計算生命科学における大規模計算の重要性

生命科学の対象は、シミュレーションであれ、情報解析であれ、常に自由度間の 相関の大きさに起因する大規模複雑系という困難さがあり、また系の著しい多様 性による個別論として扱わざるを得ない。個別論とは、系における詳細にいたる 特殊性が機能発現に与える影響を見ようというものであり、そこに考慮すべきモ デルの自由度が増大するひとつの理由、即ち大規模計算の重要性がある。

理化学研究所HPCI計算生命科学推進プログラム

木寺詔紀

2013年10月25日
独立行政法人理化学研究所

柳田 敏雄 センター長（生命システム研究センター）が文化功労者に選出

生命システム研究センターの柳田 敏雄センター長（大阪大学 特任教授）が平成25年度文化功労者に選ばれました。

柳田敏雄センター長は、タンパク質を１分子レベルで観察可能な高性能顕微鏡を開発し、筋肉の駆動力を生み出す分子モーターの動作原理を解明するなど、生命システムを構成する分子機械に関する生物物理学研究で世界をリードしてきました。その卓越した見識で、生命システム研究センターをけん引しています。また、HPCI計算生命科学推進プログラムディレクターを兼任し、新しい計算生命科学の開拓にも当たっています。さらに、大阪大学大学院生命機能研究科の特任教授として、また、情報通信研究機構/大阪大学　脳情報通信融合研究センター長として分子から個体まで広く生命現象に関わる原理を追求し、基礎研究と科学技術の発展に尽力しています。

顕彰式は11月5日都内で行われる予定です。

柳田センター長のコメント

日本の文化とも言うべき、いい加減に、ほどよく、の考え方を持ち込み、恩師大澤文夫先生と共に生命の理解に挑戦してきました。欧米文化にはあまり見られないユニークなゆらぎの概念で、長年世界の研究者と論争してきたのですが、このように評価していただいてとても嬉しいです。

The Nobel Prize in Chemistry 2013

Martin Karplus, Michael Levitt, Arieh Warshel

The Nobel Prize in Chemistry 2013

© Nobel Media AB

Martin Karplus

Photo: Keilana via Wikimedia Commons

Michael Levitt

Photo: Wikimedia Commons

Arieh Warshel

The Nobel Prize in Chemistry 2013 was awarded jointly to Martin Karplus, Michael Levitt and Arieh Warshel *"for the development of multiscale models for complex chemical systems".*

"for the development of multiscale models for complex chemical systems"

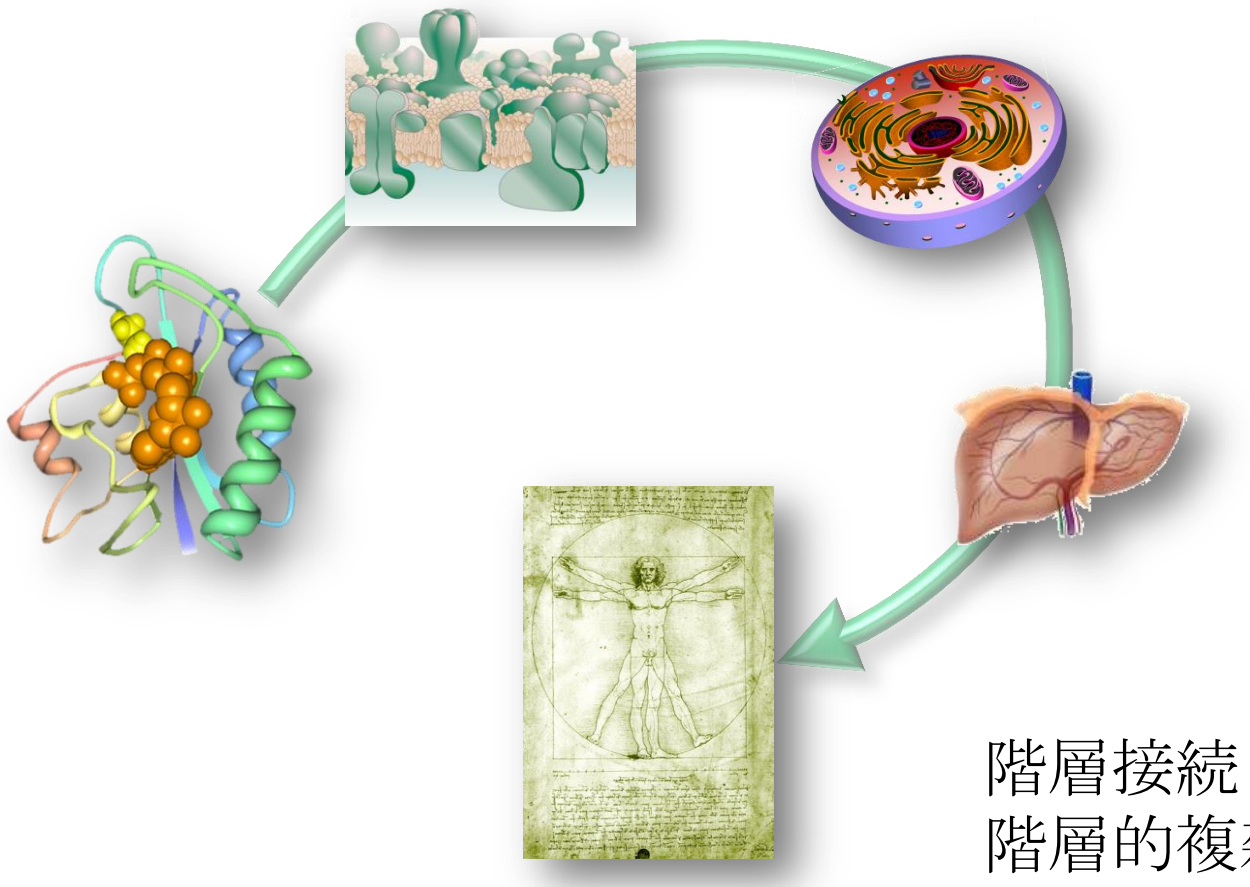
The computer – your Virgil in the world of atoms

Chemists used to create models of molecules using plastic balls and sticks. Today, the modelling is carried out in computers. In the 1970s, Martin Karplus, Michael Levitt and Arieh Warshel laid the foundation for the powerful programs that are used to understand and predict chemical processes. Computer models mirroring real life have become crucial for most advances made in chemistry today.

This year's Nobel Laureates in chemistry took the best from both worlds and devised methods that use both classical and quantum physics. For instance, in simulations of how a drug couples to its target protein in the body, the computer performs quantum theoretical calculations on those atoms in the target protein that interact with the drug. The rest of the large protein is simulated using less demanding classical physics.

