

# 「京」コンピュータで薬をつくる

京都大学大学院 薬学研究科  
奥野 恭史

# 「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

コンピュータ創薬の中心拠点形成による我が国の創薬イノベーションを牽引



**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 京都大学薬学研究科 奥野恭史

**製薬企業(11社)：**

アスピオファーマ, エーザイ, 小野薬品工業, 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, 科研製薬, 杏林製薬

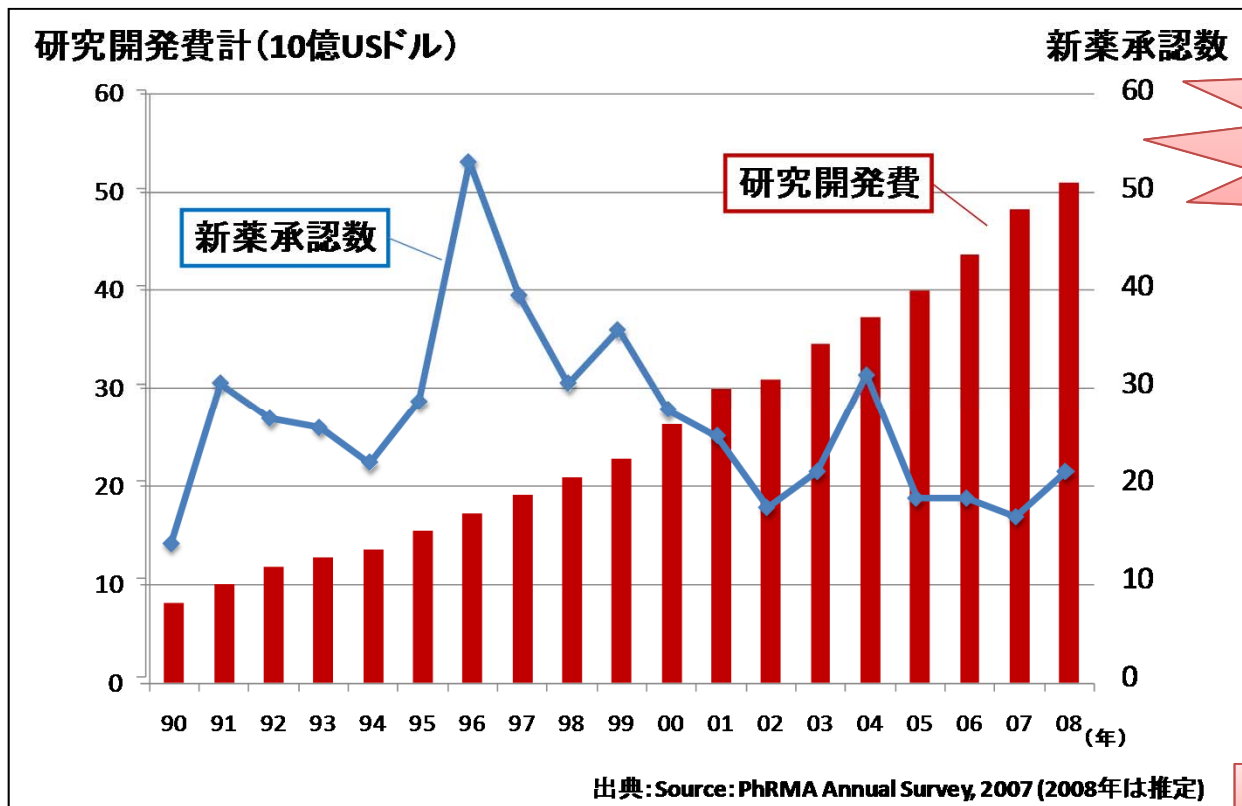
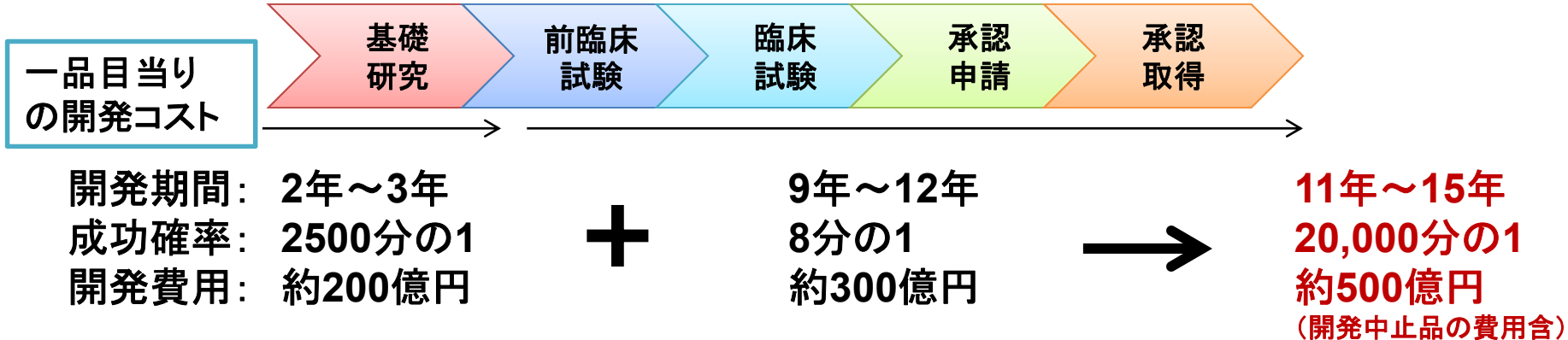
**IT企業(2社)：**

(株)京都コンステラ・テクノロジーズ, 三井情報(株)

