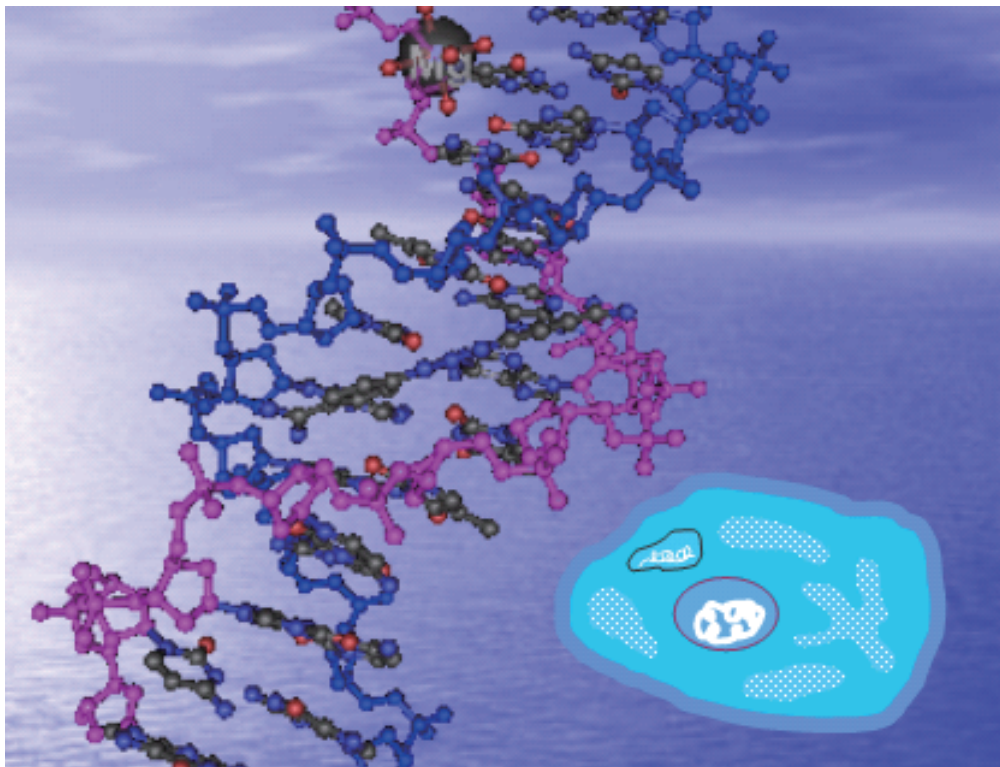


最新分子生物学即席入門



慶應義塾大学先端生命科学研究所

平成 20 年 4 月 9 日

はじめに

この冊子は分子生物学をほとんど知らない高校生や大学生、社会人の方々に手っ取り早く分子生物学の基礎と現状を理解して頂く事を目的として、慶應義塾大学先端生命科学研究所の若手研究者有志によって作成されました。

シーケンサーなどの分子生物学用の実験装置の開発が急ピッチで進められ、分子レベルでの生命現象に関する様々な知見が急速に蓄積しています。分子生物学はまさに激動の時代の中にあると言っても過言ではありません。

生物の基本原理を示しつつ、そんなエキサイティングな分子生物学の現状をこの冊子を通じて少しでも読者に伝えることができれば幸いです。

Saito R.



目 次

1	分子生物学の幕開けと遺伝子	1-4
2	遺伝子発現の中心教義	3-6
3	微生物遺伝子の基本構造	4-8
4	核酸の化学	4-8
5	タンパク質の化学	5-9
6	タンパク質コード領域とその解読	7-11
7	バイオインフォマティクス	7-11
8	遺伝子操作技術と微生物の利用	8-15
9	分子生物学と分析化学	9-16
10	これからの分子生物学	10-18

1 分子生物学の幕開けと遺伝子

「親子はなぜ容姿が類似しているのか?」「人はなぜ癌にかかるのか?」「ヒトはどのように進化してきたのか?」

生物や生命現象に関する興味深い疑問は尽きませんが、これらの疑問にミクロなレベルから答えてくれるのが分子生物学です [1, 2]。分子生物学はその名の通り、分子レベルで生命現象を理解することを目的とした学問分野です。

