

2014年10月7日 2014年度「計算生命科学の基礎」
17:00～18:30 神戸大学工学部学舎1階 C3-101(創造工学スタジオ2)

計算生命科学概論



理化学研究所

HPCI計算生命科学
推進プログラム

江口至洋

yeguchi@riken.jp

1

今日から15回にわたる講義、「計算生命科学の基礎」を始めます。

第一回目の今日の講義は、「計算生命科学とはなんだろう？」という質問にお応えする意味で「計算生命科学概論」としました。

生命科学は、主に**観察と実験**を基礎においた学問領域ですが、

ジャック・モノーはよく言っていた

**「立派な理論に支えられていない実験
なんか、決して信じてはいけないよ！」**

"Never trust an experiment that is not supported by a good theory."

と。

Cohn M (2012) Cell Mol Life Sci. 69, 405-12

ジャック・モノー： 1965年ノーベル生理学・医学賞受賞者
「酵素とウイルスの合成に関する遺伝的制御の研究」

2

生命科学が「主に観察と実験に基礎を置いた学問領域である」ことに異論はないと思います。

しかし、1965年にノーベル生理学・医学賞を受賞したジャック・モノーは、よく次のように言っていたと伝えられています。

「立派な理論に支えられていない実験なんか、決して信じてはいけないよ」

と。モノーは実験研究者ですが、直観よりも論理的思考を重視した人であるとも言われています。

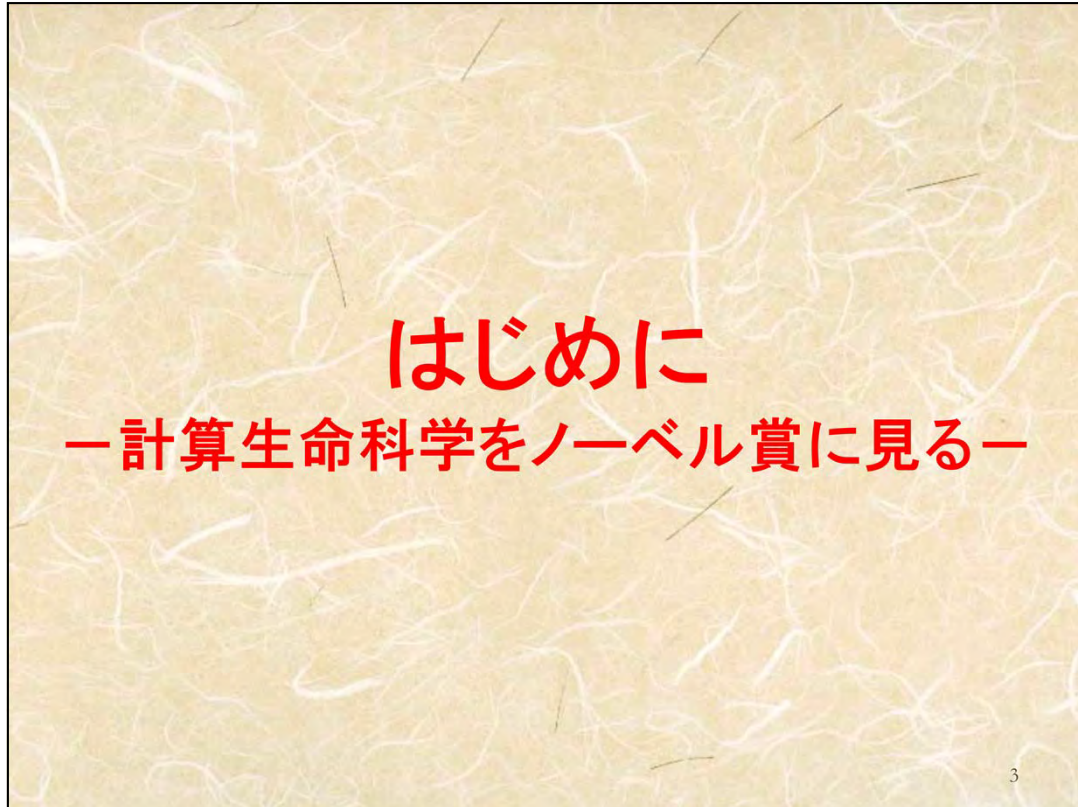
私はこの言葉を敷衍して、

「理論に基づく計算を行い、実験を支える。実験から得られるデータを計算し、解析し、理論を構築する。」

つまり、理論と実験が協調的に進められる中間点に計算があると考えています。

では、具体的に「計算生命科学とはなんだろう？」と考えていきます。

昨日ノーベル医学・生理学賞が発表されました。今週はノーベル・ウィークとも言われていますので、今日は、過去にノーベル賞を受賞した研究から、計算生命科学をみていきましょう。



生命科学の分野でも、ノーベル賞を受賞した多くの研究が計算生命科学に基づき進められてきましたし、計算生命科学が新たな発見、新たな方法論の開発を行ってきました。

1963年のノーベル生理学・医学賞は活動電位に



4

1963年のノーベル生理学・医学賞はHodgkinとHuxleyに与えられました。受賞理由は「神経細胞膜の末梢および中枢部分における興奮と抑制に関するイオン機構の発見」です。

歩いている、がびょうを踏んでしまった場合、がびょうの先端にちょっと触れただけで、多くの人は急いで足をひっこめます。

1963年のノーベル生理学・医学賞は活動電位に



5

この足を急いでひっこめる動作を、活動電位が担っています。

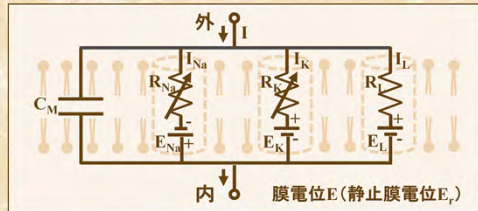
感覚ニューロンが足の先まで伸びていて、がびょうが皮膚を圧迫すると感覚ニューロンに活動電位が発生します。その活動電位の情報が脊髄にある介在ニューロンに伝えられ、一つは脳に伝達されていきます。しかし、脳で「あれ！？ 尖ったものを踏んづけたな」と判断してから足をひっこめるようだと、かなりの傷を負うこととなります。そこでもう一つの情報伝達経路があり、すぐに近くの運動ニューロンに情報を伝達し、急いで足をひっこめる動作を行います。

この一連の動作で活動電位が中心的役割を果たしています。

そのメカニズムをHodgkinとHuxleyが明らかにしたのです。

1963年のノーベル生理学・医学賞は活動電位に

A.L.Hodgkin と A.F.Huxley は1963年のノーベル生理学・医学賞を受賞した。



膜の電気回路モデル

ニューロンのチャネルの存在は、まだ直接には証明されていないにもかかわらず、「膜にゲートがある！」との仮説を立てた。

Hodgkin AL & Huxley AF (1952) A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve, J Physiol, 117, 500-544

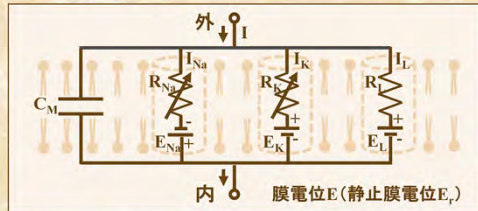
HodgkinとHuxleyは活動電位を発生させるニューロンの細胞膜と等価な電気回路モデルを作り上げました。

まずはコンデンサー(C_M)、そしてNa、K、Lで示される3つのゲートです。Lは「リーク(漏れ)」です。それぞれのゲートは電池と抵抗からなっていますが、NaゲートとKゲートの抵抗は可変抵抗となっていて、膜電位Eによって変化します。

この論文を発表した1952年当時、ニューロンのチャネル分子の存在は明らかになってはいません。HodgkinとHuxleyはその存在を仮定し、電気回路でチャネル分子の機能をモデル化したのです。

1963年のノーベル生理学・医学賞は活動電位に

A.L.HodgkinとA.F.Huxleyは1963年のノーベル生理学・医学賞を受賞した。



膜の電気回路モデル

ニューロンのチャネルの存在ははまだ直接には証明されていないにもかかわらず、「膜にゲートがある！」との仮説を立てた。

$$I = C_M \frac{dV}{dt} + \bar{g}_K n^4 (V - V_K) + \bar{g}_{Na} m^3 h (V - V_{Na}) + \bar{g}_L (V - V_L) \quad (1)$$

$$\frac{dn}{dt} = \alpha_n (1 - n) - \beta_n n \quad (2) \quad \alpha_n = 0.01(V + 10) / (\exp \frac{V + 10}{10} - 1), \quad \beta_n = 0.125 \exp \frac{V}{80}$$

$$\frac{dm}{dt} = \alpha_m (1 - m) - \beta_m m \quad (3) \quad \alpha_m = 0.1(V + 25) / (\exp \frac{V + 25}{10} - 1), \quad \beta_m = 4 \exp \frac{V}{18}$$

$$\frac{dh}{dt} = \alpha_h (1 - h) - \beta_h h \quad (4) \quad \alpha_h = 0.07 \exp \frac{V}{20}, \quad \beta_h = 1 / (\exp \frac{V + 30}{10} + 1)$$

$V = E - E_r, V_{Na} = E_{Na} - E_r, V_K = E_K - E_r, V_L = E_L - E_r$
 E_{Na}, E_K : イオンの平衡電位、 E_L : リーク電位

Hodgkin AL & Huxley AF (1952) A Quantitative Description of Membrane Current and its Application to Conduction and Excitation in Nerve, J Physiol, 117, 500-544

HodgkinとHuxleyは、イカの巨大神経を用いた膨大な実験結果をもとに、電気回路モデルの数理モデル化を行いました。

式(1)は電流 I が、コンデンサーに流れる電流、Naゲートに流れる電流、Kゲートに流れる電流、そしてLゲートに流れる電流の総和であることを示しています。

NaやKゲートの可変抵抗は、抵抗の逆数であるコンダクタンスを用いて、 $\bar{g}_K n^4$ や $\bar{g}_{Na} m^3 h$ で表現されています。この n と m 、 h が式(2)、(3)、(4)によって膜電位の影響を受け、時間的に変動していきます。

この結果、膜電位 E (数理モデルでは静止膜電位を E_r として $V = E - E_r$ が用いられています)と可変抵抗の間にフィードバック機構が働いています。

3つのゲートを仮定し、さらに膨大な実験結果から微分方程式(1)、(2)、(3)、(4)の形を決定したことが、HodgkinとHuxleyの研究の神髄だと思います。

1963年のノーベル生理学・医学賞は活動電位に

Hodgkin-Huxleyの数理モデルをMATLABで解く

```
function HH
% Hodgkin-Huxleyのモデルに従い膜電位E(mV)を求める。静止膜電位を-60mVとした。
% 列ベクトルV: V(1)=E(t), V(2)=n(t), V(3)=m(t), V(4)=h(t)
tspan=[0 50]; % シミュレーション時間 (msec)
V0=[-60, 0, 0.3, 0.05, 0.60]; % 初期値: I=0として定常状態値を近似的に求めた
[t,V]=ode15s(@HHscheme, tspan, V0); % ode15sを使い微分方程式を解く
% 計算結果を図にする
figure(1); plot(t,V(:,1));
xlabel('時間 (ミリ秒)', 'FontSize', 14);
ylabel('膜電位 (mV)', 'FontSize', 14);

function dVdt=HHscheme(t,V)
I=10; % 電流 (細胞の外から内が正: mA)
gNa=120; gK=36; gL=0.3; % イオンチャネルのコンダクタンス (mS/cm^2, mS=1/mΩ)
ENa=55; EK=-72; EL=-49.387; % 電池の起電力 (mV): 細胞外の電位を0としている
C=1; % コンデンサの静電容量 (μF)
% チャネルの開閉パラメータ
an=0.01*(V(1)+50)/(1-exp(-(V(1)+50)/10));
bn=0.125*exp(-(V(1)+60)/80);
am=0.1*(V(1)+35)/(1-exp(-(V(1)+35)/10));
bm=4*exp(-(V(1)+60)/18);
ah=0.07*exp(-(V(1)+60)/20);
bh=1/(exp(-(V(1)+30)/10)+1);
% HHモデルの微分方程式の定義
dV(1)=-1/C*(gK*(V(2)^4)*(V(1)-EK)+gNa*(V(3)^3)*V(4)*(V(1)-ENa)+gL*(V(1)-EL)-I);
dV(2)=an*(1-V(2))-bn*V(2);
dV(3)=am*(1-V(3))-bm*V(3);
dV(4)=ah*(1-V(4))-bh*V(4);
dVdt=[dV(1); dV(2); dV(3); dV(4)];
```