Biologicalな研究とは何か？＝下階層からの階層接続

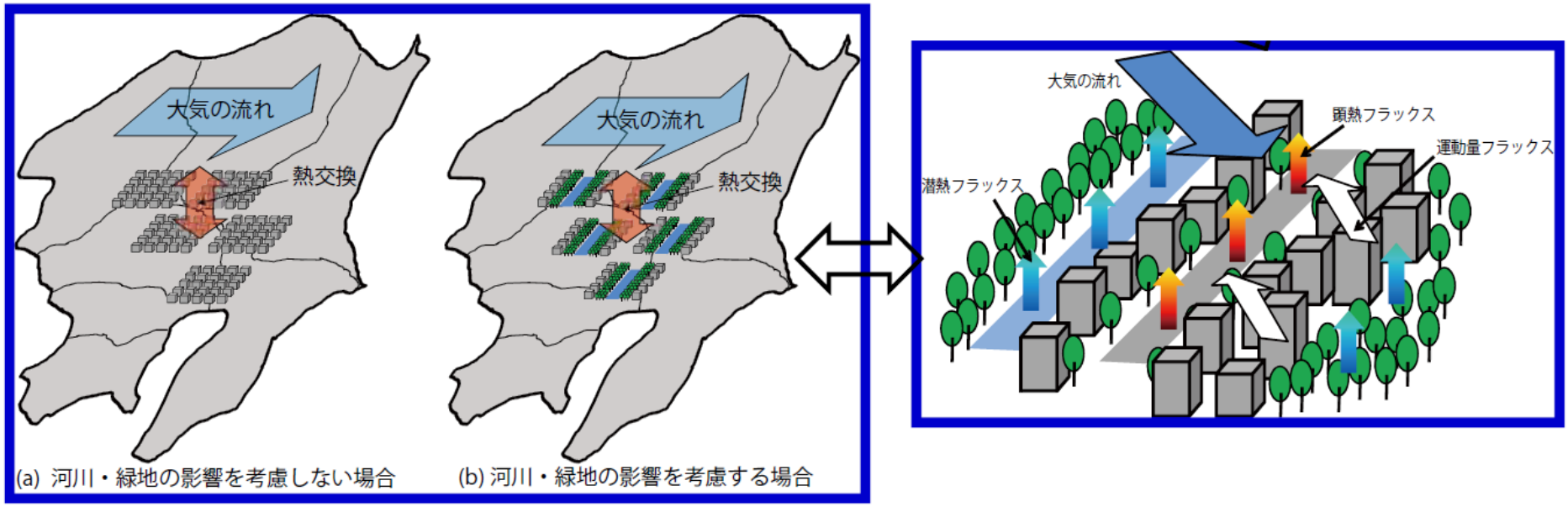
下層の情報で上層の現象を説明する



階層接続
階層的複雑系

階層接続の他計算科学分野での例
マルチスケールシミュレーション

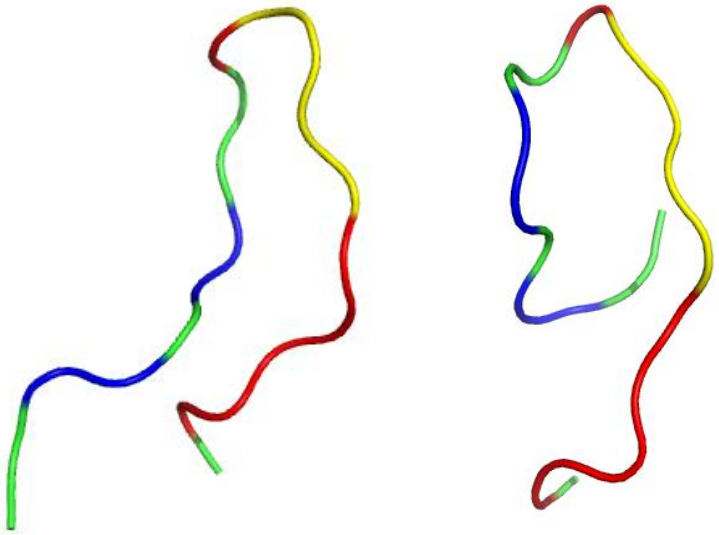
同一モデルの分解能の相違：非階層的複雑系



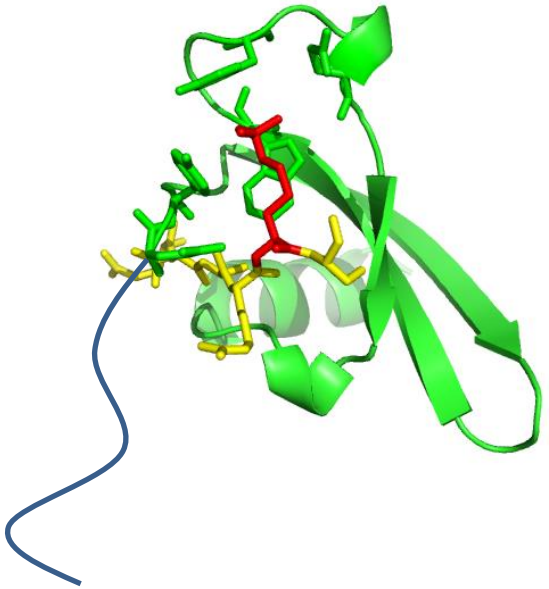
気候変動に適応可能な環境探索のためのマルチスケールシミュレーション
http://www.jamstec.go.jp/esc/projects/fy2012/2-takahashi.pdf

例

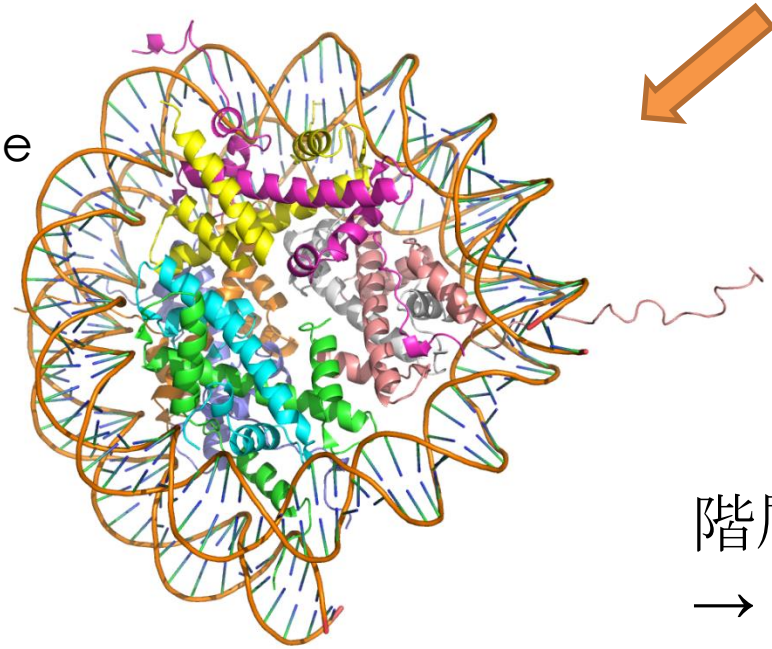
Conformational change of 20 residue peptides upon modification