**大学等：**

京都大学大学院薬学研究科, (独)産業技術総合研究所, 理研HPCI企画調整グループ

# 製薬業界に横たわる深刻な問題



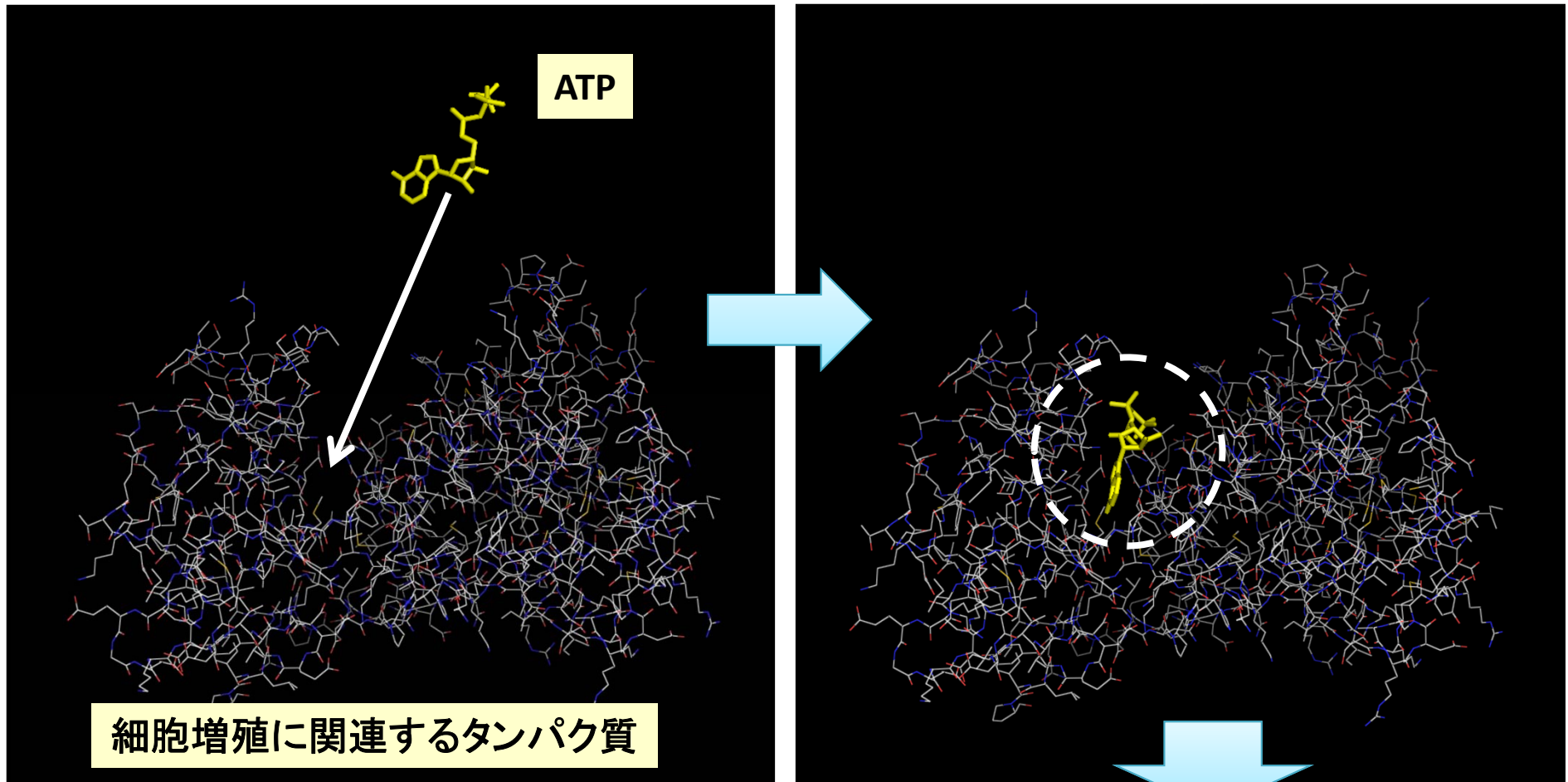
高騰し続ける開発費

- 製薬企業の経営圧迫
- 希少疾患薬の枯渇
- 医療費の高騰

医薬品開発の効率化が急務

計算で業界の課題に挑戦

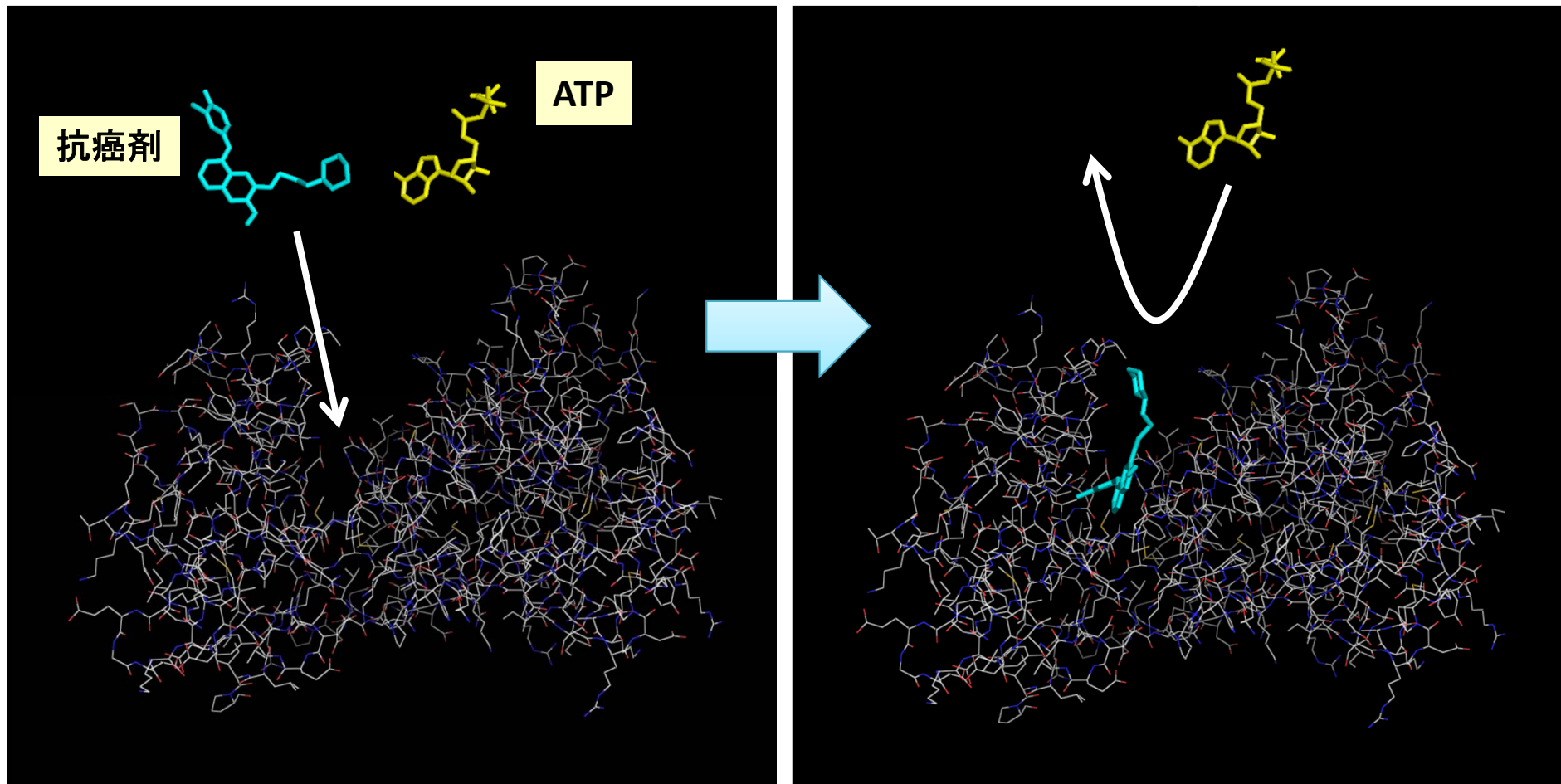
# 薬づくりの考え方



細胞増殖関連タンパク質に生体物質ATPが結合することにより、細胞が増殖する

癌は細胞が異常増殖する病気

# 薬づくりの考え方



細胞増殖関連タンパク質とATPとの結合をブロックする化合物があると、細胞増殖が止まる ⇒ 抗癌剤

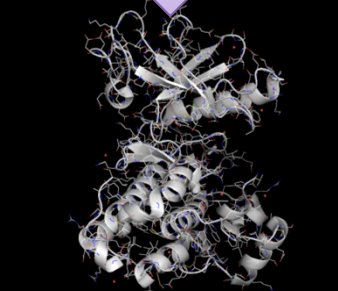
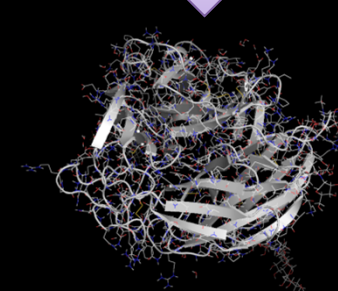
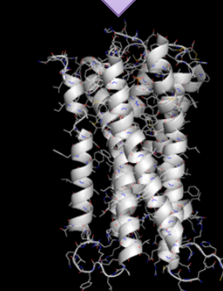
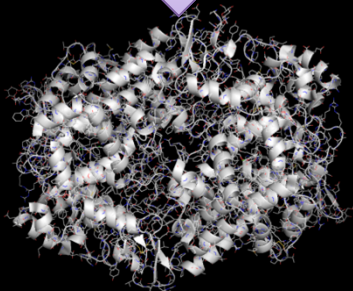
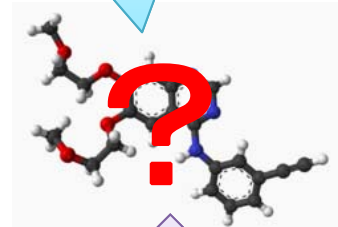
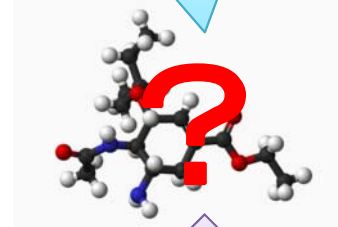
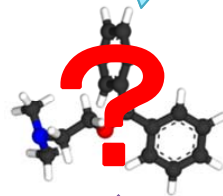
