

スパコン「京」が拓くコンピュータ創薬の未来

京都大学 大学院薬学研究科
先端医療振興財団 先端医療センター研究所
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



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)：

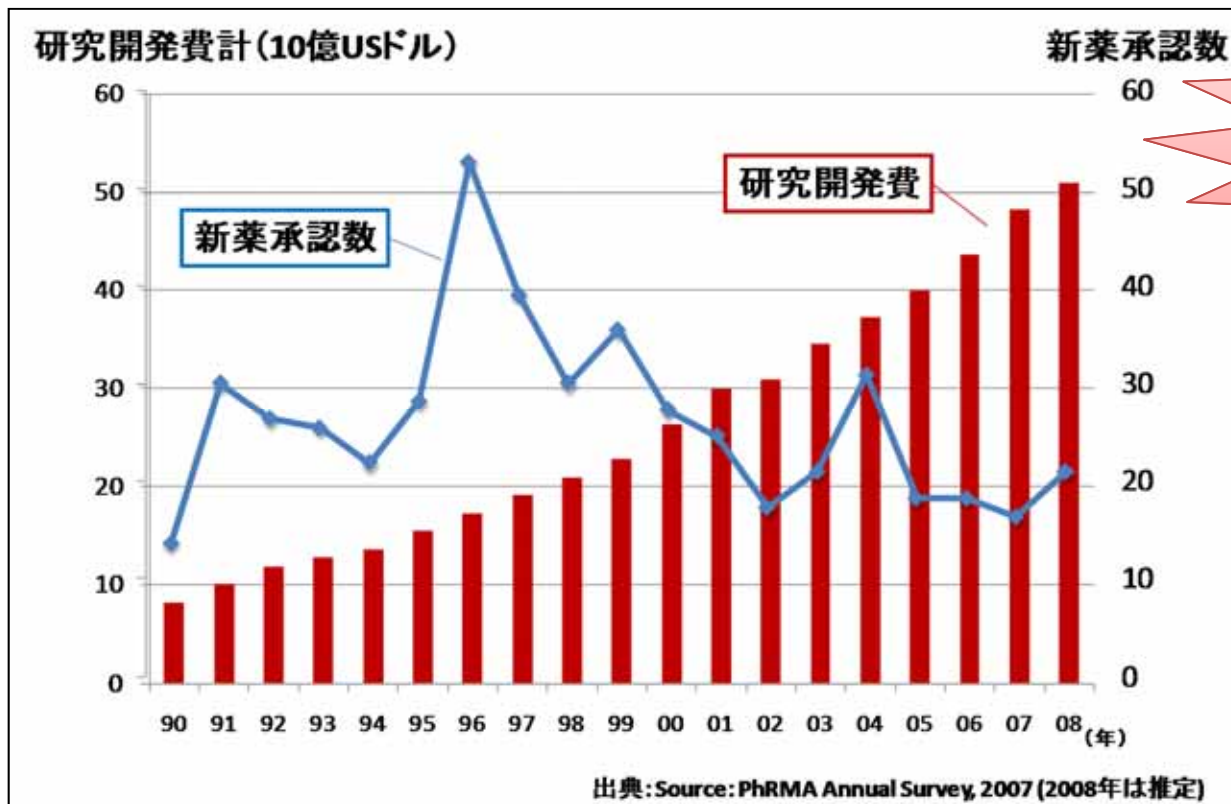
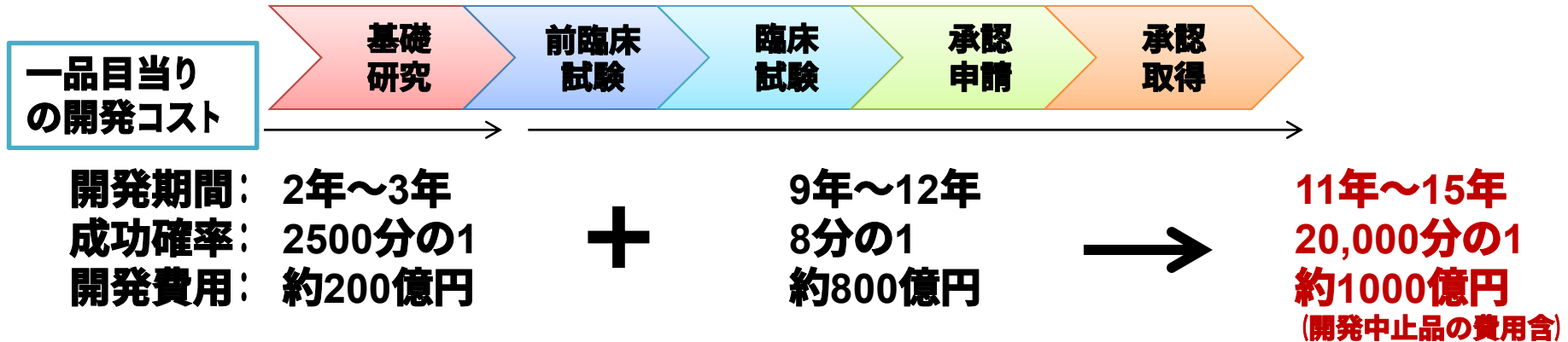
アスピオファーマ, エーザイ, 小野薬品工業, キッセイ薬品工業, 参天製薬,
塩野義製薬, 大日本住友製薬, 田辺三菱製薬, 日本新薬, 科研製薬, 杏林製薬

