

データ駆動科学と計測の協奏

高田昌樹

東北大学 総長特別補佐
多元物質科学研究所 教授

**The world's most valuable resource is no longer oil,
but data**

The Economist 2017年5月6日

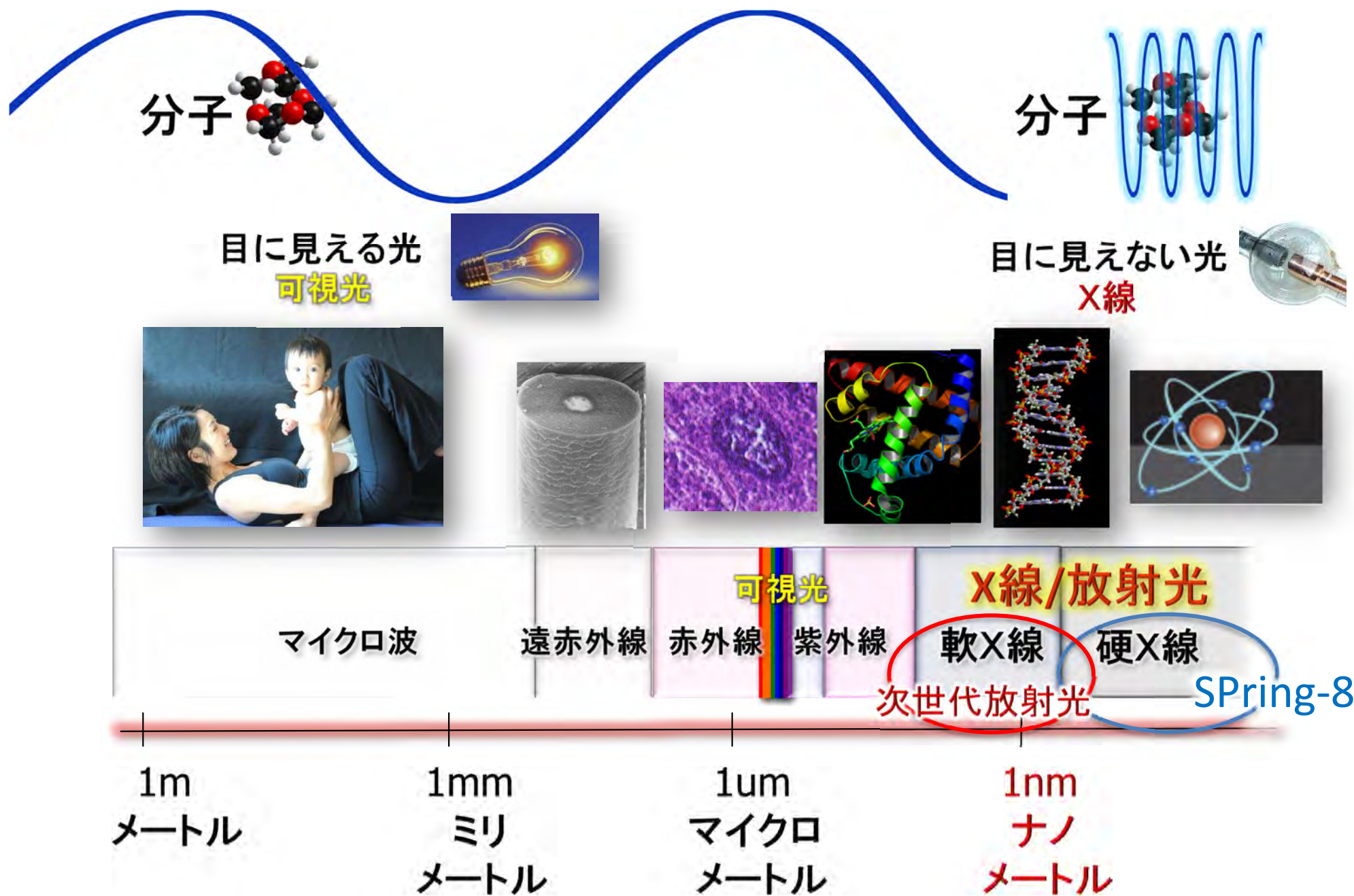
データ計測は、科学技術の基本

光は、計測するツール

本日お話しすること

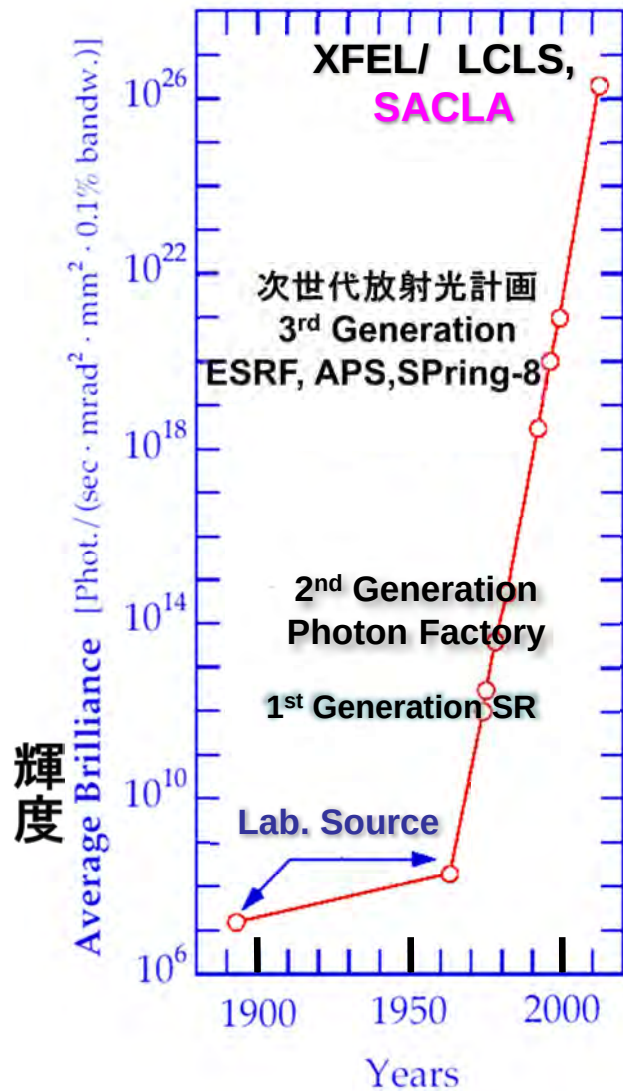
- 創薬、材料科学では、AI技術、ビッグデータ解析の応用により、材料の計測・分析データを用いて材料開発期間の短縮、低コスト化を目指す「データ駆動型開発」の取組が世界各国で加速
- 放射光(X線)、電子顕微鏡、中性子、NMR等を用いた計測は、物質構造の基本データを提供
- これまで、たんぱく質や物質の構造の最小単位となる結晶構造や電子状態のデータを、物質構造の改変による機能・物性の改良をAI・データ科学を駆使してシステマティックに予測するデータ駆動型開発のアプローチに貢献
- 次世代放射光の、光の性能の飛躍的な進歩が、明確な秩序性を持たない材料の階層性(マルチスケール)の可視化を実現
- データ同化などにより、データ駆動科学と計測の協奏に新たな展開が始まる

X線、放射光は、ナノの世界が見える光



放射光、X線レーザーの進化の歴史

X線の高輝度性の急速な発展



X線自由電子レーザー

コヒーレント光
SACLA



2011

100% コヒーレント

フェムト秒光パルス

シンクロtron放射光

カオス光

フォトンファクトリー

SPring-8



1997

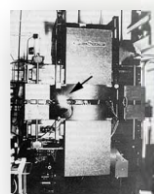
INS-SOR



1983



1974



1947

次世代放射光
2023?

より明るく

ナノへ

ピコ秒へ

高エネルギーへ



スマホの画面

放射光がささえる 科学技術



エコタイヤ



虫歯予防ガム



インフルエンザ治療薬



人工光合成

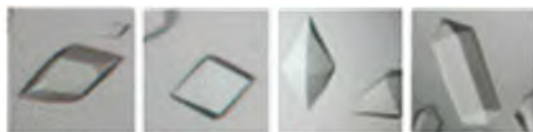
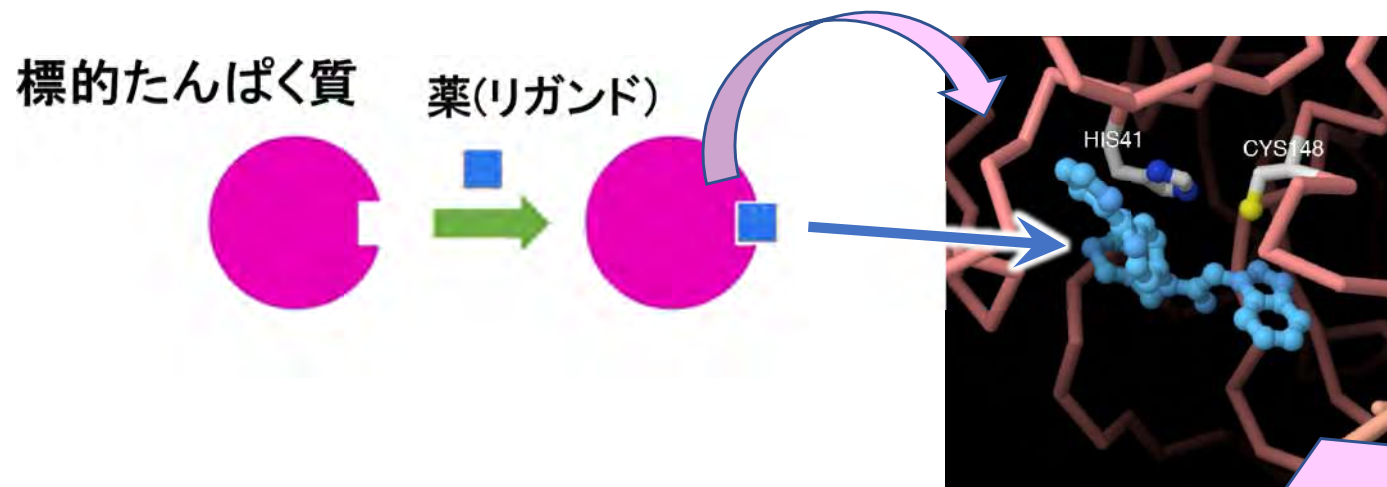


シャンプー

たんぱく質構造データベース Protein Data Bank (公的アーカイブ)



大阪大学蛋白質研究所
教授 栗栖源嗣



タンパク質の結晶

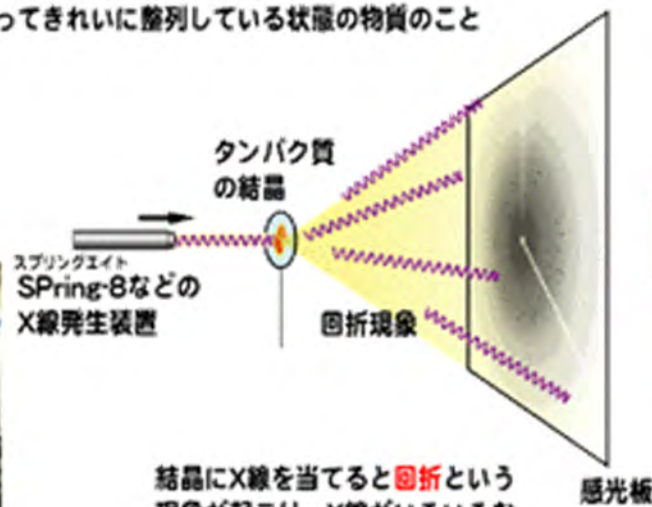
結晶とは、分子が一定の規則に従ってきれいに整列している状態の物質のこと

たんぱく質の構造決定の概略

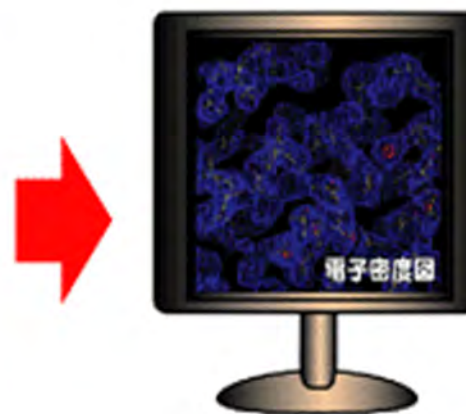
SPring-8 ホームページより



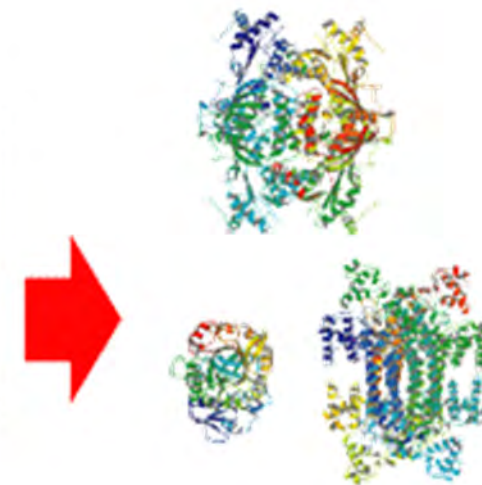
SPring-8のBL26B2ビームライン



結晶にX線を当てると回折という現象が起こり、X線がいろいろな方向に曲がる。その結果、背後の感光版に点の模様ができる。



コンピューターによって点の模様を解析する



タンパク質の立体構造

蛋白質構造データバンク (Protein Data Bank: PDB)