この分野が確立された 1950 年代当初の主な研究対象は単純で扱いやすい微生物 (microorganism) でした。微生物とは、人間の肉眼では見ることができない程度に小さな生物のことを指します。そのサイズはいろいろですが、大きくても $10\mu\text{m}$ 、ふつうはせいぜい $1\mu\text{m}$ 程度です。微生物のなかには、有名な大腸菌 (*Escherichia coli*) などの細菌 (bacteria)¹ やカビ (糸状菌)、キノコ (担子菌)、酵母などが含まれます。17 世紀後半にオランダのレーウエンフックが顕微鏡 (図 1) を発明してからこれらの微生物が実際に「小さな生物」として認識されるようになりました。そしてこれらの微生物を使って、個体の維持に必要で子孫にも受け継がれる情報、すなわち遺伝子 (Gene) がどのような仕組みで機能するのか研究が進められたのです。



図 1: レーウエンフックの顕微鏡

遺伝子という言葉が使用されたのは新しく、1933 年にヨハンセンが使用したのが最初だといわれています。しかし、遺伝という概念の成立は古く、紀元前にまでさかのぼることができます。紀元前約 4 世紀の古代バビロニアには、馬の頭やたてがみの形が子孫に代々伝わる様子が描かれた石が残されています。また紀元前約 3 世紀には、中国で交雑によってイノシシを改良し、ブタをつくっていました。そこでは、なぜ「カエルの子はカエルなのか」という素朴な疑問に対する答えとして、親の形質は遺伝という概念 (物質だとは認識されていません) によって伝えられると認識されていました。この遺伝が、遺伝子という物質によって成り立つとする概念の発端は、19 世紀のヨーロッパで生まれます。1865 年、当時オーストリアの修道士であったメンデル (図 2) は、エンドウの交配実験から、エンドウの形質の伝達頻



図 2: ヨハン・メンデル (1822-1884)

度がある一定の整数値をとることを見出し、これによって遺伝が未知の物質によって仲介されることを予測しました [3]。有名なメンデルの法則の発見です。1900 年にこのメンデルの概念が基本的に正しいものであることが証明されましたが、その遺伝子の本体が DNA (デオキシリボ核酸、deoxyribonucleic acid) という化学物質であることが証明されるには、さらに 100 年ほどの時間

¹ 細菌は原核生物 (真正細菌と古細菌がある) とも呼ばれる。他の生物は真核生物に分類される。

が必要でした。

1928年、イギリスのグリフィスは、熱殺菌処理した病原性肺炎双球菌 (*Streptococcus pneumoniae*) と、未処理の非病原性肺炎双球菌を混合することによって非病原性肺炎双球菌が病原性を獲得することを発見しました。これは形質転換と呼ばれ、その原因物質を形質転換因子と呼びました。その後、アメリカ・ロックフェラー大学のアベリーは、肺炎双球菌の形質転換が、DNAの授受によって起きることを証明し、ここに現代の遺伝学・分子生物学の黎明期がスタートしました。生物の遺伝子の本体がDNAという化学物質であるという発見は、当時の遺伝学者のみならず、化学、物理学といった一見無関係とも思える分野の研究者たちを魅了し、多くの人材がこの分野に参入するきっかけとなりました。

1953年、当時イギリス・ケンブリッジ大学のキャベンディッシュ研究所で研究していた若き物理学者であるワトソンとクリック (図3) は、この遺伝子の本体であるDNAの立体構造を明らかにすることによって遺伝現象の分子メカニズムを解明できると信じ、ついにあの美しい二重らせん構造を発見しました。彼らの発見で重要な点は、

1. DNAはアデニン (A)、グアニン (G)、シトシン (C)、チミン (T) という4種類の塩基から構成される (4節参照)
2. それらが互いに決まったペア (AとT、CとG) を作るような二重らせん構造をとる

ということです。これによってDNA複製のメカニズムが予測できるようになったと同時に、DNA mRNA タンパク質という遺伝情報発現の流れである「中心教義 (セントラルドグマ, central dogma)」がより明確に認識されるようになりました (2節参照)。

1961年、パスツール研究所のジャコブとモノーは、大腸菌のラクトース分解系酵素の研究から、有名な「オペロン説」を提唱しました。彼らは、遺伝子上には遺伝子発現を調節する領域が存在し、この作用によって、細胞外の変化に応じた応答 (例えばラクトースがたくさんある場合は、ガラクトシダーゼというラクトース分解酵素が誘導されるといった) が起きるということを、実験から予測しました。これは現在の分子生物学では拡大解釈され、遺伝子の基本単位は、タンパク質のアミノ酸配列情報が書き込まれた領域 (タンパク質コード領域) と、この情報の発現を制御する制御部位 (プロモーター, promoter) から構成される (この単位をオペロンと呼びます) という概念が出来上がりました。これらの発見から、生物の遺伝現象はこれまでに人類が築き上げてきた諸知識を集約することによって明らかにできることがわかり、分子生物学が発展を遂げ、多くの研究者によって多大な成果が得られています。アメリカのマサチューセッツ工科大学の利根川教授による抗体分子の多様性原理の解明 (1989年度ノーベル生理学・医学賞受賞) も、この分子生物学的技術と概念の上になされた仕事です。