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

- ・ コンピュータ創薬の根本課題に挑戦
- ・ 製薬会社による現場利用に耐えうる計算フロー(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

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

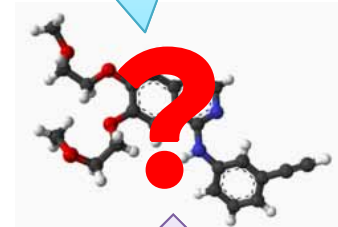
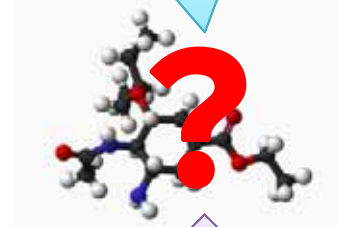
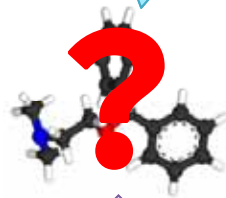
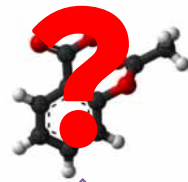
計算で業界の課題に挑戦

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

タンパク質と化合物の
膨大な組合せ

研究開発費の高騰

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、コンピュータ上で結合をシミュレーションする
⇒開発コストの削減への期待

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

Docking計算法

「京」によるコンピュータ創薬の根本課題への挑戦

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
コンピュータ創薬の根本課題の解決を目指す

「京」によるコンピュータ創薬の根本課題への挑戦

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

CGBVS法

②結合の強さを予測

予測精度が悪すぎる: 正答率: 5%程度

Docking計算法

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%)

MP-CAFEE法

「京」によるコンピュータ創薬の根本課題への挑戦

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

CGBVS法

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%)

Docking計算法

MP-CAFEE法

結合するかどうかを予測: CGBVS法 (Chemical Genomics-based Virtual Screening法)