8

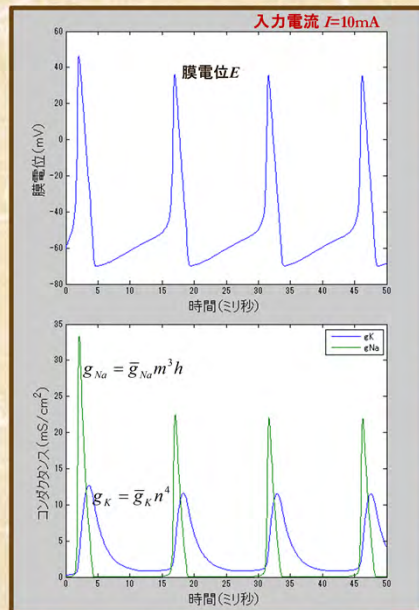
HodgkinとHuxleyは数理モデルを今でいう電子計算機ではなく、当時の機械式計算機で3週間かけて解いたと言われています。しかし、今では簡単に解くことができます。ここにMATLABで書いたプログラムを示します。

Tspan=[0 50]がシミュレーションしたい時間を示しています。V0は、膜電位と n 、 m 、 h の時刻0での値、初期値です。数理モデルにある微分方程式は、MATLABのode15sという関数で簡単に計算されます。その結果得られる膜電位の時間変化はplot(t,V(:,1))で図に描かれます。これだけです。

残りは、HodgkinとHuxleyの数理モデルにある微分方程式の具体的な表現です。入力電流Iをここでは10にしていますが、自由に変化させることができます。

1963年のノーベル生理学・医学賞は活動電位に

Hodgkin-Huxleyの数理モデルから得られる膜電位の時間変化



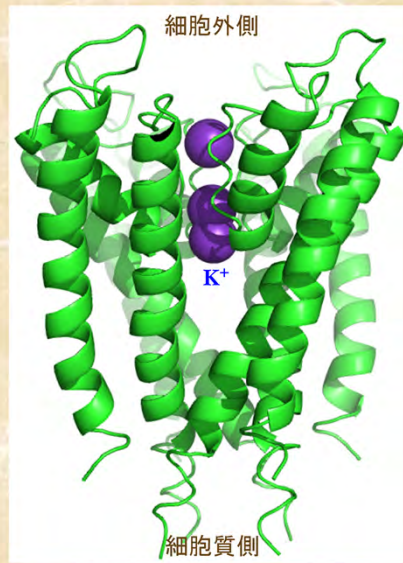
9

プログラムを実行した結果得られた膜電位 E の時間変化を示します。4つの活動電位の動きが表現されています。実験でも定常的に電流を入力すると、活動電位が次々と現れてきます。

下の図はNaゲートとKゲートのコンダクタンスの時間変化です。最初にNaゲートが開き活動電位を発生させ、Naゲートは瞬時に閉じていきます。一方、Naゲートに対して時間遅れを持ってKゲートが開き、活動電位を終わらせていきます。Kゲートのコンダクタンスを $\bar{g}_K n^4$ と n の4乗で表現したことにより、Kゲートが開くまでの時間遅れをもたらしています。またNaゲートのコンダクタンスは $\bar{g}_{Na} m^3 h$ と、Naゲート活性化のための m^3 と不活性化のための h で表現されています。これらはHodgkinとHuxleyによる計算生命科学における大発明だったと思います。

図に示されるNaゲートとKゲートの時間的な開閉の動きはHodgkinとHuxleyによってなされた実験と整合性を持っています。

1963年のノーベル生理学・医学賞は活動電位に



カリウムチャネルのMcKinnonらの論文は、Hodgkin-Huxleyらの論文(1952年)から46年後の1998年に出された。

K⁺イオンを選択的に通す**選択フィルター**の存在を確認した。

Protein Data Bank(PDB): 1BL8

Doyle DA, Morais Cabral J, Pfuetzner RA, Kuo A, Gulbis JM, Cohen SL, Chait BT, MacKinnon R (1998) Science, 280, 69-77

10

HodgkinとHuxleyらが論文発表した1952年から46年後の1998年にMcKinnonらはKチャネルの一つの構造をX線結晶構造解析によって明らかにしました。

この構造はKイオンを選択的に通す選択フィルターを明確に示しています。

McKinnonは2003年に「イオンチャネルの構造と機構の研究」でノーベル化学賞を受賞しています。

2011年のノーベル生理学・医学賞は自然免疫に

自然免疫

「微生物感染や外来の分子に対して先天的に備わっている抵抗性。

...

これらの抵抗性には現在知られている限り、対象を特異的に認識する受容体は関与しない。

これに対して後天的に成立する特異的免疫が**獲得免疫**とよばれる。」

出典：今堀和友、山川民夫 監修 「生化学辞典 (第3版)」東京化学同人(1998)

11

次に自然免疫研究で計算生命科学が果たした役割です。

1998年に発刊された「生化学辞典」での自然免疫の項を示します。

獲得免疫は特異的免疫だが、自然免疫については微生物や外来分子などを特異的に認識する受容体は知られていない、と記載されています。当時は、獲得免疫が高度な免疫機能を担い、自然免疫は補助的機能を担うと考えられていたと思います。少なくとも、私はそう考えていました。

しかしこの生化学辞典が出版された1998年前後に、自然免疫の研究は急速に進歩していたのです。

2011年のノーベル生理学・医学賞は自然免疫に

1996年9月 ショウジョウバエ成体が真菌に抵抗する自然免疫は**Toll**分子が担っている(**Hoffmann**)。

Lemaitre B et al. (1996) The Dorsventral Regulatory Gene Cassette *spätzle/Toll/actus* Controls the Potent Antifungal Response in *Drosophila* Adults, *Cell*, **86**, 973-983

1997年7月 ヒトではショウジョウバエTollのホモログ**TLR4**が真菌に抵抗する自然免疫を担っている。

Medzhitov R et al. (1997) A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity, *Nature*, **388**, 394-397

1998年12月 マウスの**TLR4**が真菌のリポ多糖(LPS)を認識し、自然免疫機構を働かせる(**Beutler**)。

Alexander P et al. (1998) Defective LPS Signaling in C3H/HeJ and C57BL/10ScCr Mice: Mutations in *Tlr4* Gene, *Science*, **282**, 2085-2088

1999年4月 マウスの**TLR4**が真菌のリポ多糖(LPS)を認識し、自然免疫機構を働かせる。

Hoshino K et al. (1999) Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the *Lps* gene product, *J Immunol*, **162**, 3749-3752

12

1996年Hoffmannらは「ショウジョウバエ成体が真菌に抵抗する自然免疫は、Toll分子が担っている」ことを明らかにしました。Tollはショウジョウバエの初期発生において形態形成に関わる遺伝子として発見されたものですが、成体では自然免疫に関係していたのです。

そして1997年にはMedzhitov(メチトフ)らが「ヒトではショウジョウバエTollのホモログTLR4が真菌に抵抗する自然免疫を担っている。」ことを明らかにします。TLRはToll-like Receptorの略です。

1998年にはBeutlerらが、「マウスのTLR4が真菌のリポ多糖(LPS)を認識し、自然免疫機構を働かせる。」ことを明らかにします。TLR4という受容体は真菌(カビ)の細胞外膜にあるLPSという分子を認識していたのです。

同様の研究成果は大阪大学の審良先生らによっても明らかにされ、その研究成果は1999年4月に論文発表されています。

これら4つの研究は、自然免疫の概念を大きく変えるもので、HoffmannとBeutlerは2011年のノーベル生理学・医学賞を受賞しています。

これら画期的な研究に計算生命科学も大きく寄与しています。

2011年のノーベル生理学・医学賞は自然免疫に

1996年9月 ショウジョウバエ成体が真菌に抵抗する自然免疫はToll分子が担っている(Hoffmann)。

Lemaitre B et al. (1996) The Dorsoventral Regulatory Gene Cassette *spätzle/Toll/actus* Controls the Potent Antifungal Response in *Drosophila* Adults, *Cell*, **86**, 973-983

1997年7月 ヒトではショウジョウバエTollのホモログTLR4が真菌に抵抗する自然免疫を担っている。

Medzhitov R et al. (1997) A human homologue of the *Drosophila* Toll protein signals activation of adaptive immunity, *Nature*, **388**, 394-397

1998年12月 マウスのTLR4が真菌のリポ多糖(LPS)を認識し、自然免疫機構を働かせる(Beutler)。

Alexander P et al. (1998) Defective LPS Signaling in C3H/HeJ and C57BL/10ScCr Mice: Mutations in *Tlr4* Gene, *Science*, **282**, 2085-2088

1999年4月 マウスのTLR4が真菌のリポ多糖(LPS)を認識し、自然免疫機構を働かせる。

Hoshino K et al. (1999) Cutting edge: Toll-like receptor 4 (TLR4)-deficient mice are hyporesponsive to lipopolysaccharide: evidence for TLR4 as the *Lps* gene product, *J Immunol*, **162**, 3749-3752

計算生命科学の役割 (配列の類似性探索の基本式)

$$D_{i,j} = \min[\min_{1 \leq k \leq l} \{D_{i-k,j} + v(k)\},$$

$$D_{i-1,j-1} + \delta(a_i, b_j),$$

$$\min_{1 \leq k \leq l} \{D_{i,j-k} + v(k)\}]$$

塩基配列のデータベースGenbankの中からショウジョウバエTollに似た配列を探索した。用いたソフトウェアはBLASTである。

マウスのゲノム塩基配列情報に含まれる遺伝子が $Tlr4$ であることをBLASTとGRAILで明らかにした。

「学生に毎日のように出されるゲノム情報から相同性の高い遺伝子を検索させて、12個のTLR遺伝子候補を探し出しました。そして、これらを片っ端からノックアウトしていったのです。こうして、LPSを認識する受容体がTLR4であるとわかったのが98年の夏でした。」
審良静男(2011)「自然免疫の点を線につなぐ」生命誌ジャーナル69号から

13

塩基配列データベースのGenbankと、相同性探索ソフトウェアBLAST、そして塩基配列データの中でタンパク質をコードしている領域を探索するソフトウェアGRAILです。

Medzhitovらの論文では、「Genbankの中からTollに似た配列を探索した。用いたソフトウェアはBLASTである」と書かれています。

BeutlerらはBLASTとGRAILを用いています。

審良先生らもおそらくBLASTを用いていたと思います。

BLASTの相同性探索で用いられる基本的な式を右上に示しておきます。式そのものは簡単ですが、直接その式を用いてGenbankのような大規模なデータベースから類似配列を探索するには時間がかかりすぎます。そのため、BLASTは高速化のための工夫をしています。

2011年のノーベル生理学・医学賞は自然免疫に

BLAST? 作るのは難しい。しかし、使うのは簡単です。



BLASTのようなプログラムを書くことはちょっと難しいです。ただし、使うのは簡単です。例えば、あるアミノ酸配列に似た遺伝子をデータベースDDBJ(これはGenbankの兄弟で日本で運用されています)に投げると、しばらくして「その配列に似た遺伝子はTLR4だよ」という答えを返してくれます。