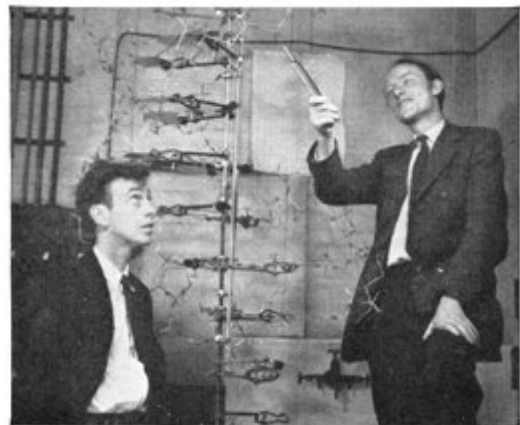


図3: DNA二重らせんモデルを囲むワトソンとクリック

役割	体の器官	構成タンパク質の名前（例）
体の形を整える	皮膚，毛髪，骨，肺	ケラチン，コラーゲン，エラスチン
感覚器	目のレンズ，角膜，網膜，ガラス体	コラーゲン，クリスタリン，ロドプシン
食物の消化	消化酵素	ペプシン，トリプシン，キモトリプシン，リパーゼ，ヌクレアーゼ
栄養の運搬	血液タンパク質	アルブミン，トランスフェリン，リポタンパク質
呼吸	肺と諸器官での酸素交換	ヘモグロビン，ミオグロビン，炭酸デヒドラーターゼ
免疫など	免疫システム	免疫グロブリン，補体， γ - マクログロブリン
体内情報伝達	細胞間連絡	成長ホルモン，インスリン，グルカゴン

表 1: 体内のタンパク質の主な働き

2 遺伝子発現の中心教義

タンパク質は分子レベルにおける生命現象の主役と言っても過言ではありません。細胞内で起きる化学反応（糖やアミノ酸の分解など）は、酵素と呼ばれるタンパク質によって触媒されていますし、また細胞の形を維持するために骨格のように張りめぐられたアクチンという物質も、その正体はタンパク質です。さらに表 1 に我々の体内で活躍するタンパク質の例を示しますが、このように、タンパク質は細胞内のさまざまな機能を司るという重要なはたらきを担っていることが分かるでしょう。

タンパク質は、20 種類のアミノ酸がさまざまな長さにつながった分子で、そのアミノ酸配列によってタンパク質の性質が決定されています。このタンパク質のアミノ酸配列の設計図が、遺伝子上に塩基配列として書き込まれています。おおまかにいうと、遺伝情報の発現というのは、実はこの遺伝子上の塩基配列を正確に読みとってアミノ酸を順序よくつないでいくことを指しています（6 節参照）。この遺伝情報の発現は図 4 に示したように、3 つの段階を経て行われます。つまり、DNA 上に書き込まれた遺伝情報は複製されて各細胞に分配され、発現する際、一度 mRNA（メッセンジャー RNA: ribonucleic acid）という分子に転写されます。この反応は RNA ポリメラーゼという酵素が触媒します。次にリボソームという大きな複合体によって、mRNA に写し取られた情報をもとにアミノ酸を連結していきます。これを翻訳と呼んでいます。この遺伝子の発現メカニズムは基本的にすべての生物で共通なので、中心教義と呼ばれています。