<https://pd bj.org>

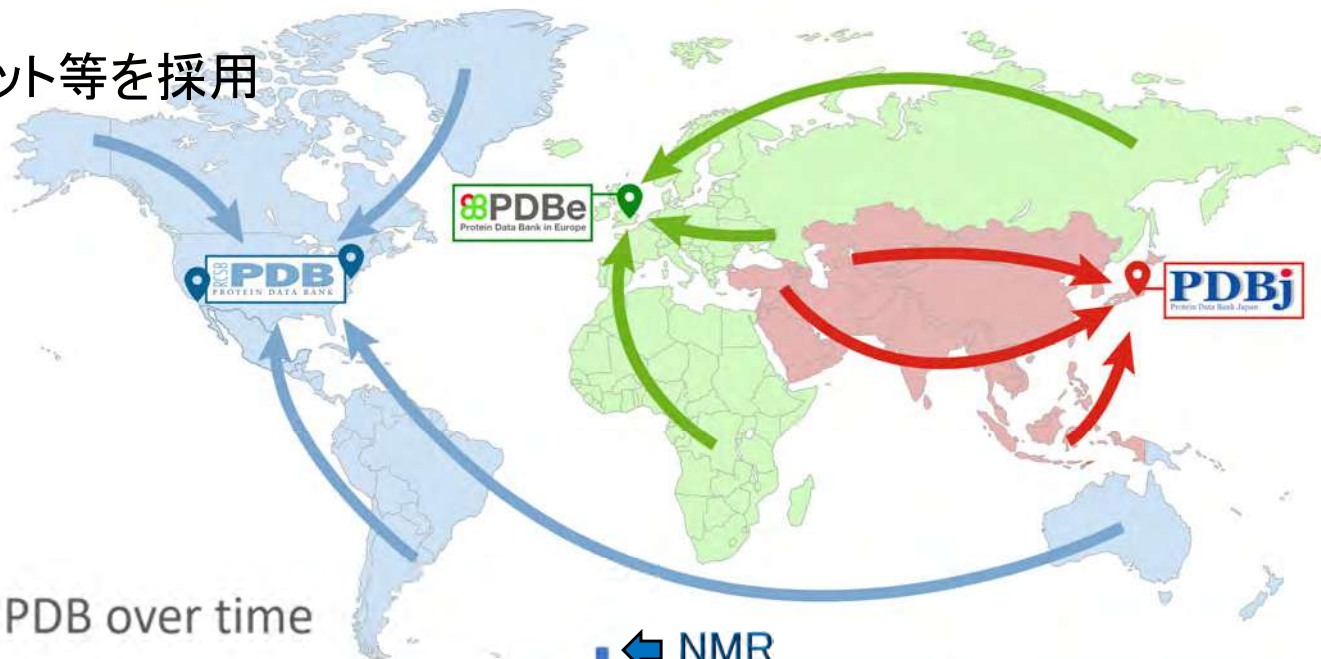
生体高分子の立体構造情報を集めた
世界で唯一のデータベース
情報は無償で利用可

1. 1971年: 米国で創設
2. 2000年: 日欧も個別に活動を開始
3. 2003年: 国際組織wwPDBを組織
共同で運営開始
日本の拠点 PDBjは創立メンバー
大阪大学蛋白質研究所

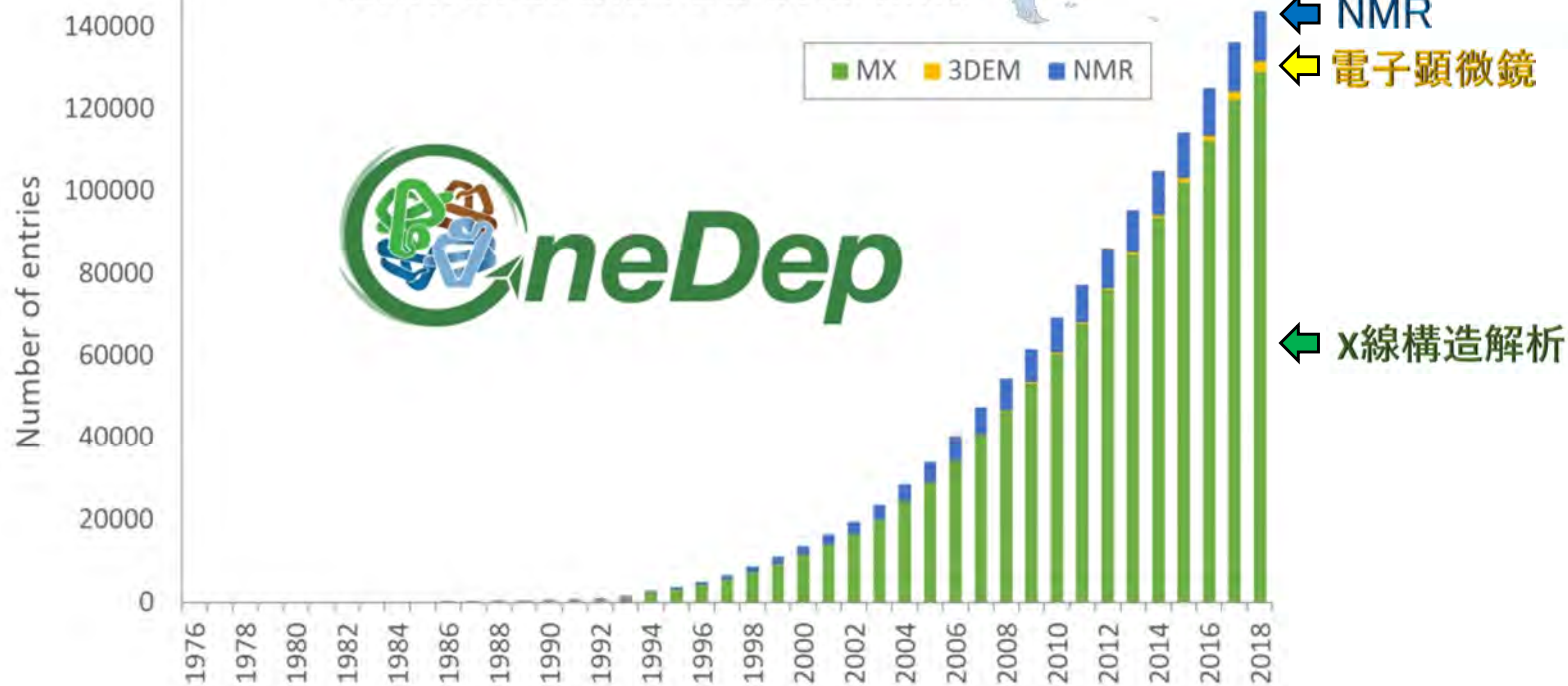


2016年からデータ処理システムを3極で共通化.

日本の提案した標準フォーマット等を採用
(PDB/RDF,BMRB/RDF)

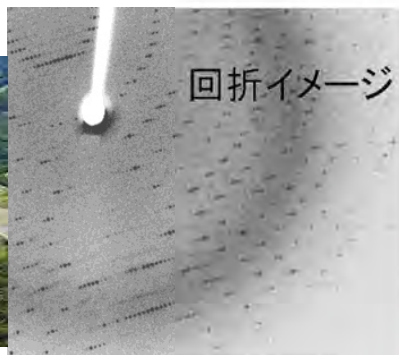


Growth of the PDB over time



計測技術の違いによるアーカイブデータの情報の違い

X線 放射光

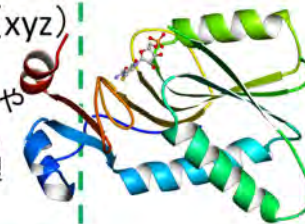


回折イメージ

→ 構造因子(hkl) → 座標データ(xyz)

配列やヘテロ化合物の情報, 実験条件や生物種の情報などmetadataを付与して, PDBアーカイブとしてwwPDBが一元管理

PDBj
Protein Data Bank Japan



NMR



Spectral データ

Chemical shifts
Coupling constants
Time-domain data
Spectral peak lists

Relaxation データ

R1/T1, R2/T2, R1rho/T1rho, NOEs



PDBj
Protein Data Bank Japan

→ 座標データ(xyz)

配列やヘテロ化合物の情報, 実験条件や生物種の情報などmetadataを付与して, PDBアーカイブへ

PDBj
Protein Data Bank Japan

電子顕微鏡



画像データ EMPIAR

欧州EBIがデータベースを運用中。ファイルサイズが大きいため、欧州のデータが中心。遠距離からの登録は長時間を要する。



→ 密度マップ → 座標データ



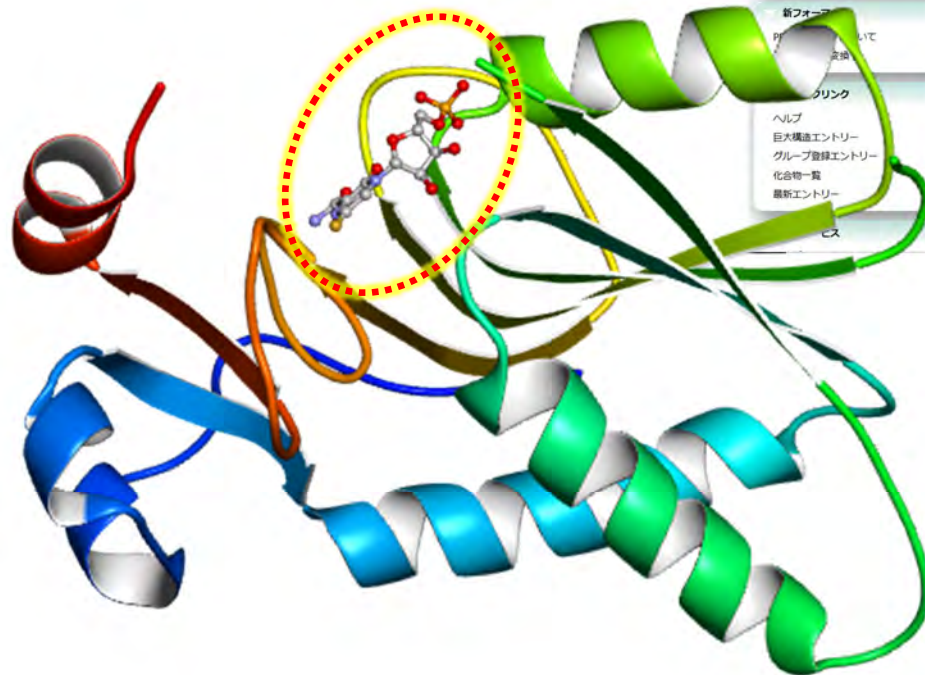
配列やヘテロ化合物の情報, 実験条件や生物種の情報などmetadataを付与して, PDBおよびEMDBアーカイブへ

PDBj
Protein Data Bank Japan

PDBj
Protein Data Bank Japan

<https://empiar.pdbj.org>

ヒトがもつ酵素とアビガンの複合体構造の登録



164174
登録日: 2020-05-20
00:00 UTC / 09:00 JST

PDBj
Protein Data Bank Japan

English 日本語 简体中文 繁體中文 한국어

検索: pdbj.org 全体を検索 (日本語on)

ホーム
トップページ
統計情報
ヘルプ
FAQ
お問い合わせ
PDBjの利用・利用規約
リンク集

データ登録 (OneDep)
ヘルプ
PDB, EMDB, BMRBへの登録

ダウンロード
PDBアーカイブからのデータダウンロード

新フォーマット
PDBx/mmCIF
PDBx/mmCIF
PDBx/mmCIF
ヘルプ
巨大構造エントリ
グループ登録エントリ
化合物一覧
最新エントリ

4KN6

Crystal structure of human hypoxanthine-guanine phosphoribosyltransferase in complex with 6-fluoro-3-hydroxy-2-pyrazinecarboxamide (T-705) ribose-5'-monophosphate

4KN6 の概要

分子名称	Hypoxanthine-guanine phosphoribosyltransferase, 6-fluoro-3-oxo-4-(5-O-phosphono-beta-D-ribofuranosyl)-3,4-dihydropyrazine-2-carboxamide (3 entities in total)
機能のキーワード	favipiravir, 6-oxopurine phosphoribosyltransferase, transferase
由来する生物種	Homo sapiens (human)
細胞内の位置	Cytoplasm P00492
ポリマー-鎖数	1
分子量合計	24779.34
構造登録者	Naesens, L., Guddat, L., Keough, D., van Kullenburg, A.B., Meijer, J., Vande Vooorde, J., Balzarini, J. (登録日: 2013-05-08, 公開日: 2013-08-14, 最終更新日: 2013-09-25)
主引用文献	Naesens, L., Guddat, L.W., Keough, D.T., van Kullenburg, A.B., Meijer, J., Vande Vooorde, J., Balzarini, J. Role of human hypoxanthine-guanine phosphoribosyltransferase in activation of the antiviral agent T-705 (favipiravir). <i>Mol. Pharmacol.</i> , 84:615-629, 2013. PubMed: 23907213 (主引用文献が同じPDBエントリ) DOI: 10.1124/mol.113.087247 Import into Mendeley
実験手法	X-RAY DIFFRACTION (2.728 Å)

ダウンロード
Sequence (fasta)
PDBx/mmCIF
PDBML (ヘッダのみ (no-atom))
PDB形式 (全ての情報)
検証レポート (PDF)
More...