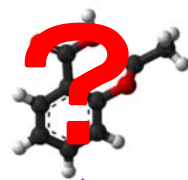
創薬とは、病気の原因タンパク質を見つけ出し、そのタンパク質に結合する(機能制御する)新規化合物を創ることと言える

# 何故、医薬品開発は難しいのか？

化合物の種類：  $10^{60}$ 以上

候補化合物の化学合成と生物活性評価などの実験を  
何度も繰り返す試行錯誤が必要

タンパク質と化合物の  
膨大な組合せ



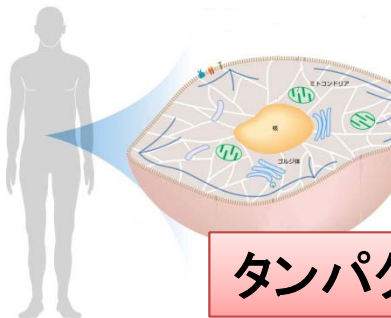
解熱鎮痛剤

花粉症の薬

抗インフルエンザ薬

抗癌剤

タンパク質の種類： 10万以上



# 何故、医薬品開発費は高騰するのか？

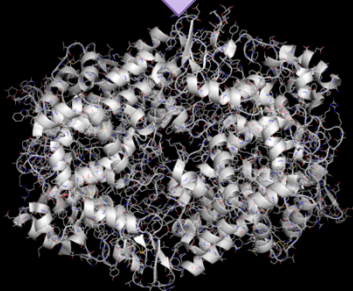
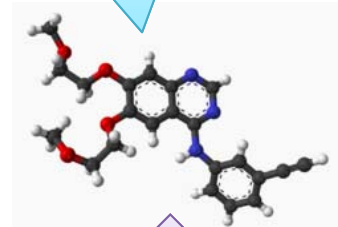
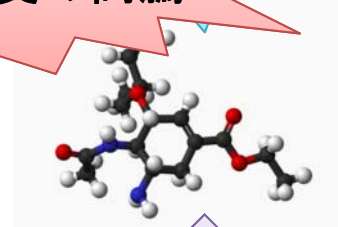
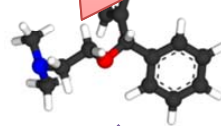
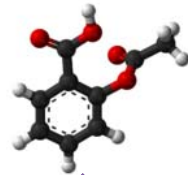
化合物の種類：  $10^{60}$ 以上

全ての化合物の薬効を実験で確かめるのは不可能

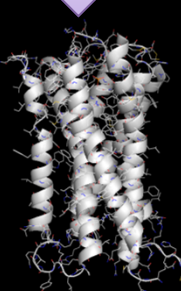
膨大な候補化合物から医薬品を探し当てることは、勘・経験・運任せ  
⇒多くの無駄な実験と失敗を伴う