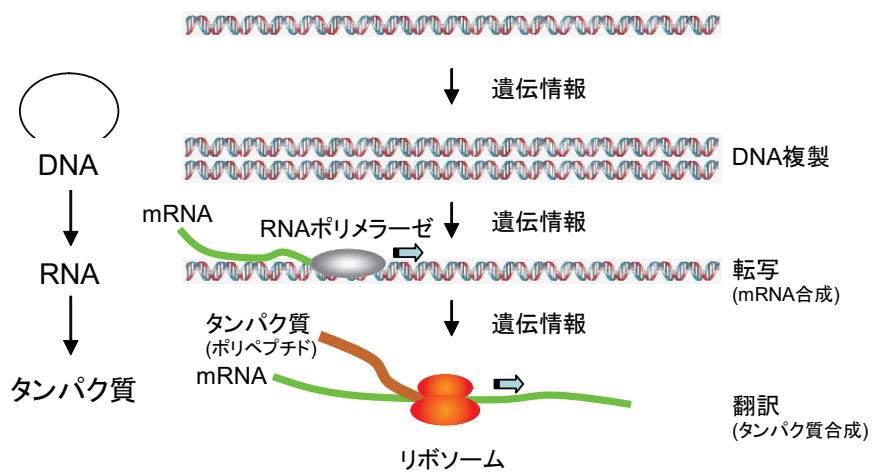


図 4: 遺伝子発現の中心教義

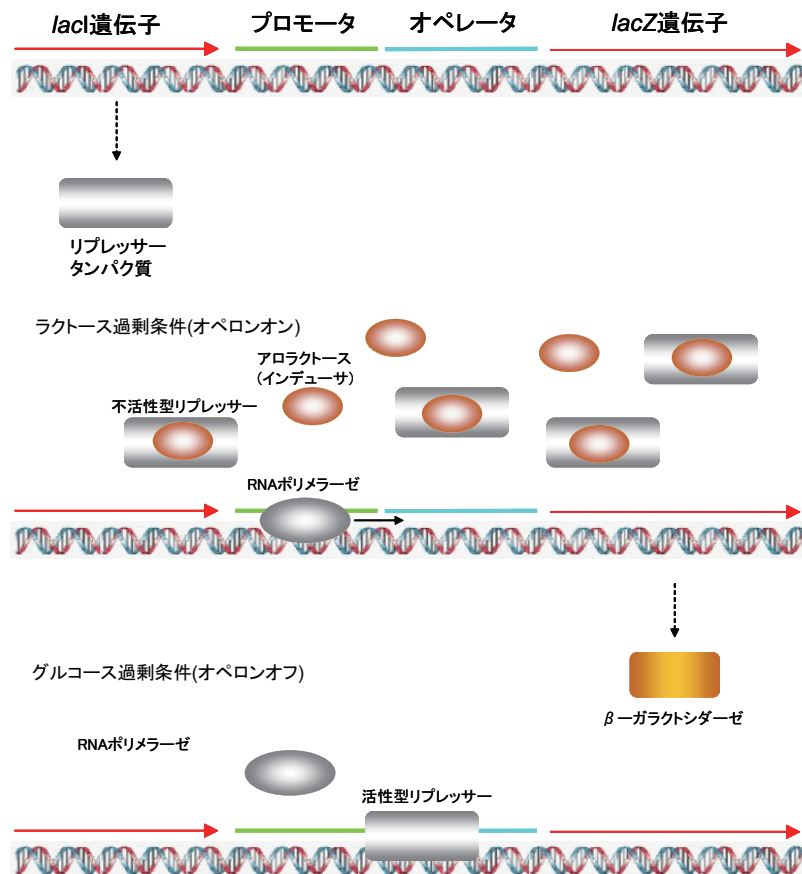


図 5: *Lac* オペロンの遺伝子制御

されます²。RNA はアデニン、グアニン、シトシン、ウラシル (U: uracyl) の 4 種類の塩基がリボースとリン酸を含むヌクレオチドとしてつながったものです。アデニンとグアニンは比較的大きな分子でプリン塩基とよばれ、チミンとシトシンは小さく、ピリミジン塩基とよばれます。DNA は二重らせん構造をとっていますが、アデニンはチミンと、またグアニンはシトシンとそれぞれ水素結合で対合しています。対合する塩基を塩基対 (base pair) とよびます。DNA の長さは、base (b) や base pair (bp) を単位とした塩基対の数で表します。

5 タンパク質の化学

タンパク質はアミノ酸が鎖状につながった構造をしています。構成するアミノ酸の種類や並び方、長さによって多様な機能を持つことができます。図 7 にタンパク質を構成する 20 種類³のアミノ酸の構造式と名称、表記法を示します。タンパク質の中のアミノ酸は、そのカルボキシル基 (-COOH) と隣のアミノ酸のアミノ基 (NH₂-) が脱水縮合したペプチド結合によって鎖状につながっています。アミノ酸はその側鎖の種類によりおおまかに以下の 4 種類に分類できます。

1. 疎水性アミノ酸 (Ara, Val, Leu, Ile, Met, Trp, Phe, Pro)
2. 極性無電荷アミノ酸 (Gly, Ser, Thr, Cys, Tyr, Asp, Gln)
3. 正電荷をもつ極性アミノ酸 (Lys, His, Arg)
4. 負電荷をもつ極性アミノ酸 (Asp, Glu)