Protein-peptide complex with a disordered region

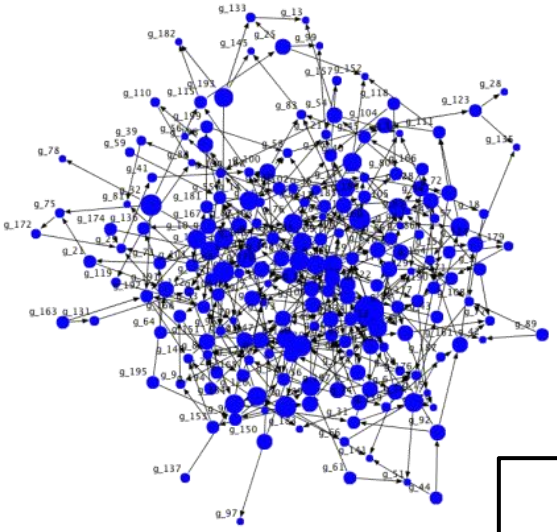


Nucleosome
Histone tail

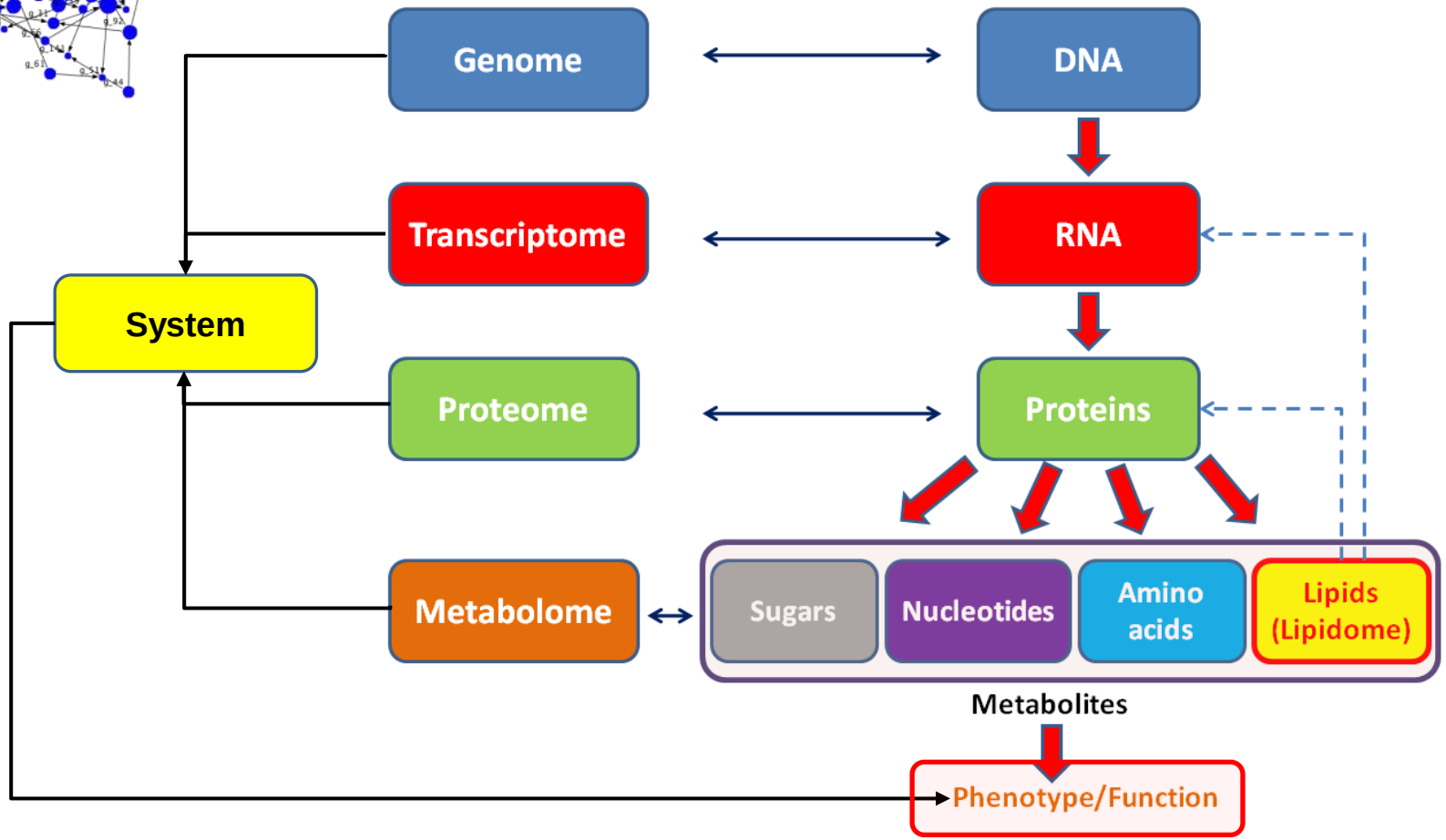


Epigenetic transcriptional regulation

階層接続
→　高次巨大複雑系への指向

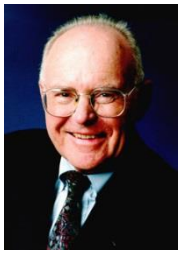


階層接続：生命情報解析
Systems : Omics



http://www.sciencebasedmedicine.org/

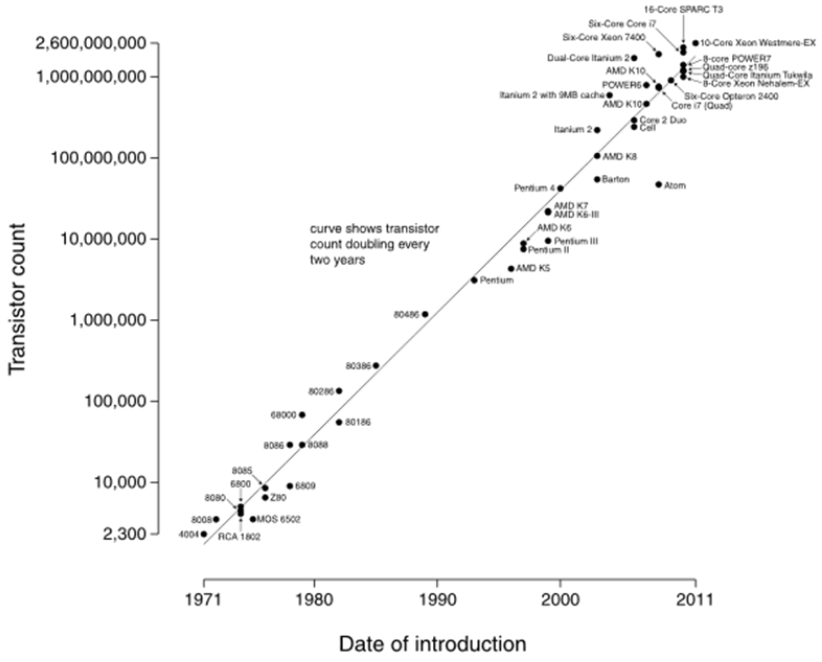
Big Data in Life Science
大規模計算の必要性

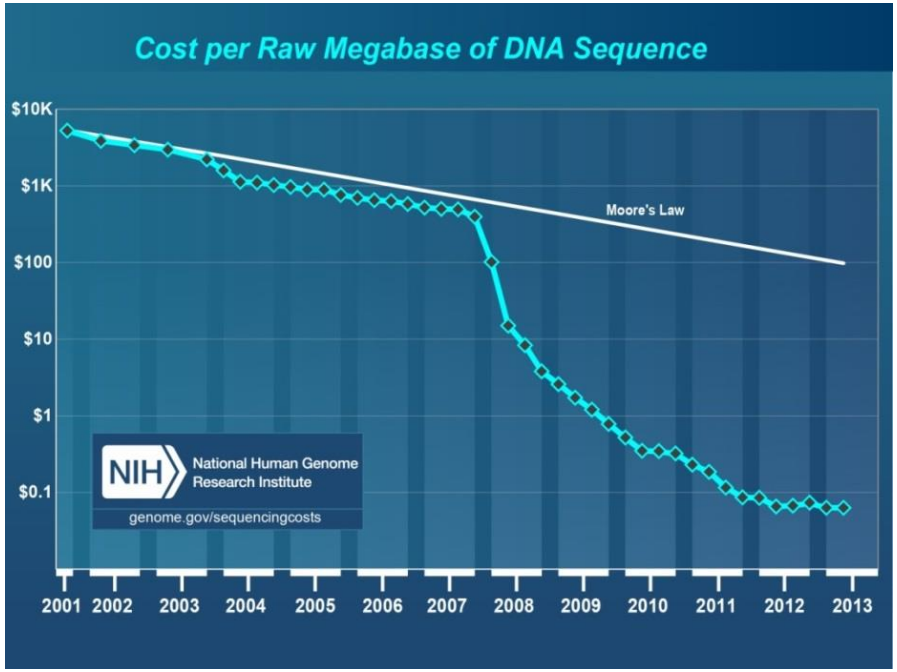


Gordon Moore

Moore's Law

HiSeq2500

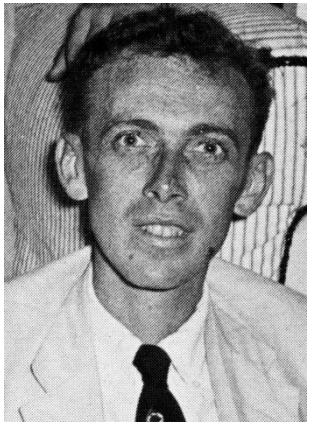




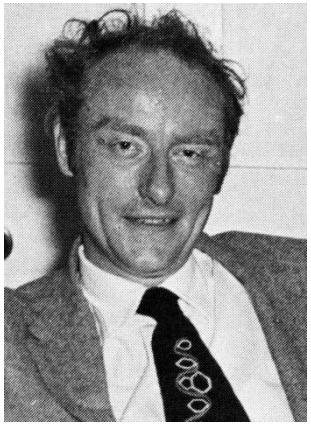
http://www.genome.gov/sequencingcosts/

枚挙の論理

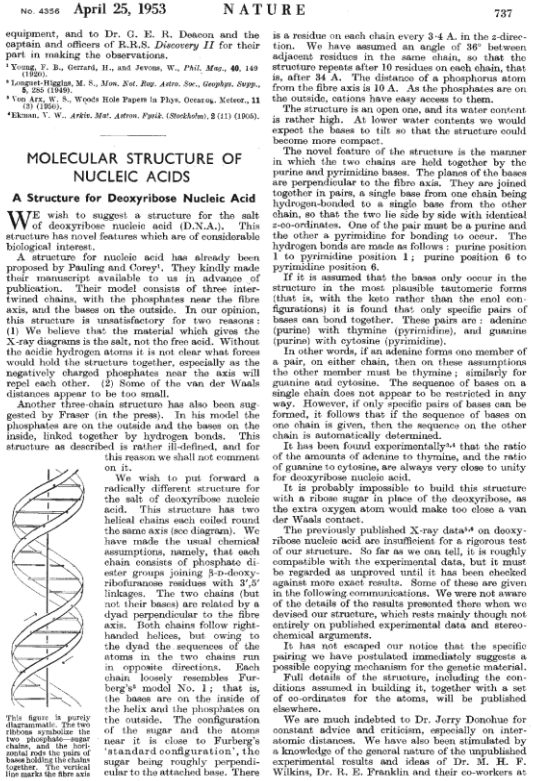
The Nobel Prize in Physiology or Medicine 1962



James Watson



Francis Crick



No. 4356 April 25, 1953 NATURE 737

equipment, and to Dr. G. E. R. Deacon and the captain and officers of R.R.S. *Discovery II* for their part in making the observations.

