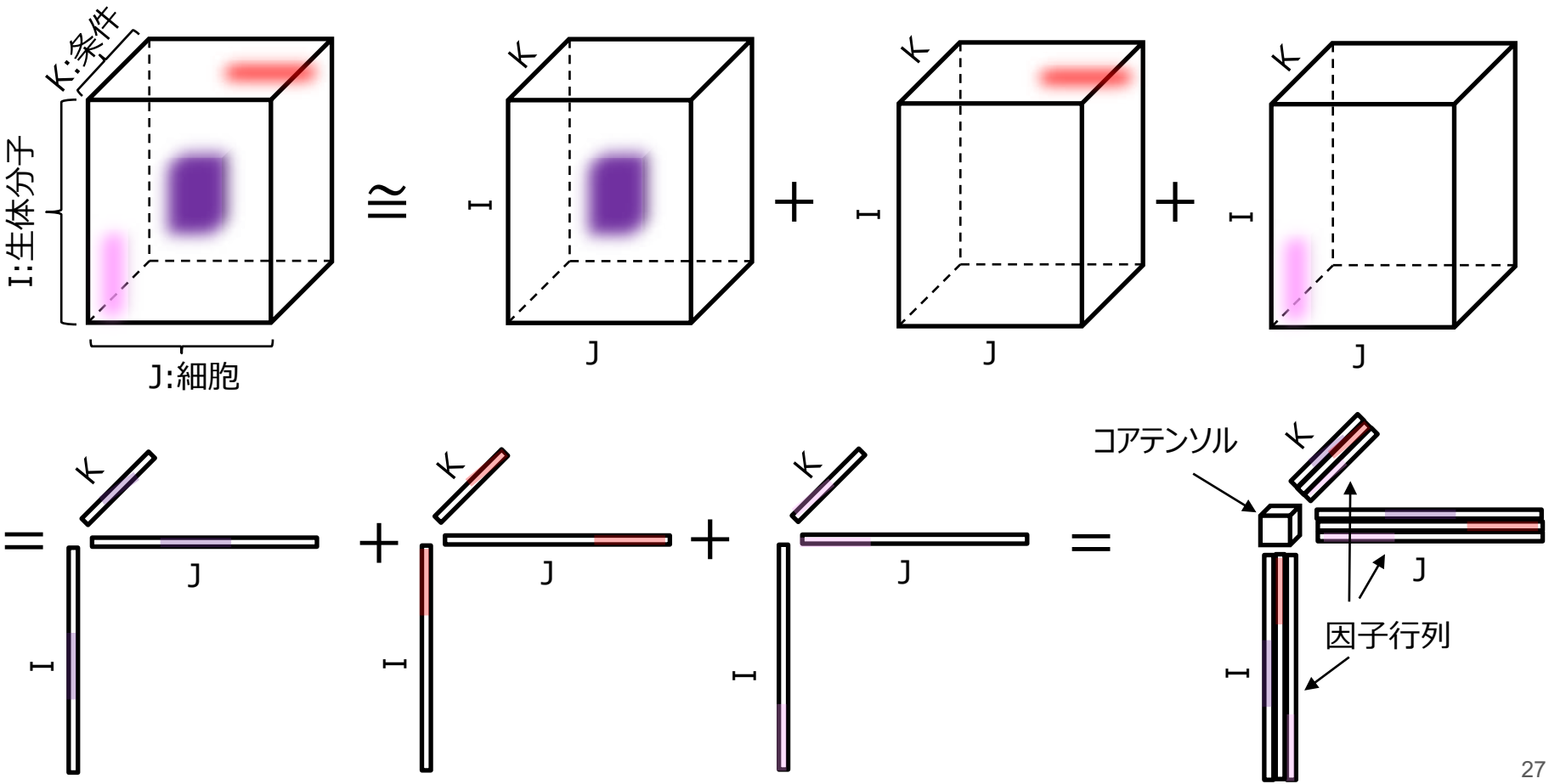
- 理研の露崎らはリガンド-受容体カプリングによる細胞間相互作用をテンソル分解を用いて解析している。→ここでは、彼らのアイデアに沿ってテンソル分解の応用例を紹介する。

  K.Tsuyuzaki, M.Ishii, I.Nikaido. Uncovering hypergraphs of cell-cell interaction from single cell RNA-sequencing data. doi: https://doi.org/10.1101/566182 (bioRxiv) 2019.