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



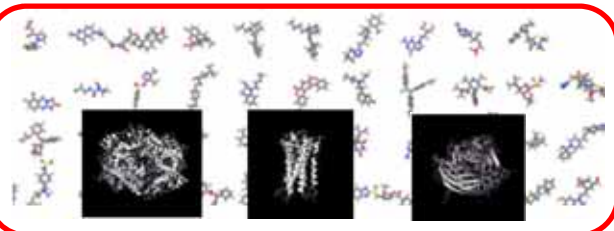
顔パターンの
統計ルール化

予測

人の顔を自動認識



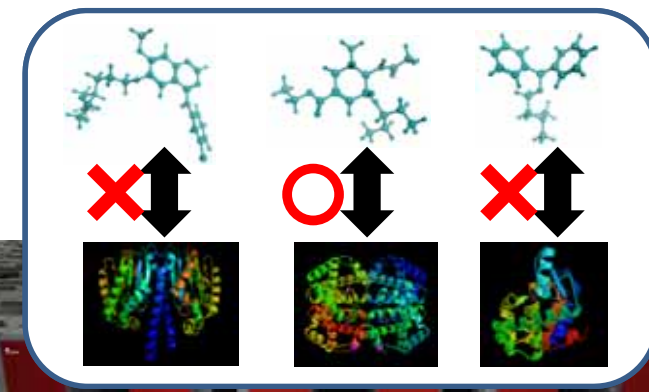
大量のタンパク質と化合物の結合データを学習



結合パターンの
統計ルール化

予測

病気の原因タンパク質に
結合する化合物を「京」で認識

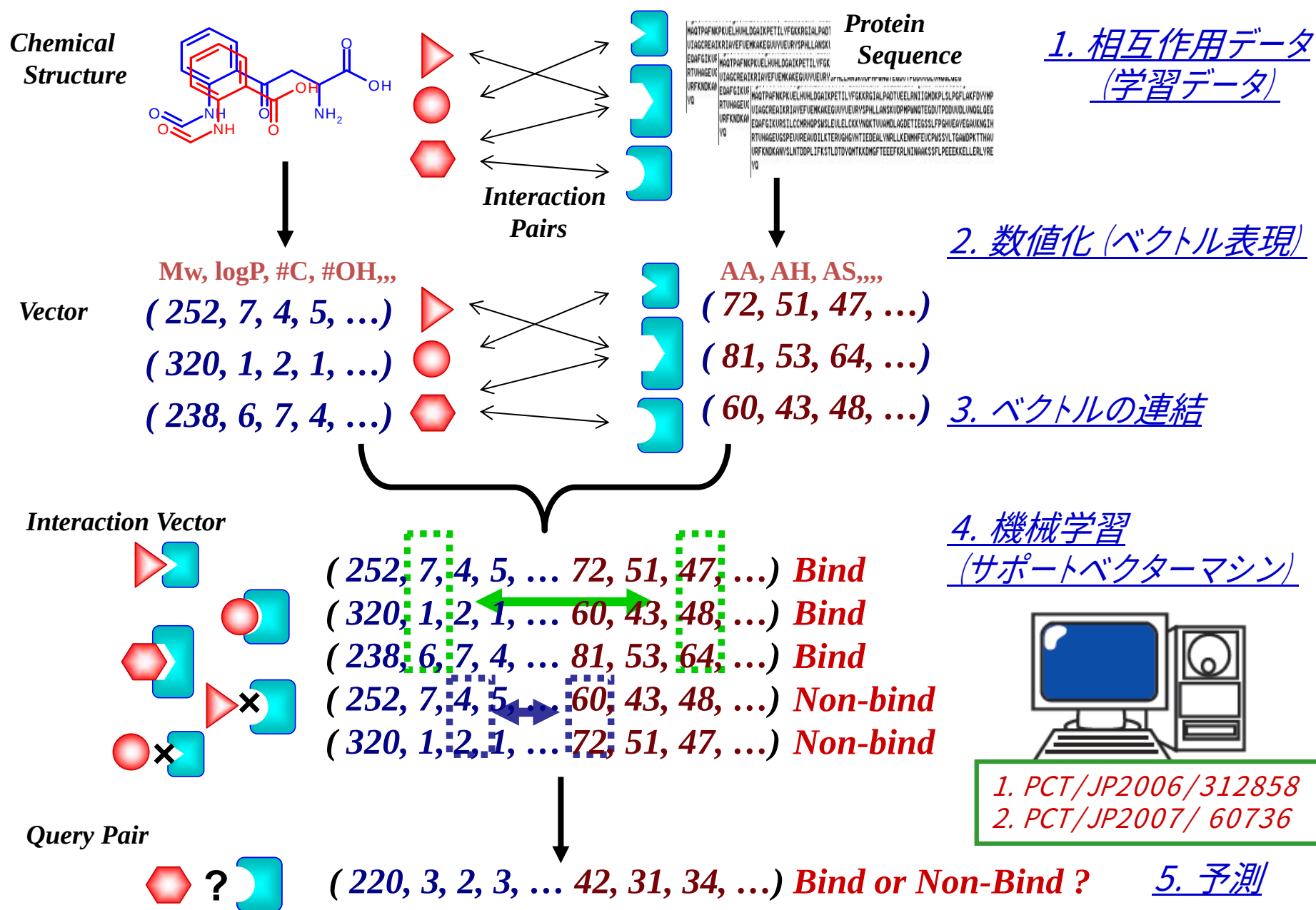


「京」により
超高速予測を実現

膨大なタンパク質と
化合物の組合せに対応

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



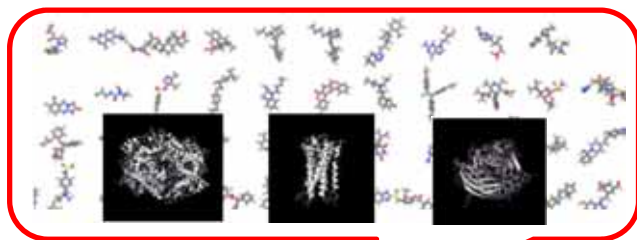
CGBVS法による世界最大規模のタンパク質－化合物結合予測

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種の疾患原因タンパク質と
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を学習