構造
非対称単位を表示

ヘルプ
Molmolで表示

PDBが創薬を加速している

2010-2016年の期間に米国FDAで承認された新薬210件のうち、PDBに登録された創薬標的タンパク質をターゲットとしているものは88%の184件である。
その標的タンパク質との複合体構造がPDBに登録されている件数は5,914件に上る。

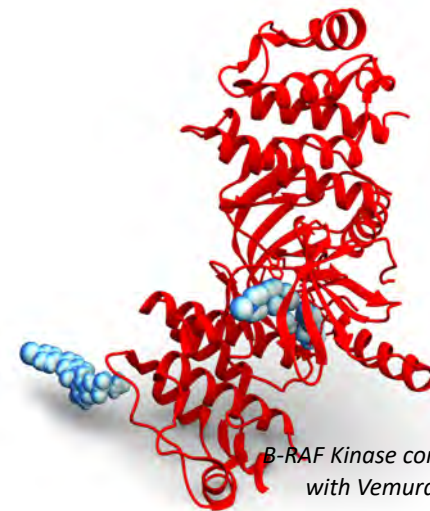
世界で1日あたり180万件ものアクセス数がある！

210 NEW DRUGS
approved
2010-2016

>\$100 BILLION

2000-2016

of NIH funding
contributed to these approvals
(20% of NIH Budget)²



B-Raf Kinase complex
with Vemurafenib
PDB ID 3og7
Bollag *et al.* (2010)
Nature 467, 596-599

5,914 PDB Structures
contributed to **184** of the
drug

1. Westbrook & Burley (2019) *Structure* 27, 211-217
2. Galkina Cleary *et al.* (2018) *PNAS* 115, 2329-2334; Value in 2016 US\$

エコタイヤの開発

放射光と計算科学の協創



住友ゴム工業株式会社
岸本浩通



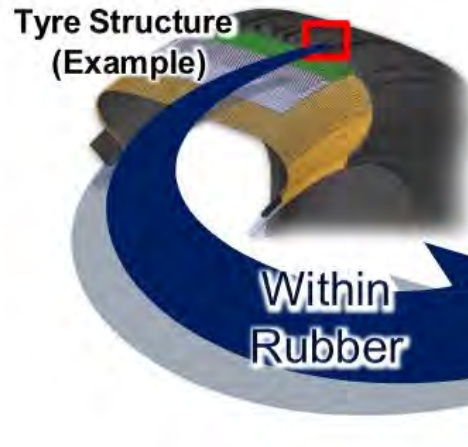
住友ゴム工業株式会社
中瀬古広三郎



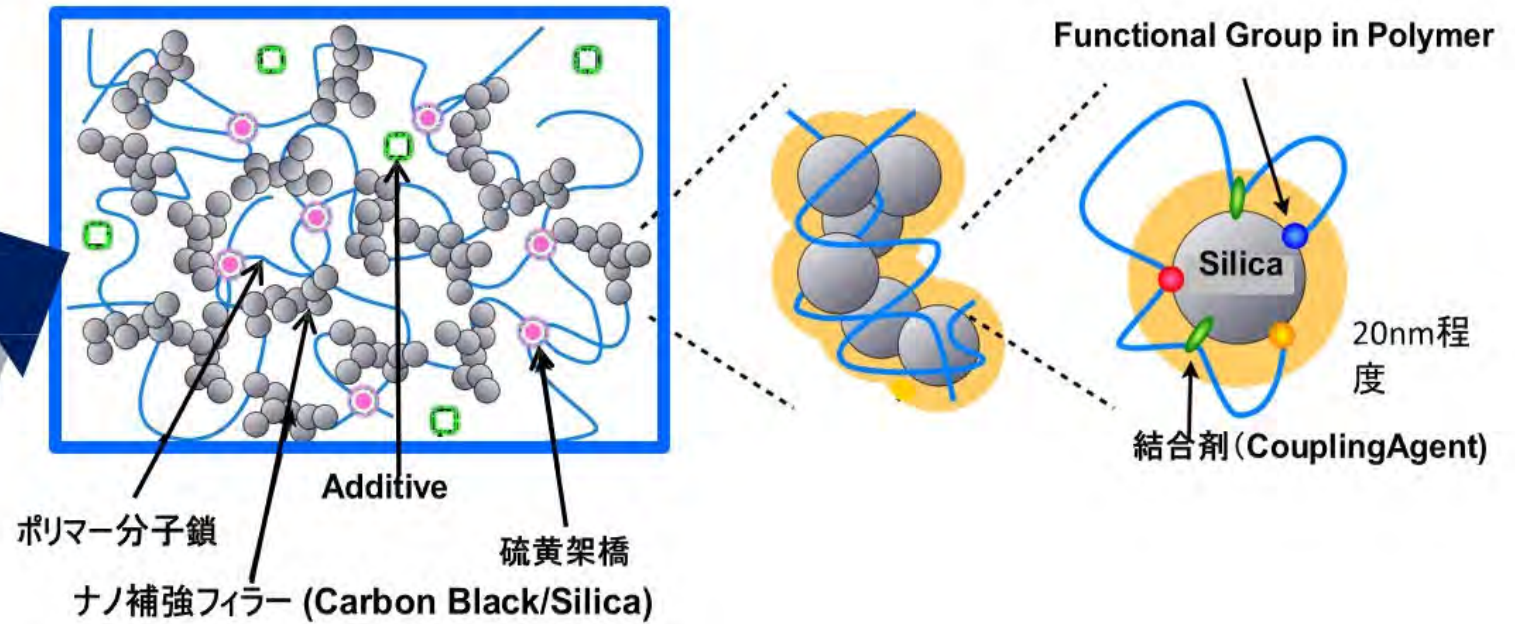
東京大学
雨宮慶幸

タイヤ用ゴムの構成と構造

様々な素材が空間スケールで多層構造的に製造され様々な性能を支配

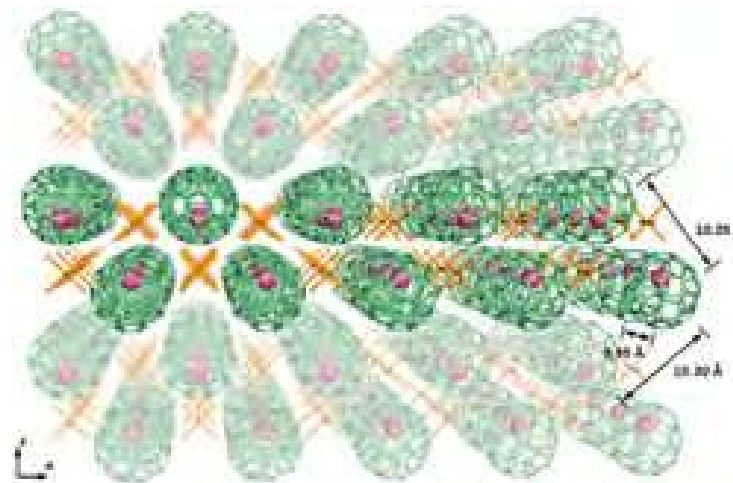


ゴム: 秩序の低い構造

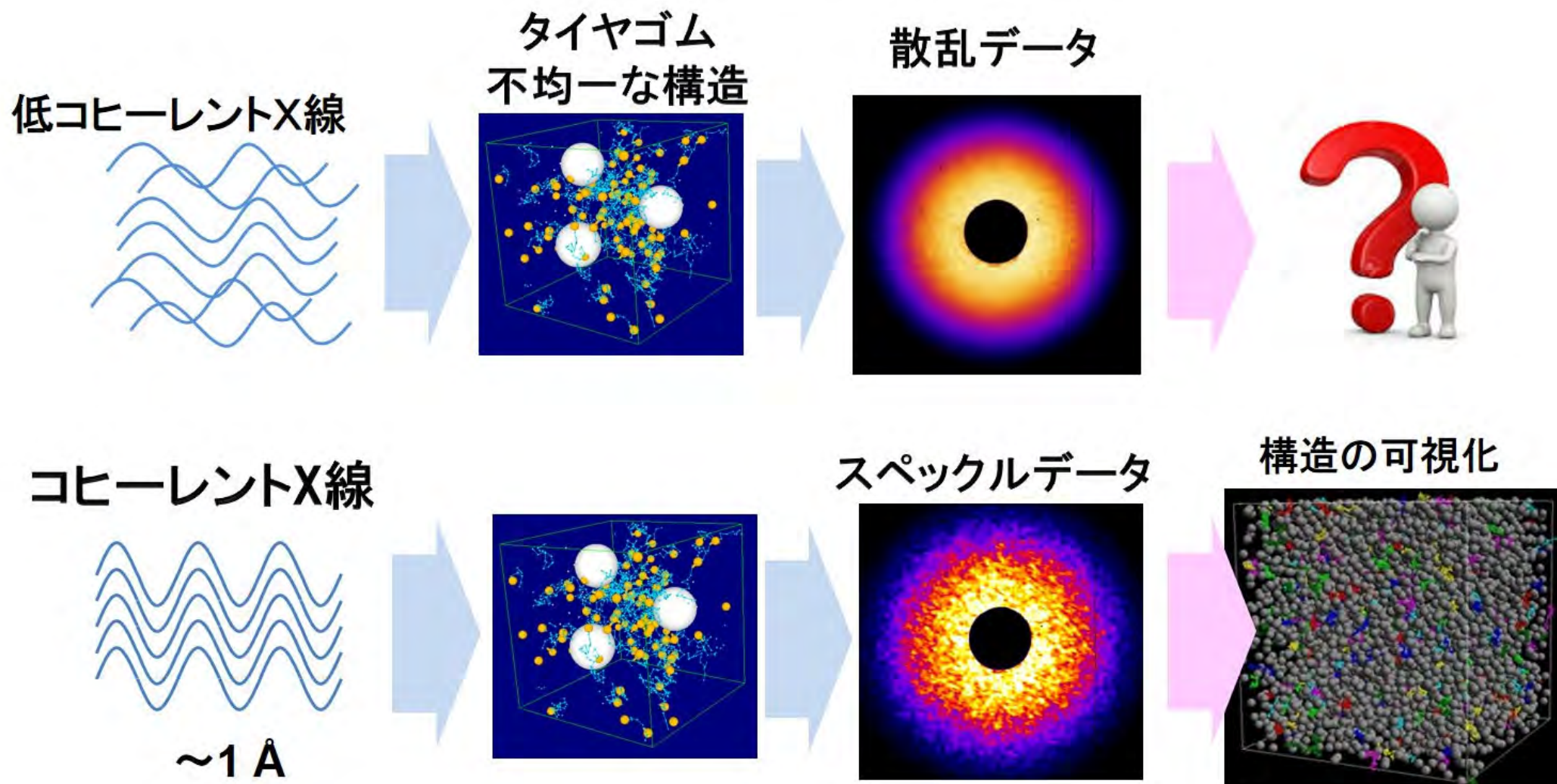


結晶の秩序構造

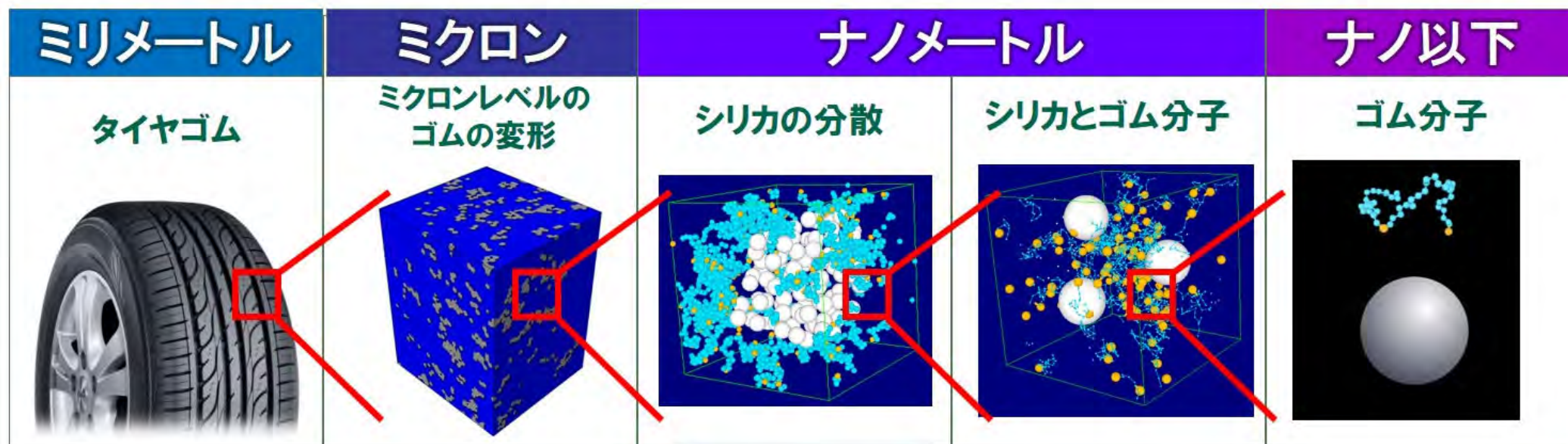
規則正しく分子が並ぶ



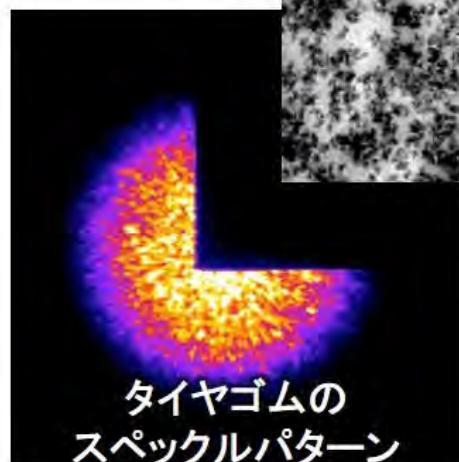
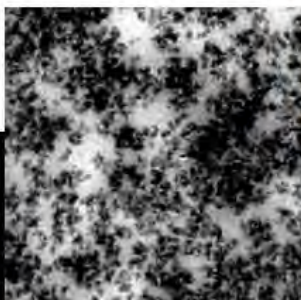
放射光のコヒーレント成分が 不均一な構造やその動きを可視化できるようになる