研究開発費の高騰

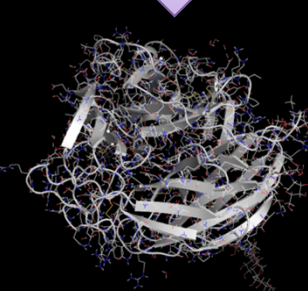
タンパク質と化合物の膨大な組合せ



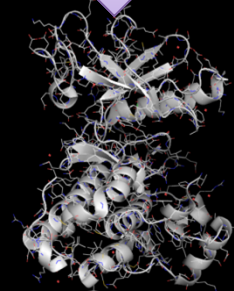
解熱鎮痛剤



花粉症の薬

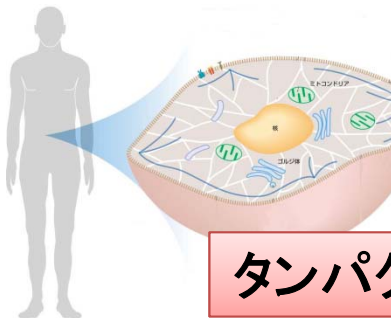


抗インフルエンザ薬



抗癌剤

タンパク質の種類： 10万以上



# コンピュータ創薬への期待

化合物の種類:  $10^{60}$ 以上

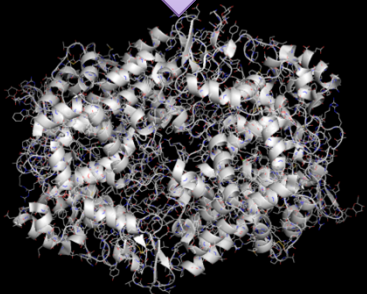
全ての化合物の薬効を実験で確かめるのは不可能

膨大な候補化合物から医薬品を探し当てることは、勘・経験・運任せ  
⇒多くの無駄な実験と失敗を伴う

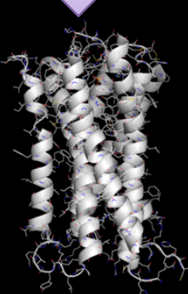
①結合するかどうかを予測

実験の代わりに、コンピュータ上で結合をシミュレーションする  
⇒開発費用の削減への期待

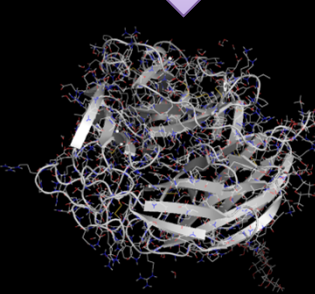
②結合の強さを予測



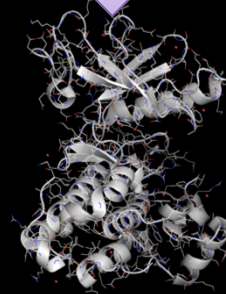
解熱鎮痛剤



花粉症の薬

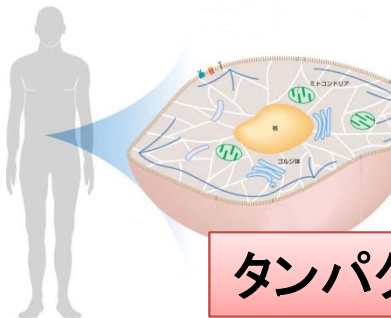


抗インフルエンザ薬



抗癌剤

タンパク質の種類: 10万以上





# コンピュータ創薬の現状と課題

化合物の種類:  $10^{60}$ 以上



全ての化合物の薬効を実験で確かめるのは不可能

実験の代わりに、コンピューター上で結合をシミュレーションする  
⇒開発費用の削減への期待

①結合するかどうかを予測

計算時間がかかるため、全ての化合物を計算するのは不可能

②結合の強さを予測

予測精度が悪すぎる: 正答率: 5%程度

計算機パワーの問題

現状のコンピュータ創薬技術では、  
医薬品開発効率の抜本的な解決になっていない

# 「京」による医薬品開発効率の抜本的解決

化合物の種類:  $10^{60}$ 以上

全ての化合物の薬効を実験で確かめるのは不可能

実験の代わりに、コンピューター上で結合をシミュレーションする  
⇒開発費用の削減への期待

①結合するかどうかを予測

計算時間がかかるため、全ての化合物を計算するのは不可能

化合物とタンパク質の結合予測の超高速化

②結合の強さを予測

予測精度が悪すぎる: 正答率: 5%程度

精密な結合シミュレーションによる  
予測精度の劇的向上(目標正答率70%)

コンピュータ創薬の  
ブレークスルー

「京」の圧倒的な計算機パワーで  
医薬品開発効率の抜本的な解決を目指す

# 「京」による医薬品開発効率の抜本的解決

化合物の種類:  $10^{60}$ 以上



全ての化合物の薬効を実験で確かめるのは不可能

実験の代わりに、コンピューター上で結合をシミュレーションする  
⇒開発費用の削減への期待

## ①結合するかどうかを予測

計算時間がかかるため、全ての化合物を計算するのは不可能

化合物とタンパク質の結合予測の超高速化

## ②結合の強さを予測

予測精度が悪すぎる: 正答率: 5%程度

精密な結合シミュレーションによる  
予測精度の劇的向上(目標正答率70%)