2011年のノーベル生理学・医学賞は自然免疫に

dTOLL

LENLESIEFGSNKLRQMPRGIFGKMPKLLKQLNLSNQLHNLTKHDFEGATSVLGDIDHDNGIEQLPHDVFAHLTNVTDINLSANLFRSLPQGLFDHNKHLNEVRLMNN
RVPLATLPSRLFANQPELQILRLRAELQSLPGDLFEHSTQITNISLGDNLLSTLPATLLEHQVNLSDLSNNRLTHLPDSLFAHTTNTDLRLEDNLLTGISGDIFS
NLGNLVTLVMSRNRLRT

hTLR4

VEVVPNITYQCMELNFYKIPDNLFPSTKNLDLSFNPLRHLGYSYFFSFPPELQVLDLRSCEIQTIEDGAYQSLSHLSTLILTCNPIQSLALGAFSGLSSLQKLVAVETN
LASLENFPIGHLKTLKELNVAHNLIQSFKLPYFNSLNLLEHLDLSSNKIQSIYCTDLRVLHQMPLLNLSLDLSLNPMMFIQPGAFKEIRLHKLTLRNNFDSLNVMT
CIQGLAGLEVHRLVLGEFRN

DDBJ が公開しているソフトウェアClustalWを用いて解析した。 <http://clustalw.ddbj.nig.ac.jp/index.php?lang=ja>

15

ここにその結果を示します。入力データであるdTOLLがショウジョウバエのToll遺伝子のアミノ酸配列です。DDBJに類似配列の探索を依頼すると、「hTLR4が似ている配列の一つにあるよ」と言ってくれます。

この2つの配列を見て、すぐに「似ているな〜！」と解りますか？通常、人の目では解りません。

2011年のノーベル生理学・医学賞は自然免疫に

dTOLL
 LENLESIEFGSNKLRQMPRGIFGKMPKQKLNLSNQLHNLTKHDFEGATSVLGDIDHDNGIEQLPHDVF AHL TNVTDINLSANLFRSLPQGLFDHNKHLNEVRLMNN
 RVPLATLPSR L FANQPELQILRLRAELQSLPGDLFEHSTQITNLSLGDNLLSTLPATLLEHQVNLLSLDLSNNR L THLPDSLFAHTTNLTDLRLEDNLLTGISGDI FS
 NLGNLVTLMVSRNRLRT

hTLR4
 VEVVFNITYQCME LNFYKIPDNLPFSTKNLDLSFNPLRHLGYSYFFSFPPELQVLDLSRCEIQTIEDGAYQSLSHLSTLILTCNPIQSLALGAFSGLSSLQKLVAVETN
 LASLENFPIGHLKTLKELNVAHNLIQSFKLP EYFSNLTNLEHLDLSSNKIQSIYCTDLRVLHQMP L LNSLDLSLNP MNFIQPGAFKEIRLHKLTLRNNFDSLNVMTK
 CIQGLAGLEVHRLV LGEFRN

```

dTOLL  LENLESIEFGSNKLRQMPRGIFGKMP-KLKQKLNLSNQLHNLTKHDFEGATSVLGDIDHD 59
      * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :
hTLR4  VEVVFNITYQCME LNFYK--IPDNLPFSTKNLDLSFNPLRHLGYSYFFSFPPELQVLDLSR 58

dTOLL  NGIEQLPHDVF AHL TNVTDINLSANLFRSLPQGLFDHNKHLNEVRLMN-NRVPLATLPSR 118
      * : : : : * : : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * :
hTLR4  CEIQTIEDGAYQSLSHLSTLILTCNPIQSLALGAFSGLSSLQKLVAVETN L ASLENFPIG 118

dTOLL  L FANQPELQILRLRAELQSLPGDLFEHSTQITNLSLGDNLLSTLPAT---LLEHQVNLL 174
      : . ** : : : : : ** : * : * : : : * . * : : : * : * : : *
hTLR4  HLKTLKELNVAHNLIQSFKLP-EYFSNLTNLEHLDLSSNKIQSIYCTDLRVLHQMP L LNL 177

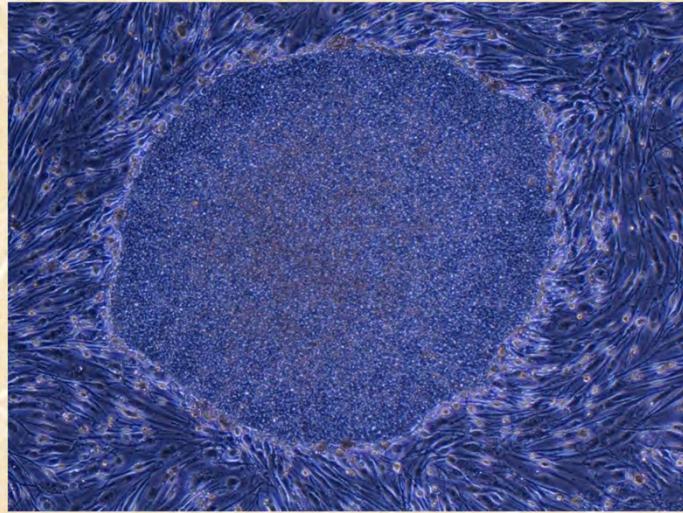
dTOLL  SLDLSNNR L THLPDSLFAHTTNLTDLRLEDNLLTG-ISGDIFSNLGNLVTLMVSRNRLRT 233
      ***** * : : : . * . * . * . * : * : : : : : * . * . * . : : : *
hTLR4  SLDLSLNP MNFIQPGAFKEIR-LHKLTLRNNFDSLNVMTKCIQGLAGLEVHRLV LGEFRN 236
  
```

DDBJ が公開しているソフトウェアClustalWを用いて解析した。 <http://clustalw.ddbj.nig.ac.jp/index.php?lang=ja> 16

そこでDDBJはdTOLLとhTLR4の配列をこのように並べた結果を出力してくれます。アラインメント(alignment)といいます。

「*」印の所が、完全に一致している所、「:」印の所が良く似たアミノ酸対の所です。計算生命科学の研究成果である、このようなデータベースとソフトウェアを20世紀の末頃、自然免疫の研究者たちは競って活用していたのです。

2012年のノーベル生理学・医学賞はiPS細胞に



線維芽細胞から樹立したヒトiPS細胞のコロニー(集合体)
(コロニーの横幅は実寸約0.5ミリメートル)

【クレジット】
京都大学教授 山中伸弥
<http://www.cira.kyoto-u.ac.jp/i/pressrelease/media/101129-134117.html>

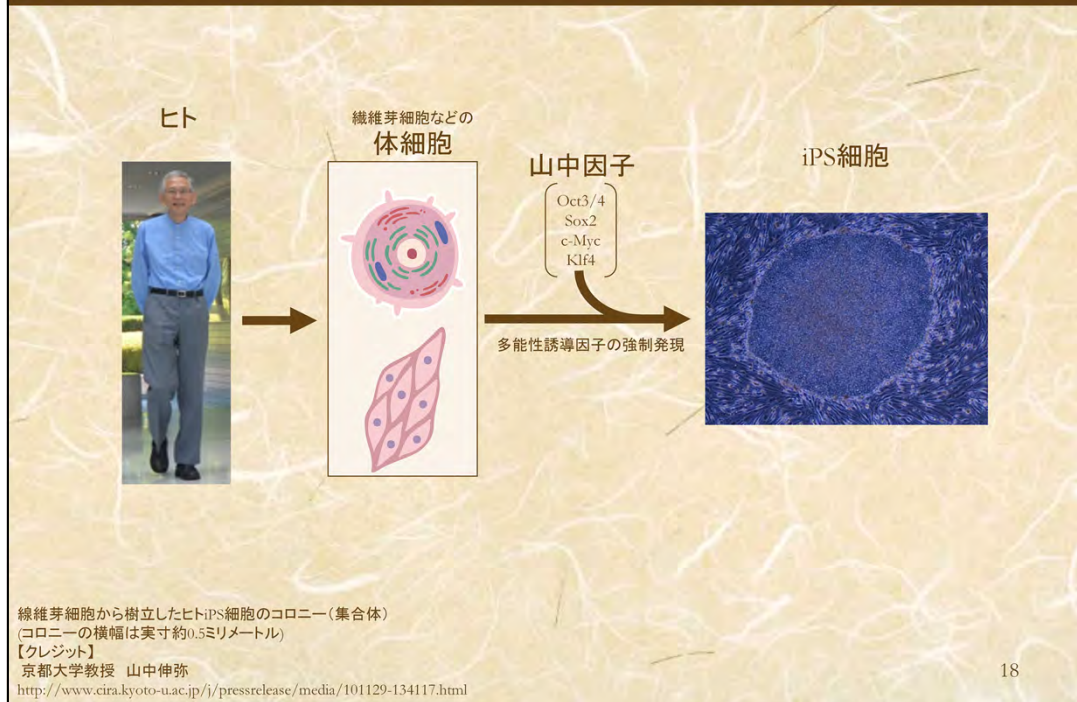
17

みなさんもよくご存じのように、2012年のノーベル生理学・医学賞はiPS細胞に与えられました。

この写真は京都大学の山中先生が作り出したiPS細胞です。

このコロニー(集合体)では、円く、くっきりとした輪郭の中に高い密度でiPS細胞が作り出されています。

2012年のノーベル生理学・医学賞はiPS細胞に



iPS細胞の作り方もすでにご存じのことでしょうが、この図のようになっています。

ヒトの体から体細胞を取り出し、山中因子と言われる4つの遺伝子を強制発現させ、iPS細胞は作られます。

では、山中因子はどのようにして見出されたのでしょうか？

研究の初期の段階で計算生命科学が活躍しました。

2012年のノーベル生理学・医学賞はiPS細胞に

高橋、山中らが思ったこと

ES細胞に特異的に発現している遺伝子を、体細胞に強制発現させると、体細胞はリプログラミングされ、多能性を獲得する。

「ES細胞に特異的に発現している遺伝子」＝「多能性獲得に必要な遺伝子」

Takahashi K and Yamanaka S (2006) Cell, 126, 663-676; Tanabe K, Takahashi S, Yamanaka S (2014) Proc. Jpn. Acad., Ser. B 90, 83-96

19

iPS細胞を生み出した山中因子の最初の論文はCellという雑誌に、高橋、山中の連名で出されています。

高橋、山中らはまず「ES細胞に特異的に発現している遺伝子を、体細胞に強制発現させると、体細胞はリプログラミングされ、多能性を獲得する。」と考えました。つまり、『「ES細胞に特異的に発現している遺伝子」＝「多能性獲得に必要な遺伝子」』です。

そのため、「ES細胞に特異的に発現している遺伝子」を探します。

2012年のノーベル生理学・医学賞はiPS細胞に

高橋、山中らが思ったこと

ES細胞に特異的に発現している遺伝子を、体細胞に強制発現させると、体細胞はリプログラミングされ、多能性を獲得する。

「ES細胞に特異的に発現している遺伝子」＝「多能性獲得に必要な遺伝子」

ES細胞の機能維持に重要な機能を果たすと考えられていた遺伝子

計算生命科学が寄与したこと

- 1) マウスの各種臓器、組織、細胞、そして初期胚やES細胞で発現している遺伝子のデータベースが構築されていた。(RIKEN、National Center for Biological Information)
- 2) データベースから、臓器、組織、細胞の間で特異的に発現している遺伝子を見つけ出すプログラムが公開されていた。(Digital Differential Display / NCBI)

山中因子の発見

Takahashi K and Yamanaka S (2006) Cell, 126, 663-676; Tanabe K, Takahashi S, Yamanaka S (2014) Proc. Jpn. Acad., Ser. B 90, 83-96

20

山中因子を見つけ出す初期の段階で計算生命科学の寄与がありましたが、山中らは次のように述べています。

1) マウスの各種臓器、組織、細胞、そして初期胚やES細胞で発現している遺伝子のデータベースが構築されていた。林崎先生らが作り上げたRIKENのデータベースと、アメリカのNCBI (National Center for Biological Information) のデータベースです。

2) データベースから、臓器、組織、細胞の間で特異的に発現している遺伝子を見つけ出すプログラムが公開されていた。NCBIのDDD (Digital Differential Display) というプログラムです。