住友ゴム：SPring-8と地球シミュレータによるエナセーブ開発 A Tailor-made Molecular Design of Energy Saving Tire



ゴム中の
カーボンブラック



ゴム分子、シリカ、カーボンの新しい立体構造

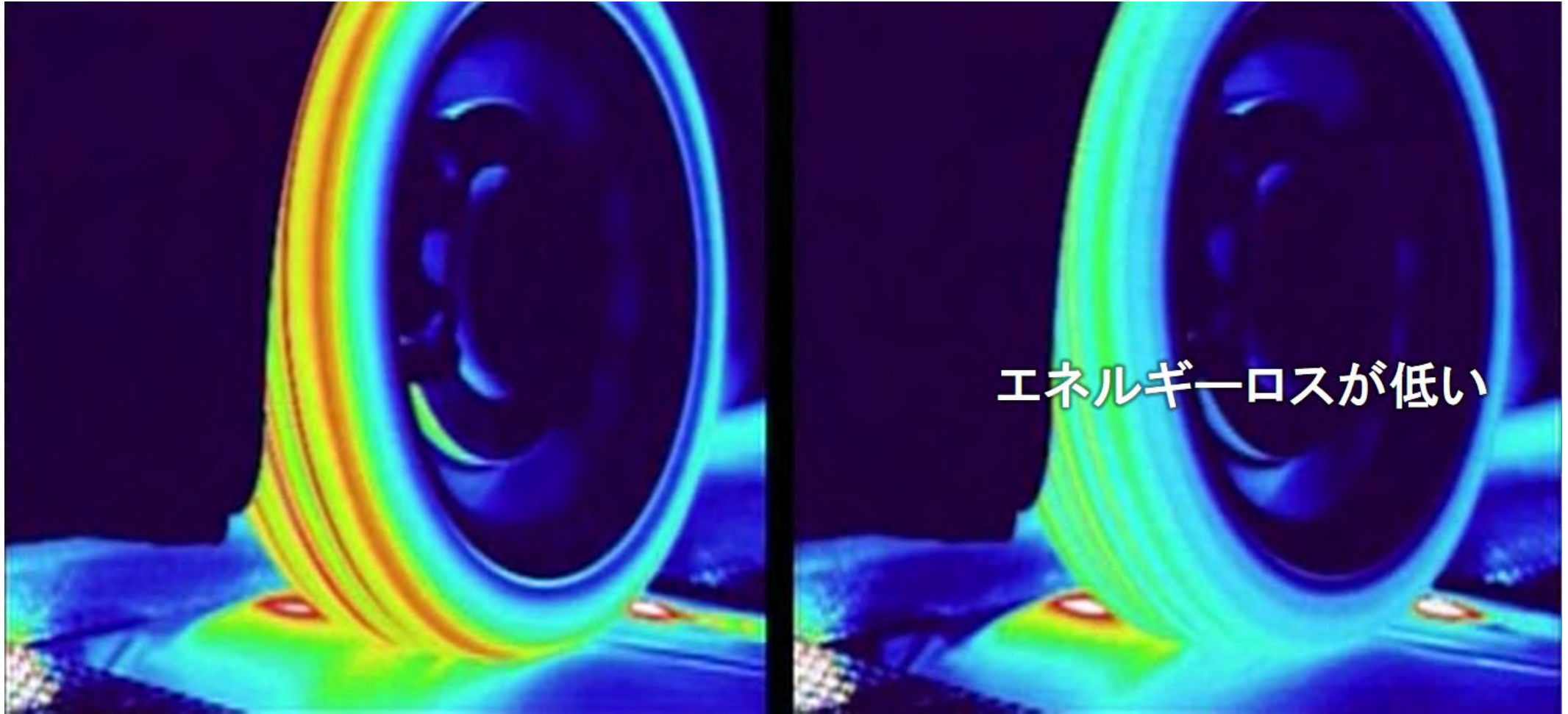
試行錯誤型開発から、テーラーメイド型開発へ



エナセーブの低燃費性

従来品

エナセーブ
PREMIUM



4Dナノデザイン

低燃費タイヤの分子デザイン



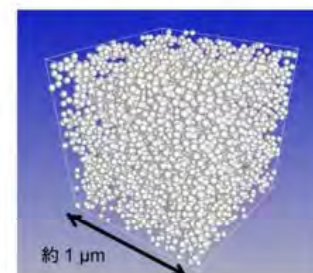
SUMITOMO RUBBER INDUSTRIES, LTD.

計算科学シミュレーション

放射光計測



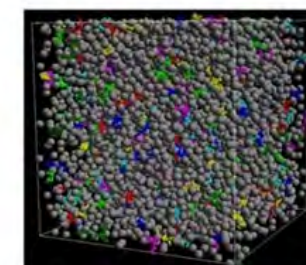
提供: (独) 海洋研究開発機構



約 1 μ m



提供: 理化学研究所



シミュレーションモデル図
※1
(灰色: シリカ粒子, カラー: ポリマー分子を一部表示)

6% 燃費向上を達成
39% 転がり抵抗を低減

有料会員限定 記事 今月の閲覧本数: 10 本中 3 本

体外式人工肺の血栓、発生原因を解明 東大教授ら

新型コ口才 医療・健康 科学&新技術

2020/5/10 17:00

保存 共有 印刷

東京大学の原田慈久教授らは、新型コロナウイルスの重症患者の治療に使われる体外式膜型人工肺（ECMO＝エクモ）の弱点であるチューブ内部での血栓の発生原因を突き止めた。血栓ができにくくなるコーティング剤の開発に取り組み、チューブに塗るだけで使えるようにする。1年後にチューブの臨床試験の開始を目指す。

エクモは新型コロナで重篤な呼吸困難に陥った患者の肺の代わりとして働く医療機器。治療率のない重症患者にとっては「最後のとりで」とされるが、長時間使い続けるとチューブ内で血が固まり、血栓ができやすい。血栓が体内に流れると心臓や脳の血管が詰まる恐れがある。



④ 画像の拡大

重症患者の呼吸機能を維持するのに利用する体外式膜型人工肺（ECMO=エクモ）

原田教授は九州大学の田中賢教授と共同で、大型放射光施設「SPring-8」を使い、血栓の原因となるたんぱく質と、チューブ内側のポリマー材料を解析。ポリマー材料の表面で壁の役割を果たす水分子が時間の経過とともにはがれてなくなると、たんぱく質がポリマー材料に吸着して固まり、血栓になることが分かった。

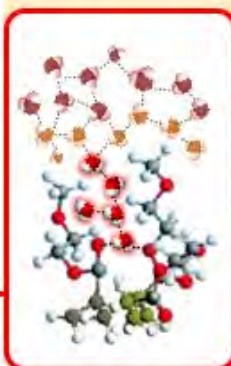
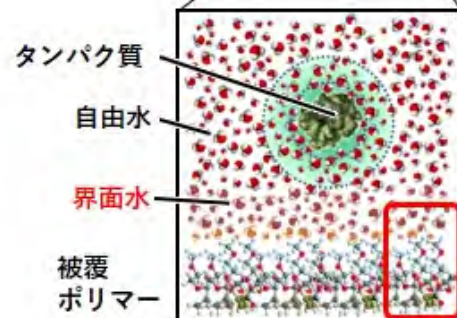
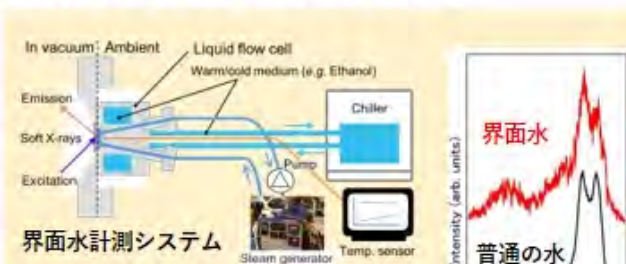
解析結果を基に水分子がなくなるいないコーティング剤を開発する。チューブに表面処理を施せば血栓の発生を抑えられるとみる。通常は数日以上たつと血栓ができやすくなるが、表面処理したチューブを使ったエクモでは血栓を発生させずに2~3週間の連続使用を可能にしたい考えだ。

集中治療室（ICU）でエクモを扱うには複数の医療スタッフが必要となる。心臓や脳を傷つける血栓の有無を医師や看護師らが数時間おきに確認しているためだ。確認作業の手間を減らせれば医療スタッフの業務に余裕が生まれ、医療の効率化につながるとの期待もある。

次世代のX線分光は界面で機能する水を見分けることができる
血栓形成を阻害する界面水をポリマーからデザインする



原田 慈久
(東大/東北大)



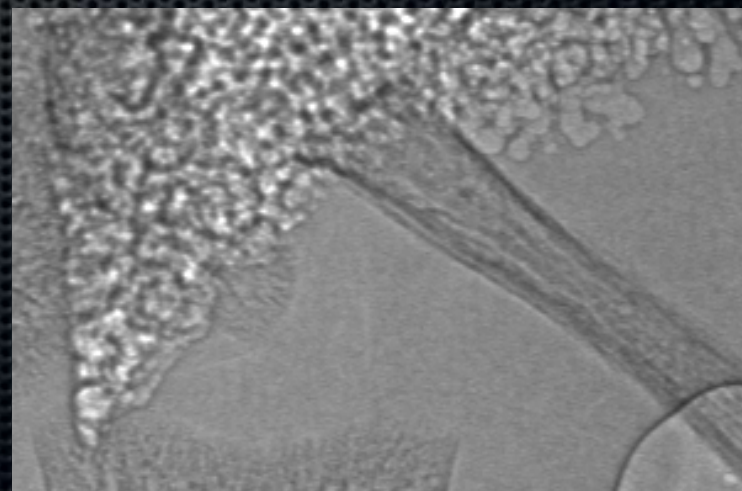
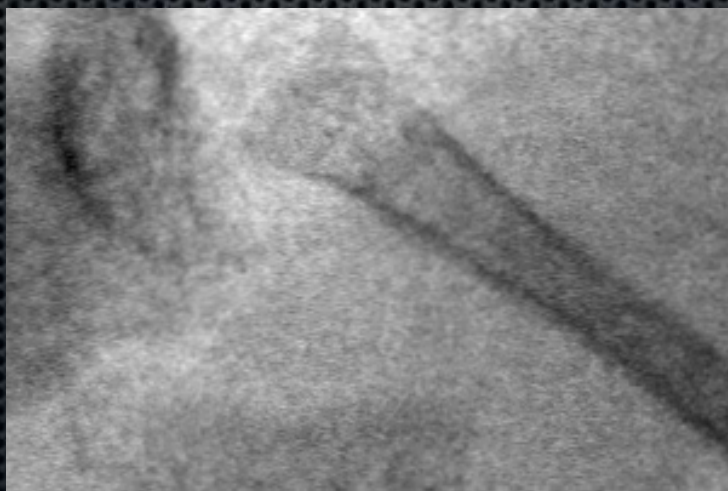
高効率・高分解能X線分光で
ミクロな水環境が可視化される

界面水
(タンパク質吸着を防ぐ水)
不凍水
(界面水を安定化させる水)



Rob Lewis教授
モナシュ大学
(オーストラリア)

