

# スパコン創薬の現状と未来

京都大学 医学研究科  
理化学研究所 生命システム研究センター  
理化学研究所 計算科学研究機構  
先端医療振興財団 先端医療センター研究所

奥野恭史

# 現代の創薬(分子標的創薬)の流れ

化合物ライブラリー  
(製薬会社 数万種)

仮想化合物  
10の60乗件

活性化合物の  
スクリーニング

上市に至るまでの成功確率: 2万分の1以下  
(開発費用1000億円、開発期間約15年)

リード最適化

前臨床試験

臨床試験  
承認

結合情報  
1千万件

生物活性情報  
12億件

遺伝子発現情報  
100万件

疾患情報  
2.2万件

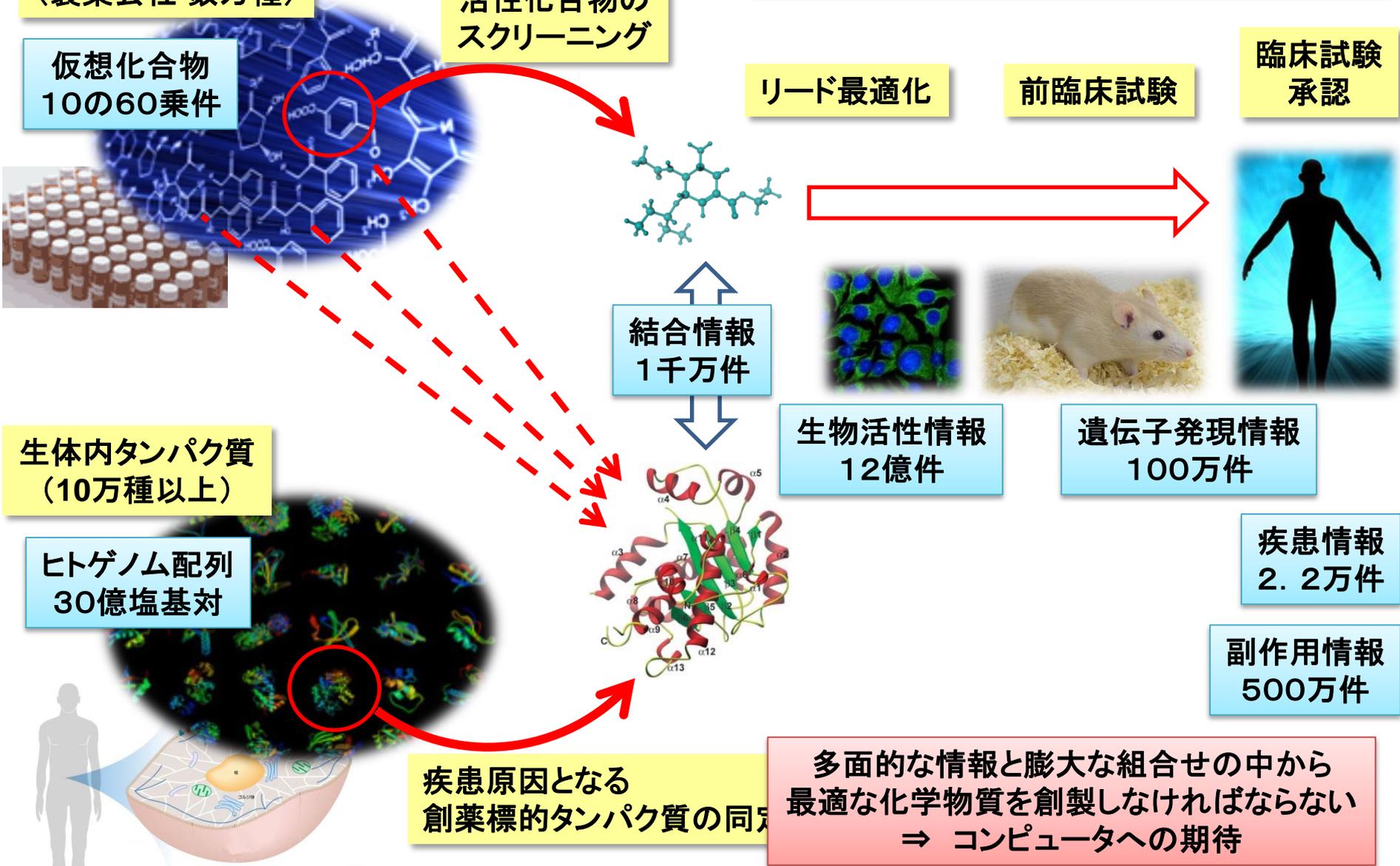
副作用情報  
500万件

生体内タンパク質  
(10万種以上)

ヒトゲノム配列  
30億塩基対

疾患原因となる  
創薬標的タンパク質の同定

多面的な情報と膨大な組合せの中から  
最適な化学物質を創製しなければならない  
⇒ コンピュータへの期待



# 「京」産業利用枠：新薬開発を加速する「京」インシリコ創薬基盤の構築

ビッグデータ創薬

シミュレーション創薬



**申請主体（事務局）：** NPO法人バイオグリッドセンター関西

研究代表 京都大学医学研究科 奥野恭史

**製薬企業（23社）：** アスピオファーマ、杏林製薬、エーザイ、小野薬品工業、科研製薬、キッセイ薬品工業、参天製薬、塩野義製薬、大日本住友製薬、田辺三菱製薬、日本新薬、協和発酵キリン、千寿製薬、大正製薬、帝人ファーマ、東レ、日産化学工業、マルホ、持田製薬、アステラス製薬、日本たばこ産業、MeijiSeikaファルマ、カネカ

**IT企業（2社）：** 京都コンステラ・テクノロジーズ, 三井情報

**大学等：** 京都大学医学研究科, 産業技術総合研究所, 理研AICS, 先端医療振興財団, 都市活研

- コンピュータ創薬の根本課題に挑戦
- 次世代の計算創薬「ビッグデータ創薬」と「シミュレーション創薬」の開拓
- 製薬会社による現場利用に耐える計算フロー（計算精度と計算時間）の構築
- 我が国のコンピュータ創薬の中心拠点形成

# 「京」を用いた機械学習による世界最大規模のタンパク質-化合物結合予測

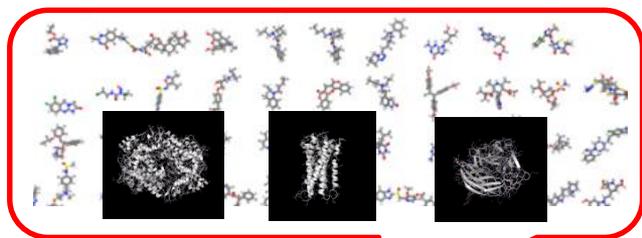
以後、予測結果を基に、各製薬会社が独自に医薬品開発を行う

論文等で結合することが分かっている  
タンパク質と化合物の結合ペア: 12万ペア

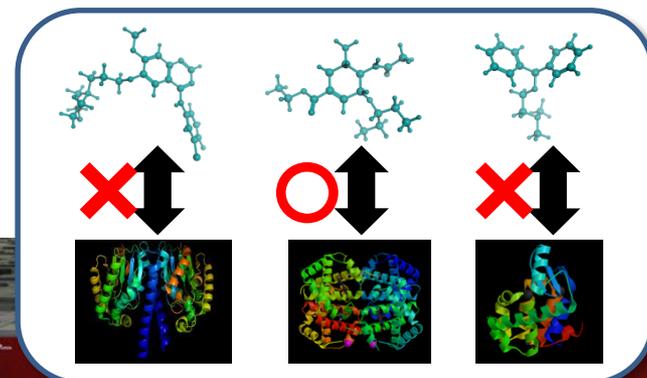
世界最大規模(189.3億ペア)の結合予測  
631種のタンパク質(GPCR、キナーゼ)と  
3000万種の化合物の全組合せ