これらデータベースとプログラムを用いて、「ES細胞に特異的に発現している遺伝子」、つまり「多能性獲得に必要な遺伝子」を絞り込んでいったのです。

これで終わりではなく、単に出発点で、それ以降、山中先生らのグループによる長い研究が積み重ねられ、ようやく山中因子が発見されたのです。

2012年のノーベル生理学・医学賞はiPS細胞に



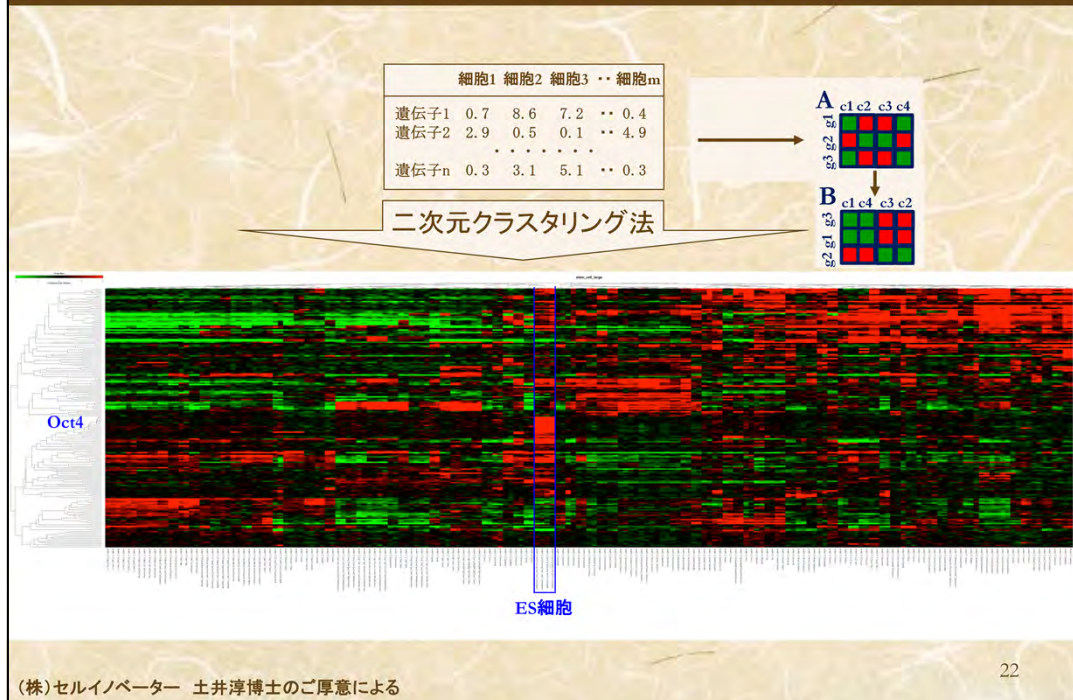
ある種の細胞に特異的に発現している遺伝子は、二次元クラスタリング法でも鳥瞰図的に探すことができます。

細胞1から細胞mで、遺伝子1から遺伝子nがこの表のように発現しているとします。0.7などの数値は遺伝子の発現量を示しています。

この表に対して二次元クラスタリング法を適用すると、下の図が出力されます。小さくて見えませんが、下の方に細胞名が書かれています。右には遺伝子名が書かれています。赤い所は発現量が多い遺伝子を、緑の所はほとんど発現していない遺伝子を示しています。

この図を良く見ると。

2012年のノーベル生理学・医学賞はiPS細胞に



ES細胞が青い線で囲った所に集まっています。中央付近の赤い領域はES細胞に特異的に発現している遺伝子の集団です。この中には山中因子にも含まれているOct4があります。

この二次元クラスタリング法は操作的には簡単で、右上にその例を示します。表の数値で例えば1.0以上の所を赤、それ以下の所を緑で描くと図Aになります。この図Aの列をそっくり入れ替えたり、行をそっくり入れ替えたりして、図Bを得ます。そうすると、細胞c2とc3では遺伝子g1とg3が特異的に発現しているのがすぐに解ります。また、細胞c1とc4では遺伝子g2が特異的に発現しています。

このような二次元クラスタリング法を用いた図を多くの論文で見ることができます。高橋、山中のCell誌の論文の中にも現れています。

計算生命科学はある意味、ノーベル賞を受賞した研究論文だけでなく、今では、広く生命科学の研究論文の中に溶け込んでいるとも言えます。

2013年のノーベル化学賞は計算生命科学に

量子力学

(20世紀初頭、1926年前後に確立された学問)

「量子力学によって物理学や化学が取り扱う多くの分野で基礎となる法則が完全に明らかになった」

$$H\Psi(x_1, x_2, \dots, x_n, X) = E\Psi(x_1, x_2, \dots, x_n, X) \quad \text{クーロン・ポテンシャル}$$
$$H = -\sum_{\mu=1}^N \frac{\hbar^2}{8\pi^2 M_{\mu}} \nabla_{\mu}^2 - \sum_{i=1}^n \frac{\hbar^2}{8\pi^2 m_i} \nabla_i^2 + \underbrace{V_{ne}(x, X) + V_{ee}(x) + V_{nn}(X)}_{\text{クーロン・ポテンシャル}}$$

「量子力学を具体的な問題に正確に適用しようとする、極めて複雑な方程式が出てきて、上手に解くには複雑すぎる。したがって、**複雑な分子でも**それほどの計算を必要としない、近似的で実際的な方法の開発が必須である」(Dirac PAM (1929) Proc R Soc Lond A, 123, 714-733)

23

自然免疫とiPS細胞の研究では、計算生命科学がノーベル賞を支えたとも言えますが、2013年のノーベル化学賞は計算生命科学そのものに与えられています。

20世紀初頭、物理学においてパラダイムの転換があり、量子力学が生まれたのです。量子力学の誕生に多大な貢献をしたDiracは1929年に次のように高々と宣言しています。

「量子力学によって物理学や化学が取り扱う多くの分野で基礎となる法則が完全に明らかになった」と。

生命科学で多く表れてくる原子や分子を取り扱う場合の基本式がここに示されています。 $H\Psi=E\Psi$ 。Schrödinger方程式です。 H の最初の2つの項は、古典力学でいう運動エネルギーに対応しています。残りの項は、みなさんよく御存じのクーロン・ポテンシャルです。ただこれだけの、簡単な形をしています。問題は波動関数 Ψ (プサイ)で、電子の数 n だけの座標と、ここでは X とだけ示していますが、原子核の数 N だけの座標を含む多変数関数です。

そのため通常はSchrödinger方程式を解くことはできません。そのため、Diracをして

「量子力学を具体的な問題に正確に適用しようとする、・・・、近似的で実際的な方法の開発が必須である」

と嘆かせています。

2013年のノーベル化学賞は計算生命科学に

量子力学

(20世紀初頭、1926年前後に確立された学問)

「量子力学によって物理学や化学が取り扱う多くの分野で基礎となる法則が完全に明らかになった」

$$H\Psi(x_1, x_2, \dots, x_n, X) = E\Psi(x_1, x_2, \dots, x_n, X) \quad \text{クーロン・ポテンシャル}$$
$$H = -\sum_{\mu=1}^N \frac{\hbar^2}{8\pi^2 M_{\mu}} \nabla_{\mu}^2 - \sum_{i=1}^n \frac{\hbar^2}{8\pi^2 m_i} \nabla_i^2 + \underbrace{V_{ne}(x, X) + V_{ee}(x) + V_{nn}(X)}_{\text{クーロン・ポテンシャル}}$$

「量子力学を具体的な問題に正確に適用しようとすると、極めて複雑な方程式が出てきて、上手に解くには複雑すぎる。したがって、**複雑な分子でも**それほどの計算を必要としない、近似的で実際的な方法の開発が必須である」(Dirac PAM (1929) Proc R Soc Lond A, 123, 714-733)

複雑な分子？

物理や化学の世界では、水素原子や、ヘリウム原子、水素分子までは、「**まあ、簡単かな**」という感じ。

生命科学で主に扱うタンパク質など巨大な分子は、「**超複雑**」という感じ。さらに、酵素反応など生命の本質的機能を担う現象の解明には、「**発想の転換**」が必要でしょう。

24

Diracの言う複雑な分子とは、何でしょうか。

物理や化学の世界では、HやHe、水素分子は「まあ、簡単かな」という感じですが。

生命科学が取り扱うタンパク質分子は「超複雑」、酵素反応は、飛んでいますね。「発想の転換」が求められます。その「発想の転換」に2013年のノーベル賞が与えられたのです。

2013年のノーベル化学賞は計算生命科学に

量子力学(QM)

$$H\Psi(x_1, x_2, \dots, x_n, X) = E\Psi(x_1, x_2, \dots, x_n, X)$$

$$H = -\sum_{\mu=1}^N \frac{\hbar^2}{8\pi^2 M_{\mu}} \nabla_{\mu}^2 - \sum_{i=1}^n \frac{\hbar^2}{8\pi^2 m_i} \nabla_i^2 + V_{ne}(x, X) + V_{ee}(x) + V_{nn}(X)$$

物理学や化学が取り扱う多くの分野で基礎となる法則。
ただ、具体的な問題に適用するには複雑すぎる。

量子力学の近似的だが 実際の解法

$$\Psi(x_1, x_2, \dots, x_n, X) \Rightarrow \phi_1(x_1), \phi_2(x_2), \dots, \phi_n(x_n)$$

$$\Psi(x_1, x_2, \dots, x_n, X) \Rightarrow \rho(x)$$

Hartree-Fock法、LCAO-MO-SCF法
密度汎関数法
FMO法 など

LCAO-MO-SCF: linear-combination-of-atomic-orbitals molecular-orbital self-consistent-field, FMO: Fragment Molecular Orbital

25

(n+N)個の多変数からなる波動関数 Ψ をそのまま取り扱おうとすると Schrödinger 方程式を解くことは極めて困難です。そこでまず、取り扱う分子に含まれるN個の核の変数 X を Born-Oppenheimer 近似により定数としてしまい、 Ψ を残されたn個の電子についての多変数関数とします。それでも Schrödinger 方程式を解くことはできませんので、大きくは2つの実際的な取り扱い方法が開発されています。

一つは1電子の分子軌道近似です。右上の式では $\phi_n(x_n)$ と示されています。この分子軌道の掛算や足し算、引き算の和でもってn変数の波動関数 Ψ を近似的に表現します。この分子軌道という考えは分子を取り扱う化学にとっては極めて受け入れやすい考えです。

2つ目は波動関数 Ψ を明示的に取り扱うのを諦めてしまいます。そして、分子の存在する空間座標 x での電子密度 $\rho(x)$ でもって、対象とする分子のエネルギーなどを計算しようとする試みです。

Hartree-Fock法やLCAO-MO-SCF法は1電子の分子軌道近似を採用しています。

密度汎関数法は2つ目の試みを基礎にしています。ただ、計算の段階で軌道という概念を完全に捨て去ったわけではありませんが。なお、Kohnは密度汎関数法の開発で、1998年のノーベル化学賞を受賞しています。

FMO法は、タンパク質などの高分子を取り扱うのに適した方法として開発されました。例えば、タンパク質分子をまとめて取り扱うのではなく、アミノ酸ごとをフラグメントとし、その総体としてタンパク質分子を取り扱います。軌道という考えの中に密度汎関数法的な考えを取り入れた方法とみるができます。

2013年のノーベル化学賞は計算生命科学に

量子力学(QM)

$$H\Psi(x_1, x_2, \dots, x_n, X) = E\Psi(x_1, x_2, \dots, x_n, X)$$

$$H = -\sum_{\mu=1}^N \frac{\hbar^2}{8\pi^2 M_{\mu}} \nabla_{\mu}^2 - \sum_{i=1}^n \frac{\hbar^2}{8\pi^2 m_i} \nabla_i^2 + V_{ne}(x, X) + V_{ee}(x) + V_{nn}(X)$$