[1] Young, F. B., Gerrard, H., and Jevons, W., *Phil. Mag.*, **40**, 149 (1920).
[2] Longuet-Higgins, M. S., *Mon. Not. Roy. Astro. Soc., Geophys. Supp.*, **5**, 285 (1949).
[3] Von Arx, W. S., Woods Hole Papers in Phys. Oceanog. Meteor., **11** (3) (1950).
[4] Ekman, V. W., *Arkiv. Mat. Astron. Fysik.* (Stockholm), **2** (11) (1905).

MOLECULAR STRUCTURE OF NUCLEIC ACIDS

A Structure for Deoxyribose Nucleic Acid

We wish to suggest a structure for the salt of deoxyribose nucleic acid (D.N.A.). This structure has novel features which are of considerable biological interest.

A structure for nucleic acid has already been proposed by Pauling and Corey[1]. They kindly made their manuscript available to us in advance of publication. Their model consists of three inter-twined chains, with the phosphates near the fibre axis, and the bases on the outside. In our opinion, this structure is unsatisfactory for two reasons: (1) We believe that the material which gives the X-ray diagrams is the salt, not the free acid. Without the acidic hydrogen atoms it is not clear what forces would hold the structure together, especially as the negatively charged phosphates near the axis will repel each other. (2) Some of the van der Waals distances appear to be too small.

Another three-chain structure has also been suggested by Fraser (in the press). In his model the phosphates are on the outside and the bases on the inside, linked together by hydrogen bonds. This structure as described is rather ill-defined, and for this reason we shall not comment on it.

We wish to put forward a radically different structure for the salt of deoxyribose nucleic acid. This structure has two helical chains each coiled round the same axis (see diagram). We have made the usual chemical assumptions, namely, that each chain consists of phosphate di-ester groups joining β-D-deoxy-ribofuranose residues with 3',5' linkages. The two chains (but not their bases) are related by a dyad perpendicular to the fibre axis. Both chains follow right-handed helices, but owing to the dyad the sequences of the atoms in the two chains run in opposite directions. Each chain loosely resembles Furberg's[2] model No. 1; that is, the bases are on the inside of the helix and the phosphates on the outside. The configuration of the sugar and the atoms near it is close to Furberg's 'standard configuration', the sugar being roughly perpendicular to the attached base. There is a residue on each chain every 3·4 A. in the *z*-direction. We have assumed an angle of 36° between adjacent residues in the same chain, so that the structure repeats after 10 residues on each chain, that is, after 34 A. The distance of a phosphorus atom from the fibre axis is 10 A. As the phosphates are on the outside, cations have easy access to them.

The structure is an open one, and its water content is rather high. At lower water contents we would expect the bases to tilt so that the structure could become more compact.

The novel feature of the structure is the manner in which the two chains are held together by the purine and pyrimidine bases. The planes of the bases are perpendicular to the fibre axis. They are joined together in pairs, a single base from one chain being hydrogen-bonded to a single base from the other chain, so that the two lie side by side with identical *z*-co-ordinates. One of the pair must be a purine and the other a pyrimidine for bonding to occur. The hydrogen bonds are made as follows: purine position 1 to pyrimidine position 1; purine position 6 to pyrimidine position 6.

If it is assumed that the bases only occur in the structure in the most plausible tautomeric forms (that is, with the keto rather than the enol configurations) it is found that only specific pairs of bases can bond together. These pairs are: adenine (purine) with thymine (pyrimidine), and guanine (purine) with cytosine (pyrimidine).

In other words, if an adenine forms one member of a pair, on either chain, then on these assumptions the other member must be thymine; similarly for guanine and cytosine. The sequence of bases on a single chain does not appear to be restricted in any way. However, if only specific pairs of bases can be formed, it follows that if the sequence of bases on one chain is given, then the sequence on the other chain is automatically determined.

It has been found experimentally[3,4] that the ratio of the amounts of adenine to thymine, and the ratio of guanine to cytosine, are always very close to unity for deoxyribose nucleic acid.

It is probably impossible to build this structure with a ribose sugar in place of the deoxyribose, as the extra oxygen atom would make too close a van der Waals contact.

The previously published X-ray data[5,6] on deoxy-ribose nucleic acid are insufficient for a rigorous test of our structure. So far as we can tell, it is roughly compatible with the experimental data, but it must be regarded as unproved until it has been checked against more exact results. Some of these are given in the following communications. We were not aware of the details of the results presented there when we devised our structure, which rests mainly though not entirely on published experimental data and stereo-chemical arguments.

It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material.

Full details of the structure, including the conditions assumed in building it, together with a set of co-ordinates for the atoms, will be published elsewhere.

We are much indebted to Dr. Jerry Donohue for constant advice and criticism, especially on inter-atomic distances. We have also been stimulated by a knowledge of the general nature of the unpublished experimental results and ideas of Dr. M. H. F. Wilkins, Dr. R. E. Franklin and their co-workers at

This figure is purely diagrammatic. The two ribbons symbolize the two phosphate–sugar chains, and the horizontal rods the pairs of bases holding the chains together. The vertical line marks the fibre axis

"for their discoveries concerning the molecular structure of nucleic acids and its significance for information transfer in living material"

by an increase in protein content, while the amount of desoxyribonucleic acid remains unchanged.

Acknowledgments.—This work was supported by research grants from the University of California Board of Research. We are greatly indebted to Professor A. W. Pollister, Dept. of Zoology, Columbia University, for allowing the senior author use of his laboratory facilities to conduct the measurements described herein.

[1] Salvatore, C. A., *Biol. Bull.*, **99**, 112–119 (1950).
[2] Caspersson, T., *Skand. Arch. Physiol.*, **73**, Suppl. 8 (1936).
[3] Pollister, A. W., and Ris, H., *Cold Spring Harbor Symp. Quant. Biol.*, **12**, 147–157 (1947).
[4] Swift, H. H., *Physiol. Zool.*, **23**, 169–198 (1950).
[5] Swift, H. H., these PROCEEDINGS, **36**, 643–654 (1950).
[6] Ris, H., and Mirsky, A. E., *J. Gen. Physiol.*, **33**, 125–146 (1949).
[7] Leuchtenberger, C., Vendrely, R., and Vendrely, C., these PROCEEDINGS, **37**, 33–37 (1951).
[8] Alfert, M., *J. Cell. Comp. Physiol.*, **36**, 381–410 (1950).
[9] Schrader, F., and Leuchtenberger, C., *Exp. Cell Res.*, **1**, 421–452 (1950).
[10] Pollister, A. W., and Leuchtenberger, C., these PROCEEDINGS, **35**, 66–71 (1949).
[11] Leuchtenberger, C., *Chromosoma*, **3**, 449–473 (1950).
[12] Mirsky, A. E., and Ris, H., *Nature*, **163**, 666–667 (1949).

THE STRUCTURE OF PROTEINS: TWO HYDROGEN-BONDED HELICAL CONFIGURATIONS OF THE POLYPEPTIDE CHAIN

BY LINUS PAULING, ROBERT B. COREY, AND H. R. BRANSON*

GATES AND CRELLIN LABORATORIES OF CHEMISTRY,
CALIFORNIA INSTITUTE OF TECHNOLOGY, PASADENA, CALIFORNIA†

Communicated February 28, 1951

During the past fifteen years we have been attacking the problem of the structure of proteins in several ways. One of these ways is the complete and accurate determination of the crystal structure of amino acids, peptides, and other simple substances related to proteins, in order that information about interatomic distances, bond angles, and other configurational parameters might be obtained that would permit the reliable prediction of reasonable configurations for the polypeptide chain. We have now used this information to construct two reasonable hydrogen-bonded helical configurations for the polypeptide chain; we think that it is likely that these configurations constitute an important part of the structure of both fibrous and globular proteins, as well as of synthetic polypeptides. A letter announcing their discovery was published last year.[1]

The problem that we have set ourselves is that of finding all hydrogen-bonded structures for a single polypeptide chain, in which the residues are