大量のタンパク質と化合物の結合データの機械学習

病気の原因タンパク質に  
結合する化合物を「京」で認識



予測



結合パターンの  
統計ルール化

「京」により  
超高速予測を実現

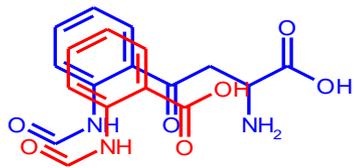
膨大なタンパク質と  
化合物の組合せに対応



# CGBVS法の計算フロー

バイオインフォマティクスとケモインフォマティクスの統合により実現

Chemical Structure



Interaction Pairs



Protein Sequence

```

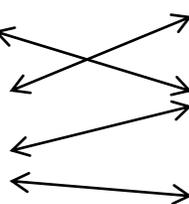
MAQTPAFNPKPKUELHHLGDGAKPETLLVFGRKRGIALPAD
VIAGCREATKRIAYEFVEMKAKEGVVEIRYSPLLANSKI
EDAFGZIKV
RTUNHAGEV
URFKNDKAI
VQ
MAQTPAFNPKPKUELHHLGDGAKPETLLVFGR
VIAGCREATKRIAYEFVEMKAKEGVVEIRYSPLLANSKI
EDAFGZIKV
RTUNHAGEV
URFKNDKAI
VQ
MAQTPAFNPKPKUELHHLGDGAKPETLLVFGRKRGIALPADTUEELRNTIGMOKPLSPGLAKFDYVNP
VIAGCREATKRIAYEFVEMKAKEGVVEIRYSPLLANSKI
EDAFGZIKV
RTUNHAGEV
URFKNDKAI
VQ
MAQTPAFNPKPKUELHHLGDGAKPETLLVFGRKRGIALPADTUEELRNTIGMOKPLSPGLAKFDYVNP
VIAGCREATKRIAYEFVEMKAKEGVVEIRYSPLLANSKI
EDAFGZIKV
RTUNHAGEV
URFKNDKAI
VQ
MAQTPAFNPKPKUELHHLGDGAKPETLLVFGRKRGIALPADTUEELRNTIGMOKPLSPGLAKFDYVNP
VIAGCREATKRIAYEFVEMKAKEGVVEIRYSPLLANSKI
EDAFGZIKV
RTUNHAGEV
URFKNDKAI
VQ

```

1. 相互作用データ (学習データ)

Vector

Mw, logP, #C, #OH, ...  
 (252, 7, 4, 5, ...)  
 (320, 1, 2, 1, ...)  
 (238, 6, 7, 4, ...)

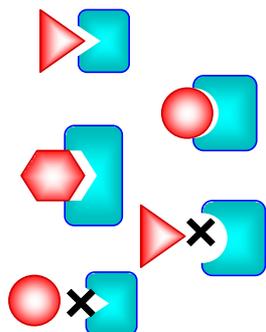


AA, AH, AS, ...  
 (72, 51, 47, ...)  
 (81, 53, 64, ...)  
 (60, 43, 48, ...)

2. 数値化(ベクトル表現)

3. ベクトルの連結

Interaction Vector



(252, 7, 4, 5, ... 72, 51, 47, ...) Bind  
 (320, 1, 2, 1, ... 60, 43, 48, ...) Bind  
 (238, 6, 7, 4, ... 81, 53, 64, ...) Bind  
 (252, 7, 4, 5, ... 60, 43, 48, ...) Non-bind  
 (320, 1, 2, 1, ... 72, 51, 47, ...) Non-bind

4. 機械学習 (サポートベクターマシン)



Query Pair



(220, 3, 2, 3, ... 42, 31, 34, ...) Bind or Non-Bind ?

5. 予測

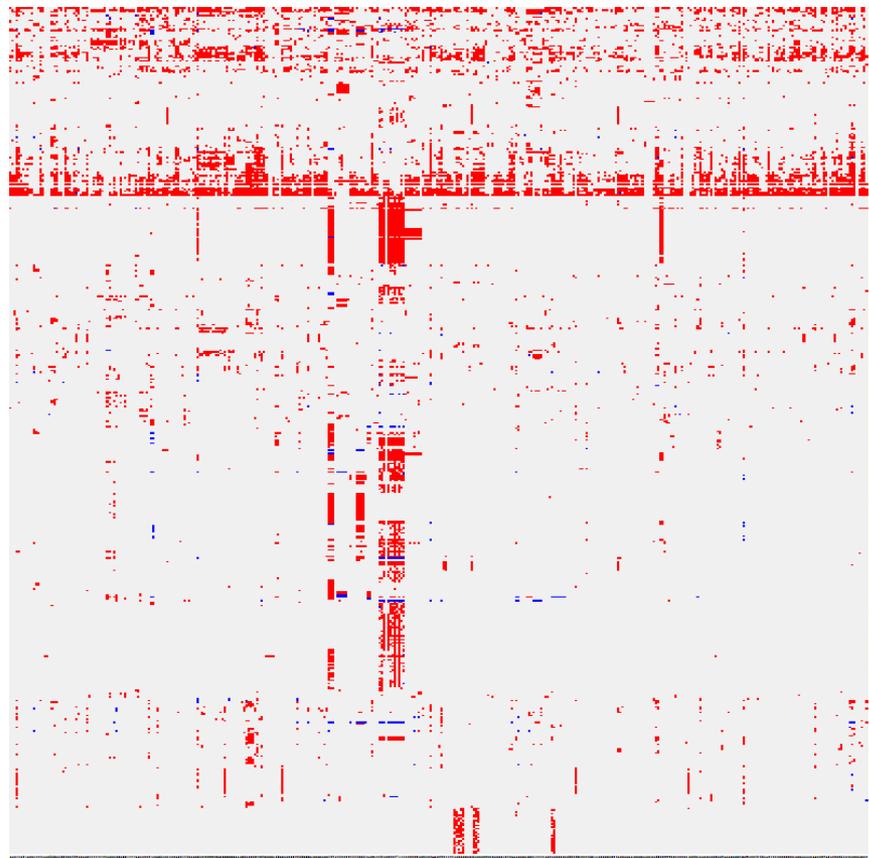
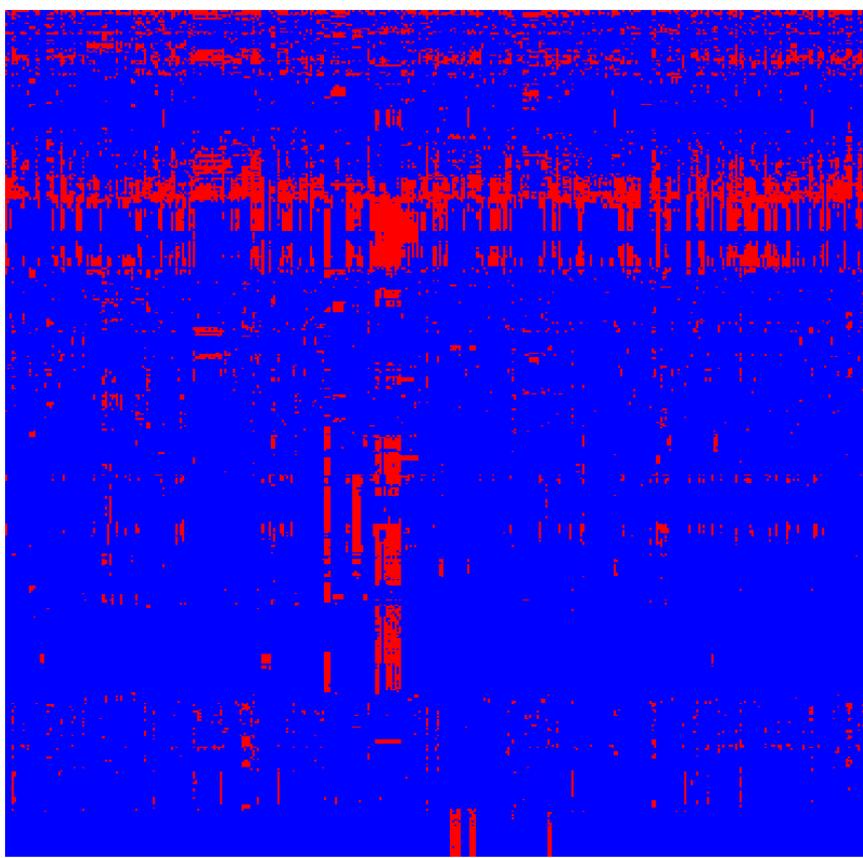
# 「京」を用いたCGBVS法による相互作用予測(Kinaseの例)