創薬の  
ワークスルー

「京」の圧倒的な計算機パワーで  
医薬品開発効率の抜本的な解決を目指す

# 結合するかどうかを予測:「京」によるタンパク質と化合物の結合予測

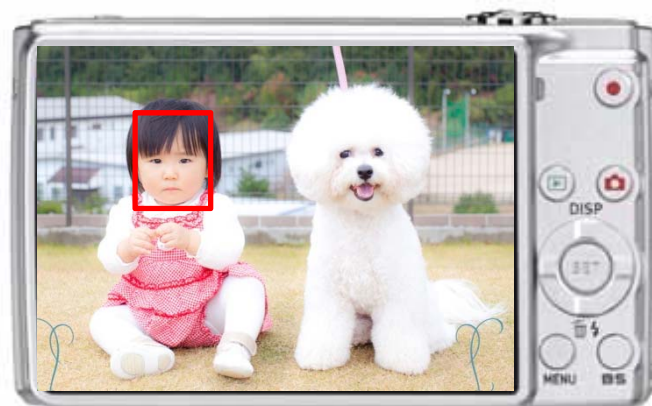
膨大なタンパク質と化合物の組合せを高速計算するために、パターン認識技術を適用

大量の人の顔画像を学習



顔パターンの  
統計ルール化

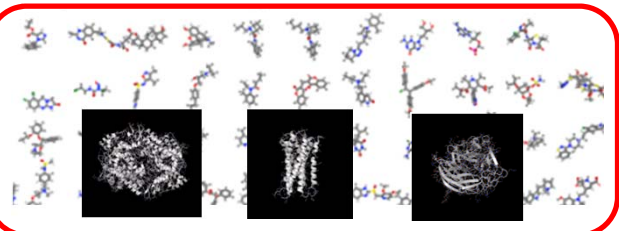
人の顔を自動認識



予測

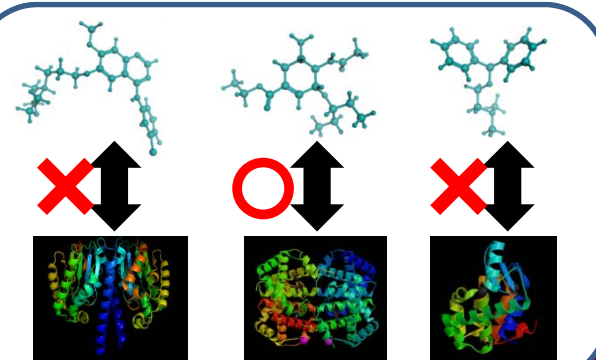


大量のタンパク質と化合物の結合データを学習



結合パターンの  
統計ルール化

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



「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

# 結合するかどうかを予測:「京」によるタンパク質と化合物の結合予測

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

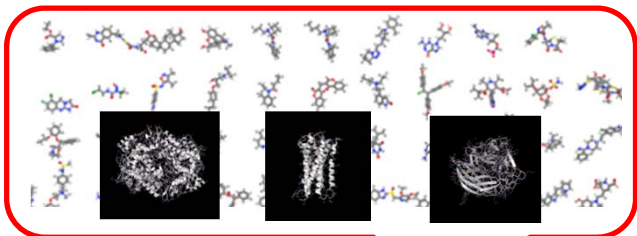


世界最大規模(189.3億ペア)の結合予測  
631種の疾患原因タンパク質と  
3000万種の化合物の全組合せ

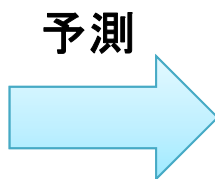
論文等で結合することが分かっている  
タンパク質と化合物の結合ペア:12万ペア



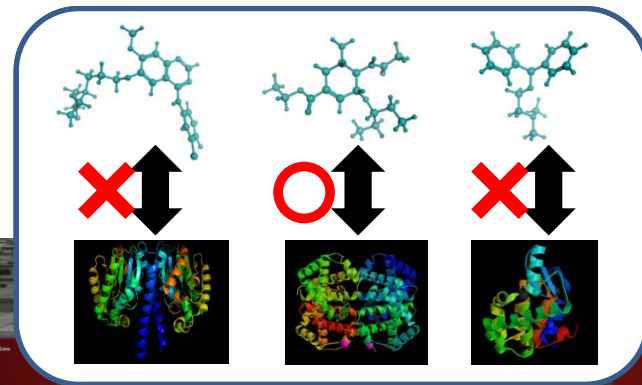
大量のタンパク質と化合物の結合データを学習



結合パターンの  
統計ルール化



病気の原因タンパク質に  
結合する化合物を「京」で認識



「京」により  
超高速予測を実現

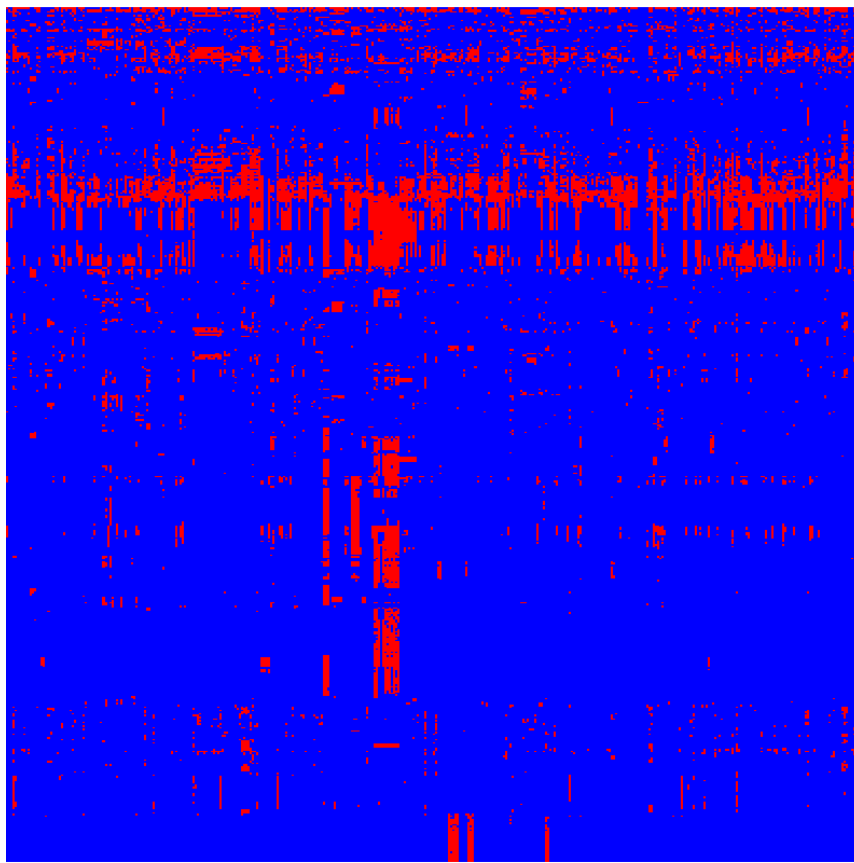


膨大なタンパク質と  
化合物の組合せに対応

# 予測の実例: 「京」によるタンパク質と化合物の結合予測

タンパク質-化合物の全組合せ(189.3億ペア)を計算するのに、「京」をフルに使用したら、5時間45分で計算が終了する。\*汎用計算機(16ノード使用)では約2年かかる。

「京」が予測した結合パターン(予測結果)