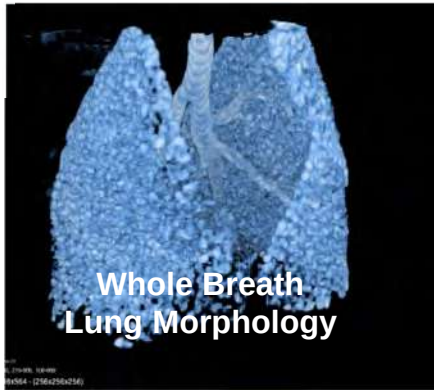
新生児の呼吸開始のメカニズムを SPring-8の先端イメージングで ウサギの新生児の動的観察により可視化



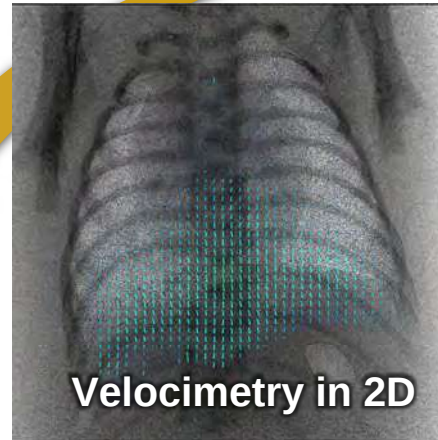


Stuart B
Hooper

放射光による観測から計算機シミュレーションへ



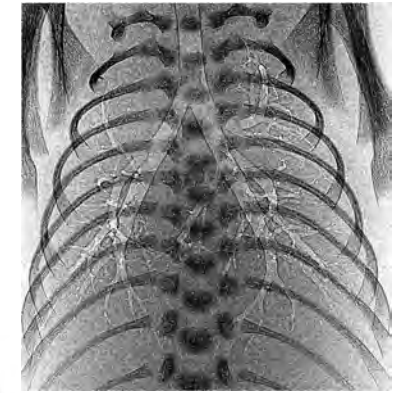
Whole Breath
Lung Morphology



Velocimetry in 2D



人工呼吸 + PEEP



人工呼吸

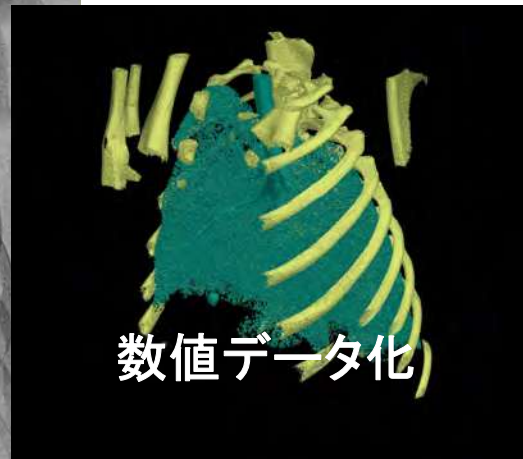


データ同化？

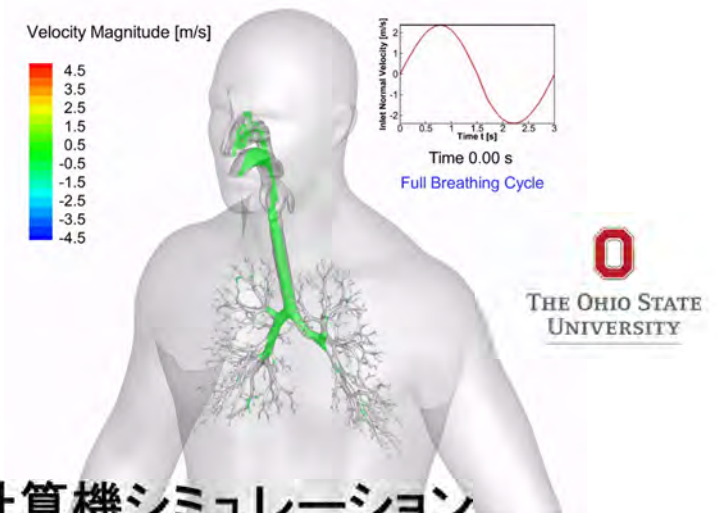
MONASH University
Centre for Synchrotron Science



放射光による観察
ウサギの肺呼吸の開始



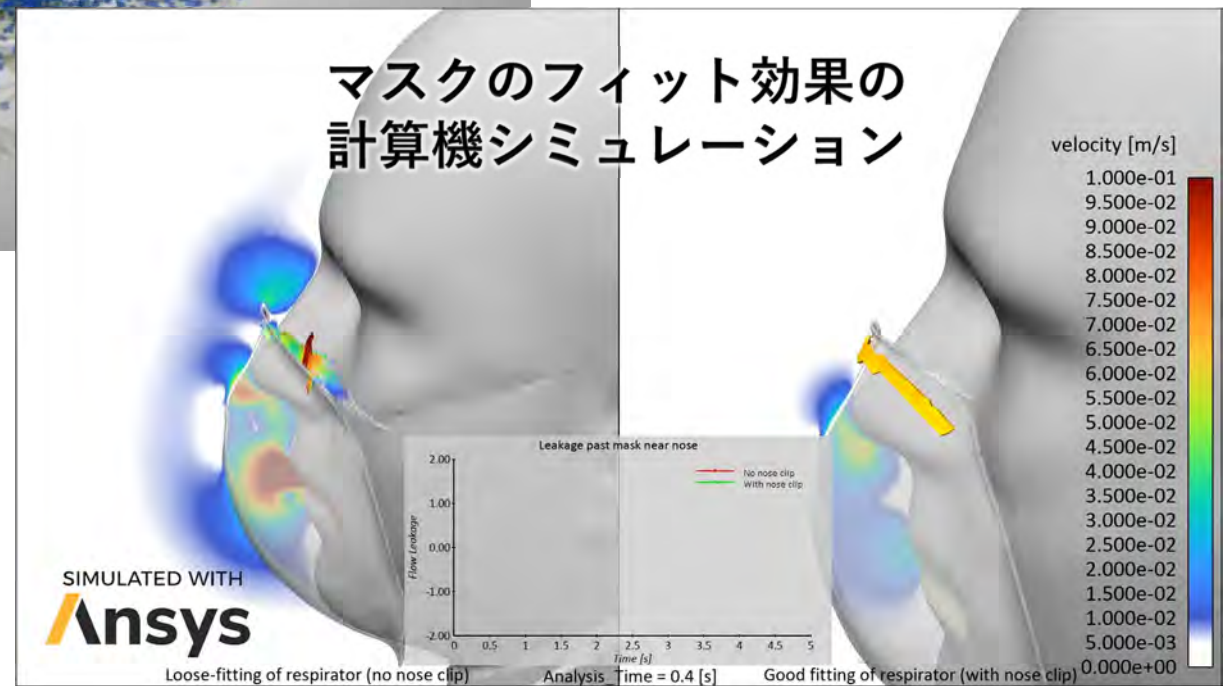
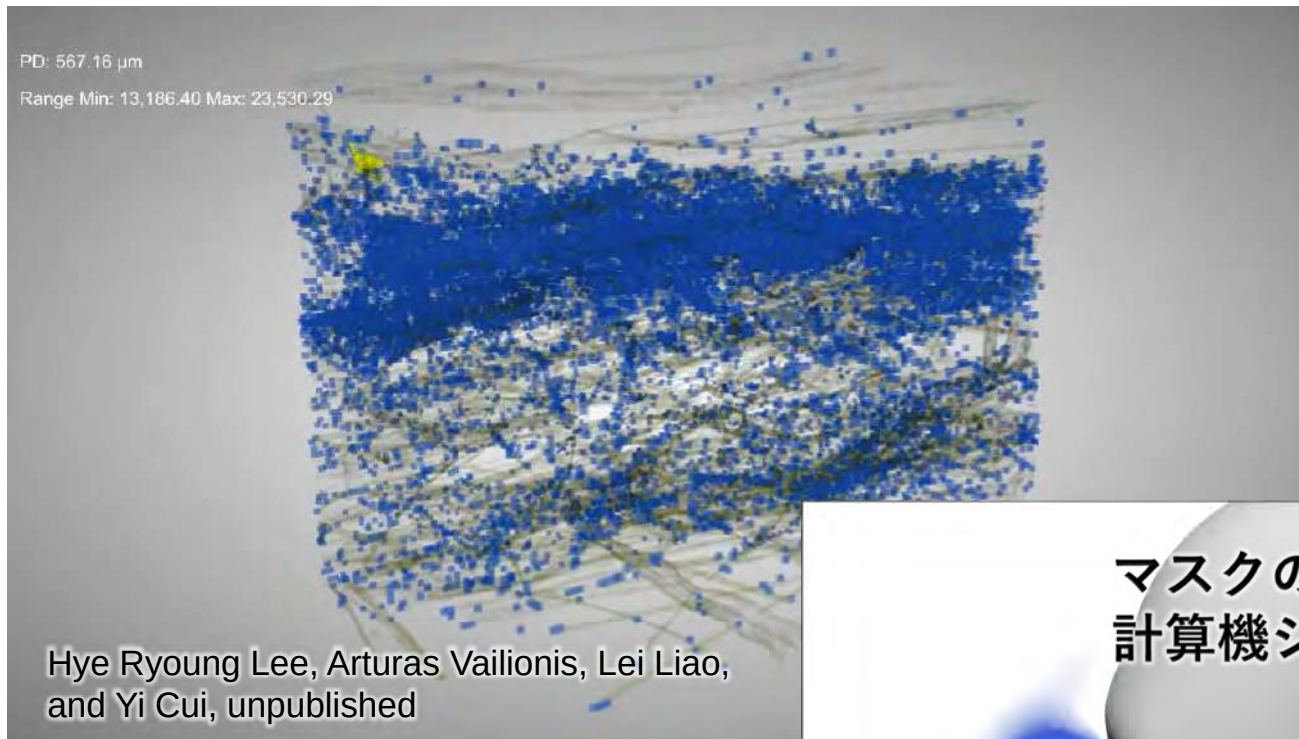
数値データ化



計算機シミュレーション

Transient Airflow Patterns in an Elastic Lung Model.
By Yu Feng, October 11, 2018
Computational Biofluidics and Biomechanics
Laboratory (CBBL@OSU)