タンパクー化合物の全組合せ(189.3億ペア)を計算するのに汎用計算機(16ノード使用)では約2年かかる。

「京」8万ノードを使った大規模並列計算をすれば、5時間45分で計算終了

「京」が予測した結合パターン(予測結果)

実際の結合パターン(実験結果)



Kinase(388種) ⇒ 631種

Kinase(388種)

# 機械学習によるタンパク質-化合物結合予測

以後、予測結果を基に、各製薬会社が独自に医薬品開発を行う

論文等で結合することが分かっている  
タンパク質と化合物の結合ペア: 12万ペア

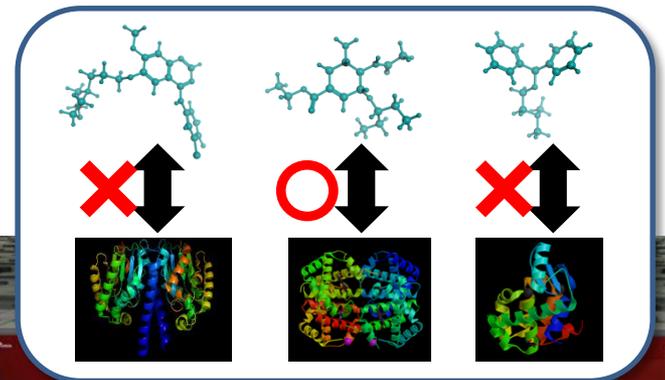
世界最大規模(189.3億ペア)の結合予測  
631種のタンパク質(GPCR、キナーゼ)と  
3000万種の化合物の全組合せ

大量のタンパク質と化合物の結合データの機械学習

病気の原因タンパク質に  
結合する化合物を「京」で認識

サポートベクターマシン  
学習データ件数: 12万相互作用

予測



結合パターンの  
統計ルール化

「京」により  
超高速予測を実現

膨大なタンパク質と  
化合物の組合せに対応



# 機械学習によるタンパク質-化合物結合予測

以後、予測結果を基に、各製薬会社が独自に医薬品開発を行う

論文等で結合することが分かっている  
タンパク質と化合物の結合ペア: 12万ペア

世界最大規模(189.3億ペア)の結合予測  
631種のタンパク質(GPCR、キナーゼ)と  
3000万種の化合物の全組合せ

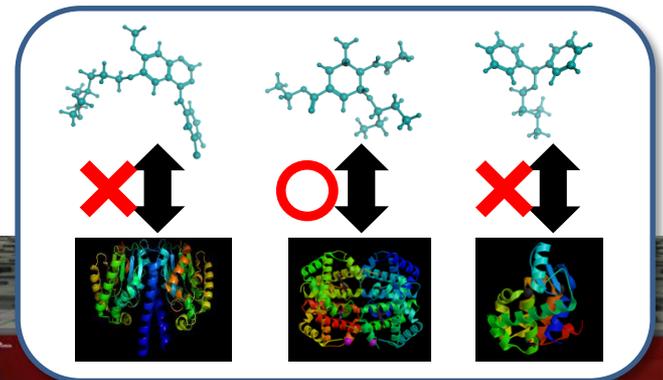
大量のタンパク質と化合物の結合データの機械学習

病気の原因タンパク質に  
結合する化合物を「京」で認識

Deep Learning  
学習データ件数: 数百万相互作用

測

結合パターンの  
統計ルール化



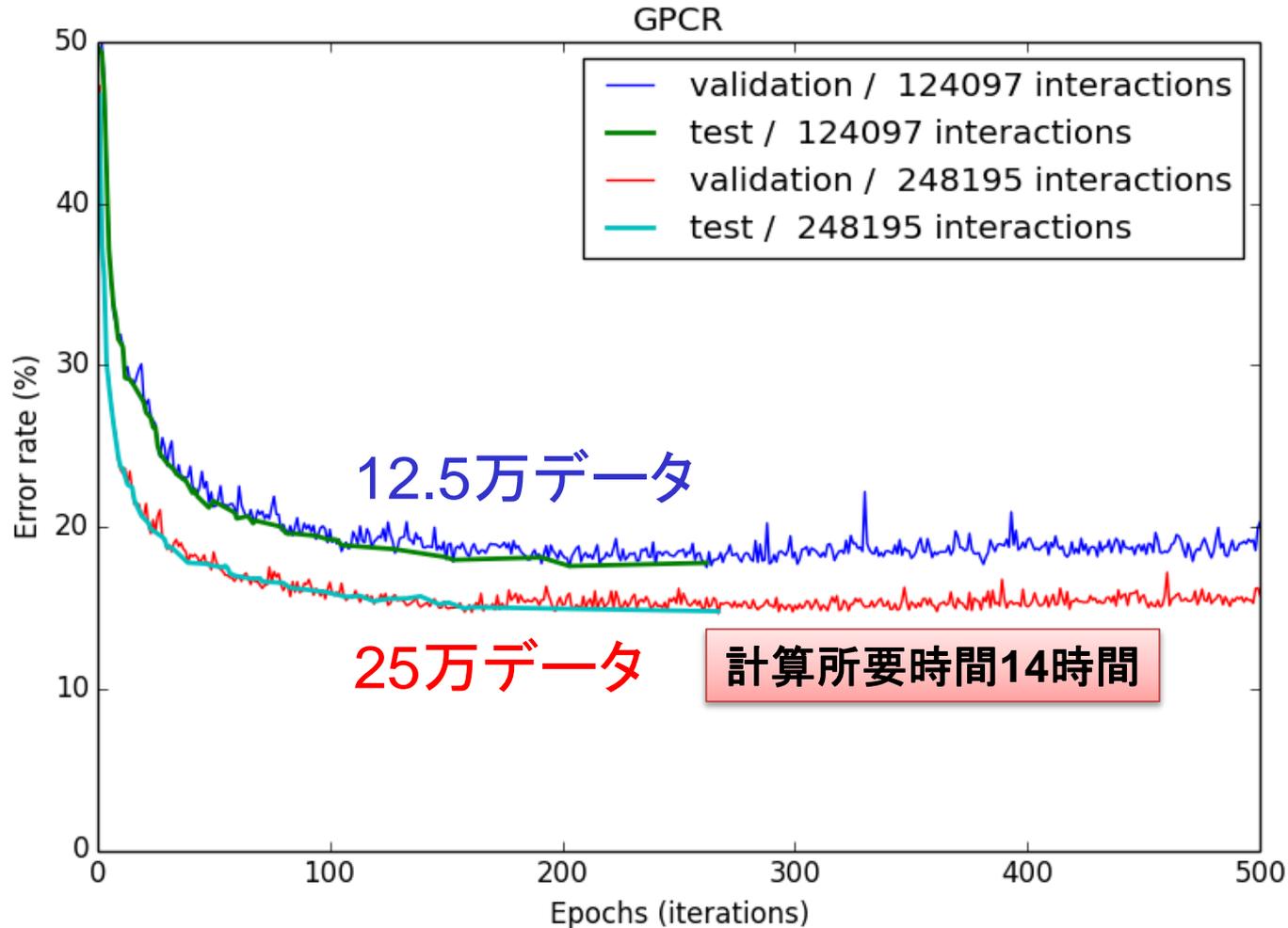
「京」により  
超高速予測を実現

膨大なタンパク質と  
化合物の組合せに対応



# 学習データのスケールビリティ(予測率と学習時間)

## 学習データ件数25万と12.5万での予測性能の比較



学習データの件数が多いほうが効率的に学習できる

# 「京」産業利用枠：新薬開発を加速する「京」インシリコ創薬基盤の構築

ビッグデータ創薬

シミュレーション創薬



**申請主体（事務局）：** NPO法人バイオグリッドセンター関西

研究代表 京都大学医学研究科 奥野恭史

**製薬企業（23社）：** アスピオファーマ、杏林製薬、エーザイ、小野薬品工業、科研製薬、キッセイ薬品工業、参天製薬、塩野義製薬、大日本住友製薬、田辺三菱製薬、日本新薬、協和発酵キリン、千寿製薬、大正製薬、帝人ファーマ、東レ、日産化学工業、マルホ、持田製薬、アステラス製薬、日本たばこ産業、MeijiSeikaファルマ、カネカ