物理学や化学が取り扱う多くの分野で基礎となる法則。
ただ、具体的な問題に適用するには複雑すぎる。

量子力学の近似的だが 実際的解法

$$\Psi(x_1, x_2, \dots, x_n, X) \Rightarrow \phi_1(x_1), \phi_2(x_2), \dots, \phi_n(x_n)$$

$$\Psi(x_1, x_2, \dots, x_n, X) \Rightarrow \rho(x)$$

Hartree-Fock法、LCAO-MO-SCF法
密度汎関数法
FMO法 など

古典力学 (電子はあきらめ、原子を見よう)

分子力学(MM: Molecular Mechanics)

$$V = \sum_{\text{bonds}} \frac{1}{2} K_{ij} (r_{ij} - b_{ij})^2 + \sum_{\text{angles}} \frac{1}{2} K_{ijk} (\theta_{ijk} - \theta_{ijk}^0)^2$$

$$+ \sum_{\text{dihedrals}} K_{\phi} (1 + \cos(n\phi - \phi_0))$$

$$+ \sum_{i \neq j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i \neq j} \frac{q_i q_j}{\epsilon_0 r_{ij}}$$

分子動力学(MD: Molecular Dynamics)

$$m_i \alpha_i = F_i = -\nabla_i V$$

LCAO-MO-SCF: linear-combination-of-atomic-orbitals molecular-orbital self-consistent-field, FMO: Fragment Molecular Orbital

26

量子力学の世界から離れ、古典力学の世界でタンパク質などの生体分子を取り扱おうとする方法が、分子力学(MM)や分子動力学(MD)です。つまり、電子の取り扱いをあきらめ、原子核と電子からなる原子の構成体として生体分子をみます。化学でいう分子構造式(例:水素分子のH-H)を思い浮かべて頂けば、いいと思います。例えば、水素分子では2つの水素原子がフックの法則に従ったバネで繋がっているとして取り扱います。

運動エネルギーの項は簡単ですので、問題はポテンシャル項です。量子力学ではクーロン・ポテンシャルだけが現れており、なじみやすかったのですが、分子力学では左下のVという関数が一般には用いられています。最後の項にクーロン・ポテンシャルがありますが、それまでの4つの項は、量子力学をあきらめたために生じてきた項です。ただ、このような関数型でタンパク質分子を取り扱えるというのは画期的なことです。

一度ポテンシャル関数Vが決められると、分子動力学はニュートンの運動方程式を解くことに専念できます。

2013年のノーベル化学賞は計算生命科学に

量子力学 (QM)

$$H\Psi(x_1, x_2, \dots, x_n, X) = E\Psi(x_1, x_2, \dots, x_n, X)$$

$$H = -\sum_{\mu=1}^N \frac{\hbar^2}{8\pi^2 M_{\mu}} \nabla_{\mu}^2 - \sum_{i=1}^n \frac{\hbar^2}{8\pi^2 m_i} \nabla_i^2 + V_{nc}(x, X) + V_{ec}(x) + V_{nn}(X)$$

物理学や化学が取り扱う多くの分野で基礎となる法則。
ただ、具体的な問題に適用するには複雑すぎる。

量子力学の近似的だが 実際の解法

$$\Psi(x_1, x_2, \dots, x_n, X) \Rightarrow \phi_1(x_1), \phi_2(x_2), \dots, \phi_n(x_n)$$

$$\Psi(x_1, x_2, \dots, x_n, X) \Rightarrow \rho(x)$$

Hartree-Fock法、LCAO-MO-SCF法
密度汎関数法
FMO法 など

古典力学 (電子はあきらめ、原子を見よう)

分子力学 (MM: Molecular Mechanics)

$$V = \sum_{\text{bonds}} \frac{1}{2} K_{ij} (r_{ij} - b_{ij})^2 + \sum_{\text{angles}} \frac{1}{2} K_{ijk} (\theta_{ijk} - \theta_{ijk}^0)^2$$

$$+ \sum_{\text{dihedrals}} K_{\phi} (1 + \cos(n\phi - \phi_0))$$

$$+ \sum_{i \neq j} 4\epsilon_{ij} \left[\left(\frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] + \sum_{i \neq j} \frac{q_i q_j}{\epsilon_0 r_{ij}}$$

分子動力学 (MD: Molecular Dynamics)

$$m_i \alpha_i = F_i = -\nabla_i V$$

量子力学と古典力学を マルチスケール化

QM/MM
QM/MM MD
など

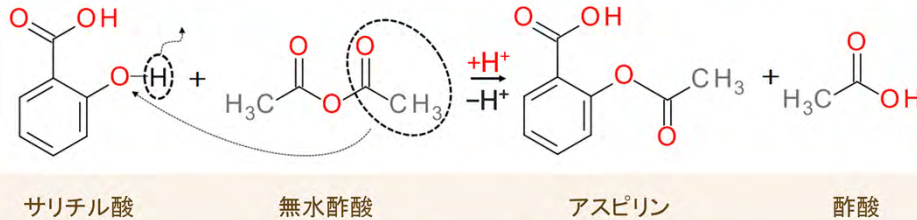
LCAO-MO-SCF: linear-combination-of-atomic-orbitals molecular-orbital self-consistent-field, FMO: Fragment Molecular Orbital

27

2013年のノーベル化学賞は量子力学と古典力学を連携させ、マルチスケール化した方法、QM/MM法に与えられています。

2013年のノーベル化学賞は計算生命科学に

酸触媒下での、サリチル酸のアセチル化によるアスピリンの合成反応を見ると、



量子力学：分子の電子状態の計算はできる。しかし、膨大な計算資源が必要となる。

古典力学：少ない計算資源で計算できる。しかし、電子状態の変化する過程は記述できない。

28

なぜ、量子力学と古典力学のマルチスケール化が必要なのかを、古くから解熱鎮痛薬として使用され続けているアスピリンの合成を例に示します。

上段は、酸触媒下でサリチル酸と無水酢酸からアスピリンが合成される反応式です。この反応では、酸触媒によって、無水酢酸のアセチル基 $\text{CH}_3\text{CO}-$ が活性化され、その $\text{CO}-$ の炭素原子がサリチル酸の水酸基($-\text{OH}$)の酸素原子を攻撃することから始まります。攻撃したアセチル基は水酸基の H を水素イオンとして放り出し、酸素原子と結合し、アスピリンが生成されます。

この反応の段階でアセチル基やその近くの電子の状態、さらにはサリチル酸の水酸基の電子状態は大きく変化し、電子の移動が起こります。

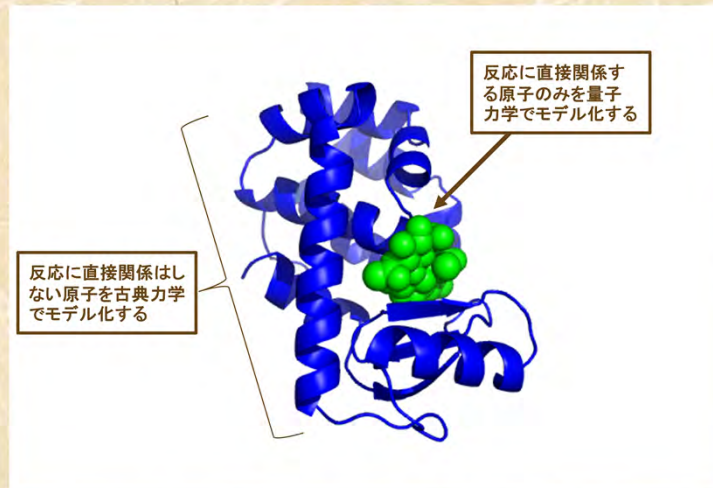
量子力学は、原理的には、この反応に伴う電子状態を正確に計算することができます。しかし、膨大な計算資源が必要です。

一方、古典力学は少ない計算資源で計算することができますが、電子状態の変化を記述することはできません。

特に、巨大な分子である酵素の反応など生命科学で重要な反応を解析しようとする、量子力学と古典力学の「いいところ」を組み合わせ、マルチスケール化することが必須になります。

2013年のノーベル化学賞は計算生命科学に

古典力学と量子力学を統合した研究
「複雑な化学システムのためのマルチスケール・モデル」



PDB: 148Lを用いて作成した。

29

2013年のノーベル化学賞は「複雑な化学システムのためのマルチスケール・モデル」に与えられています。

そのイメージ図を示します。青色は酵素で、緑色は酵素反応の基質だとします。「複雑な化学システムのためのマルチスケール・モデル」では、反応に直接関係する原子(複数の原子からなりますがそれほど多くはありません)のみを量子力学でモデル化し、「反応に直接関係はしない原子を古典力学でモデル化する」という意味でマルチスケール化して酵素反応の解析を可能としたのです。

2013年のノーベル化学賞は計算生命科学に

古典力学と量子力学を統合した研究
「複雑な化学システムのためのマルチスケール・モデル」

授賞理由の中で述べられていること
「もっとも重要なことは、マルチスケール・モデルは
**理論と実験の実りある協力関係を作り上げ、
他の方法では解けない問題をも解く道を切り拓いた**」

30

ノーベル賞の授賞理由の中で「マルチスケール・モデルは理論と実験の実りある協力関係を作り上げた」点と、その協力関係により「他の方法では解けない問題をも解く道を切り拓いた」点が最も重要なことだと、述べられています。

計算生命科学はその協力関係の結節点にいるのだと考えています。ある意味、理論と手を結び、実験とも手を結んでいる、つまり両手が必要だと思います。

生命科学での 計算生命科学の位置付け

31

これまでノーベル賞受賞研究をとりあげて、計算生命科学を実例でもって説明してきました。ここではこれまでの説明をとりまとめてみます。

計算生命科学は生命科学に必須です



32

生命科学の研究の流れは、行きつ戻りつしたり、飛び越したり、多様かつ複雑で、図示することは困難ですが、思い切って単純化するとこのようになります。

まず、「研究対象と目的/仮説」を立てます。研究目的に最善の研究対象を選ぶことは、困難ですが、重要です。HodgkinとHuxleyは実験のし易さもイカの巨大神経を研究対象としています。

つぎに、「観察と実験」によるデータ収集が行われます。HodgkinとHuxleyは電位固定法という実験手法で膜を通過する電流を測定しています。

得られた実験データと知識ベース/データベースとからモデル化、および数理モデルを作ります。山中先生らのiPS細胞の研究例では、ES細胞を含む多くの種類の細胞で計測された遺伝子の発現量データベースが用いられています。モデルと数理モデルはDDDでした。HodgkinとHuxleyは等価回路モデルとn、m、hを含んだ数理モデルを作成しています。

数理モデルをもとに計算が行われ、モデル(仮説)の検証作業がなされます。ここで検証結果が否定的な場合には、再度、図のサイクルを回ることになります。

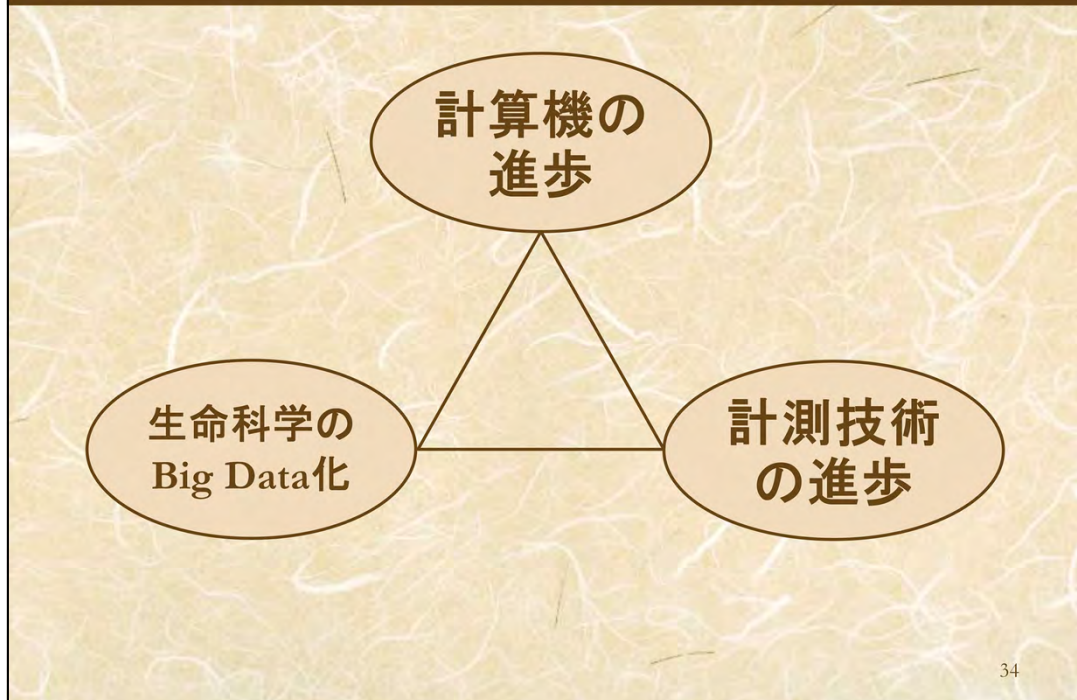
図示された各段階で計算生命科学はその役割を果たしています。「観察と実験」では見えにくいですが、現在実験室で使われている多くの計測機器には計算機が組み込まれていて、独自のモデルと数理モデルに従い、データを生成しています。モノーの言った「立派な理論に支えられていない実験なんか、決して信じてはいけませんよ！」に従えば、「計測機器をBlack Box化してはいけませんよ！」とも言えます。