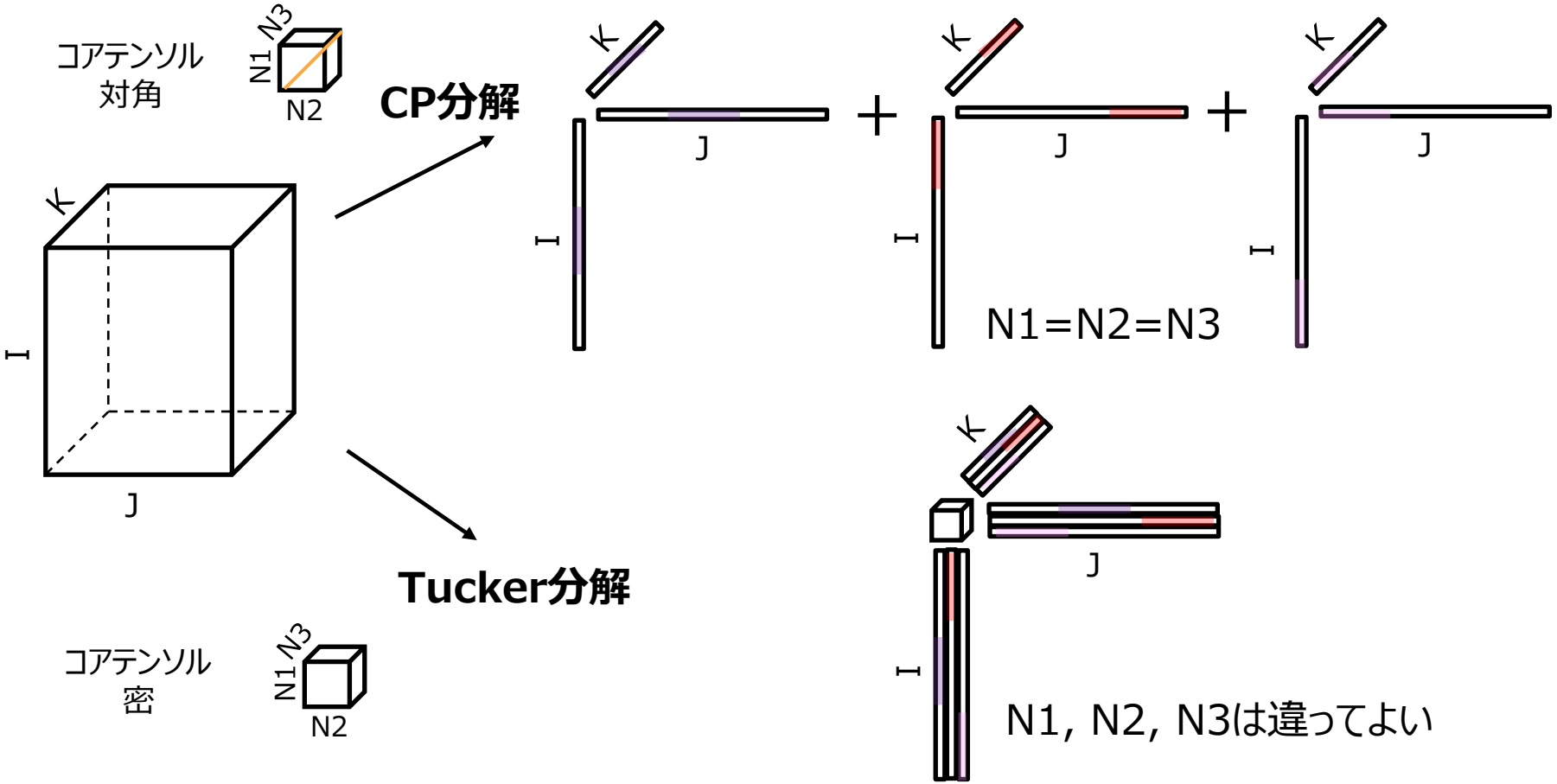
# Data Dry: バイオデータに相性のよいテンソル表現



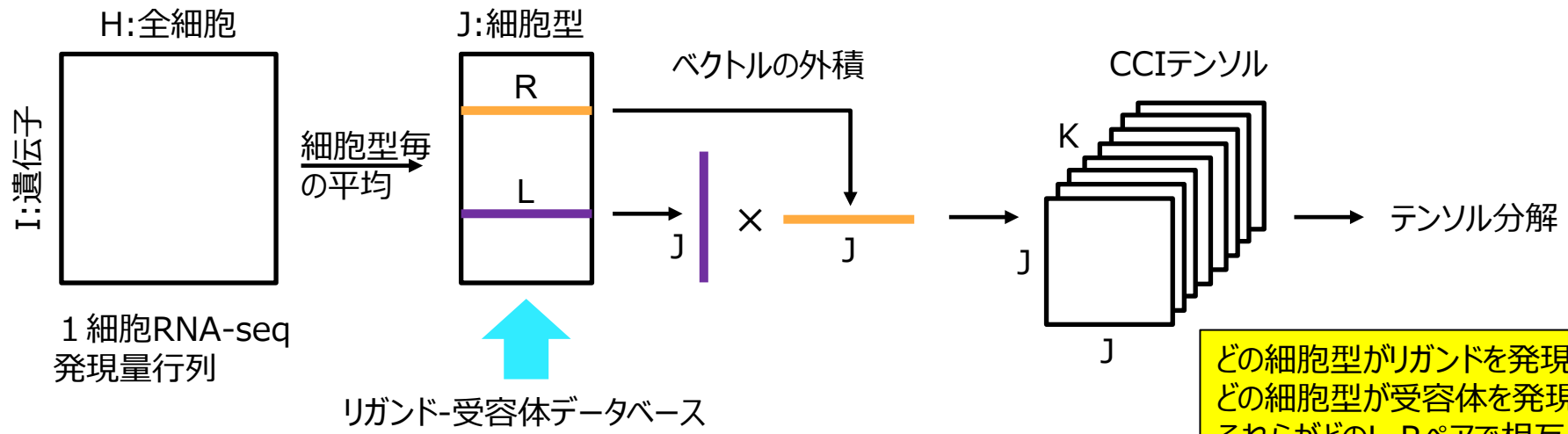
# Data Dry: テンソル分解



# Data Dry: CP分解とTucker分解

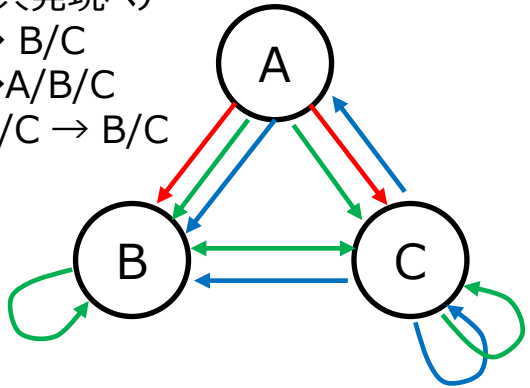


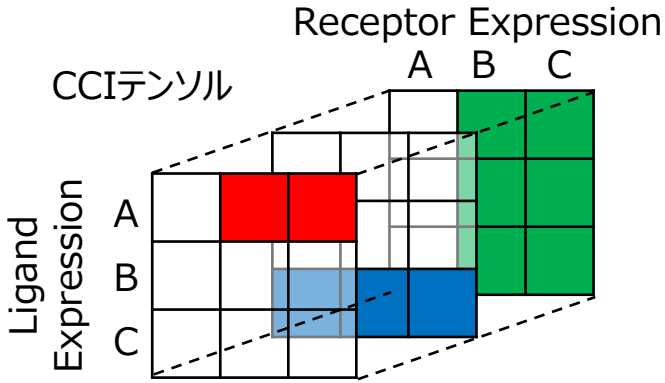
# Data Dry: 細胞間相互作用解析への応用




どの細胞型がリガンドを発現し、どの細胞型が受容体を発現し、それらがどのL-Rペアで相互作用しているか。

L-R共発現ペア
A → B/C
C →A/B/C
A/B/C → B/C




一部改変して引用
https://www.slideshare.net/antiplastics/20191204-mbsj2019

# Data Dry: 露崎さんたちの関連サイト＆論文

- scTensor
  https://bioconductor.org/packages/release/bioc/html/scTensor.html
- nnTensor
  https://cran.r-project.org/web/packages/nnTensor/index.html

- K. Tsuyuzaki, M. Ishii, I. Nikaido. Uncovering hypergraphs of cell-cell interaction from single cell RNA-sequencing data. doi: https://doi.org/10.1101/566182 (bioRxiv) 2019.
- 露崎弘毅．テンソル分解 -ヘテロなバイオデータを繋ぐ次世代型データ解析技術-．分子生物学会年会2019（フォーラム２F-09). https://www.slideshare.net/antiplastics/20191204-mbsj2019

# 3つのDry

- Synthetic
- Data
- **Modeling**

# 良いモデルとは？

- もちろん生物学的な現象を定性的／忠実に模擬できなくてはならない。その上で、他の条件で生起する現象を予測できること。その構造が現象を引き起こすメカニズムの推定に役立つこと。
- あらかじめ想定されたメカニズムの確認ではなくて、それまで考えもしなかったような洞察をもたらしてくれること。