**IT企業（2社）：** 京都コンステラ・テクノロジーズ, 三井情報

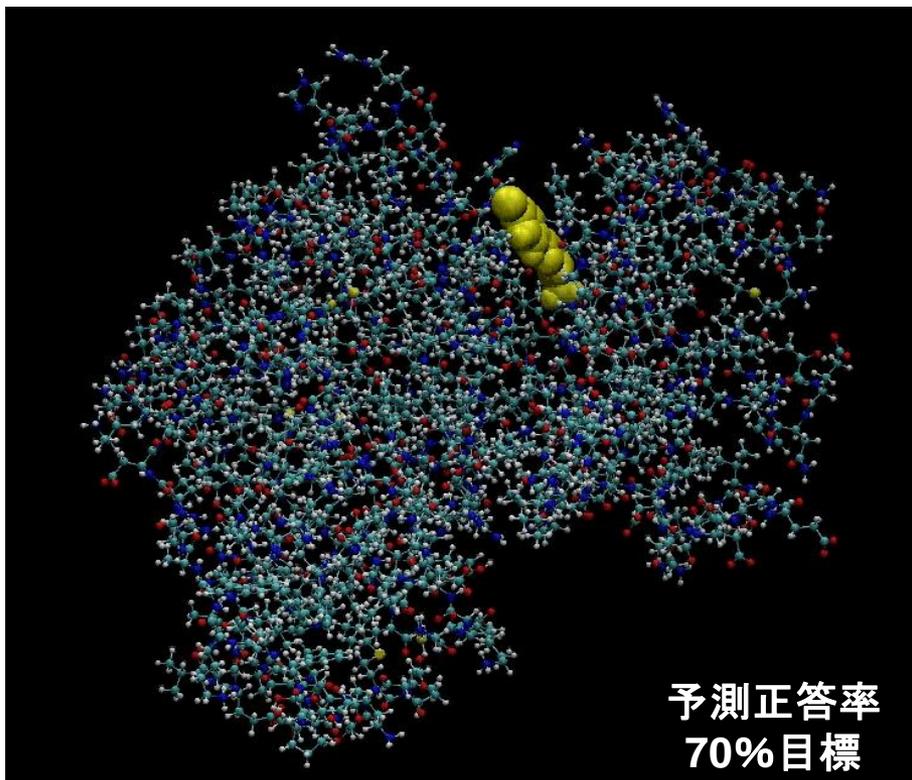
**大学等：** 京都大学医学研究科, 産業技術総合研究所, 理研AICS, 先端医療振興財団, 都市活研

- コンピュータ創薬の根本課題に挑戦
- 次世代の計算創薬「ビッグデータ創薬」と「シミュレーション創薬」の開拓
- 製薬会社による現場利用に耐える計算フロー（計算精度と計算時間）の構築
- 我が国のコンピュータ創薬の中心拠点形成

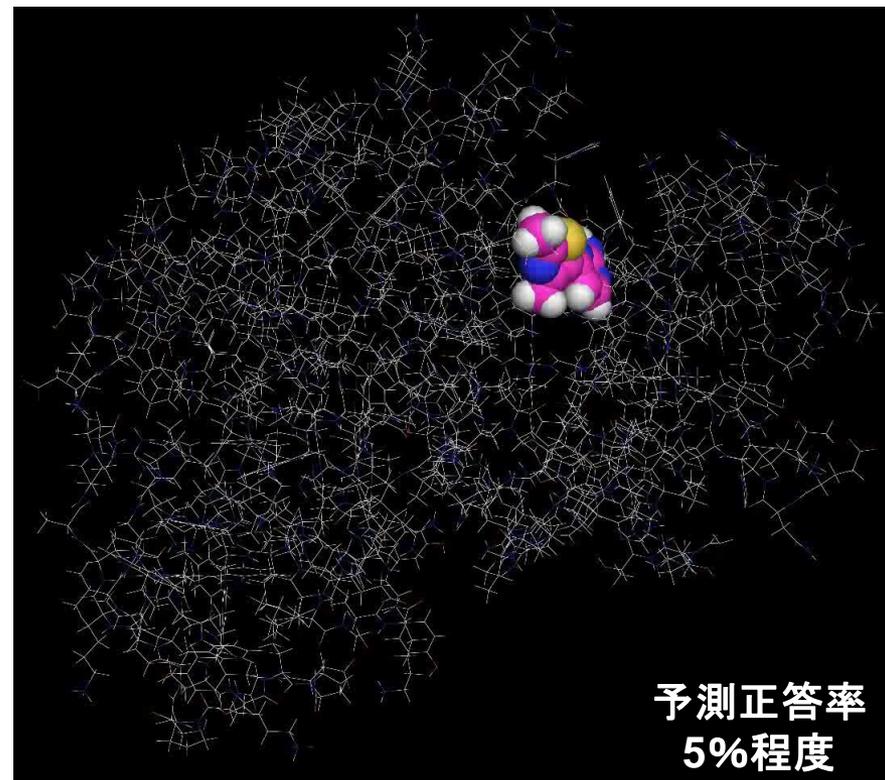
# 結合親和性予測:MP-CAFREE法 (Massively Parallel Computation of Absolute binding Free Energy)

アンサンブルシミュレーションによって、  
正確かつ頑強にタンパク質と化合物の結合の強さ(結合自由エネルギー)を求める。

## 「京」による結合シミュレーション(MP-CAFREE)



## 従来型の結合シミュレーション(Docking)



結合の強さを正確かつ頑強に計算をするには、分子の動きや溶媒(水分子)も含めた大規模シミュレーションが必要であるが、これまでは計算機パワーの問題で非現実だった。

# GROMACS: 分子動力学(MD)シミュレーションのフリーソフト

理研 計算科学研究機構  
「京」コンピュータ



原子数42,088  
10ノード(80プロセス)  
13.3 ns / day

大阪大学 サイバーメディアセンター  
大規模可視化対応PCクラスター(VCC)