X線トモグラフによる映像データ N95マスクによる食塩微粒子の捕獲

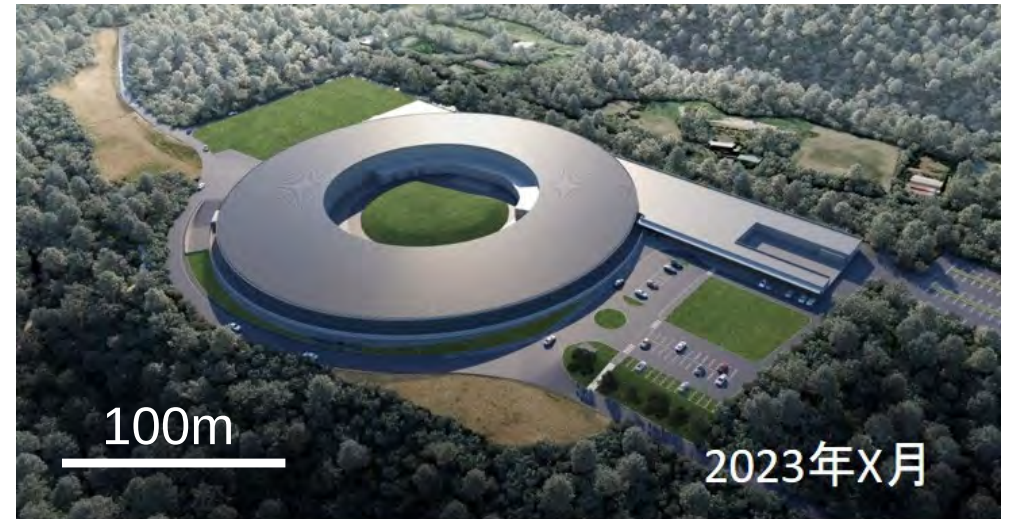


次世代放射光

SPring-8の100倍の輝度とコヒーレンスで
「モノの見え方」が変わる



土地造成の現状
2020年4月



完成予想図

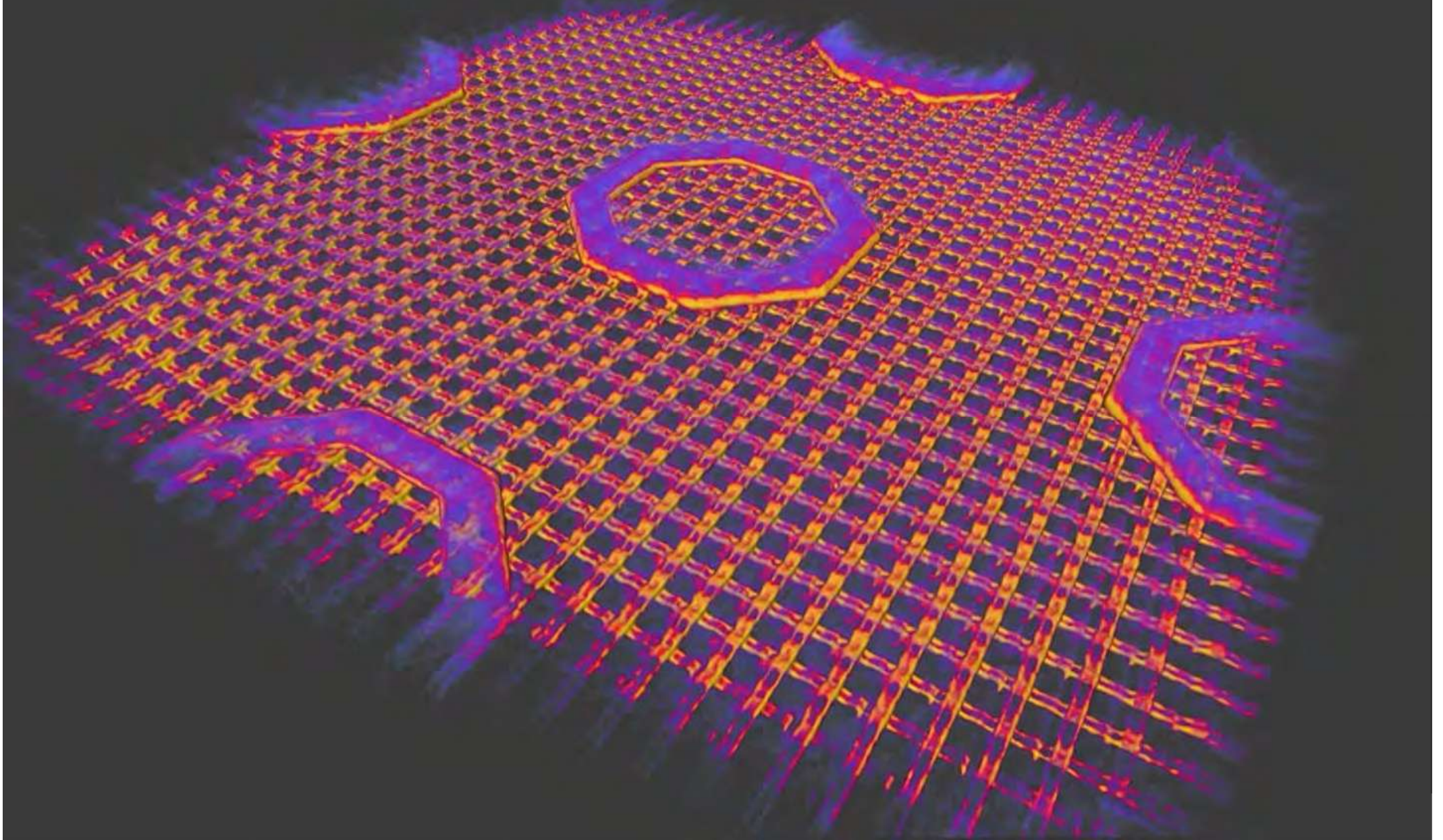
2023年完成予定
東北大学 青葉山新キャンパス



次世代放射光施設が加速する可視化

デバイスの見え方が変わる

スイス放射光： 16nm FinFETデバイスのマルチスケール可視化





脳の見え方が変わる

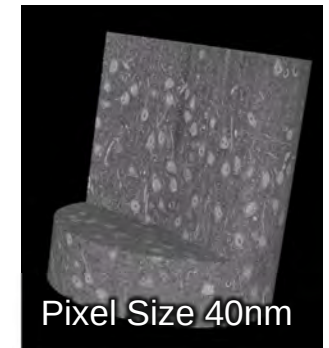
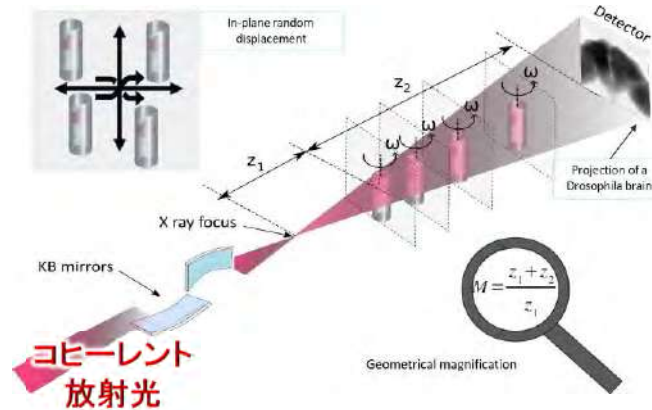
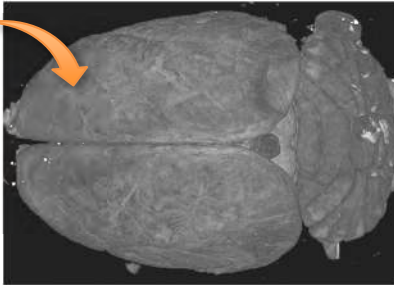
Holographic X-ray nano-Tomography(XNH)

ヨーロッパ共同体放射光施設



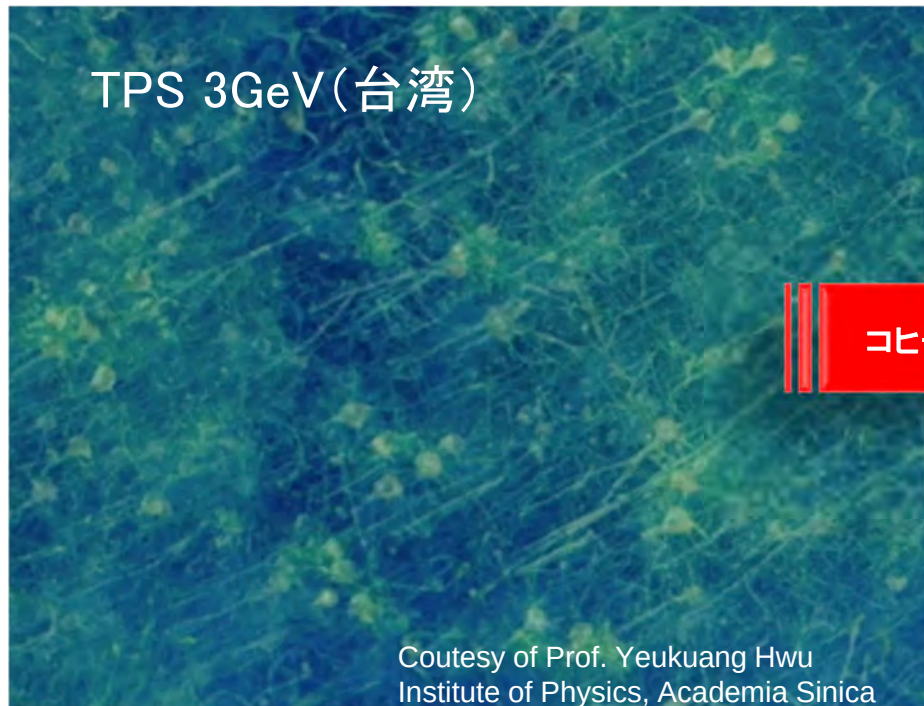
A. Pacureanu

SPring-8(8GeV)



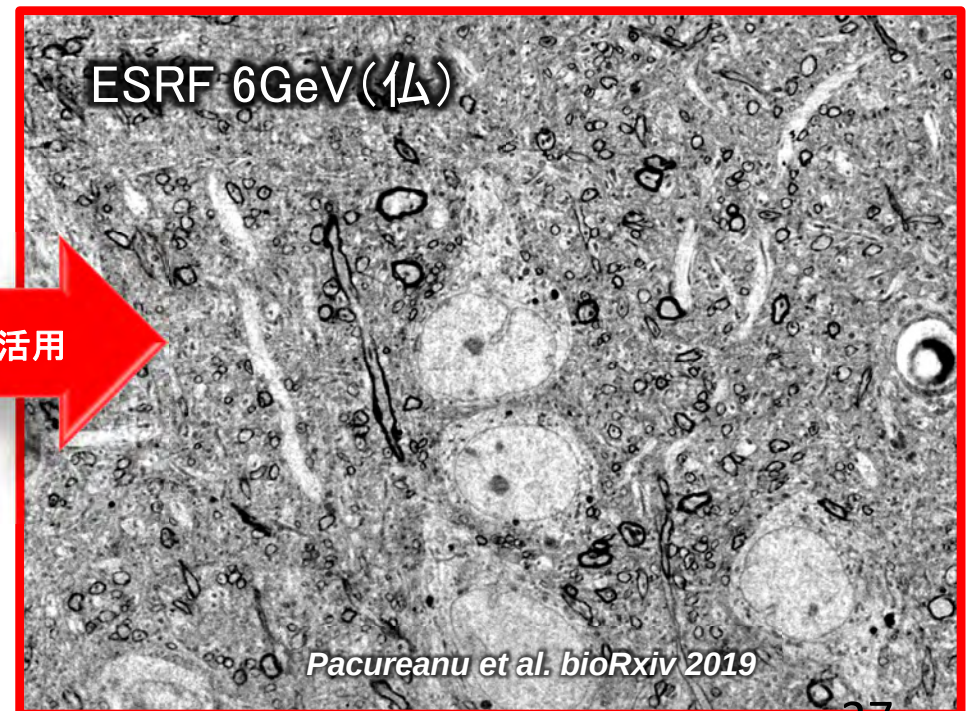
Holographic X-ray nano-omography(XNH)

TPS 3GeV(台湾)



Coutesy of Prof. Yeukuang Hwu
Institute of Physics, Academia Sinica

ESRF 6GeV(仏)