各分類に属するアミノ酸同士は、側鎖の大きさに違いがありますが、化学的な性質は似ているということができます。

タンパク質の構造は 4 つの段階によって決定されています。1 次構造とはアミノ酸の配列のことを指します。2 次構造とは、連続するアミノ酸が局所的にとる立体構造を指し、 α -ヘリックス (らせん) や β -シート構造などが知られています。3 次構造は、それらの 2 次構造が組み合わさって一本のペプチド鎖としてどのような全体構造をしているかを示します。そして 4 次構造とは、3 次構造をとったペプチド鎖がさらに高次に組み合わさって複合体をつくるときの、その組み合わせ構造をさします。このようにアミノ酸が連結してできたタンパク質は、非常に複雑な構造を持っていて (図 8)、様々な細胞内反応の触媒機能やその他の機能を司っています。

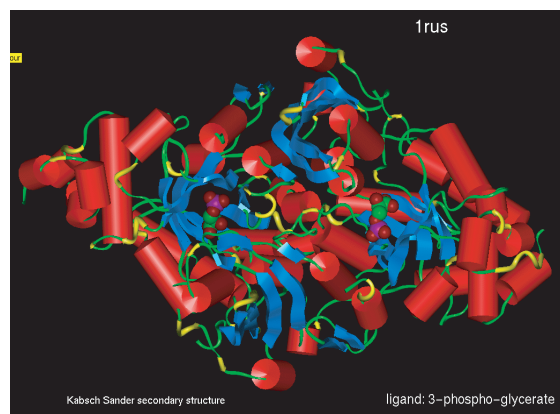


図 8: タンパク質の立体構造の例

²最近では sRNA や miRNA など様々な RNA が多数発見されており、未解明の生命現象や分子進化を解き明かす分子として注目を集めている。

³近年ではセレノシステインやピロリシンなど細胞内で合成される新たなアミノ酸も発見されている。

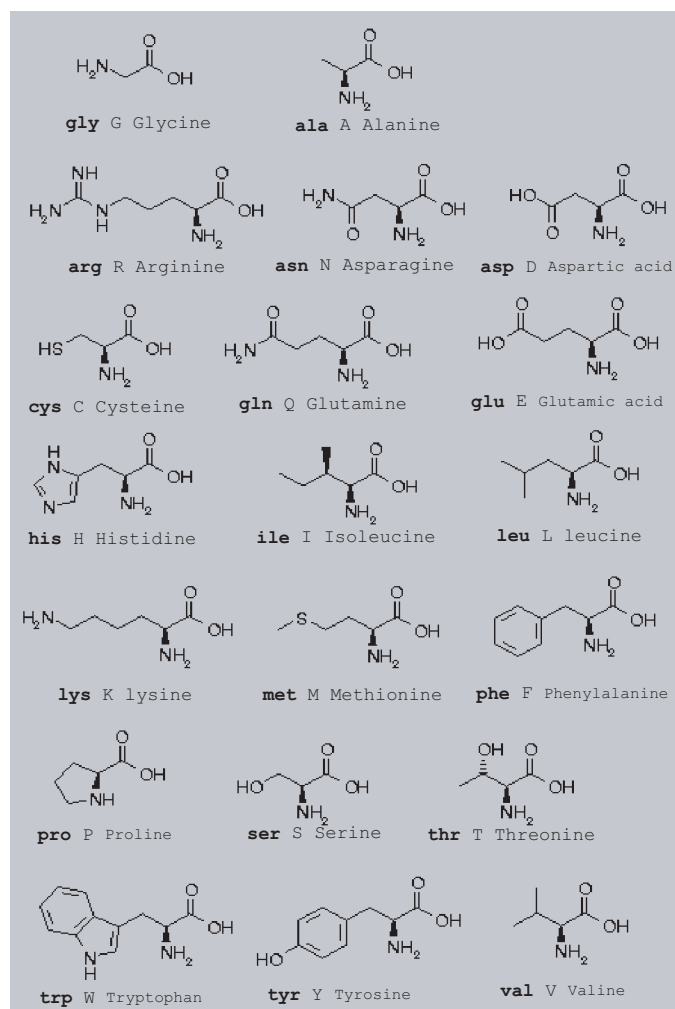


図 7: タンパク質を構成する 20 種のアミノ酸の構造式と表記法

タンパク質の構造と機能の研究は、現在多くの研究者が参入している分野の一つで、最近の 10 年ほどで立体構造の解かれたタンパク質の数は指数関数的に多くなっています⁴。しかし、タンパク質の構造から機能（ある反応を触媒したりある物質に結合したりする能力）を予測することは難しく、未知の分野となっています。タンパク質の立体構造からその機能を予測しようという試みが、日本を中心にした研究グループによって進められており、物理学や化学などの研究者が参加しています。