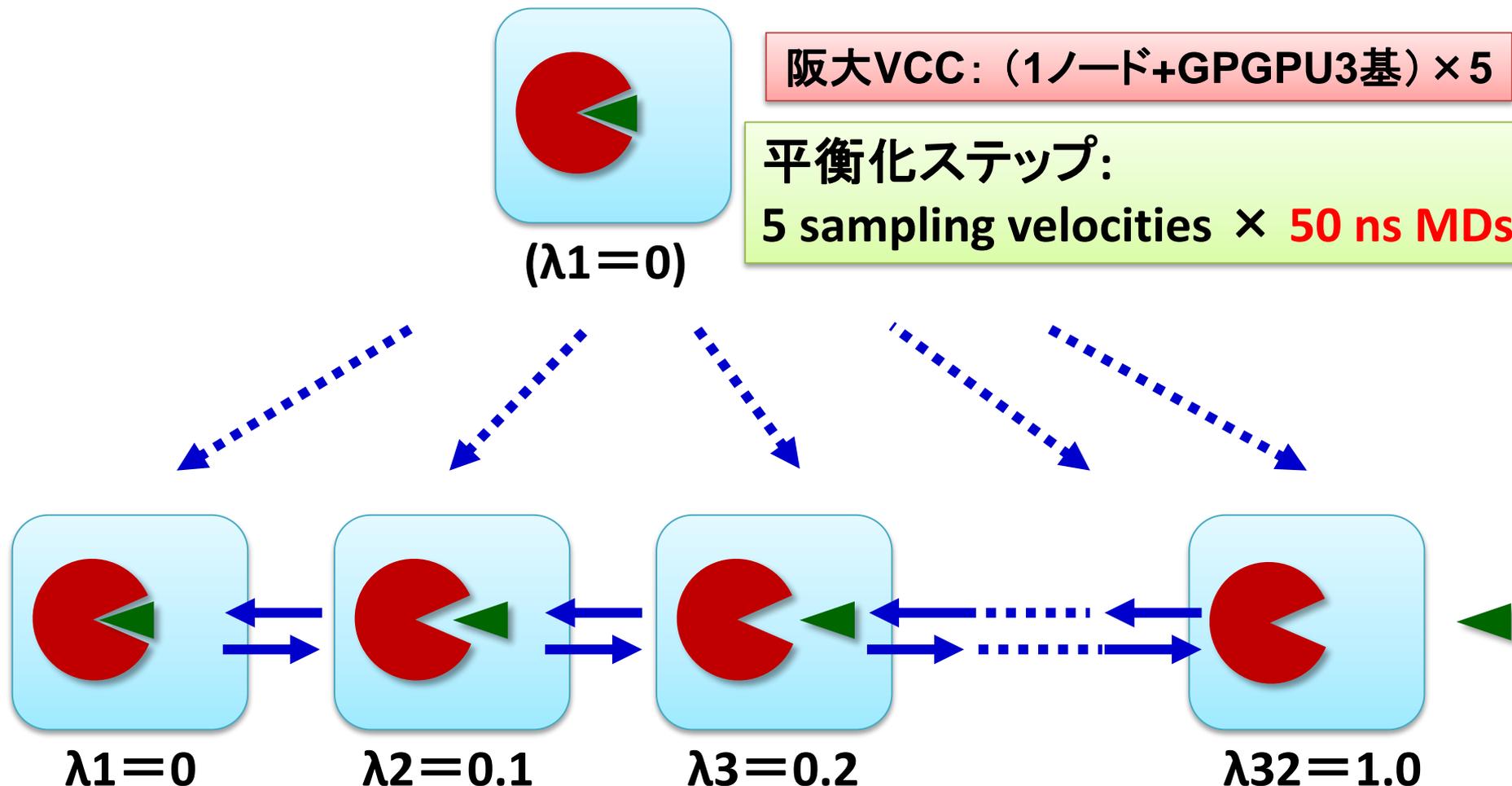


原子数42,088  
1ノード+GPGPU3基  
51.7 ns / day

約4倍

GROMACS 4.6.2 (「京」へ最適化あり)

# MP-CAFEE法の計算プロセス: 平衡化計算と非平衡過程計算



阪大VCC: (1ノード+GPGPU3基) × 5

平衡化ステップ:

5 sampling velocities × 50 ns MDs

$(\lambda_1=0)$

$\lambda_1=0$

$\lambda_2=0.1$

$\lambda_3=0.2$

$\lambda_{32}=1.0$

非平衡化ステップ:

6 sampling velocities × 32  $\lambda$  points × 2 ns MDs

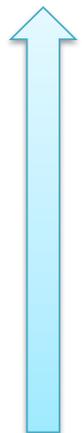
「京」: 192 × 10ノード

# MP-CAFEE法による予測と実験結果比較

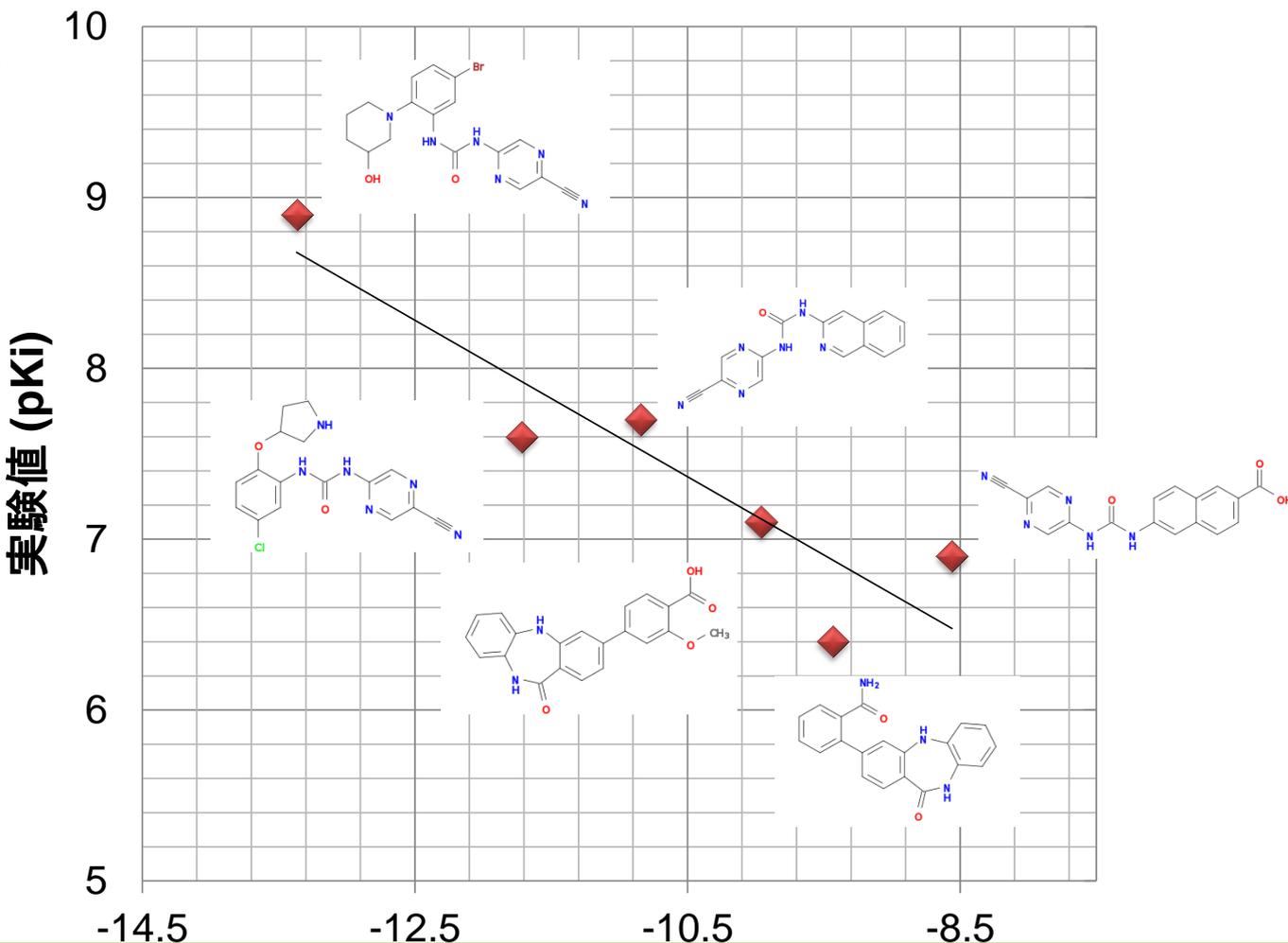
150個の化合物についてタンパクとの結合の強さを計算するのに、通常の汎用機では20年かかるところが、「京」をフルに利用したら1週間程度で計算が可能