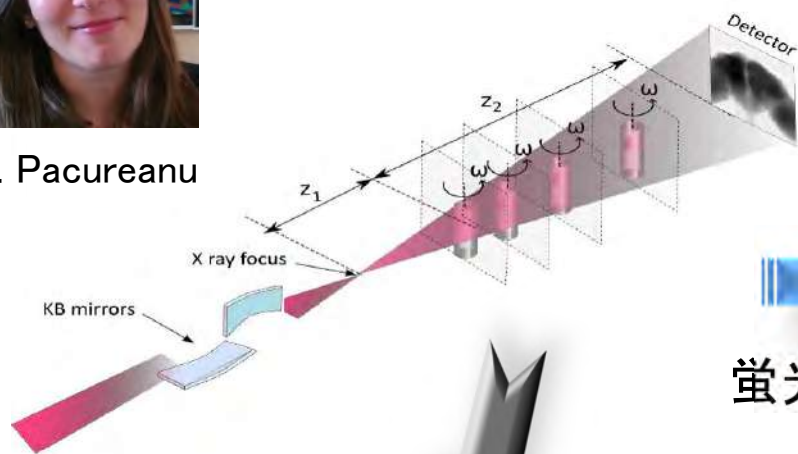
Pacureanu et al. bioRxiv 2019

コヒーレント光活用

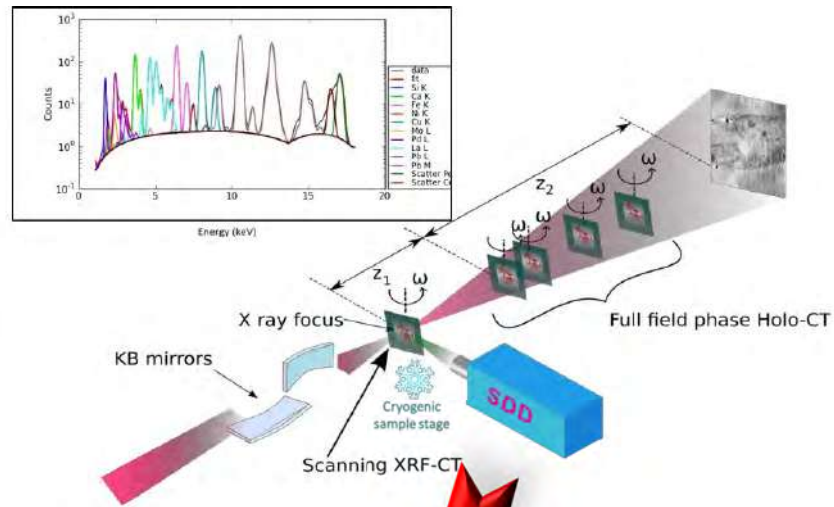


A. Pacureanu

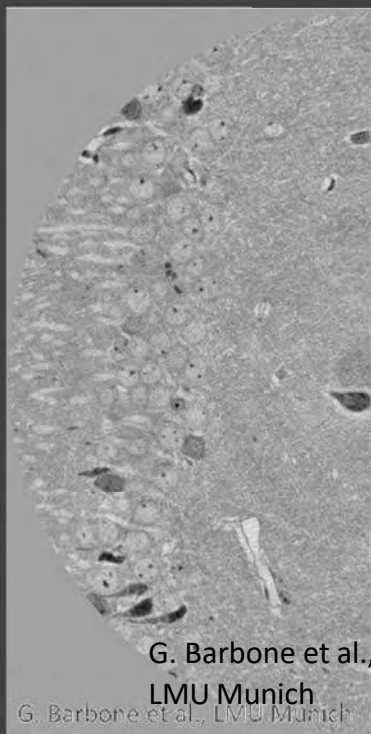
蛍光分析との統合が、海馬への鉄の蓄積を可視化



蛍光分析と統合



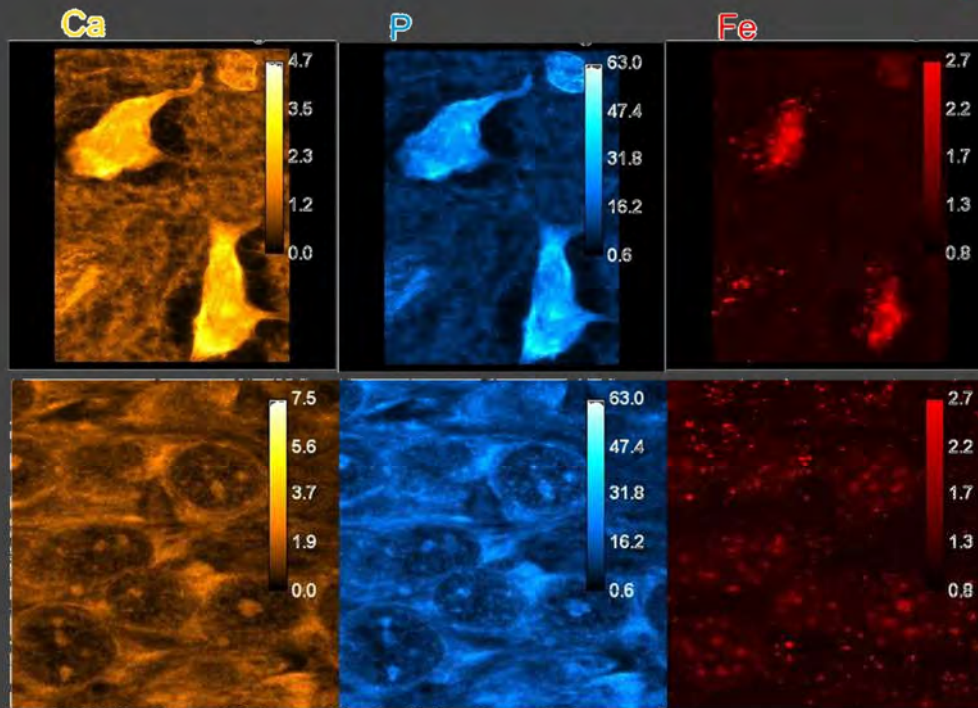
Sub-ppmレベル



G. Barbone et al.,

LMU Munich

G. Barbone et al., LMU Munich



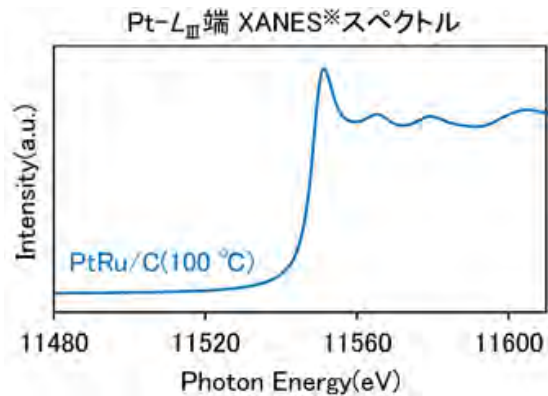
アルツハイマー病
のラット

鉄が細胞核
に蓄積

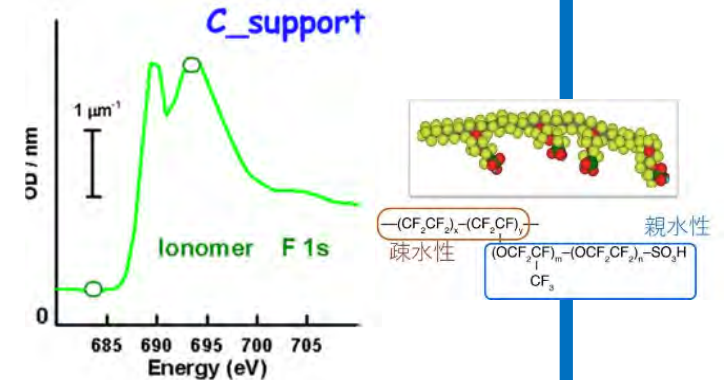
正常なラット

触媒の機能失活の見え方が変わる

従来データ



放射光計測
X線吸収スペクトル

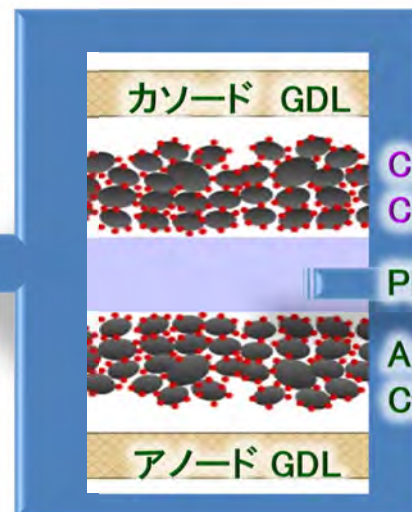


燃料電池の機能可視化



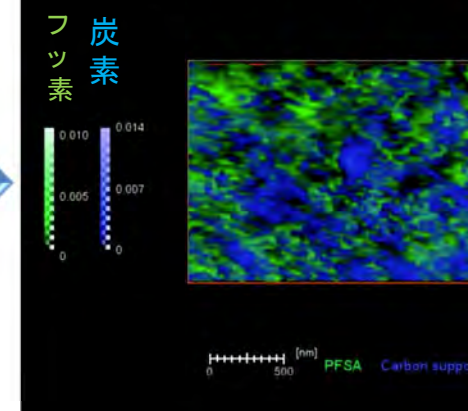
ミクロン分解能で
全体を可視化

赤: Pt触媒
青 酸化したPt触媒
緑: 電極等



セパレータ(PEM)中のフッ素の可視化:

乾湿サイクル下での伸縮に対する
膜の疲労耐久性へのソリューション



不均一性の可視化が、データの価値を変える



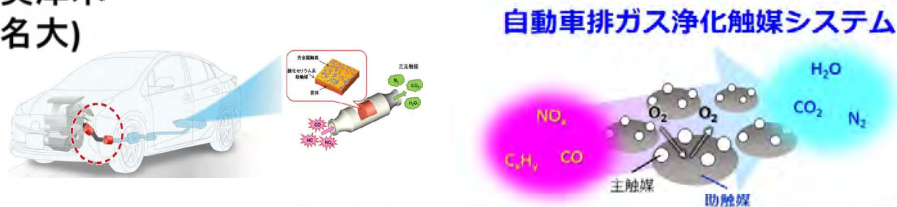
唯美津木
(名大)

13ナノメートルの分解能

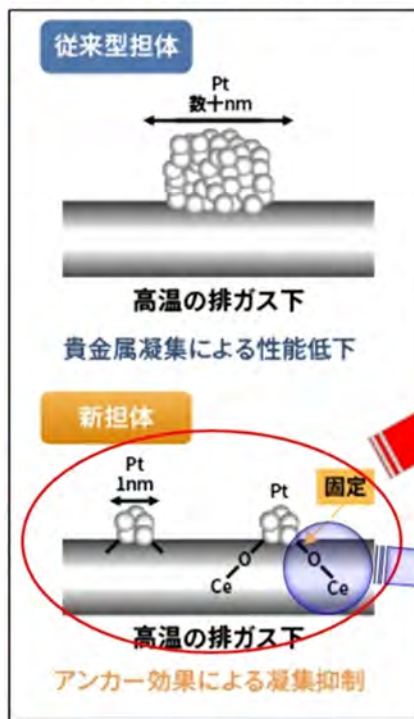
反応の中間過程: 不均一な9万点の情報
様々な条件下での反応を教えてくれるビッグデータ



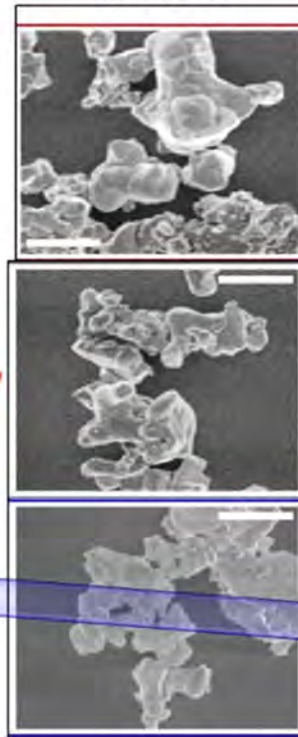
高橋幸生
(阪大→東北大)



貴金属凝集を抑制する
新しいメカニズム



電子顕微鏡
(形を見る)



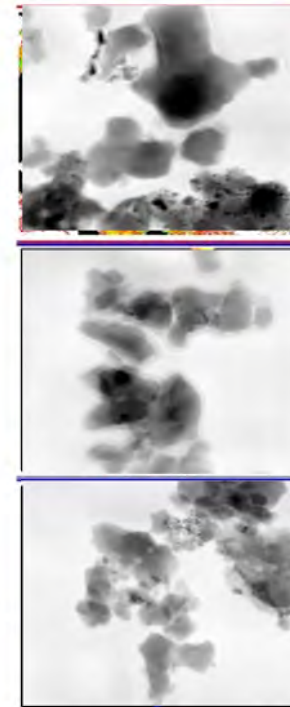
酸素放出

酸素吸蔵

放射光

(Ce原子の酸化反応を見る)

Ce⁺³ Ce⁺⁴



人の流れを可視化するKomprenoとの比較

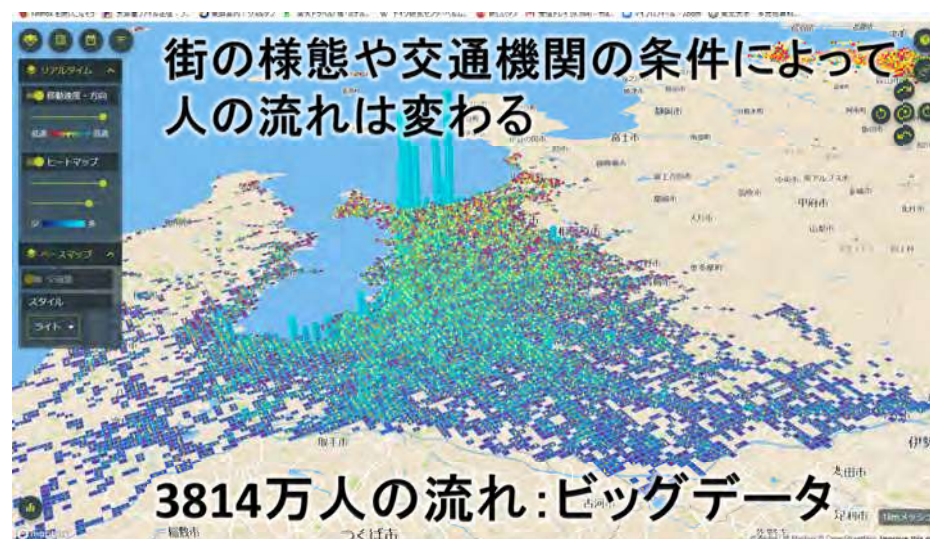
Kompreno https://www.agoop.co.jp/kompreno_public/



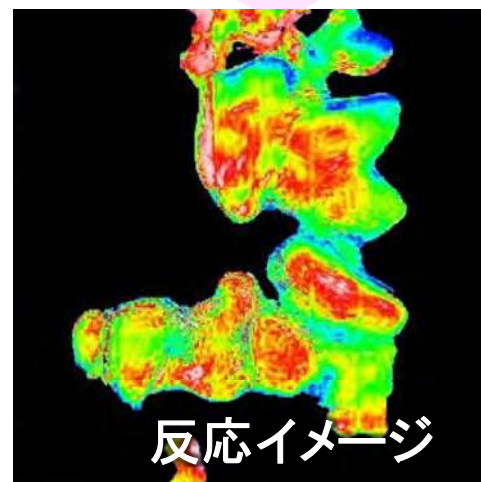
地図イメージ



次世代の計測



人の流れ(相互作用)