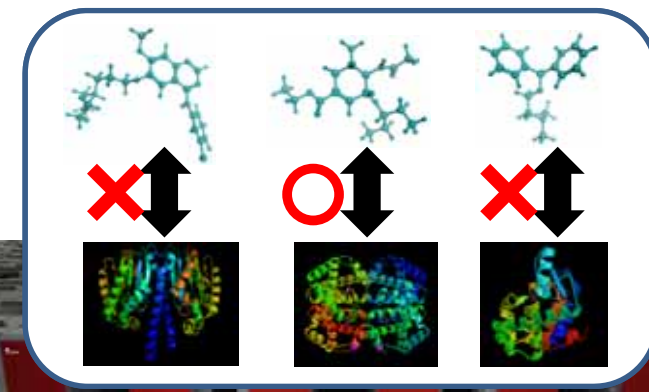
結合パターンの
統計ルール化



予測



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

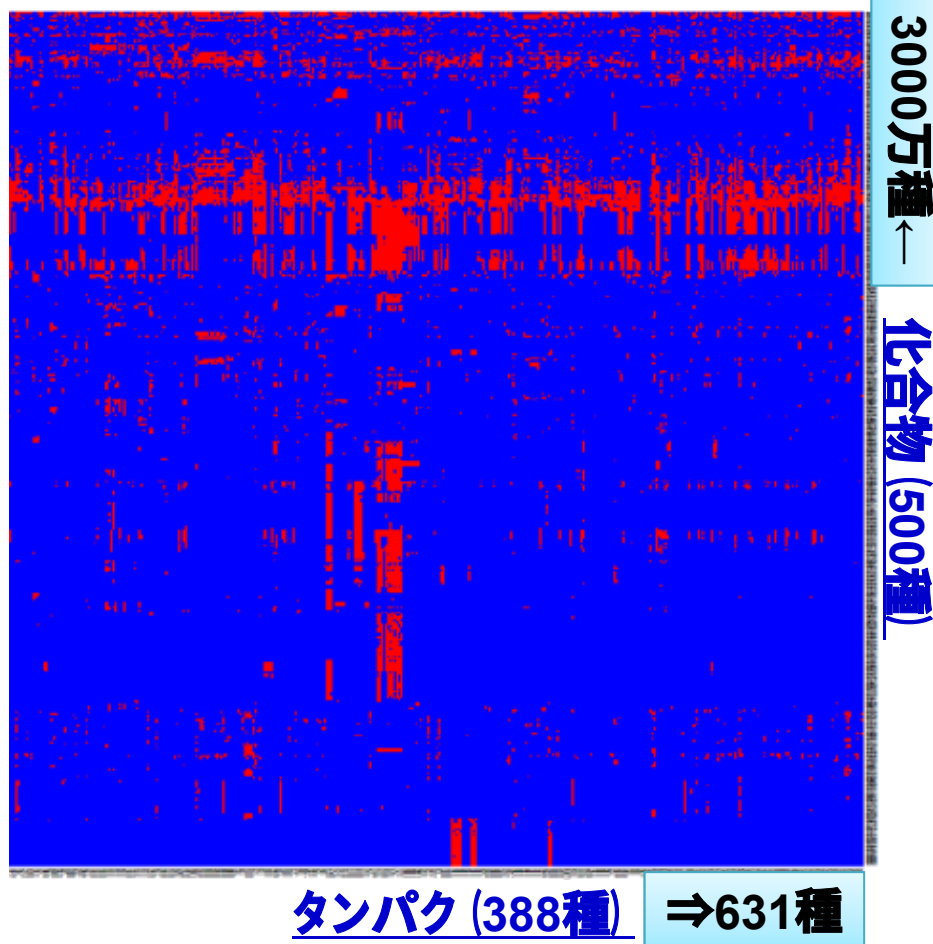