## CHK1

強



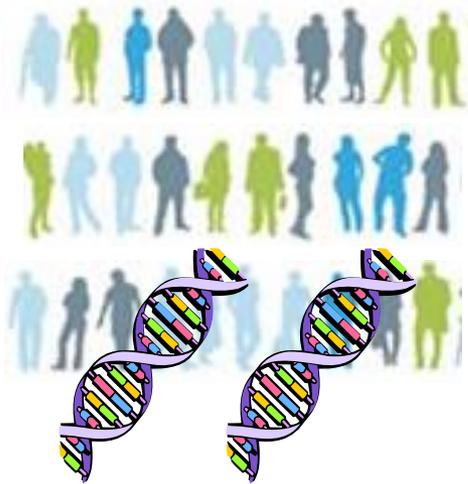
弱



実験では、新規な化合物を化学合成し生物活性評価するのに2ヶ月程度かかるのに対し「京」では、3日程度で人間の労働力無しで、結果の取得が可能。

# 個別化薬剤治療のための分子シミュレーション

## パーソナルゲノム情報



## 薬理ゲノミクス

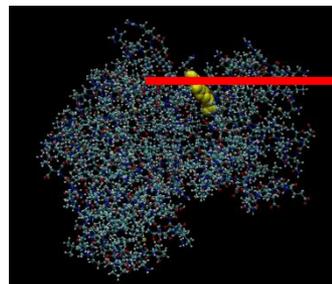
京大病院がんセンター  
院内でのフロー構築中

## 個別化薬剤治療

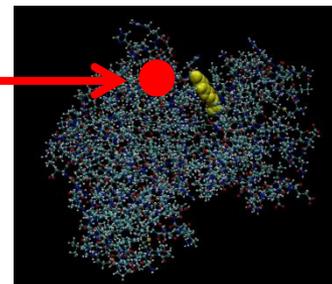
個人個人の遺伝子タイプから  
薬剤反応性や副作用危険性を  
判定する。

既報のガイドライン・研究実績  
に基づいて、遺伝子タイプによ  
る最適な薬剤を選択。

- がん患者の場合、がん細胞の変異により抗がん剤の薬剤耐性が生じる
- 遺伝子タイプと薬剤反応性の分子メカニズムは十分理解されていない

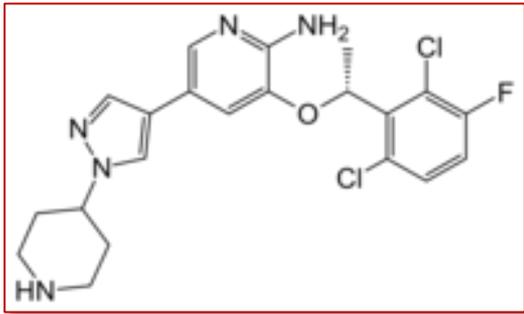


計算機上で  
変異

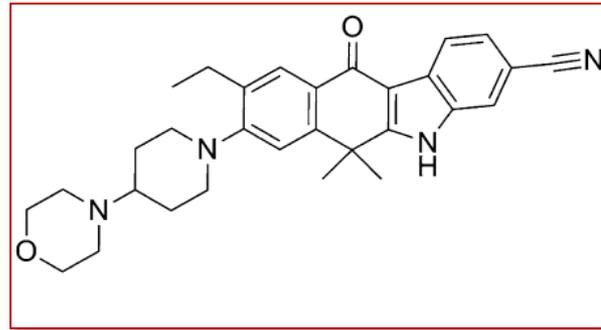


# 非小細胞肺癌治療薬の薬剤耐性

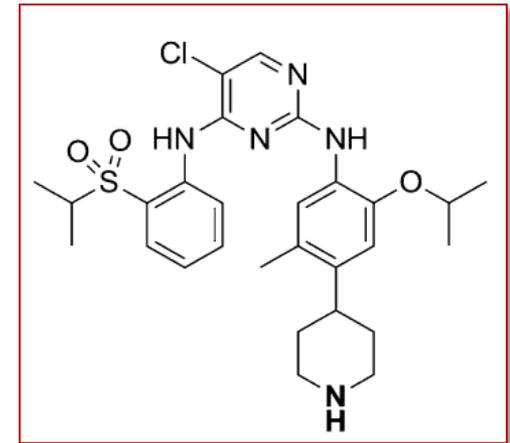
## Crizotinib



## Alectinib



## Ceritinib



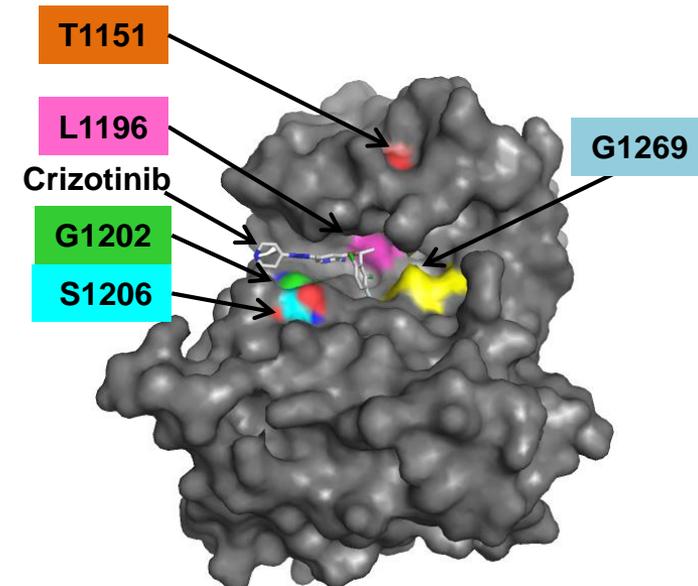
## Crizotinib Resistance



1 year on crizotinib



2 years on crizotinib

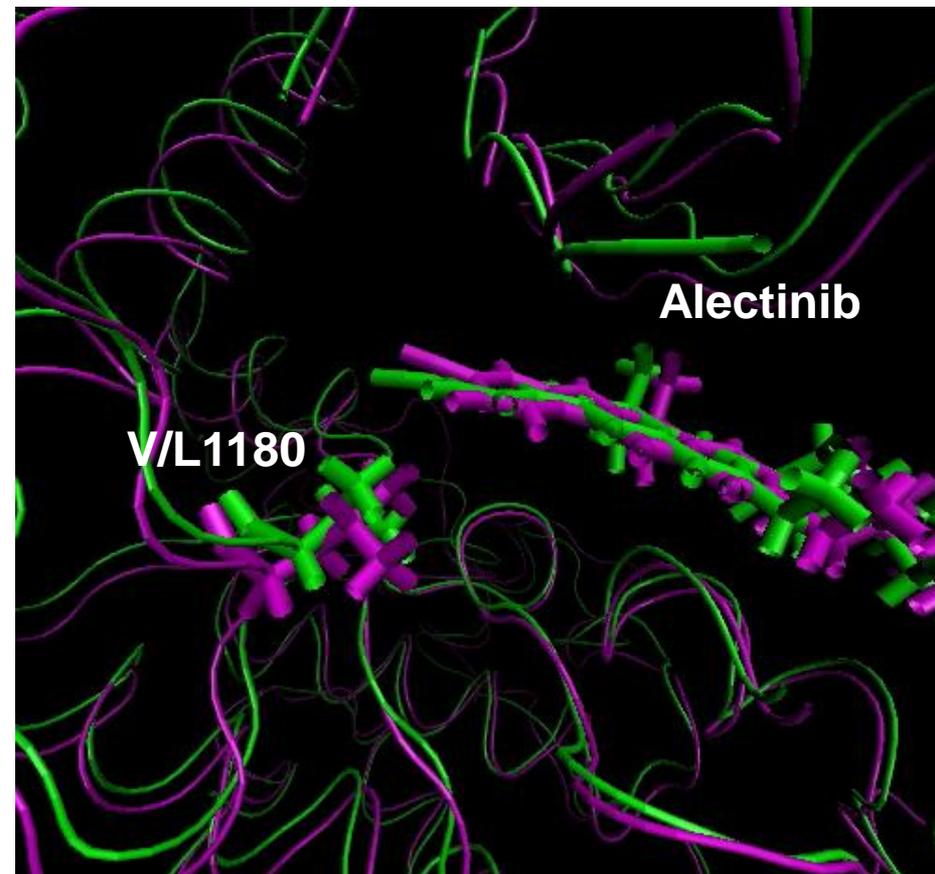
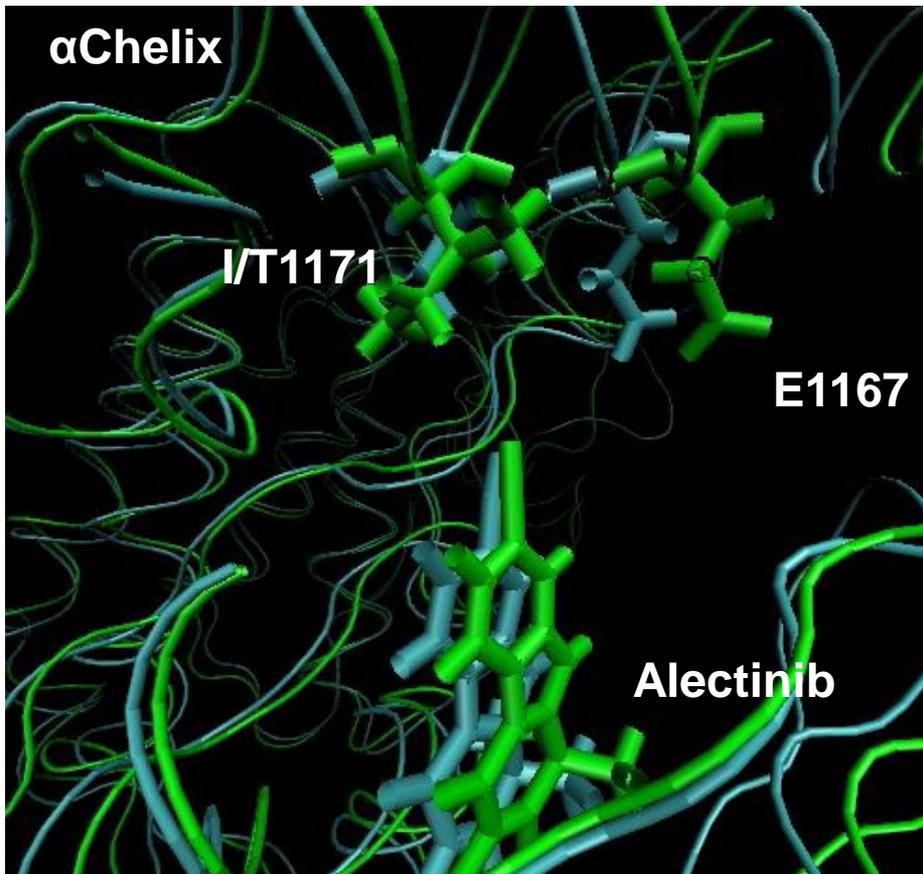