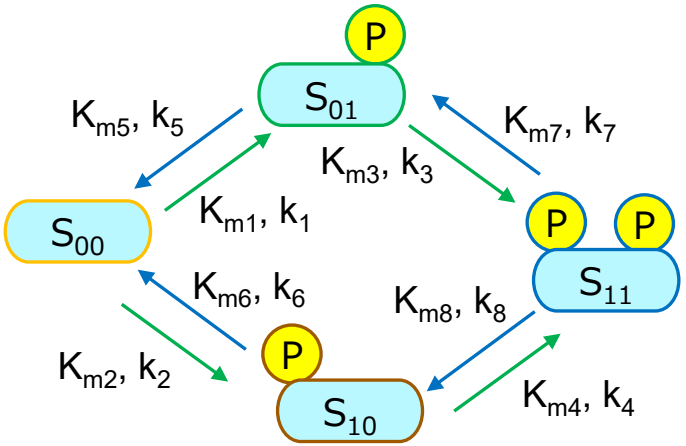
モデリングそのものが医学・生物学的な「リアリティ」を感じさせてくれることが大事

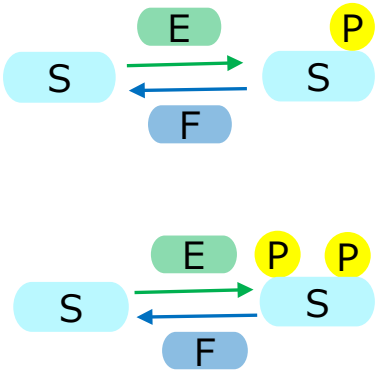
# Modeling Dry: データ駆動型？モデリング

- 「モデルの構造が現象を引き起こすメカニズムの推定に役立つこと」、「それまで考えもしなかったような洞察をもたらしてくれること」といっても、簡単にそんなモデルは降ってこない。
- 一つの参考例としてUedaらのアプローチを紹介する。

C.C. Jolley, K.L. Ode, H.R. Ueda. A design principle for posttranslational biochemical oscillator. Cell Report 2, 938-950, 2012.

- 概日リズムを生出すことができるようなminimal posttranslationalメカニズムを探索する。11変数微分方程式系と３つの関係式からなるモデル。



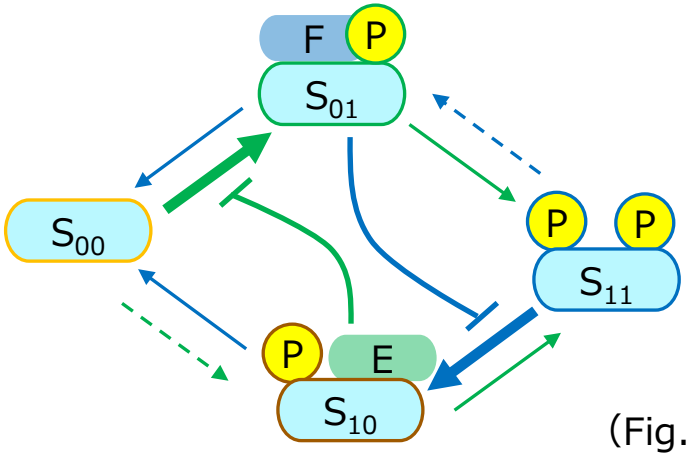




（Fig.1を改変して引用）

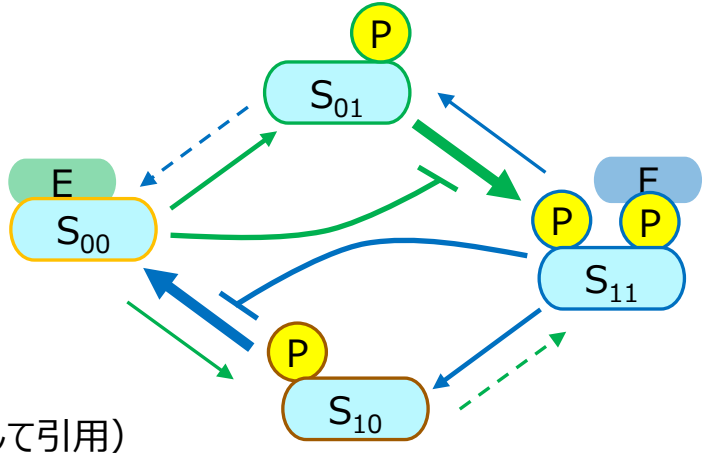
# Modeling Dry: データ駆動型？モデリング

- 16個ある反応係数の組をランダムに1,280,000,000組生成。そのうち、リミットサイクルを生じた組は1,201,874組（ヒット率0.094%）だった。
- 生成されたダイナミクスをクラスタリングし、2つの代表的なクラスターを得ている。
- さらに、得られたモデルに対して、振動において支配的なパラメータや温度補償性などの生物学的に深い検討を行っている。

クラスター 1




クラスター 2




（Fig.2を改変して引用）

# コロナの流行に左右されないライフスタイル

## ウエットとドライの往還

ドライな生活にもいろいろある。

- Synthetic Dry
- Data Dry
- Modeling Dry

この際、ドライなスキルも身に着けて、ウェットな生活を充実させよう。

ツールやe-learningコンテンツのリンクを充実させていきます。
http://gp-ds.tohoku.ac.jp/research/education_link.html