濃度、構造の
不均一性、
化学反応、
相互作用、
の膨大な
情報
ビッグデータ

3.0 Cerium Valence 4.0

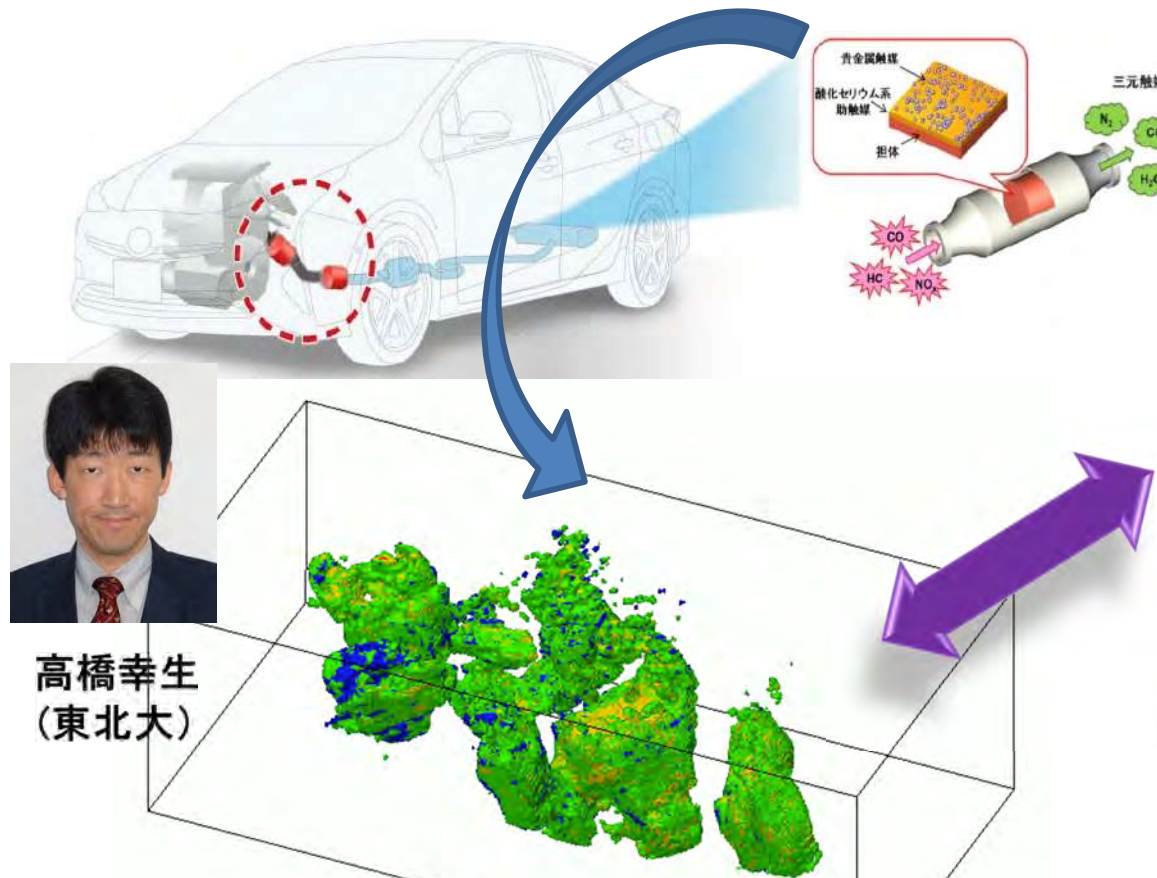
機能のナノ画像が、ビッグデータになる

試行錯誤のモノづくりを、AIが変革する

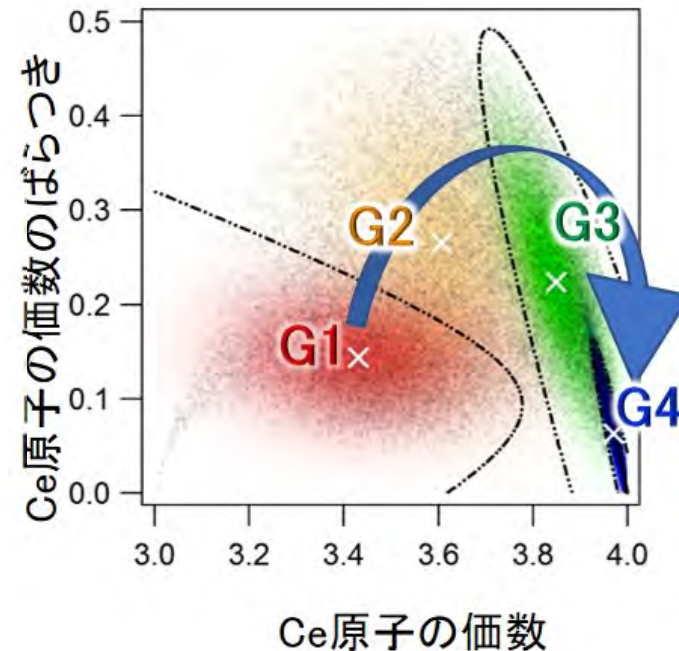
コヒーレント光が、10ナノメートル分解能で、触媒の酸化する様子を画像化する。そして、ビッグデータ解析が、複雑な触媒材料開発にインフォマティクスを導入し、仮説検証サイクルの高速化をもたらす。



ダム ヒョウ チ
Dam Hieu Chi
北先大



これまでわからなかった、触媒の材料の中で
金属原子のCeの酸化が進む様子を、
ビッグデータ解析が
(G1-G2-G3-G4)であると、教えてくれる



高橋幸生
(東北大)

約2700万点(13ナノメートル分解能)で画像化された
触媒の酸素の拡散の様子を色で示したもの。

ナノの磁石機能可視化ー＞ビッグデータ同化



宝野和博
(NIMS)



中村哲也
(東北大学)



磁石機能の可視化
N極S極のナノでの挙動

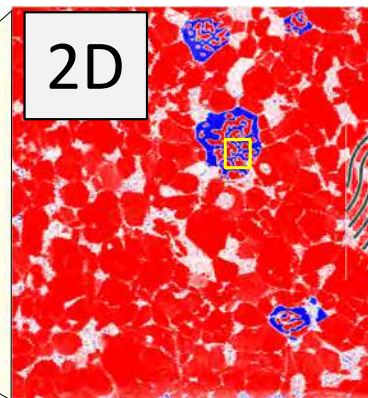
+

計算機シミュレーション
データベース

=

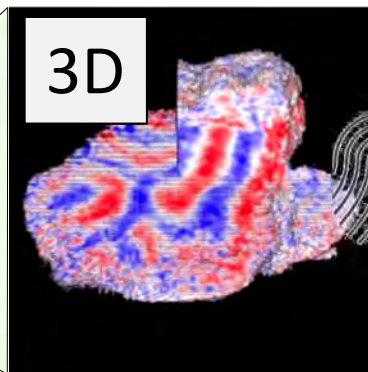
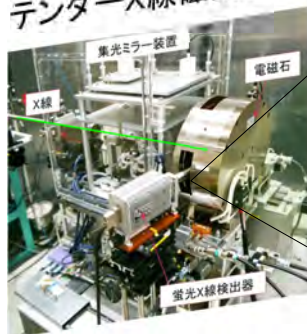
次世代磁石
のデザイン

軟X線ナノビーム顕微分光

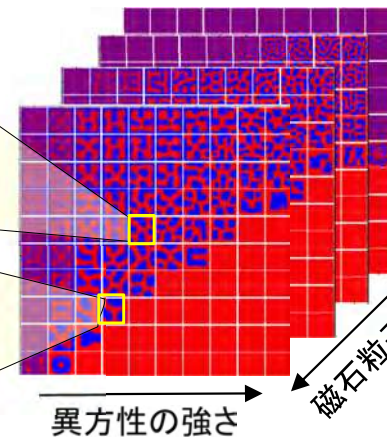
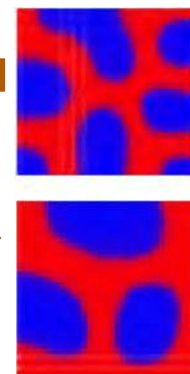


2D

テンダーX線磁気CT



3D



磁気相互作用

磁石粒子厚さ

異方性の強さ



宮下精二
(東京大学)

企業が、建設資金を拠出
学術研究とマッチング(コアリション)を形成し利用

Coalition





渡邊真史
東北大学

産学連携の市場開拓への機械学習の応用

建設資金を拠出する企業を予測、開拓する

【交渉履歴データ】

2017年12月時点でコンタクト済みの企業のうち
119社を、前向き、中立、**難色**の3つに分類。
データ数不足解消の為、ダミーデータを追加

【企業プロフィール情報】

日経会社情報DIGITALから、無料公開企業
データ(当時)を取得
資本金、従業員数、業界分類、主力事業、事業領域等

【機械学習】

ロジスティック回帰によるクラス判別関数の生成
(数式処理ソフトMathematica)

データ整形以外の学
習部分はわずか1行!
`Classify[trainingData, Method ->
"LogisticRegression"]`

【企業情報2233社】

日経会社情報DIGITAL
無料公開企業データ(当時)

判別関数

加入に**前向き**な
る確率のリスト

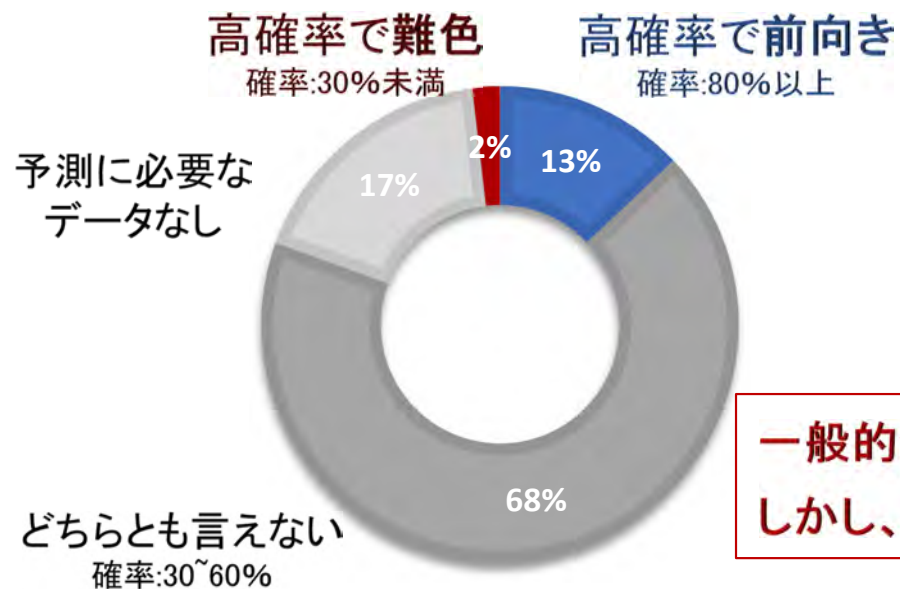
【結果例】

■ 予測では加入の確率が低いにも関わらず、加入した企業

T社	前向き確率=0.24
M会	前向き確率=0.18

→ 解釈: 地元企業の振興、地元大学との産学連携など日経データに含まれない企業戦略が大きな要素に

■ その後、2年間で加入意向を表明した企業の事前予測はどうだったか？



◆ 実際に参加した企業では、

事前予測: **前向き** > **難色**

◆ 予想確率30-80%の企業は、会社情報に含まれない個別の事情の影響大

→ 会って見ないと分からない

一般的企業情報では全てが決まらない

しかし、一般的情報から一定の感触を掴むのに有効

まだまだ、学習途上

次世代放射光は、AI・ビッグデータ時代に欠かせない ツールとなる

□ 次世代放射光： ナノの形だけでなく、機能の情報も、高精細に可視化

モノクロTV/カラーTVの時代から、5G/8K放送を見る時代へ変わるような、非連続な変化

□ 「何が起きているか」から、「なぜ起きているか」を見るツールへ進化

□ その結果、AIやデータ科学との融合が進み、仮説検証サイクルを加速

□ 非専門家も利用できるツールへとなり、分野融合、産学連携が広がる

16GB:スマートホンの平均的なメモリー

1TB: 50000本の本に相当する印刷物

2PB: フェイスブックの全ての写真の容量

20PB: 毎日グーグルによって生み出されるデータ量

1EB: ネットフリックスで1000回ストリーム再生された
総カタログ

9PB: 2018年に海外の既存放射光施設が生み出したデータ

60PB: 放射光施設が1年間生み出すデータ

60PB=60,000TB=60,000,000GB

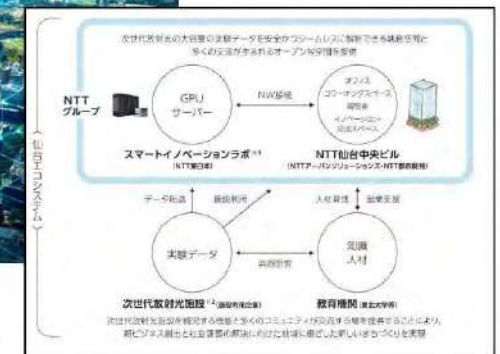
=スマホ 375万台

Society 5.0のロールモデル

データ戦略 75社以上の大企業が参画する次世代放射光施設は
先端科学技術データの必須通過点となる。

NTTアーバンソリューションズ 仙台エコシステム

<https://www.ntt-us.com/topics/2019/12/topics-191205-01.html>



ご清聴ありがとうございました

謝辞：（一財）光科学イノベーションセンター
東北大学 国際放射光イノベーションスマート研究センター
理化学研究所放射光科学研究センター
（公財）高輝度光科学研究センター
東京大学物性研究所



石川哲也
理研



村松淳司
東北大



桑田耕太郎
首都大東京