Target Protein: ALK

# AlectinibとALKタンパクとの分子シミュレーション

Wild type vs I1171T mutant

Wild type vs V1180L mutant

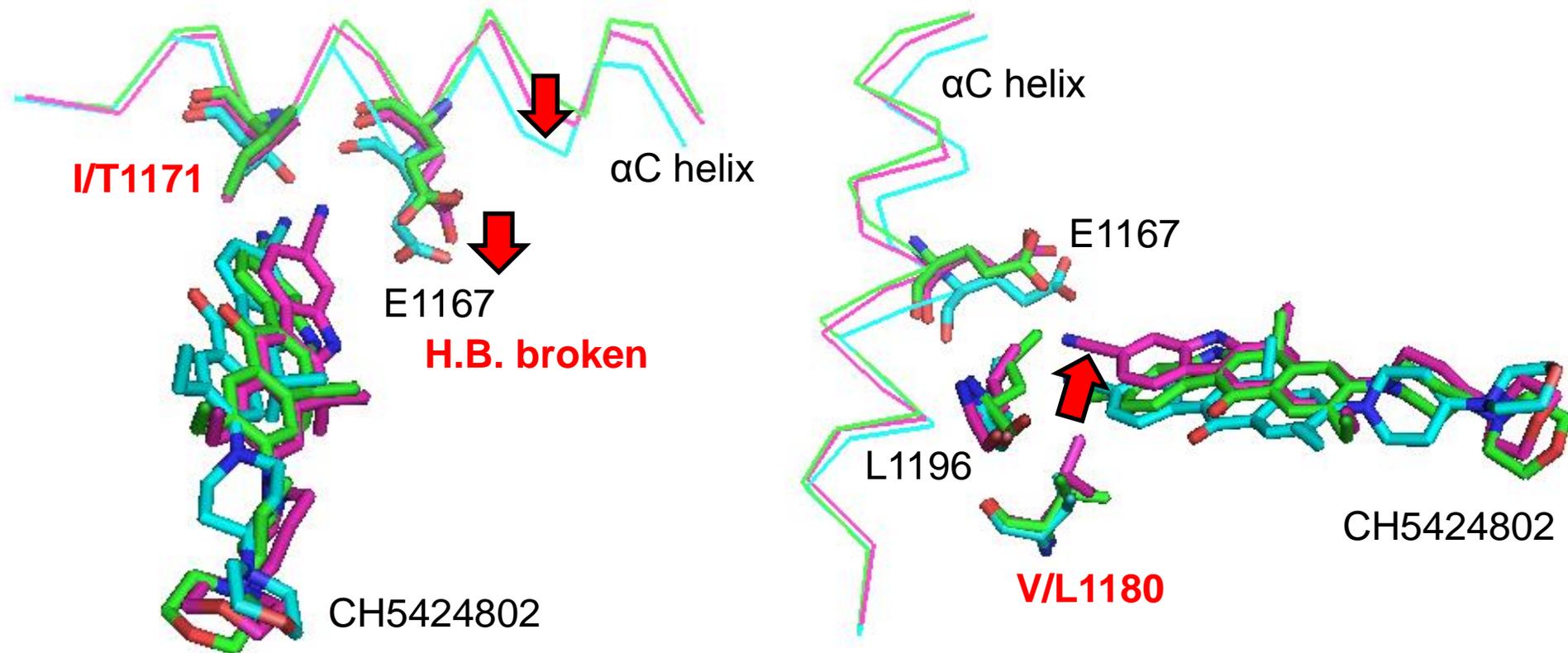


# ALK変異によるAlectinibとの結合親和性低下のメカニズム

## MP-CAFEEによる相互作用エネルギー

	$\Delta G$	coulomb	vdw
wild-CH5424802 (green)	- 77.74	- 60.41	- 17.33
I1171T-CH5424802 (cyan)	- 75.06	- <b>58.04</b>	- 17.02
V1180L-CH5424802(magenta)	- 74.55	- 60.01	- <b>14.54</b>

(kcal/mol)



# 個別化薬剤治療のための分子シミュレーション

## パーソナルゲノム情報



## 薬理ゲノミクス

京大病院がんセンター  
院内でのフロー構築中

## 個別化薬剤治療

個人個人の遺伝子タイプから  
薬剤反応性や副作用危険性を  
判定する。

既報のガイドライン・研究実績  
に基づいて、遺伝子タイプによ  
る最適な薬剤を選択。

- がん患者の場合、がん細胞の変異により抗がん剤の薬剤耐性が生じる
- 遺伝子タイプと薬剤反応性の分子メカニズムは十分理解されていない

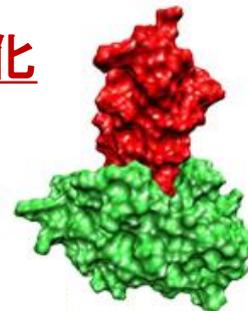
- ・ 薬剤とタンパクとの分子シミュレーションによる薬剤応答性や副作用危険性の予測
- ・ 薬剤耐性や副作用発現の分子メカニズムの解明
- ・ 薬剤耐性を回避する新薬の合理的分子デザイン