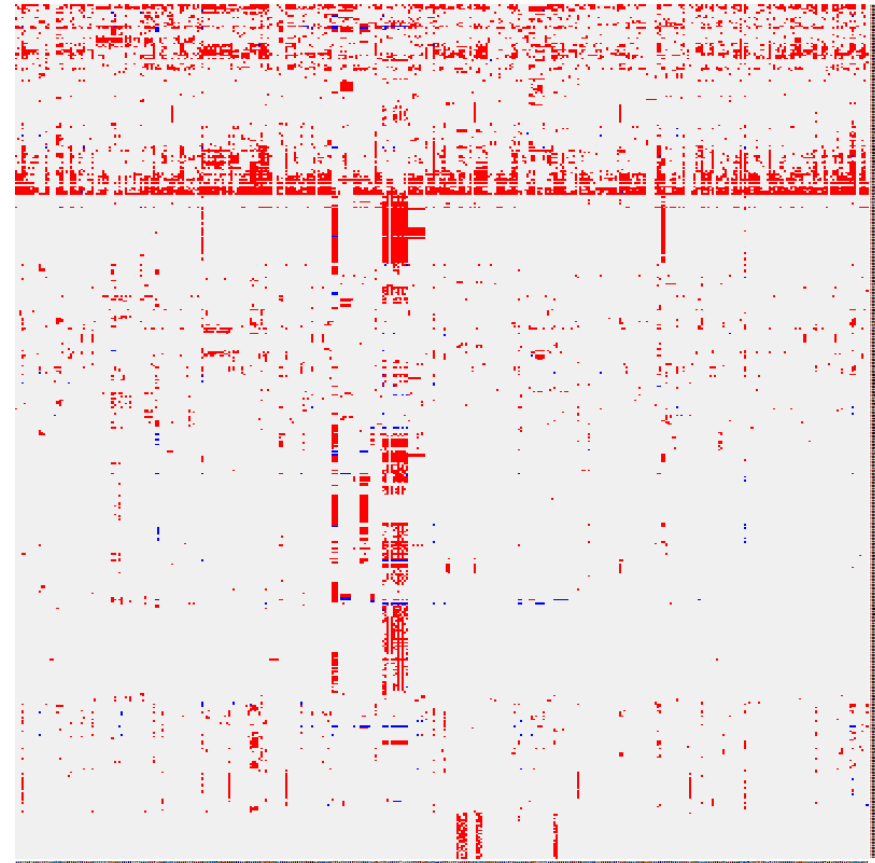
3000万種 ←

化合物(500種)

タンパク(388種) ⇒ 631種

赤: 結合する(スコア > 0.8)  
青: 結合しない(スコア < 0.8)

実際の結合パターン(実験結果)



化合物(500種)

タンパク(388種)

赤: 結合する(<30 $\mu$ M)、青: 結合しない(>80 $\mu$ M)、灰: 実験データ無し

# 「京」による医薬品開発効率の抜本的解決

化合物の種類:  $10^{60}$ 以上

全ての化合物の薬効を実験で確かめるのは不可能

実験の代わりに、コンピューター上で結合をシミュレーションする  
⇒開発費用の削減への期待

①結合するかどうかを予測

計算時間がかかるため、全ての化合物を計算するのは不可能

化合物とタンパク質の結合予測の超高速化

コンピュータ創薬の  
ブレークスルー

②結合の強さを予測

予測精度が悪すぎる: 正答率: 5%程度

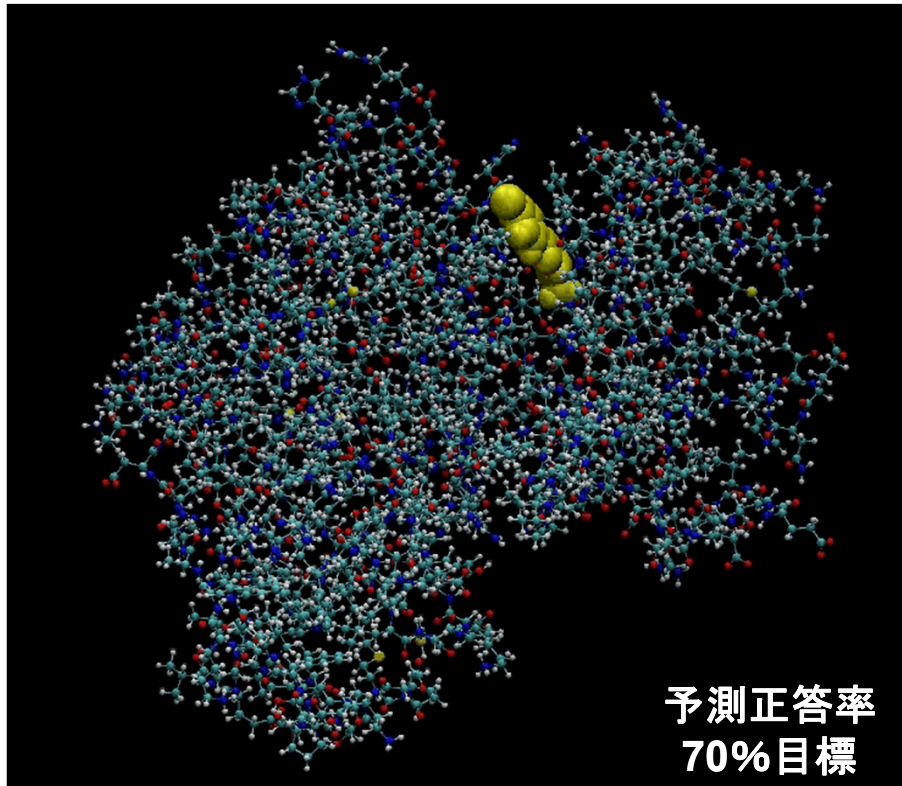
精密な結合シミュレーションによる  
予測精度の劇的向上(目標正答率70%)