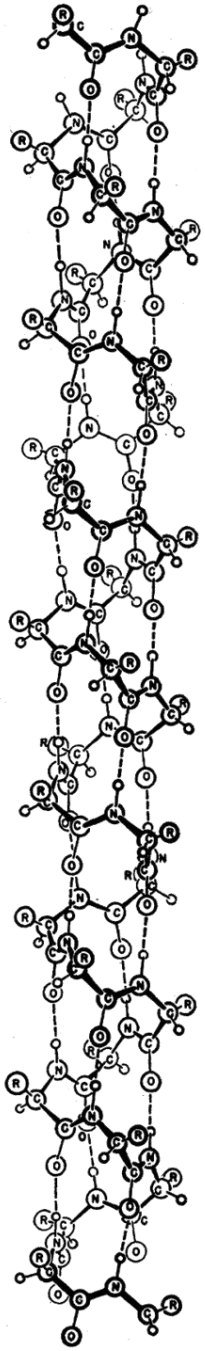
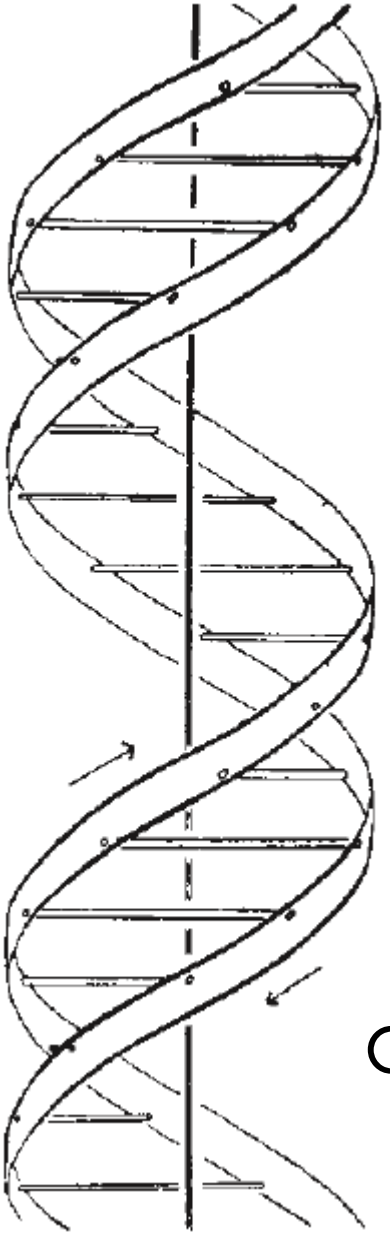


FIGURE 2
The helix with 3.7 residues per turn.

The Nobel Prize in Chemistry 1954

Linus Pauling Robert Corey

"for his research into the nature of the chemical bond and its application to the elucidation of the structure of complex substances"



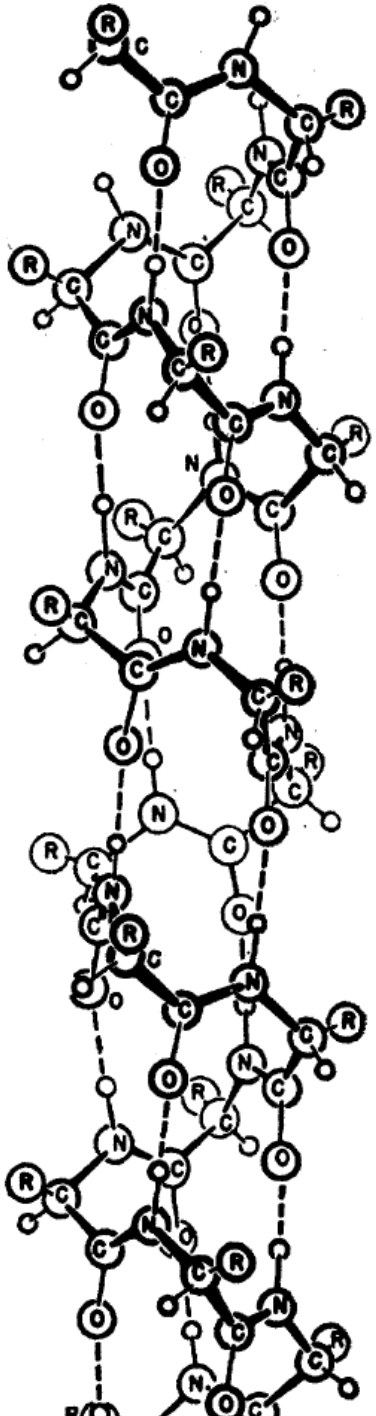
In 1950's

Principle of Life is

Symmetry

Periodicity

appeared in biomolecules



A THREE-DIMENSIONAL MODEL OF THE MYOGLOBIN MOLECULE OBTAINED BY X-RAY ANALYSIS

By Drs. J. C. KENDREW, G. BODO, H. M. DINTZIS, R. G. PARRISH and H. WYCKOFF
Medical Research Council Unit for Molecular Biology, Cavendish Laboratory, Cambridge
AND
D. C. PHILLIPS
Davy Faraday Laboratory, The Royal Institution, London

MYOGLOBIN is a typical globular protein, and is found in many animal cells. Like hæmoglobin, it combines reversibly with molecular oxygen; but whereas the role of hæmoglobin is to transport oxygen in the blood stream, that of myoglobin is to store it temporarily within the cells (a function particularly important in diving animals such as whales, seals and penguins, the dark red tissues of which contain large amounts of myoglobin, and which have been our principal sources of the protein). Both molecules include a non-protein moiety, consisting of an iron–porphyrin complex known as the hæm group, and it is this group which actually combines with oxygen; hæmoglobin, with a molecular weight of 67,000, contains four hæm groups, whereas myoglobin has only one. This, together with about 152 amino-acid residues, makes up a molecular weight of 17,000, so that myoglobin is one of the smaller proteins. Its small size was one of the main reasons for our choice of myoglobin as a subject for X-ray analysis.

In describing a protein it is now common to distinguish the primary, secondary and tertiary structures. The *primary structure* is simply the order, or sequence, of the amino-acid residues along the polypeptide chains. This was first determined by Sanger using chemical techniques for the protein insulin[1], and has since been elucidated for a number of peptides and, in part, for one or two other small proteins. The *secondary structure* is the type of folding, coiling or puckering adopted by the polypeptide chain: the α-helix and the pleated sheet are examples. Secondary structure has been assigned in broad outline to a number of fibrous proteins such as silk, keratin and collagen; but we are ignorant of the nature of the secondary structure of any globular protein. True, there is suggestive evidence, though as yet no proof, that α-helices occur in globular proteins, to an extent which is difficult to gauge quantitatively in any particular case. The *tertiary structure* is the way in which the folded or coiled polypeptide chains are disposed to form the protein molecule as a three-dimensional object, in space. The chemical and physical properties of a protein cannot be fully interpreted until all three levels of structure are understood, for these properties depend on the spatial relationships between the amino-acids, and these in turn depend on the tertiary and secondary structures as much as on the primary.

Only X-ray diffraction methods seem capable, even in principle, of unravelling the tertiary and secondary structures. But the great efforts which have been devoted to the study of proteins by X-rays, while achieving successes in clarifying the secondary (though not yet the tertiary) structures of fibrous proteins, have hitherto paid small dividends among the metabolically more important globular, or crystalline, proteins. Progress here has been slow because globular proteins are much more complicated then the organic molecules which are the normal objects of X-ray analysis (not counting hydrogens, myoglobin contains 1,200 atoms, whereas the most complicated molecule the structure of which has been completely determined by X-rays, vitamin B_{12}, contains 93). Until five years ago, no one knew how, in practice, the complete structure of a crystalline protein might be found by X-rays, and it was realized that the methods then in vogue among protein crystallographers could at best give the most sketchy indications about the structure of the molecule. This situation was transformed by the discovery, made by Perutz and his colleagues[2], that heavy atoms could be attached to protein molecules in specific sites and that the resulting complexes gave diffraction patterns sufficiently different from normal to enable a classical method of structure analysis, the so-called 'method of isomorphous replacement', to be used to determine the relative phases of the reflexions. This method can most easily be applied in two dimensions, giving a projection of the contents of the unit cell along one of its axes. Perutz attached a *p*-chloromercuri-benzoate molecule to each of two free sulphydryl groups in hæmoglobin and used the resulting changes in certain of the reflexions to prepare a projection along the *y*-axis of the unit cell[3]. Disappointingly, the projection was largely uninterpretable. This was because the thickness of the molecule along the axis of projection was 63 A. (corresponding to some 40 atomic diameters), so that the various features of the molecule were superposed in inextricable confusion, and even at the increased resolution of 2·7 A. it has proved impossible to disentangle them[4]. It was clear that further progress could only be made if the analysis were extended to three dimensions. As we shall see, this involves the collection of many more observations and the production of three or four different isomorphous replacements of the same unit cell, a requirement which presents great technical difficulties in most proteins.