膨大なタンパク質と
化合物の組合せに対応



「京」によるCGBVS法による予測の実例

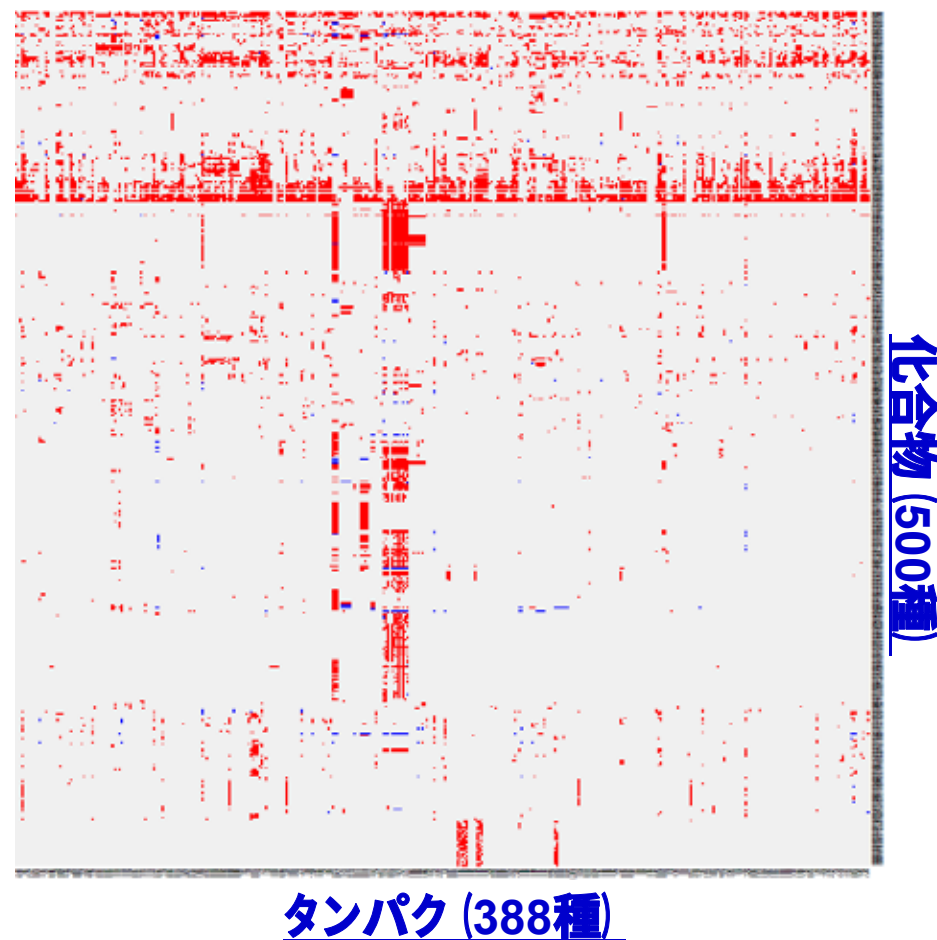
タンパク質 (キナーゼ) 388種と化合物500個の相互作用の
CGBVS法による予測と実験結果との比較