21世紀に なぜ計算生命科学が 求められているのですか？

33

古く起源をたどれば、計算生命科学は1952年のHodgkinとHuxleyの論文にまで遡ることができますが、生命科学の中で計算生命科学が注目され出したのは21世紀に入ってからでしょう。

3つの要因



34

その理由として、ここでは3つの要因を上げます。相互に強く結びついています。「計算機の進歩」と「計測技術の進歩」、そして「生命科学のBig Data化」でしょう。

計算機の進歩



35

初期の有名な電子計算機にENIACがあります。1945年にアメリカの陸軍とペンシルバニア大学とが共同開発したもので、1946年に公開されています。その場に参加したニューヨーク・タイムズの記者は「多くの計算がまばたきをする間もなく終わってしまう」と驚きを持って記事にしています。

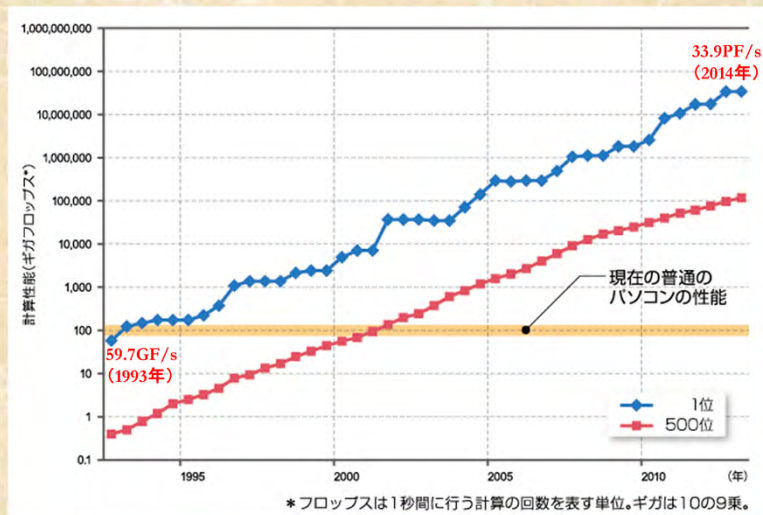
2011年にスーパーコンピュータ京が開発されましたが、その性能はニューヨーク・タイムズの記者が驚いたENIACの10兆倍以上とも言われています。京で研究している計算生命科学者からは「大学のスパコンを使っている時と違い、見える世界が格段に広がった。研究もより広い視野で考えることができるようになった。」との意見が出されています。

ENIACも京も国が主導して開発した計算機で、誰もが手軽に使えるものではありません。私の経験では、この図にある機械式計算機は1960年代までは使われていました。この取っ手を時計回りや反時計回りにくるくると回転させ、四則演算を行います。

HodgkinとHuxleyはその数理モデルを解くために、彼らがいたケンブリッジ大学の電子計算機EDSACを用いて解こうと計画していました。しかし、その時ちょうどEDSACの大規模な改修作業が入り、6か月ほど使用できない状態が続くことになったのです。仕方なく、彼らは当時の機械式計算機を用いて3週間ほどもの大変な力仕事で計算を行ったのです。

先ほども示しましたが、同じ数理モデルをみなさんが普段使っているパソコンで容易にかつ瞬時に解くことができます。HodgkinとHuxleyが用いた機械式計算機とは比べ物にならないほど高速の計算機をみなさんは容易に使える時代にいるのです。

計算機の進歩



<http://www.top500.org/>

36

世界最高速の計算機の計算速度の推移を示します。縦軸はlogスケールです。それにもかかわらず計算速度は線形に速くなってきています。この傾向は2020年頃までは続くと考えられています。それ以降、現在の集積回路の技術で高速化が進むかどうかは、はっきりとは言い切れません。

一方、みなさんが普通に使っているパソコンの性能は、だいたい20年ほど前の世界最高速の計算機の性能と同等です。そして、10年ほど前の世界で500番目に速い計算機と同等の性能です。

このことは、日々の実験室でHodgkinやHuxleyらが想像もできなかった計算速度の計算機を使える環境ができていることを意味しています。

これからもしばらくはこの傾向は続いていくでしょう。

計測技術の進歩

- 1979年 コンピュータ断層撮影法の開発 (Hounsfield G, Cormack AM)
- 1980年 DNA塩基配列の決定法の開発 (Sanger F, Gilbert W)
- 1993年 polymerase chain reaction (PCR)法の開発 (Mullis KB)
- 2002年 生体高分子の同定および構造解析のための手法の開発 (質量分析技術の開発) (田中耕一)
「溶液中での生体高分子の3次元構造の決定に関する核磁気共鳴分光法の開発」 (Wüthrich K)
- 2003年 核磁気共鳴画像法に関する発見 (Lauterbur P, Mansfield P)
- 2008年 緑色蛍光タンパク質 (GFP) の発見とその応用 (下村脩)



Copyright © 2014 Life Technologies Corporation. Used under permission
http://www.spring8.or.jp/ja/about_us/manage_structure/contacts/imageterms/.

37

ここでは、計測技術の進歩をノーベル賞にみてみます。

1979年、Computer Tomographyの開発にノーベル賞が授与されています。みなさんが大きな病気をした時に、病院で受けるCTといわれている検査です。組織や臓器のイメージング技術として用いられています。

1980年にはDNAシーケンス技術を開発したSangerらにノーベル賞が与えられています。この技術は2007年頃まで主に用いられてきました。

1993年にはPCR法にノーベル賞が与えられています。PCR法によってごく少量のDNAから大量のDNAをcopyして得ることができるようになりました。タンパク質の場合にはPCR法のようなcopy技術はありません。あればいいのですが。

2002年には生体高分子の計測技術としての質量分析技術とNMR技術にノーベル賞が与えられています。どちらも低分子の計測技術としては確立された技術でしたが、「生体高分子でも可能なの？」という驚きがありました。

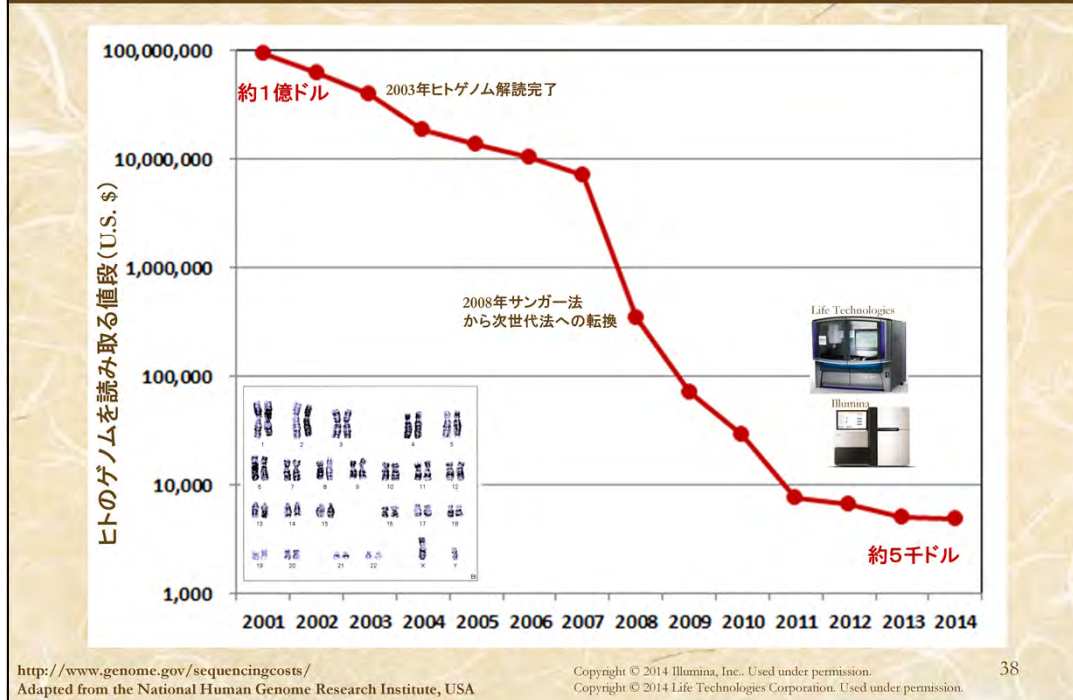
2003年にはMRI技術にノーベル賞が与えられています。この技術も、みなさんが大きな病気をした時に、病院で受ける検査項目に入ってくる場合があります。組織や臓器のイメージング技術です。

2008年には緑色蛍光タンパク質の発見にノーベル賞が与えられています。この技術により、タンパク質などの分子イメージングが急速に進展しました。

このようにDNAやタンパク質など生命科学の基本的な分子の構造と機能を解析するための計測技術が20世紀末頃に進歩してきた様子がうかがえます。また、生体のイメージング技術も、組織・臓器イメージングから分子イメージングまで、多階層で進歩してきました。

これら計測技術の進歩の速さをDNAシーケンス技術についてみてみます。

計測技術の進歩 - ヒトゲノムの塩基配列を読み取る値段の推移 -



この図は、2001年以降30億文字からなるヒトゲノムの塩基配列を読み取るのにかかる費用の推移をみたものです。

2001年は、約10年にわたる国際的な共同プロジェクトの成果としてヒトゲノムのドラフト配列が公開された年です。その2001年時点では約1億ドルが必要でした。2003年にはヒトゲノムの塩基配列を詳細に解読したとして、一応の終了宣言がなされています。

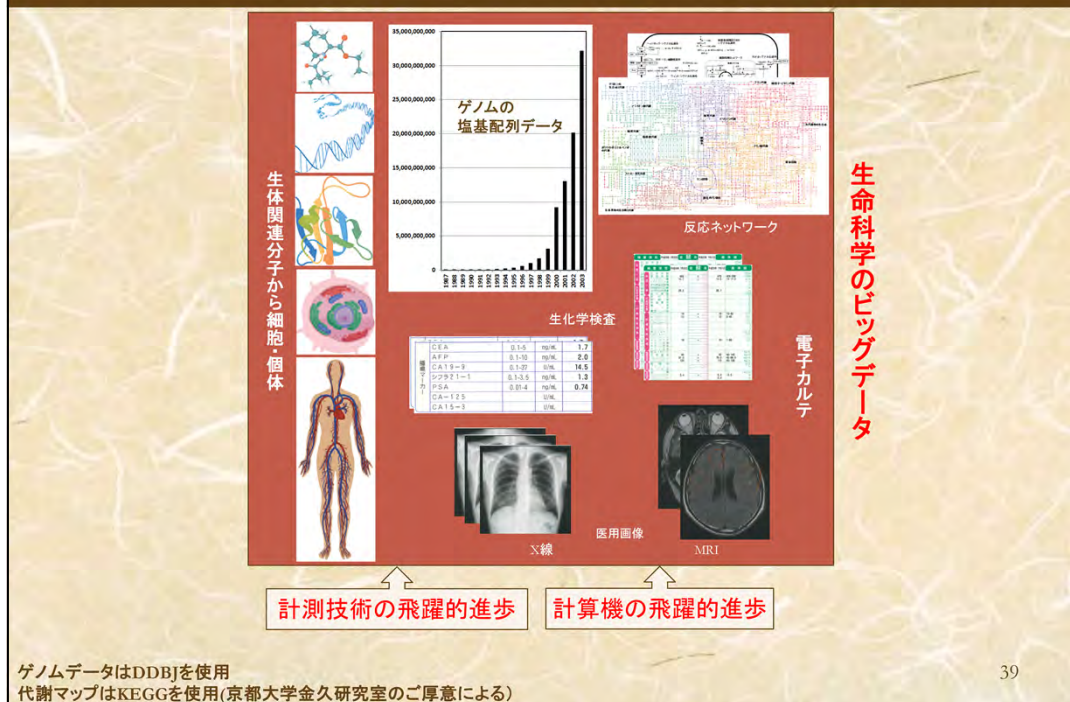
2001年から2007年までの推移をみると、徐々にではありますが、「より安く読み取ることができる」ようになってはいます。