The present article describes the application, at low resolution, of the isomorphous replacement method in three dimensions to type *A* crystals of sperm whale myoglobin[5]. The result is a three-dimensional Fourier, or electron-density, map of the unit cell, which for the first time reveals the general nature of the tertiary structure of a protein molecule.

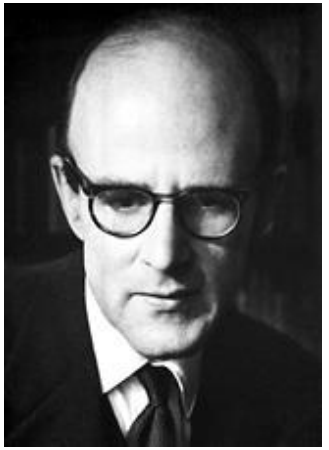
Isomorphous Replacement in Myoglobin

No type of myoglobin has yet been found to contain free sulphydryl groups, so that the method of

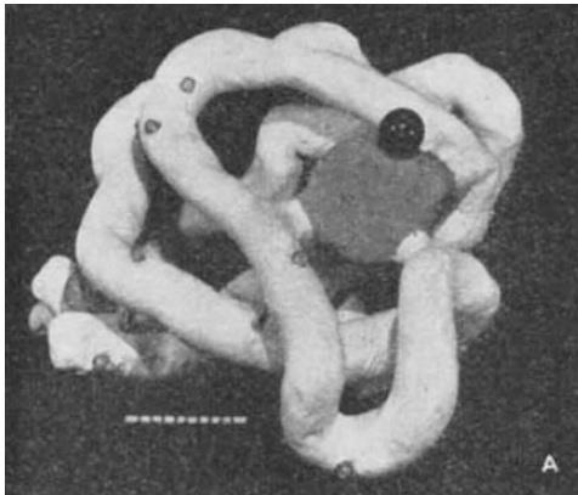
The Nobel Prize in Chemistry 1962

"for their studies of the structures of globular proteins"

John Kendrew



Max Perutz



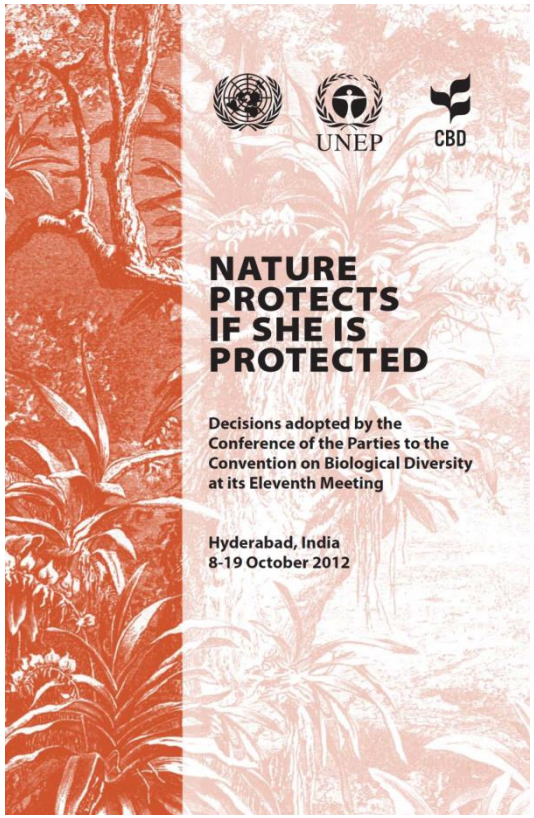
Perhaps the most remarkable features of the molecule are its complexity and its lack of symmetry. The arrangement seems to be almost totally lacking in the kind of regularities which one instinctively anticipates, and it is more complicated than has been predicated by any theory of protein structure. Though the detailed principles of construction do not yet emerge, we may hope that they will do so at a later stage of the analysis.

Symmetric
Periodic

Non-symmetric
Non-periodic

Uniform
Homogeneous

Diverse
Heterogeneous



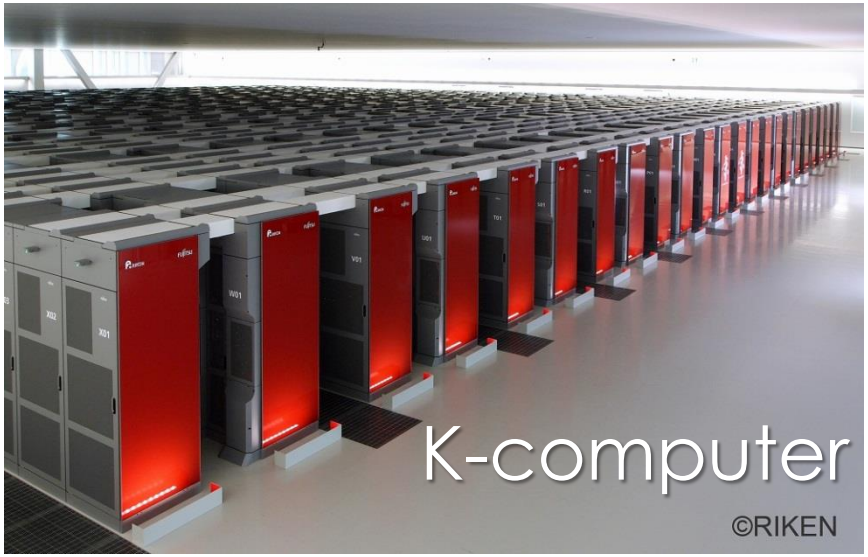
Biological Diversity

Conscious of the intrinsic value of biological diversity….
Conscious also of the importance of biological diversity for evolution and for maintaining life sustaining systems of the biosphere,

Full Diversity has to repeatedly become the target

Gene A, B,…	Homolog, Variant A, B,…, Normal/Disease
Protein A, B,…	Homolog, Variant A, B,…, Environment, Time, Localization…
Cell A, B,…	Environment, , Time, Localization, Normal/Disease
Human A, B,…	Patient A, B,…

HPC in Japan

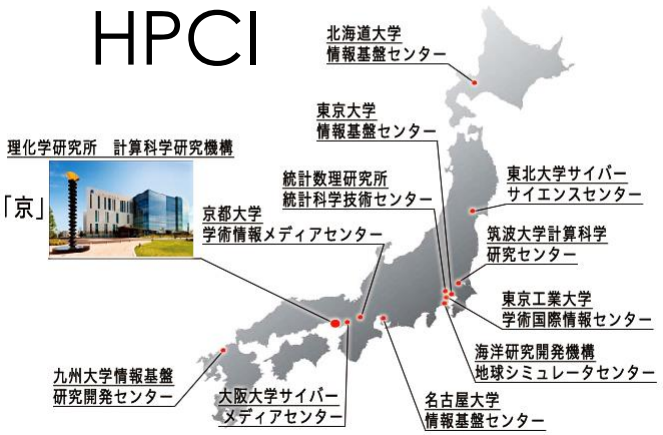


10 PFlops

▼ HPCIシステム

HPCIシステムを構成する計算資源を提供する機関（HPCI資源提供機関）

「京」と全国の大学や研究機関に設置されたスパコンを高速ネットワーク（SINET4）で結び、多様なユーザーニーズに応える革新的な共用計算環境を実現しています。



~3 PFlops

HPCI戦略プログラム 戦略分野1

「予測する生命科学・医療および創薬基盤」

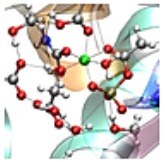
戦略機関
独立行政法人理化学研究所

SCLS
SUPERCOMPUTATIONAL LIFE SCIENCE

はじめに

柳田 敏雄
統括責任者

私たちは、HPCI戦略プログラム「分野１ 予測する生命科学・医療および創薬基盤」の戦略機関として、スーパーコンピュータ「京（けい）」を用いた研究開発及び計算科学技術推進体制の構築を実施しています。



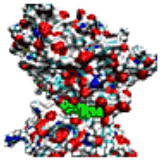
課題1

細胞内分子ダイナミクスのシミュレーション

代表：杉田 有治
理化学研究所

細胞質中の分子混雑、生体膜環境、膜を介した物質及び信号伝達など細胞環境を強く意識した分子および細胞スケールシミュレーションの実現を目指し、細胞内信号伝達経路の1分子粒度計算、膜タンパク質による物質輸送の解明、核内DNAタンパク質複合体の構造予測と機能解明を行う。