# 「京」からポスト「京」へ



ポスト「京」

高分子化



バイオ医薬の合理的設計

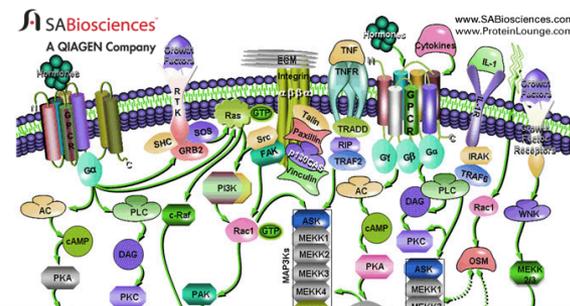
タンパク間相互作用制御

アンサンブルMD



結合自由エネルギーの高速計算  
⇒ 1週間で1万個の計算が可能

システムレベルへの拡張



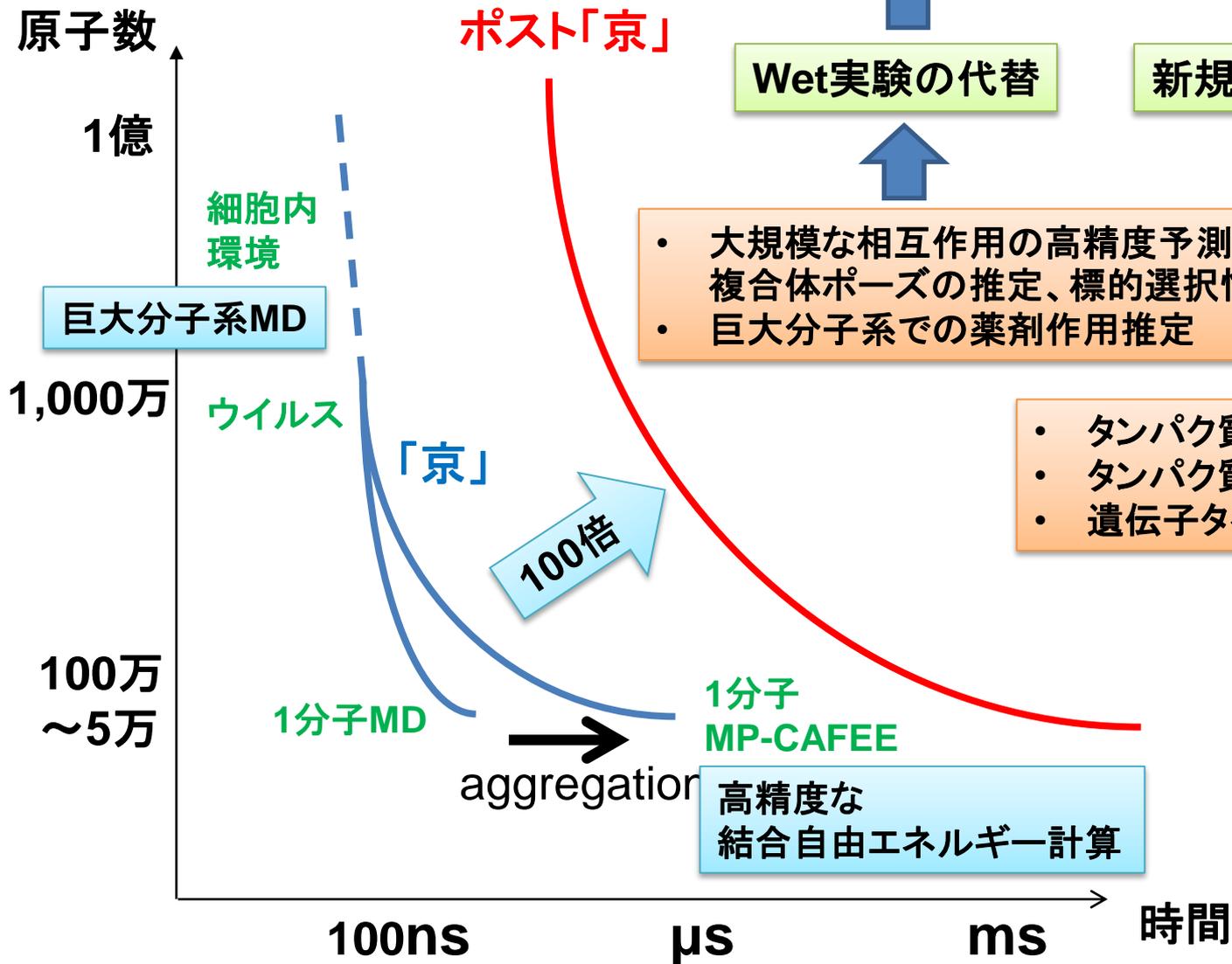
分子レベルから  
ネットワーク、細胞レベルへの橋渡し

長時間(フルタイム)MD



タンパク質の動きを制御する薬剤設計、時間を制御する薬剤設計  
⇒ ミリ秒レベルの現象が計算可能

# 「京」からポスト「京」へ



開発プロセスの効率化

薬のつくり方を革新する

Wet実験の代替

新規創薬ターゲットの創出

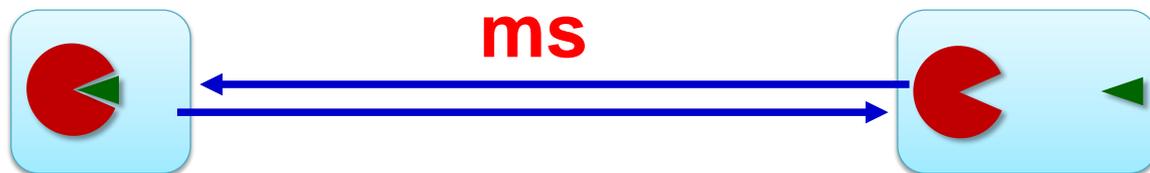
- 大規模な相互作用の高精度予測(正確な複合体ポーズの推定、標的選択性予測)
- 巨大分子系での薬剤作用推定

- タンパク質の動的機能制御
- タンパク質間相互作用の制御
- 遺伝子タイプを考慮した分子設計

高精度な結合自由エネルギー計算

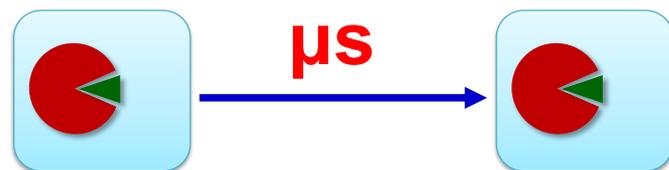
# ポスト「京」時代のシミュレーションによるデータ爆発

実際の分子動力学計算では2psごとの構造を保存しつづけている



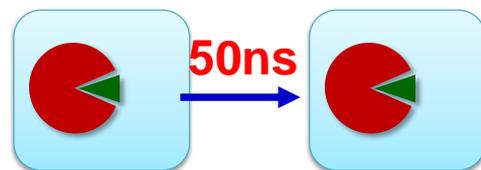
ポスト京でのデータ規模

~5億構造⇒約100TB



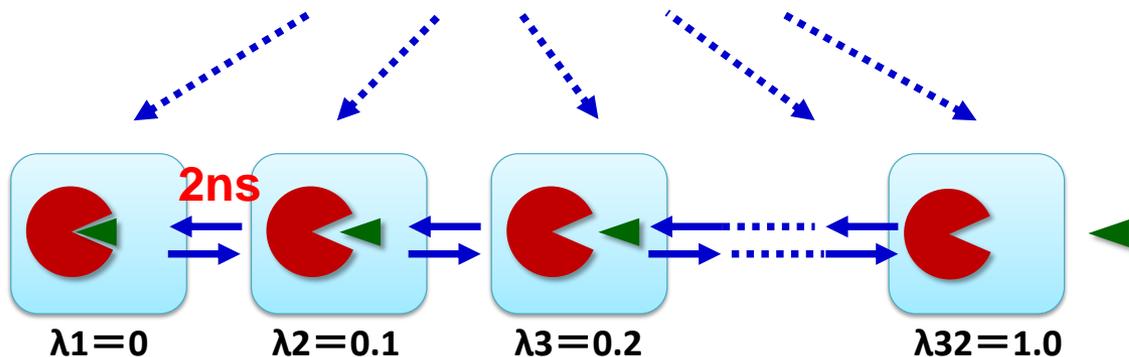
「京」でのデータ規模

~50万構造⇒約100GB



「京」でも結合親和性を算出できる近似計算法

平衡化計算：  
5初速度 × 50 ns MDs



非平衡計算：  
6初速度 × 32条件 × 2 ns MDs

タンパク質-化合物の  
1ペアあたり約50GB

# Acknowledgements

- 中外製薬株式会社
- 三井情報株式会社
- CREST「ビッグデータ応用領域」
- JST COIプログラム
- 研究教育拠点(COE)形成推進事業研究助成
- ポスト「京」重点課題1
- KBDDコンソーシアムメンバー
- 理研・生命システム研究センター(QBiC)
- 理研・計算科学研究機構(AICS)
- 大阪大学・サイバーメディアセンター
- 高度情報科学技術研究機構(RIST)
- 先端医療振興財団

## Special thanks

京大病院がんセンター

武藤 学教授

がん研究会がん化学療法センター

片山量平先生

先端医療振興財団

井村裕夫理事

鍋島陽一センター長

理研・AICS

米澤明憲副機構長

泰地真弘人先生

江口至洋先生

荒木望嗣先生

京大医・臨床システム腫瘍学講座

浜中雅俊先生

種石慶先生

大阪大学情報科学研究科

下條真司教授

伊達進先生

寺前勇希先生

神戸大学工学部

中津井雅彦先生

KBDDコンソーシアムの皆様