「京」が予測した結合パターン (予測結果)



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

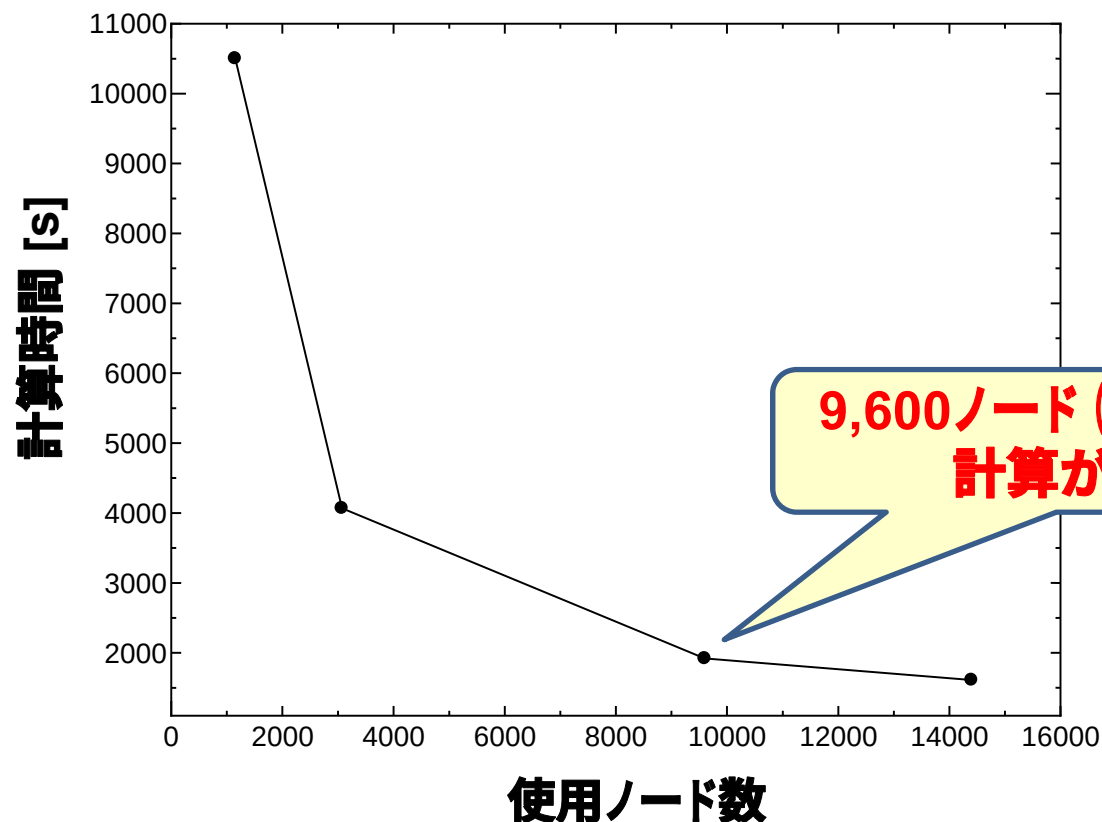
実際の結合パターン (実験結果)



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

「京」によるCGBVS法の計算速度

233タンパクと100万化合物の2億3,300万ペア相互作用の計算時間で評価



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、「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」によるコンピュータ創薬の根本課題への挑戦

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

CGBVS法

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%)