「京」の圧倒的な計算機パワーで  
医薬品開発効率の抜本的な解決を目指す

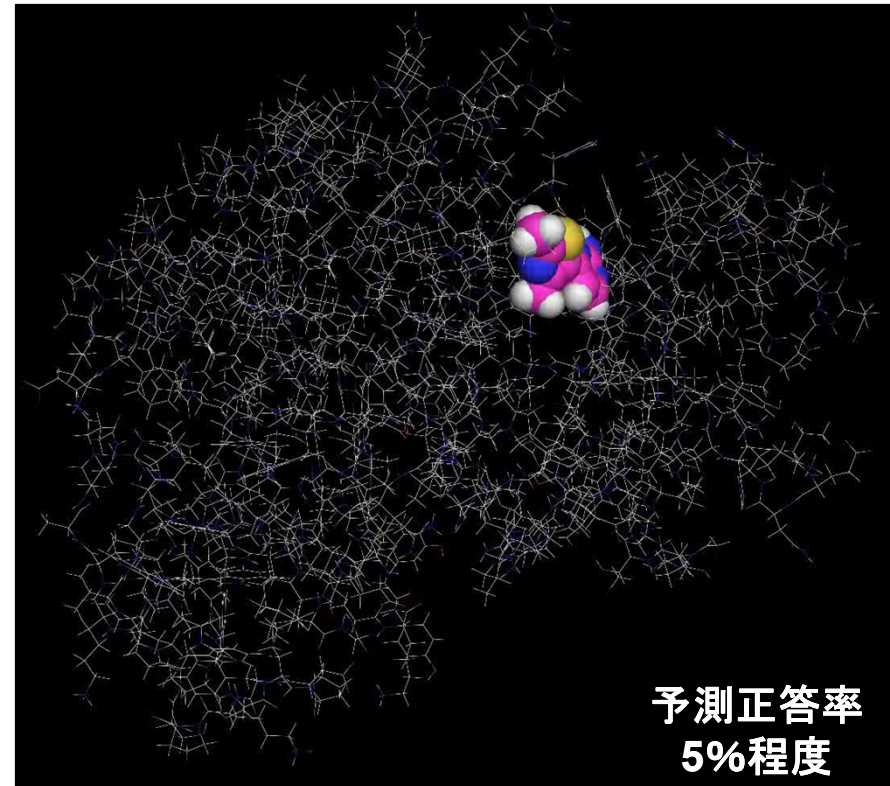
# 「京」によるタンパク質と化合物の結合の強さの予測

病気の原因タンパク質にだけ強く結合する化合物が、医薬品として適していると考えられるため、タンパク質と化合物の強さを求めることは重要。

## 「京」による結合シミュレーション



## 従来型の結合シミュレーション

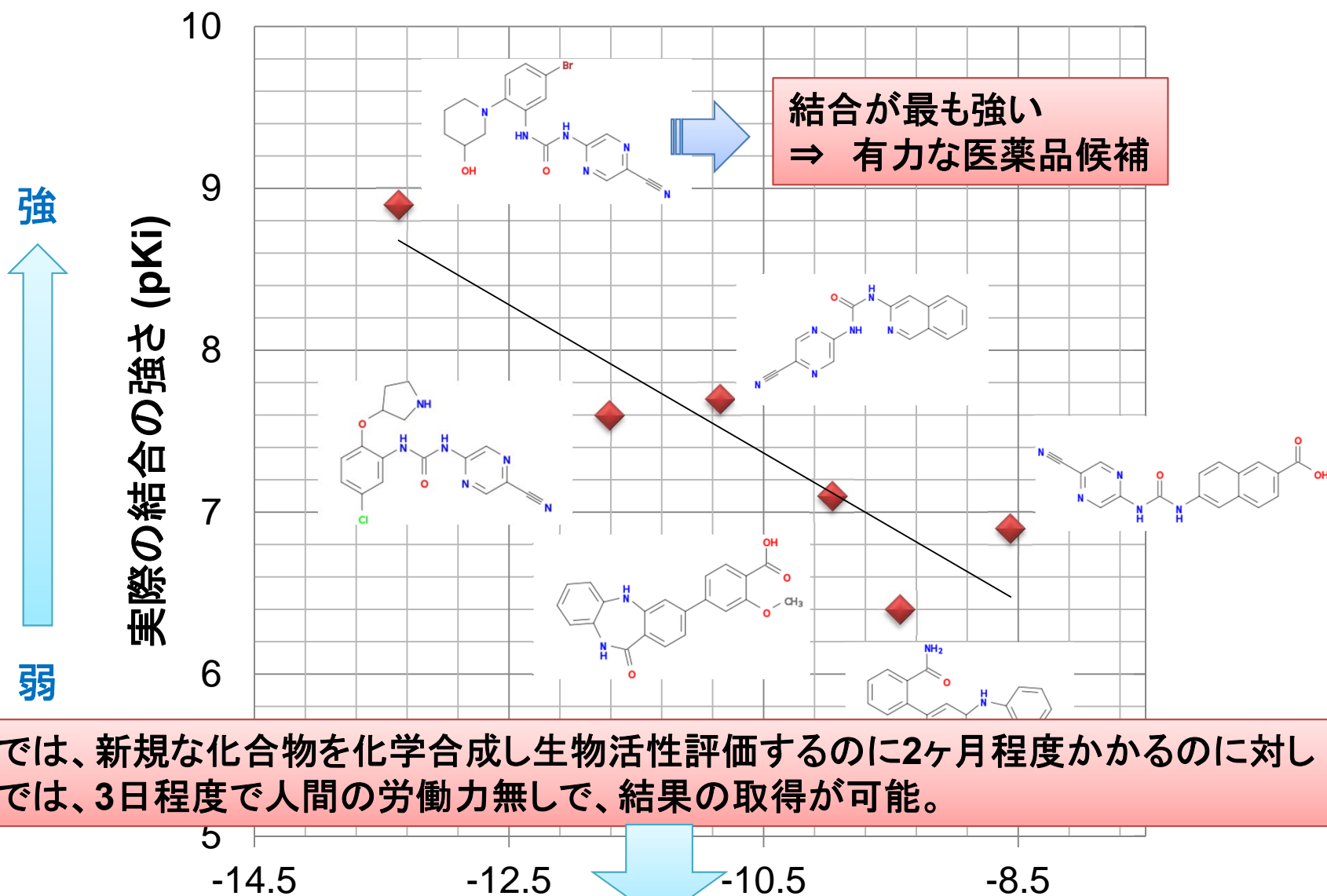


結合の強さを精密に計算するには、分子の動きや溶媒(水分子)も含めたシミュレーションが必要。これまでは計算機パワーの問題で非現実だった。

15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



# 予測の実例:「京」によるCHK1タンパク質と化合物の結合の強さの予測



数多くの化合物を計算するためには、もっと高速な計算機が必要  
⇒ ポスト「京」への強い期待！

# 「京」がもたらす製薬業界の意識改革:オープンイノベーション

## KBDDコンソーシアム: “K” supercomputer-based drug discovery consortium

現場ニーズに即した計算プロセスの評価  
研究成果の実用化・医薬製品化

### 【製薬企業】

アスピオファーマ、エーザイ、小野薬品工業、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、科研製薬、杏林製薬