6 タンパク質コード領域とその解読

タンパク質のアミノ酸配列（1 次構造）は、それをコードしている遺伝子の塩基配列によって決定されています。塩基は 3 つで 1 つの組すなわちコドンと呼ばれる単位を作っており、1 つのコドンで 1 つのアミノ酸を指定しています。この対応関係は遺伝暗号（genetic code）と呼ばれています。表 2 に遺伝暗号をまとめたものを示します。

図 9 に示すように核酸上のタンパク質の情報を持っている部分すなわちコード領域の初めの部分には開始コドンがあり、ここからアミノ酸を指定するコドンが始まります。開始コドンは AUG なのでこれはメチオニン (M) に翻訳され、次は AGC なのでセリン (S) に翻訳されます。一方、コード領域の終わりを指定しているのが終止コドンです（表 2 では赤色で示してあります）。終止コドンは 3 種類あります。ここでアミノ酸の伸長反応が停止して、ペプチド（タンパク質）が遊離します。このようにして、コード領域の中の塩基配列は翻訳されてアミノ酸配列へと変換されます⁵。

さて、このようにして DNA 中のコード領域の情報から、アミノ酸配列を知ることができますが、それを眺めただけでは、そのタンパク質のもつ機能を予測することは不可能です。機能を知るためにはそのタンパク質が相互作用する相手の分子を調べたりする実験的な検証が必要になります。ところが最近公共のデータベースに大量の DNA 情報やタンパク質の情報が蓄積してきたおかげで、コンピュータを使ってアミノ酸配列からその機能を予測することが可能になってきたのです（7 節参照）。

7 バイオインフォマティクス

生物はどれくらいの量の遺伝情報によって成り立っているのでしょうか？大腸菌の遺伝子は約 4,000 個存在し、約 4,000,000 塩基対の DNA にその情報が書き込まれています。ヒトに至っては約 30,000 個の遺伝子⁶が大腸菌の 1,000 倍弱に当たる約 3,000,000,000 塩基対の DNA にコードされています。この生物の持つ遺伝情報のセットのことをゲノム（genome: gene と染色体をあらわす chromosome とを合わせた造語）とよびますが、最近このゲノムに記された遺伝情報をすべて読んでしまおうという「ゲノムプロジェクト」が成果をあげています（図 11）。

1995 年に、アメリカの TIGR（The Institute for Genomic Research）によってマイコプラズマ菌（*Mycoplasma genitalium*）という比較的小さな染色体をもつ細菌の遺伝子解読が終了したのを

⁴近年日本でも理化学研究所でタンパク 3000 など大規模なタンパク質構造決定プロジェクトが進められた。

⁵問題：大腸菌の *thr* operon leader peptide という代謝に関わるタンパク質のコード領域は catcccATGaaacgcatt...ggtgcgggcTGAcgcgt のような配列になっている。但し大文字で示した部分は左は開始コドン、右は終止コドンである。このとき図 9 にならうと、(1) DNA 上のこの配列の逆鎖の配列は？(2) コードされるアミノ酸配列の先頭と終わりの 3 文字は？

⁶ゲノム配列がまだ読まれていない時代はヒトの遺伝子数は 100,000 くらいと見積もられていた。

	U			C			A			G		
U	UUU	Phe	F	UCU	Ser	S	UAU	Tyr	Y	UGU	Cys	C
	UUC	Phe	F	UCC	Ser	S	UAC	Tyr	Y	UGC	Cys	C
	UUA	Leu	L	UCA	Ser	S	UAA	Ochre	終止	UGA	Opal	終止
	UUG	Leu	L	UCG	Ser	S	UAG	Amber	終止	UGG	Trp	W
C	CUU	Leu	L	CCU	Pro	P	CAU	His	H	CGU	Arg	R
	CUC	Leu	L	CCC	Pro	P	CAC	His	H	CGC	Arg	R
	CUA	Leu	L	CCA	Pro	P	CAA	Gln	Q	CGA	Arg	R
	CUG	Leu	L	CCG	Pro	P	CAG	Gln	Q	CGG	Arg	R
A	AUU	Ile	I	ACU	Thr	T	AAU	Asn	N	AGU	Ser	S
	AUC	Ile	I	ACC	Thr	T	AAC	Asn	N	AGC	Ser	S
	AUA	Ile	I	ACA	Thr	T	AAA	Lys	K	AGA	Arg	R
	AUG	Met	M	ACG	Thr	T	AAG	Lys	K	AGG	Arg	R
G	GUU	Val	V	GCU	Ala	A	GAU	Asp	D	GGU	Gly	G
	GUC	Val	V	GCC	Ala	A	GAC	Asp	D	GGC	Gly	G
	GUA	Val	V	GCA	Ala	A	GAA	Glu	E	GGA	Gly	G
	GUG	Val	V	GCG	Ala	A	GAG	Glu	E	GGG	Gly	G