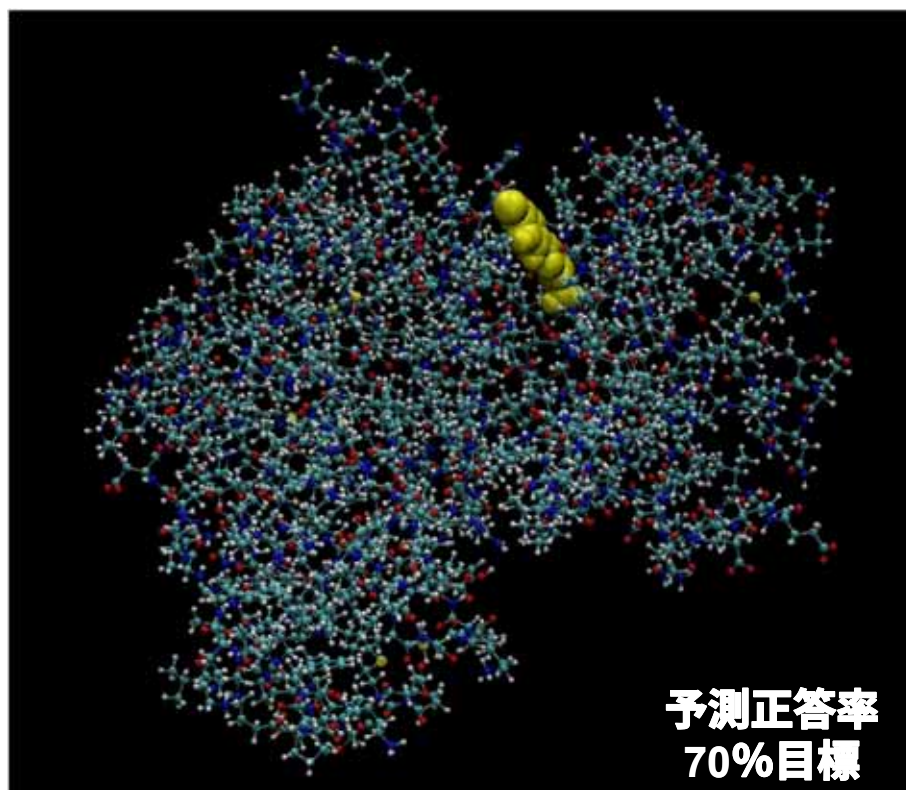
Docking計算法

MP-CAFEE法

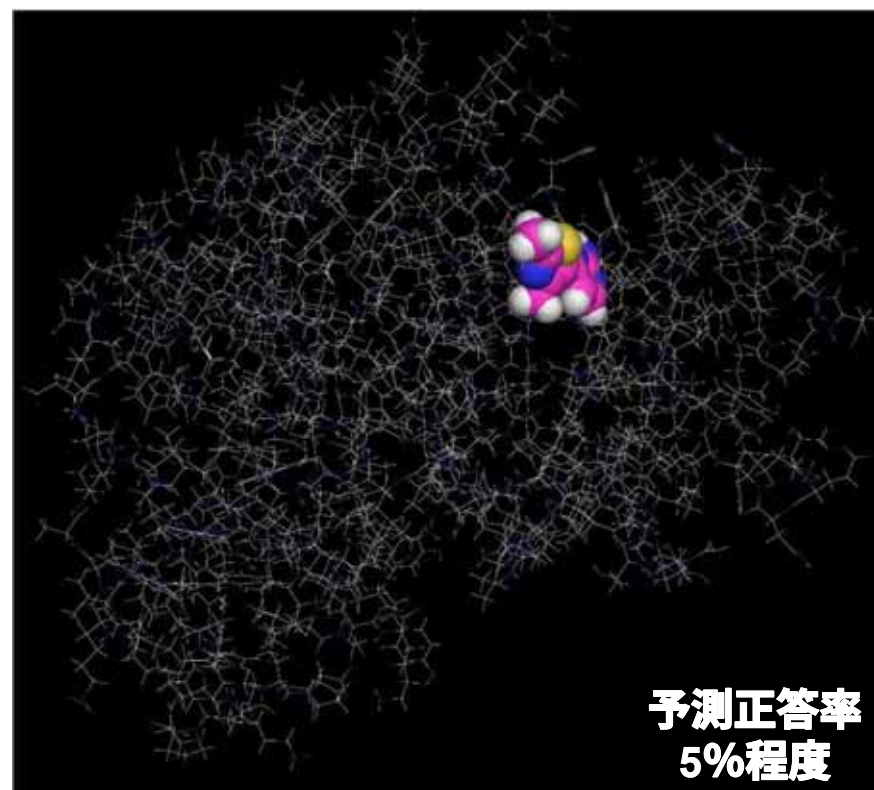
結合の強さを予測: MP-CAFEE法 (Massively Parallel Computation of Absolute binding Free Energy)