詳細ページへ



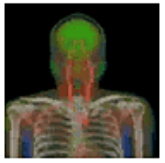
課題2

創薬応用シミュレーション

代表：藤谷 秀章
東京大学 先端科学技術研究センター

分子動力学を用いた生体高分子解析のために、「京」やHPCIの計算能力を活用するとともに、最新の計算アルゴリズムによる創薬プロセスの革新を目指し、革新的な薬の活性予測シミュレーションを行う。

詳細ページへ



課題3

予測医療に向けた階層統合シミュレーション

代表：高木 周
東京大学大学院 工学系研究科

これまで別々に開発が進められてきた各種生体シミュレータ(血栓症、心臓、筋骨格、脳神経系等)を統合し、心筋梗塞やパーキンソン病等、様々な疾患に対してより複雑なプロセスを再現する。そのために、基盤ツールを整備するとともに、「京」やHPCIを活用することで病態予測と治療支援を目指す。

詳細ページへ

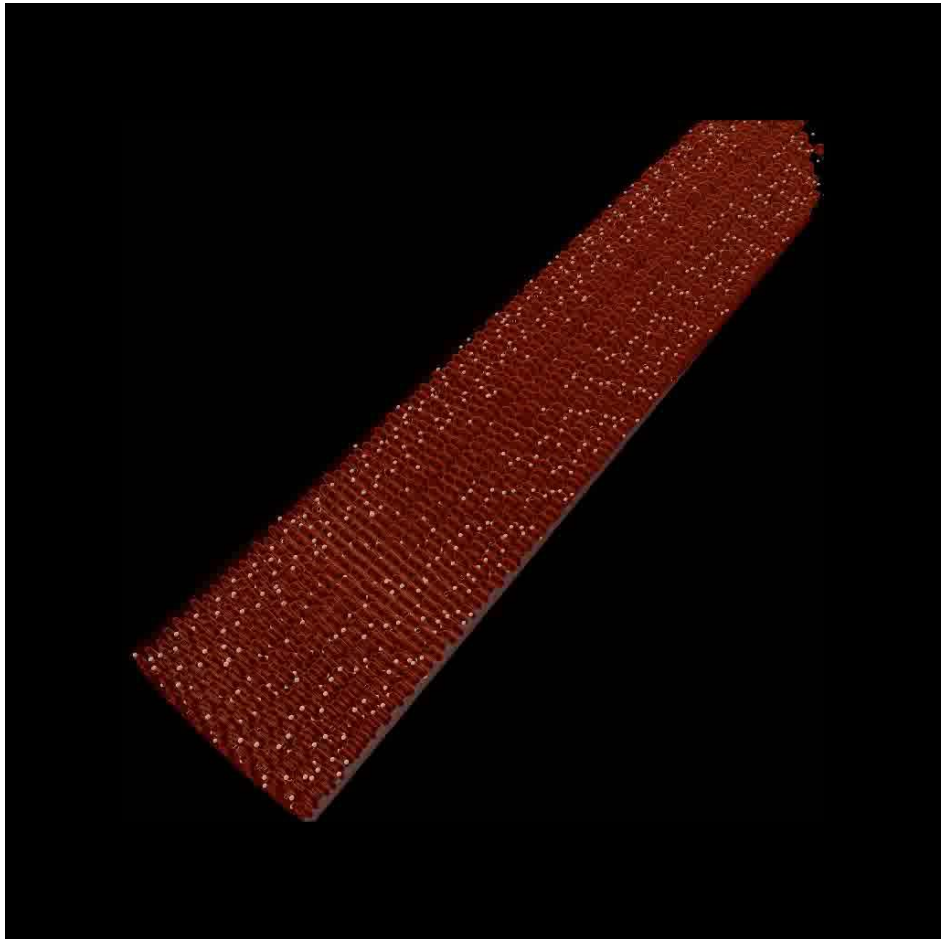
課題4

大規模生命データ解析

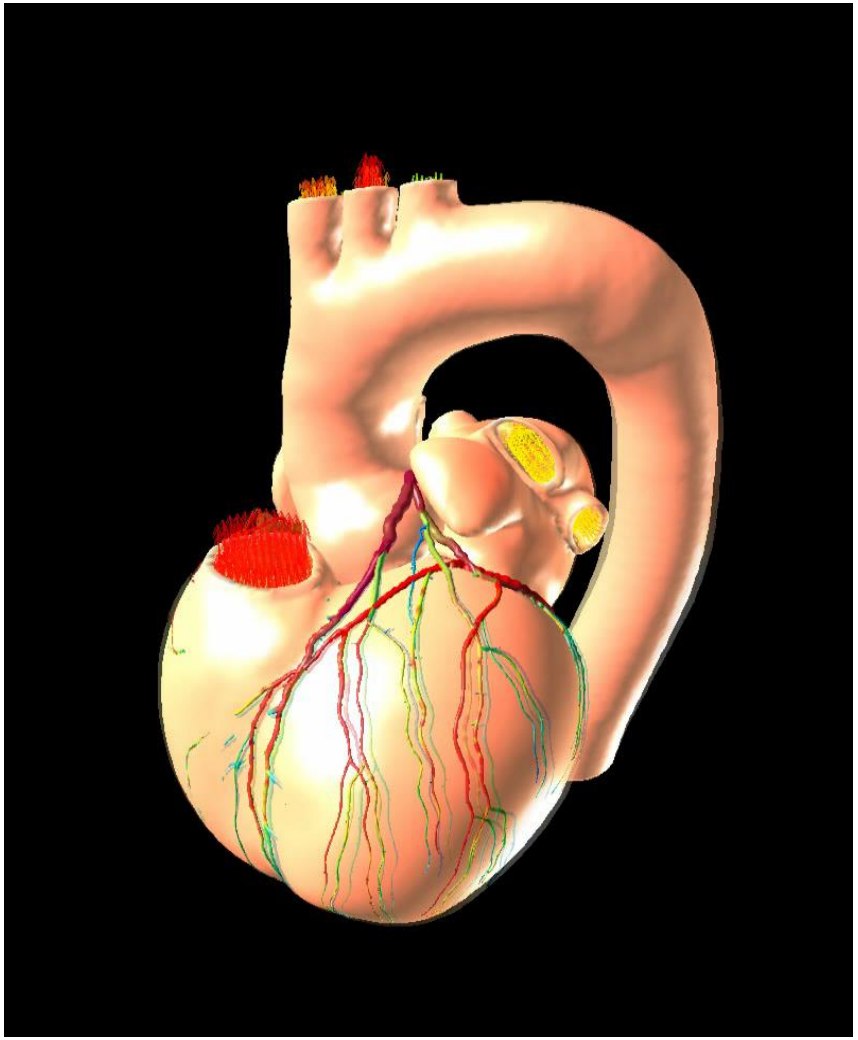
代表：宮野 悟
東京大学 医科学研究所

「京」やHPCIに最適化した最先端・大規模シークエンスデータ解析基盤を整備した上で、生命プログラムの複雑性・多様性や進化をゲノムによって理解する研究と同時に、ゲノムを基軸とした生体分子ネットワーク解析研究を行う。それにより、薬効・副作用予測、毒性の原因の推定、オーダーメイド投薬、予後予測などへの応用に貢献することを目指す。