それが、2008年には急激な価格低下が生じています。この原因はDNAシーケンシング技術の変化にあります。2007年までは主に、1980年にノーベル賞を取ったSanger法が用いられていました、それが2008年には「次世代シーケンサー」という新しい技術を用いたシーケンサーが現れたのです。

その結果、価格は今や3,000ドルまでにも低下しています。

今、次々世代とか第三世代とか言われるシーケンサーが現れてきており、ヒトゲノムの塩基配列を読み取る価格が1,000ドル以内になるのも時間の問題でしょう。このような時代になると、生命科学の研究や医療現場、製薬現場に大きな変革が生じるでしょう。アメリカのある研究者は「未来はすでに始まっている」という言葉を発信しています。

生命科学でのビッグデータの蓄積



このような計算機の進歩や計測技術の進歩に伴って、生命科学の分野で巨大なデータの蓄積が進んでいます。20世紀末から21世紀にかけて、自然免疫やiPS細胞の研究で生命科学のデータベースが基礎的な役割を果たしてきたことはすでに述べましたが、その量的な巨大化、質的な高度化、すなわちビッグデータ化が今急速に進んでいます。

その量的な面の典型例をゲノムの塩基配列データに見ることができます。真ん中、上の図は塩基配列のデータベース化がはじめられた頃、1987年から、ヒトゲノムの塩基配列の解読宣言がなされた2003年までの「データベース化され全世界の研究者が自由に閲覧し、使用できる塩基配列の数」の伸びを示しています。急激な増加をみることができます。

右上には生体内反応ネットワークの知識データベースであるKEGGの一部を示しています。代謝マップだけでも、このように複雑で、かつこの図の背景には巨大な知識が詰め込まれています。

また、左は、低分子から生体高分子、細胞、個体に至る各階層で「生命科学のデータがデータベース化されていますよ」という主張です。

個体レベルでは医療データのビッグデータ化がみられます。医学の分野で最も影響力のある雑誌New England Journal of Medicineでも2014年6月5日号に「Big Health Care Data」についての特集記事が出されるようになってきています。

一枚の図では示しきれませんが、今後「生命科学のビッグデータ」があらゆる生命科学の分野で基礎的研究基盤となっていくでしょう。

ある生命科学者の問題意識

「現在、がん生物学およびがん治療は細胞生物学や遺伝学、組織病理学、生化学、免疫学、薬理学からなる継ぎはぎ細工である。

しかし、いつの日にか、われわれは科学としてのがん生物学およびがん治療を想像することができるだろう。

そこには、現在の化学や物理学に匹敵するほどの概念構造と論理的な統一性、首尾一貫性を持った科学としてのがん生物学およびがん治療がある。」

Hanahan D, Weinberg RA (2000) The hallmarks of cancer, Cell, 100, 57-70

40

これら生命科学のビッグデータ化に対するとまどいもあります。

がん生物学の大御所ともいえるアメリカMITのWeinbergは2000年、雑誌Cellの論文を
発表し、このように述べています。

「現在、・・・(slide参照)・・・がある」と。

私はこのことは広く生命科学に言えることであり、下線で示した所を生命科学と読み替
えることができると考えています。

「現在、生命科学は継ぎはぎ細工である。

しかし、いつの日にか、われわれは科学としての生命科学を想像することができるら
う。

そこには、現在の化学や物理学に匹敵するほどの概念構造と論理的な統一性、首尾
一貫性を持った科学としての生命科学がある。」

これは私の夢でもあります。私はこの夢の実現には「理論－計算－観察と実験」という
3つの結節点にある計算生命科学が鍵になると考えています。この夢が21世紀に実現
することを願っています。

計算生命科学の研究例

41

ここから数枚のスライドで、京だけではありませんが、スーパーコンピュータを用いた計算生命科学の研究例を紹介します。

ヒトゲノムの塩基配列(30億文字)を全て読み取る

ヒトゲノム

HiSeq2500

イルミナ(株)のご厚意による

ヒトの場合30億塩基対をばらばらに読む

ATCCATCTCC...TCCTTCTCTG
GAGACTTGAC...GGAGCCTGAATC
.....
GAACCATGCT...AAGTACGAGCCG

例えば、
9億断片のジグ
ソーパズルを
スーパーコン
ピュータが解く

```
gtgccaccctc ctggagacct gccctcgggg tcccgcgatc cgaccasaca agtaagagaa cctctgtgaa gctcgtacct caactactct ctaaatassa gaatttcca gagaaggaaa gagagaagaa caaatattaa gaectctaga
ccctctgaga agactaaagt ctaccctctc aaatagcttc ttctcaatgt tgaigtgaga gttgaatgaa cttagaatct cactccaaag ttgacttasa agatctaa agatttgaag agagaagat attaatctct gatgcaggaa
tctgacagag agacatagaa caacagtctc atgatcaaa atgaaatgag caactttagc agccattag atctgagag caagaaatag agataaatg agataaatg gtagaccac cttcaagta agcacatag agctctgca tctgacaaa
cagcagcttc tccagcagc agacaacct ttgtaccag agacaacag tgaattgaaa tctctaaag caatgaaaa gacagaccca ctgcaatgc cagtctcgt gacctgatg atcccacag tgaaccacc tcaagaaag
cagcagattc ttacagaaa gttctctct cctcagcag tccagccct cctccagcag cagcaagct tgaattgca gacagcaaaa ctacaagat ttacaagaa acagcaagaa cagttaact tcaactttt gacacagcag
caacagcagc agcagcaaaa gcagcaacag cagcagcagc agcagcaaaa acagcagcag cagcagcaac agcaacagca gcagcagca cagcagcagc aacagcaaaa gcatctgaa agcaagcaaa agagcagca gcagcaaacg
cagcagcagc agttgagcag gcagcagctc gttctccagc agcagcttct ccaatgcaag cagctaccag agcaacaaa tctctcagc cttcagcgc aggcctcat ctccatcca cccgcacag cagcctctc tctcagctg
ctgctcaag ctgctttag tctctgagc attcagcaac tatgaaaga agtgactgaa gtcctatga tgaagaaaa cggcatcag cattgagcgc tagacctca gactaacat tctctcga ctacctctc caacagctc
aaagataca caaccatac acaatctcc atagtgaag gacagcttc agttctgaat gcaagcggg acagctatc acatgagag acgggacct cgcacctct ctatgcaat ggaactgca agtgccagc ctgtgaagc
atatgtgaag atttgaaa gttttgaa caacttaaa atgacatgc attgatgac cgaagcact cccantctc agtcaaatg cagttgtac agcagttga aatacngtt tcaagaaac gcaagctct tcaagcagc
atgaccact tgcacatgag acctctagag cccaaacct ctccaaacc tctaaatctg gtgtctagtg tcaactgtc gaagacatg ctggagact ccccaagag cttaactaa acccttaca caacaagc cccagctacc
ccagctacc agagacctc tgaatcacc cagcagctg tcccantgt gggagcata cgaagagagc attcaagaaa atcaaacat ccaatgtat cagagatgc cccaatcag gaatttata naagcaaga tgttaacct
ccctcaact atgcaacct cattaagcag gctatctag agctactgag caggcaatta accttaatg aattttag ctgctitaca cggacattg ctactttag gggatgccc gcaacttga agaatgagt gctctaat
cttagctca acaaatgttt tttctgata gaanaatga naagcagct gtgactgtg gacaaatgc aatacagaa gcaagctca caaagata cagaagtc acctttaga aanaatca ctcaatgt agctatga
gcagcttaa atccagctt gcagctgagc ttggcaaaa gcagcttgc ttgctaat aaccttgc tgaactaaa cctccagat gacctctgc aggcctcca cgaagacct aatgtccc tgaacacct cgaacaaat
gggaacagca ctccgggtg ctccctcag ccacacatac atcaatcca tctcaagaa gaacctgta ttgagagaa tgaagactg ccaatgcat tagtacaac agcaatcag agctcagat tagaagata cagagaaat
gaggaagc cttatctga gcaactgaa tgaagcaaa cttgtgac ctactgtga agggacat cactgacct cataacct ccaaacct gaatttga caaatitaa ctgtactat ttataga tgaataag
```

www.riken.jp/r-world/info/release/press/2004/040803/index.html

42

先ほども述べましたが、ヒトゲノムの塩基配列の読み取り技術は急速に進んでいます。ここにヒトゲノム解読方法の一例を示します。

ヒトゲノムの塩基配列を読み取るために、まずDNAをcopyし、一つ一つシュレッターにかけるようにして、ランダムに断片化します。それら断片は100文字(塩基対base pair)ほどで、30億文字のヒトゲノムの30倍ほどの文字数を読み取っています。つまり9億断片の文字列(一つ一つは100文字ほど)がヒトゲノムをランダムに覆い尽くしています。

その9億断片をジグソーパズルのように並べていきます。元の絵は2003年に発表されたヒトゲノムの塩基配列などですが、個人個人で塩基配列は異なりますので、「元の絵に似通った絵」が完成図です。その完成図が図の下に示すような連続した文字列です。

ジグソーパズルを実行し、完成図を作る段階はどうしても高性能計算機が必要になります。9億ピースのジグソーパズルを想像してみてください。「ちょっとやってみよう」というものではありません。

ヒトゲノムの塩基配列(30億文字)を全て読み取る

ヒトゲノム

イルミナ(株)のご厚意による

ヒトの場合30億塩基対をばらばらに読む

ATCCATCTCC...TCCTTCTCTG
GAGACTTGAC...GGAGCCTGAATC
.....
GAACCATGCT...AAGTACGAGCCG

例えば、
9億断片のジグ
ソーパズルを
スーパーコン
ピュータが解く

現在健康診断で行われている生化学検査と同じ程度に「私の今のゲノム」が手軽に検査される時が、近く来そうです。

www.riken.jp/r-world/info/release/press/2004/040803/index.html

43

このようにしてヒトゲノムの塩基配列を読み取りますが、さきほども「計測技術の進歩」のところで述べましたように、読み取る値段はどんどん下がってきています。

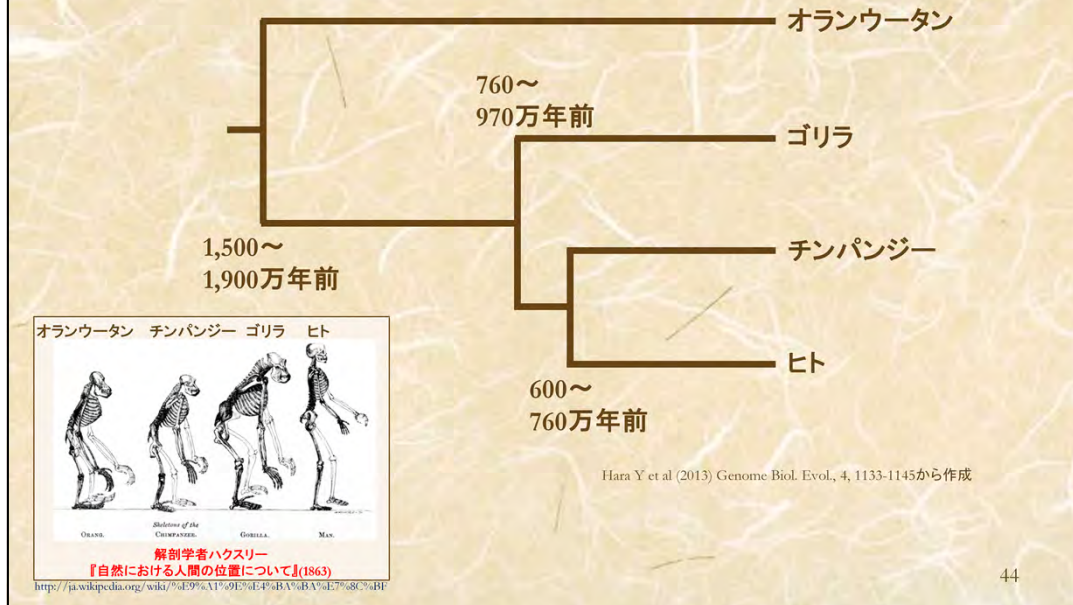
将来のことになりますが、今、クリニックで検診を受けると出される生化学検査結果と同じような感覚で「私の今のゲノム」が読み取られ、よりよい医療を受けることができるようになると思います。