スパコンの産業界解放は  
画期的！

### 【IT企業】

三井情報  
京都コンステラテクノロジーズ

現場利用可能な計算基盤の構築

### 【学・官】

京都大学  
理化学研究所AICS・RIST  
産業技術総合研究所  
NPOバイオグリッドセンター関西

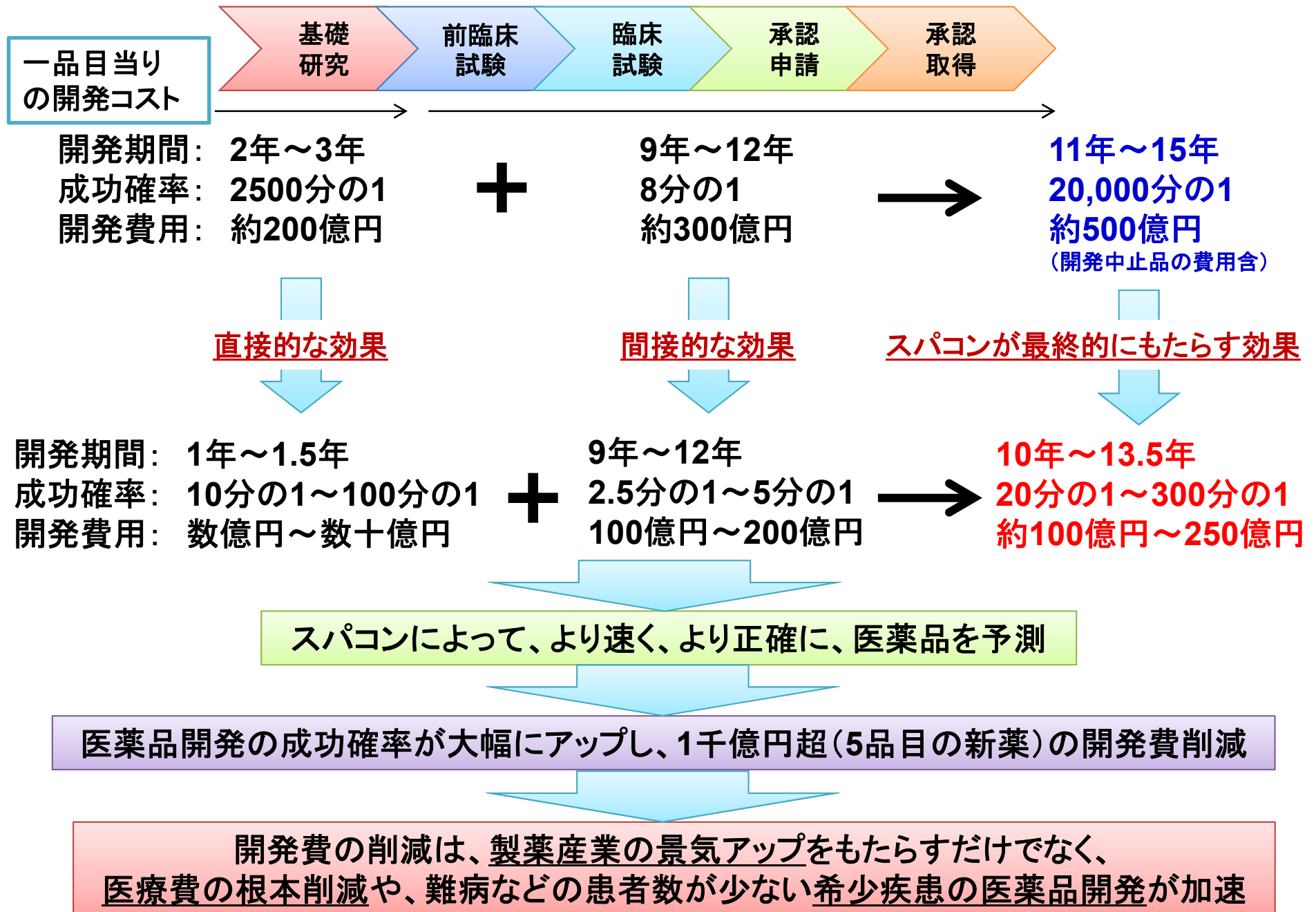
技術的ノウハウの指導・提供  
勉強会・講習会の実施



「脱」秘密主義:オープンイノベーション  
今の汎用計算機は15年前のスパコン性能であること  
から、「京」を利用できる我々は15年先の創薬計算  
技術を、今、手にしていることなる。

スパコン開発は継続に意味がある

# まとめ: スパコン(「京」/ポスト「京」)による創薬イノベーション



# 謝辞

- KBDDコンソーシアムメンバー
- 理研・計算科学研究機構
- 高度情報科学技術研究機構
- 最先端次世代研究開発支援プログラム
- NEDO若手研究 Grant

## JSTサイエンスチャンネルにて放送中

<http://sc-smn.jst.go.jp/playprg/index/6864>

The screenshot shows the JST Science Channel website interface. At the top, there is a navigation bar with the channel logo and the text 'サイエンスチャンネル (SCIENCE CHANNEL)'. Below this, there are several tabs for navigation, including 'HOME', 'サイエンスチャンネルとは', and 'ダウンロード・DVD・放送用メディアご希望の方へ'. A search bar is located in the middle of the page, and there are links for 'ログイン' and '新規ユーザー登録'. The main content area features a video player with a thumbnail image showing a server room. The video title is 'スパコン「京」によるIT創業 100億通りの創業データベース開発'. To the right of the video player, there is a '概要' (Summary) section and a '出演者名・所属機関名および協力機関名' (Cast and Affiliated Institutions) section. The summary text describes a research project that started in autumn 2012, aiming to improve the efficiency of drug development using supercomputers. The cast list includes Prof. Akio Ono from Kyoto University and other researchers from various institutions.