詳細ページへ

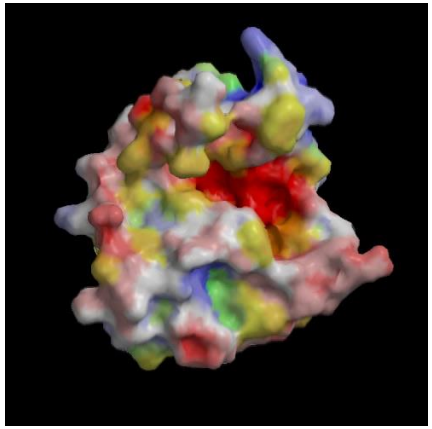


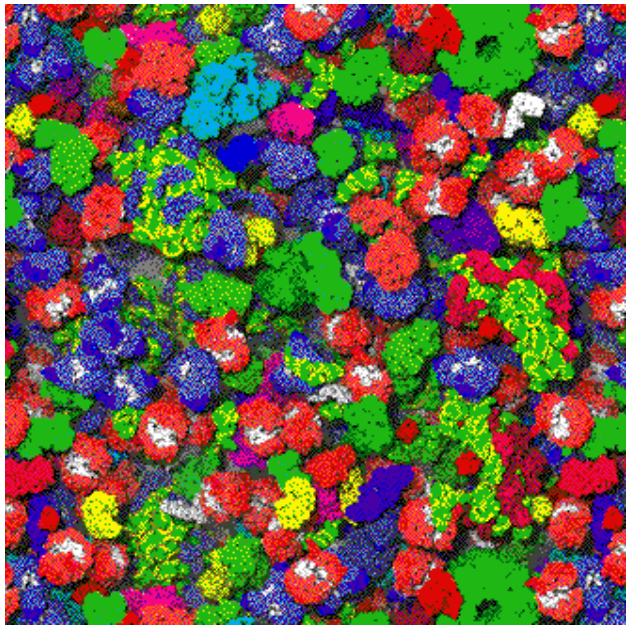
杉山和靖 (理研)
高木周（東大）



久田俊明（東大）

From Molecule to Cell



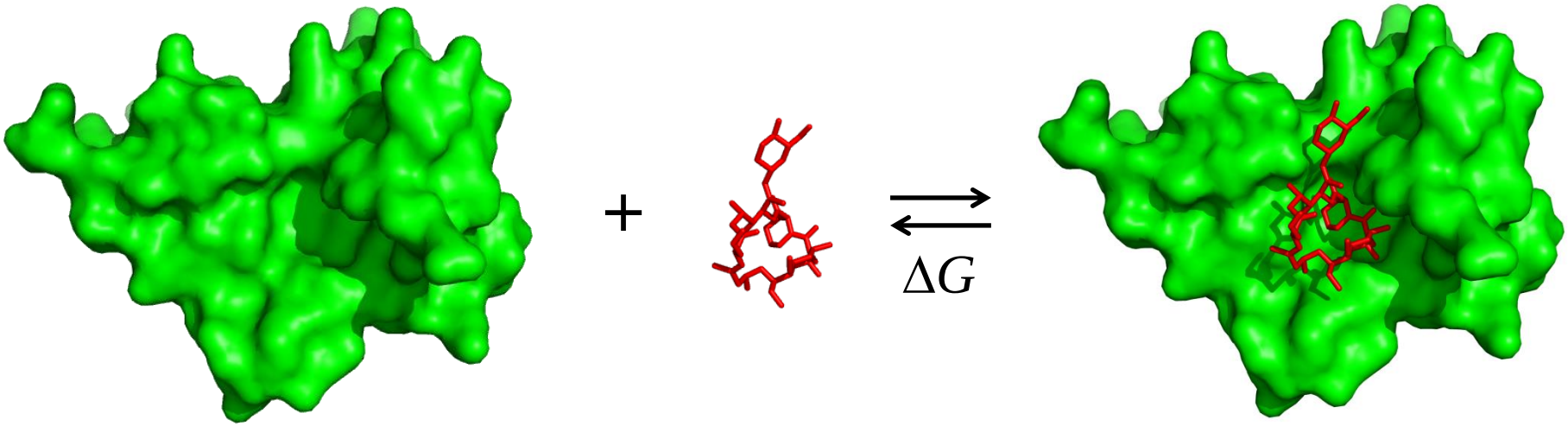


McGuffee SR, Elcock AH (2010)
PLoS Comput Biol 6(3): e1000694

細胞内混雑
1億原子系の分子動力学計算

Chromatinのマルチスケール
シミュレーション

Binding Free Energy



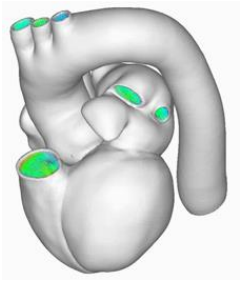
24年度自由エネルギー計算をした化合物数：300

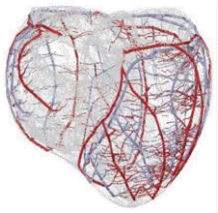
計算時間：　～19,000 時間/ノード(8コア) /化合物×300
= 5,700,000時間/ノード = ~660年/ノード

36,500化合物/365日/10PFLOPS

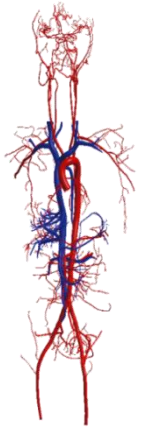
予測医療に向けた階層統合シミュレーション　高木 周（東大）

心臓
シミュレータ

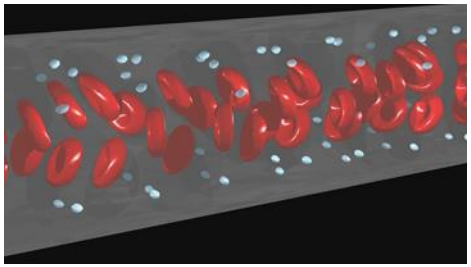


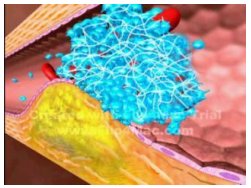


心筋梗塞

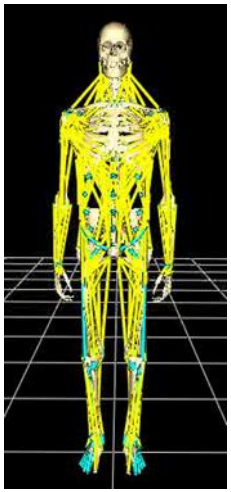


血管、血栓
シミュレータ

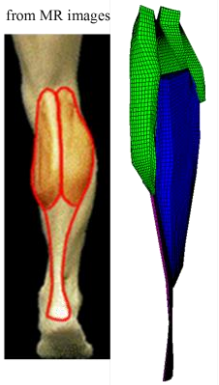


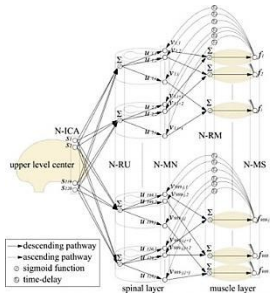


階層接続
Simulatorの統合

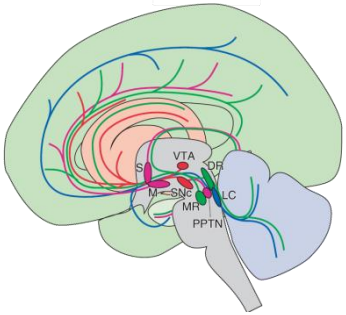


筋骨格系
シミュレータ

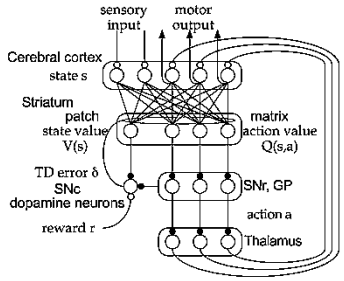




パーキンソン病



脳神経系
シミュレータ



大規模生命データ解析　宮野 悟（東大）

TCNG The Cancer Network Galaxy 0.13

Database of Cancer Gene Networks from Public Gene Expression Data

Home | About | Browse | Help | Download | Publications | Contact

search query [search]

TCNG Database Status

The Number of Gene Sets	512
The Number of Array Sets	256
The Number of Networks	512
The Number of Genes	22820
The Number of Nodes	24896
The Number of Edges	11610582

がん：薬剤感受性、転移、経時変化
脂肪細胞：褐色細胞

2013年10月25日
独立行政法人理化学研究所

柳田 敏雄 センター長（生命システム研究センター）が文化功労者に選出

生命システム研究センターの柳田 敏雄センター長（大阪大学 特任教授）が平成25年度文化功労者に選ばれました。

柳田敏雄センター長は、タンパク質を１分子レベルで観察可能な高性能顕微鏡を開発し、筋肉の駆動力を生み出す分子モーターの動作原理を解明するなど、生命システムを構成する分子機械に関する生物物理学研究で世界をリードしてきました。その卓越した見識で、生命システム研究センターをけん引しています。また、HPCI計算生命科学推進プログラムディレクターを兼任し、新しい計算生命科学の開拓にも当たっています。さらに、大阪大学大学院生命機能研究科の特任教授として、また、情報通信研究機構/大阪大学　脳情報通信融合研究センター長として分子から個体まで広く生命現象に関わる原理を追求し、基礎研究と科学技術の発展に尽力しています。

顕彰式は11月5日都内で行われる予定です。

柳田センター長のコメント

日本の文化とも言うべき、いい加減に、ほどよく、の考え方を持ち込み、恩師大澤文夫先生と共に生命の理解に挑戦してきました。欧米文化にはあまり見られないユニークなゆらぎの概念で、長年世界の研究者と論争してきたのですが、このように評価していただいてとても嬉しいです。