表 2: 遺伝暗号表

各枠の中は左から、コドン、対応するアミノ酸 1 文字表記、アミノ酸 3 文字表記を表す。開始コドンは黄色、終止コドンは赤で示してある。実際にアミノ酸を連結するときには mRNA を鋳型とするので、ここではチミン (T) の代わりにウラシル (U) でコドンを示している。

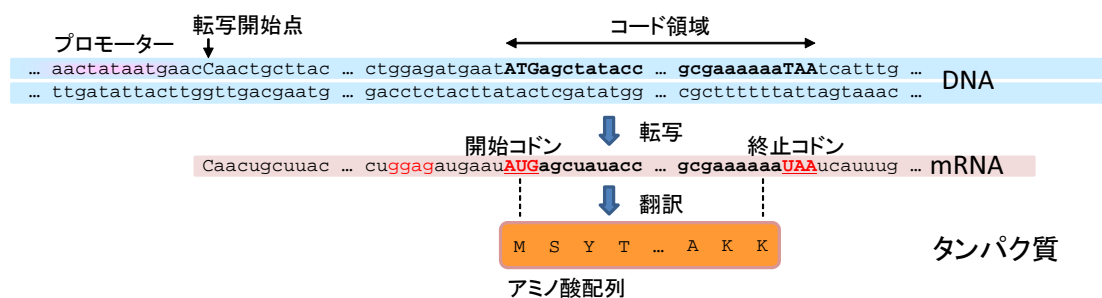


図 9: DNA 配列から遺伝暗号によってタンパク質ができる過程

大腸菌の細胞中で生産される有毒な遊離基 (ラジカル) を分解する酵素の遺伝子 *sodA* のコード領域とその周辺を示した。

「最新分子生物学即席入門」サンプル版について

2008/04/12

この冊子はサンプル版のため、一部ページが省かれております。完全版は慶應義塾大学環境情報学部先端生命科学系列の授業や、先端生命科学研究所で開催されている以下の高校生向けのイベントで配布されております。

- ・ 慶應サマーバイオキャンプ (KSBC)
- ・ サマーバイオカレッジ (SBC)
- ・ スプリングサイエンスキャンプ (SSC)

また生命科学に対して興味のある社会人向けには以下のイベントがあります。

- ・ バイオファイナンスギルド (BFG)