がんは「ゲノムの異常によって生じる疾患」とも言われています。この図に示す技術は将来、がん治療にとっても福音をもたらしてくれると思います。

Vincentらが”New England Journal of Medicine”の2012年6月7日号に発表した論文”Two Hundred Years of Cancer Research”では、そのうちに100ドルほどで個々人のゲノムの塩基配列が読み取られ、がんをもたらすゲノム設計図の複雑さが明らかにされ、今まで想像もできなかった遺伝子間のネットワークが明らかにされるだろうと、述べられています。そしてさらに、新しいがん治療への期待が表明されています。

塩基配列やアミノ酸配列から進化系統樹！

ヒトを含む類人猿の進化の歴史も、姿・形からではなく、塩基配列やアミノ酸からも推定されています。



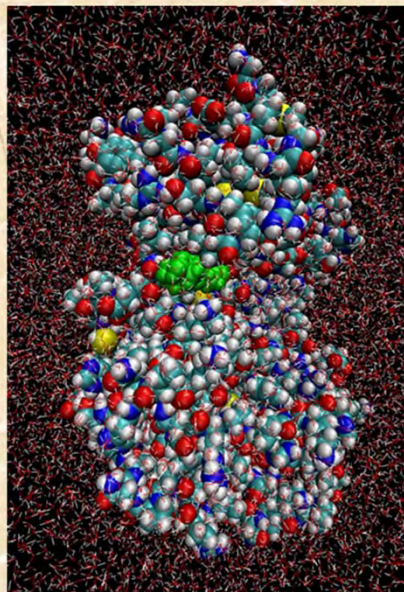
進化研究の分野でも計算生命科学が活躍しています。

左下の図は解剖学者であるハクスレーが書いた「自然における人間の位置について」という本に記載されている図です。この図から、ハクスレーが「オランウータンからチンパンジー、ゴリラ、そしてヒトへと進化してきた」と考えていたと断言はできませんが、そのように想像することはできます。

今では、進化も計算生命科学によって研究される時代になっています。ゲノムの塩基配列を用いたHaraらの研究によると、まずオランウータンが枝分かれます。それは1,500万年～1,900万年前と予想されています。次に、ゴリラが760万年～970万年前に枝別れし、ごく最近、600万年～760万年前にヒトとチンパンジーがそれぞれ違った進化の道を歩み始めています。

このように、今では進化の過程は、それぞれの生物の比較形態学的な研究、古い地層から掘り出される化石の研究とともに、ゲノムDNAによる研究が加わって解明されつつあります。なかなか信じにくいのですが、ネアンデルタール人のゲノムDNAの解読も進められており、現代人との比較研究が進められています。

キナーゼ阻害剤の働きを見る計算機顕微鏡



$$F = m \alpha$$

力Fは、フックの法則やクーロンの法則、ファンデルワールス力など物理や化学の教科書にある法則に従って導かれています。

タンパク質ERK2とその阻害剤

薬の開発のかなり初期の段階でも計算生命化学の活躍の場があります。

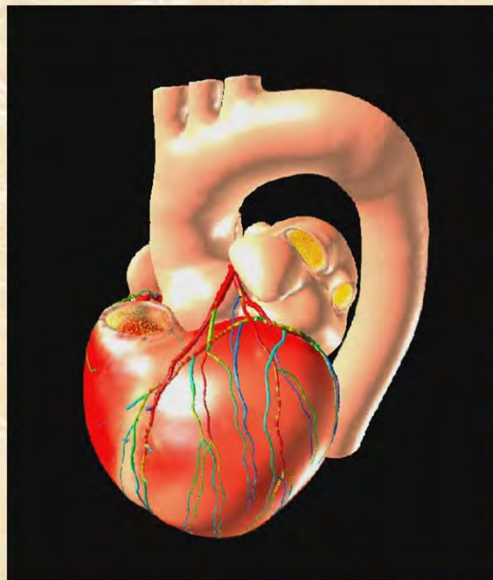
この図はタンパク質とその阻害剤が水の中で結合している状態を分子動力学計算でシミュレーションしている図です。

分子動力学計算はニュートンの運動方程式を解きますが、その方法で、いまだ合成はされていませんが新たに設計された薬剤が、標的とするタンパク質とどの程度強く結合し、薬としての機能を発揮するのかを定量的に評価することができます。

ただ、そのために必要とする計算資源は莫大で超並列高性能計算機が登場してようやく「分子動力学計算で具体的に新薬を開発してみようか」と多くの方々が考える時代になってきています。

スーパーコンピュータによって「見える世界が格段に広がった。研究もより広い視野で考えることができるようになった。」と言われるよい例と考えています。

細胞モデルからの心臓シミュレーション 久田・杉浦(東大)



心電図



バーチャル心像超音波



- ・心筋細胞内のたんぱく質の確率的運動から細胞の収縮、心拍動、血液駆出、冠循環までを一貫してシミュレート。
- ・シミュレーションから超音波エコー、流速ドップラー、心電図、カテーテル検査などの精緻なデータが再現される。そのデータを基に病態の解析が可能に。

東京大学 久田、杉浦、鷺尾、岡田研究室 協力:富士通株式会社

これは、心臓シミュレータUT-HEARTから出力される動画の一例です。

UT-HEARTは原子・分子の世界から、細胞、組織、そして臓器という多くのレベルを統合した心臓シミュレータ、すなわちマルチスケール・マルチフィジックス心臓シミュレータです。そのため、心臓の機械的な動きや血流だけでなく、右上に示す心電図をも計算することができます。

現在、この心臓シミュレータを臨床の現場で活用していくため、各地の病院と連携した研究が進められています。

計算生命科学の講義紹介

47

これまでの講義で「計算生命科学とはなんだろう?」という問に答えることができました。

では、「計算生命科学の基礎」としてこれから始まる14回の講義の概略を説明します。

計算生命科学の基礎

第1編 ゲノムから見る生命科学

第2編 タンパク質からみる生命科学

第3編 医療・創薬における計算生命科学

48

講義は、ここに示しますように、大きく3つに分かれています。
Organizerの先生方とこの詳細な構成を考えてきました。

計算生命科学の基礎

第1編 ゲノムから見る生命科学

1.1 ゲノムに記された大規模生命情報の解析

担当: 奥野恭史(京都大学大学院医学研究科)

1.2 バイオメディカル・インフォマティクス

担当: 奥野恭史(京都大学大学院医学研究科)

1.3 遺伝子ネットワーク解析

担当: 土井淳(株式会社 セルイノベータ)

1.4 細胞のシステム生物学

担当: 江口至洋(理化学研究所)

49

1. 1は、はじめに述べました自然免疫やiPS細胞の研究で用いられていた計算生命科学、特にバイオインフォマティクスといわれますが、その講義です。

1. 2は、バイオインフォマティクスを医療分野で活用していくもので、講義のタイトルをバイオメディカル・インフォマティクスとしています。

1. 3は、iPS細胞の研究で紹介した二次元クラスタリングをより精緻化した研究ともいえる、遺伝子間の相互関係をネットワーク表現する方法を紹介します。遺伝子ネットワーク解析は遺伝子の機能解析や、創薬ターゲットの探索でも利用されています。

1. 4は、はじめにも述べましたHodgkinとHuxleyの研究に代表される細胞内反応ネットワークを解析する手法と研究事例の紹介です。

計算生命科学の基礎

第2編 タンパク質からみる生命科学

2.1 計算生命科学のための量子化学基礎

担当: 佐藤文俊(東京大学生産技術研究所)

2.2 タンパク質の量子化学計算

担当: 福澤薫(日本大学松戸歯学部)

2.3 分子動力学計算と生体高分子の機能解析

担当: 中津井雅彦(神戸大学大学院工学研究科)

2.4 分子動力学計算における拡張サンプリング シミュレーション

担当: 中津井雅彦(神戸大学大学院工学研究科)

2.5 タンパク質の生物学的機能と化学反応

担当: 林重彦(京都大学大学院理学研究科)

50

2.1では、はじめに紹介しましたが、ノーベル賞を受賞した「複雑な化学システムのためのマルチスケール・モデル」の一つの柱である量子化学、その基礎を講義します。

2.2では、量子化学をタンパク質に適用していく手法と研究例を紹介します。

2.3と2.4では、古典力学、特に分子動力学計算の講義を行います。

2.5では、それら量子化学と古典力学を統合したQM/MMの最新の研究成果を「生物学的機能と化学反応」に焦点を当てて、紹介していきます。

計算生命科学の基礎

第3編 医療・創薬における計算生命科学

3.1 製薬におけるビッグデータおよびその解析

担当: 都地昭夫、北西由武(塩野義製薬株)

3.2 創薬における計算生命科学: 分子動力学計算を中心に

担当: 広川貴次(産業技術総合研究所)

3.3 創薬における計算生命科学: 量子化学計算を中心に

担当: 福澤薫(日本大学松戸歯学部)

3.4 医療におけるビッグデータ

担当: 田中博(東京医科歯科大学大学院疾患生命科学研究所)

3.5 医療における計算生命科学: 不整脈における

心臓興奮伝播現象を中心に

担当: 中沢一雄、稲田慎(国立循環器病研究センター研究所)

51

この第3編での一つのキーワードとして「ビッグデータ」を上げています。

3. 1と3. 4が製薬と医療の分野でのビッグデータを直接取り扱っています。

3. 2と3. 3は計算生命科学の創薬応用ですが、3. 2は分子動力学計算、3. 3は量子化学計算に基づいた研究を紹介します。

3. 5は計算生命科学で優れた研究成果をあげてきた電気生理に焦点を当てています。

これら14回の講義で、「計算生命科学の基礎」を学んで頂ければ幸いです。

この講義「計算生命科学の基礎」により、理論、計算、観察・実験の協力関係が一層深まって行くことを願っています。



理化学研究所 HPCI計算生命科学推進プログラム
のみなさんに感謝します

52

最後になりますが、この講義は実験生命科学に携わっている方々とともに、広く理工系の方々をも対象として組まれています。

生命科学を志す皆さんは、活動電位や自然免疫、iPS細胞などの研究で紹介したように、意識するとしないにかかわらず、研究室にある実験・計測装置から生み出される大量のデータや、世界的に公開されている膨大な生命科学データベースを基盤に研究を進めることが求められてくると思います。

理学、工学を志す皆さんには、その基礎的な理論を生命科学に生かしていく場がますます広がってきているとともに、生命科学の方々から連携を強く求められてくると思います。

この講義により理学、工学、生物学などの枠を超えて、「理論、計算、観察・実験の協力関係がより一層深まって行くことを願っています。」

ありがとうございました。