詳しくは、

<http://www.iab.keio.ac.jp>

をご覧ください。お問い合わせは、

scicamp_ttck@googlegroups.com

までお願いします。

参考文献

- [1] Alberts B et al. Molecular Biology of the Cell (4th edition) Garland Pub (2002/03)
- [2] Alberts B et al., 中村桂子、松原謙一監訳「Essential 細胞生物学 (原著第2版)」南江堂 (2005/09)
- [3] G. エドリン, Gordon Edlin(著), 伊藤 文昭, 大竹 英樹, 蓑島 伸生, 井口 義夫, 清水 淑子, 清水 信義 (翻訳)「ヒトの遺伝学」東京化学同人 (1992/04)
- [4] Fraser CM et al.(1995) The minimal gene complement of *Mycoplasma genitalium*. Science 270(5235):397-403.
- [5] Lander ES et al.(2001) Initial sequencing and analysis of the human genome. Nature. 409(6822):860-921.
- [6] Venter JC et al.(2001) The sequence of the human genome. Science. 291(5507):1304-51.
- [7] 富田勝 (監修) 斎藤輪太郎 (著)「バイオインフォマティクスの基礎～ゲノム解析プログラミングを中心に～」サイエンス社 (2005/07)
- [8] Altschul SF et al.(1990) Basic local alignment search tool. J Mol Biol 215(3):403-10
- [9] Kent WJ.(2002) BLAT–the BLAST-like alignment tool. Genome Res. 12(4):656-64.
- [10] Arakawa K, Mori K, Ikeda K, Matsuzaki T, Kobayashi Y, Tomita M(2003) G-language Genome Analysis Environment: a workbench for nucleotide sequence data mining. Bioinformatics 19(2):305-306 <http://www.g-language.org/>
- [11] 富田勝、西岡孝明 (編)「メタボローム研究の最前線」シュプリンガー・フェアラーク東京 (2003/11)
- [12] 志田保夫、笠間健嗣、黒野定、高山光男、高橋利枝「これならわかるマスマススペクトロメトリー」化学同人 (2001/03)
- [13] 日本分析機器工業会編「よくわかる分析化学のすべて」日刊工業新聞社 (2001/10)
- [14] Ishii N, Nakahigashi K, Baba T, Robert M, Soga T, Kanai A, Hirasawa T, Naba M, Hirai K, Hoque A, Ho PY, Kakazu Y, Sugawara K, Igarashi S, Harada S, Masuda T, Sugiyama N, Togashi T, Hasegawa M, Takai Y, Yugi K, Arakawa K, Iwata N, Toya Y, Nakayama Y, Nishioka T, Shimizu K, Mori H, Tomita M(2007) Multiple High-Throughput Analyses Monitor the Response of *E. coli* to Perturbations Science 316(5824):593-7
- [15] DeRisi JL, Iyer VR, Brown PO.(1997) Exploring the metabolic and genetic control of gene expression on a genomic scale. Science 278(5338):680-6.
- [16] Uetz P et al.(2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. Nature 403(6770):623-7.

- [17] Raïs B, Chassagnole C, Mazat JP.(1995) Control of threonine pathway in *E. coli*. Application to biotechnologies. *Acta Biotheor.* 43(4):285-97.
- [18] Tomita M, Hashimoto K, Takahashi K, Shimizu TS, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter JC, Hutchison CA 3rd.(1999) E-CELL: software environment for whole-cell simulation. *Bioinformatics* 15(1):72-84. <http://www.e-cell.org/>
- [19] Takahashi K, Kaizu K, Hu B, Tomita M(2004) A multi-algorithm, multi-timescale method for cell simulation. *Bioinformatics.* 20(4):538-46
- [20] 高木利久、富田勝 (編集) 「ゲノム情報生物学」中山書店 (2000/10)
- [21] Nakayama Y, Kinoshita A, Tomita M.(2005) Dynamic simulation of red blood cell metabolism and its application to the analysis of a pathological condition. *Theor Biol Med Model* 2(1):18.
- [22] Negishi Y., Nakamura H., Yachie N., Saito R., Tomita M. (2007) eXpanda: an integrated platform for network analysis and visualization. *In Silico Biology* 7, 0013. <http://medcd.iab.keio.ac.jp/expand/>
- [23] Arakawa K, Yachie N, Tomita M.(2007) E-Cell Simulation Environment 3D: 3-dimensional visualization of cellular simulation results. *ISMB 2007*. <http://ecell3d.iab.keio.ac.jp/>

索引

DNA, 1-4, 4-8

mRNA, 3-6

RNA, 4-8

アミノ酸, 3-6, 5-9

遺伝暗号, 7-11

遺伝子, 1-4

オペロン, 2-5

核酸, 4-8

ゲノム, 7-11, 10-20

酵素, 3-6

コード領域, 7-11

コドン, 7-11

システム生物学, 10-19

シミュレーション, 10-20

制限酵素, 9-16

セントラルドグマ, 2-5

相同性検索, 7-14

代謝, 10-19

大腸菌, 1-4

タンパク質, 3-6, 5-9

中心教義, 2-5, 3-6

転写, 3-6, 4-8

内部標準法, 10-18

バイオインフォマティクス, 7-14, 10-20

複製, 3-6

プラスミド, 9-16

プロモーター, 2-5, 4-8

分子生物学, 1-4

分析化学, 9-16

翻訳, 3-6

メタボローム, 10-20

メンデル, 1-4

Committee

Molecular Biology *Ohashi Y.*

E-Cell Simulation *Kikuchi S.*

Remaining Parts *Saito R.(Editor), Ito T., Saito N.*

Manuscript Proofreading *Ishii N., Sato A., Tsuge K., Masuda T., Otomo K., Kochiwa H.,
Otani N., Kuroki A.*



慶應義塾大学先端生命科学研究所
SBC・KSBC・SSC・BFG 実行委員会 (2008)