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-CAFEE)



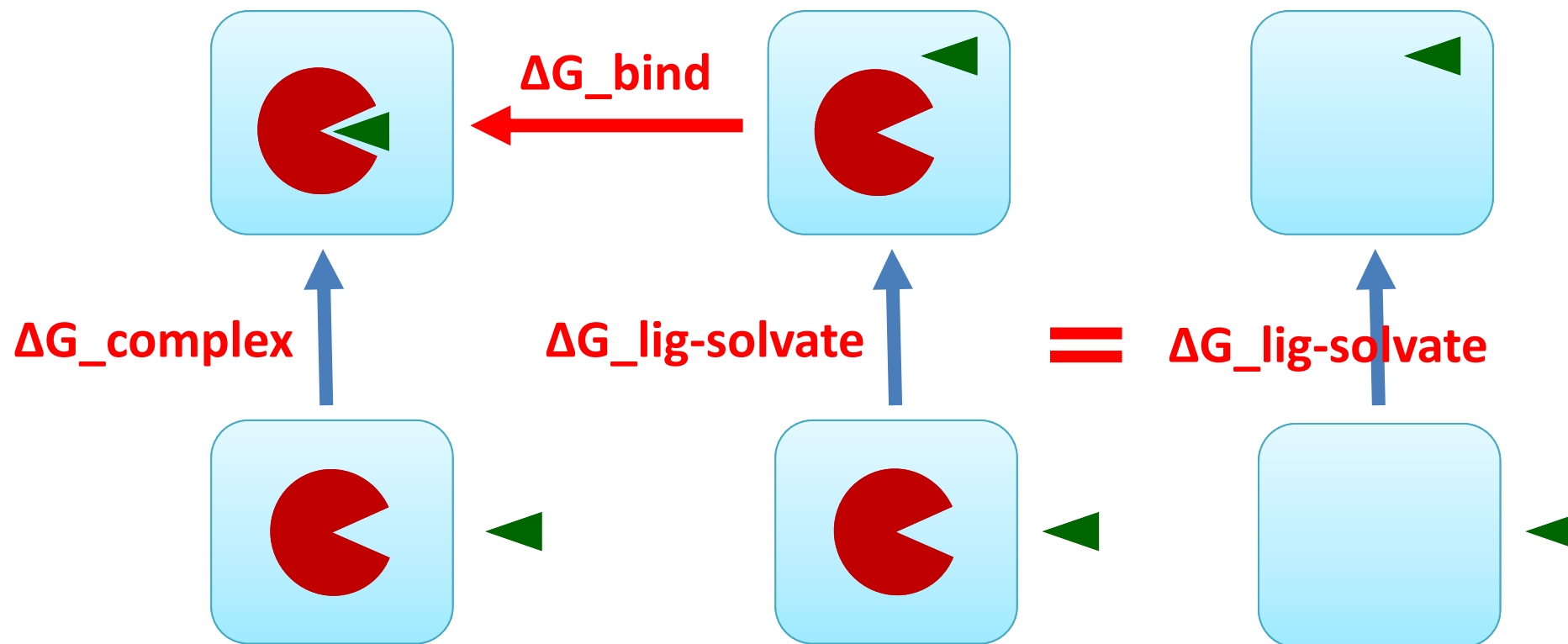
従来型の結合シミュレーション (Docking)



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) も含めた長時間シミュレーションを必要であるが、これまでは計算機パワーの問題で非現実だった。

タンパク質と化合物の結合自由エネルギーの算出

タンパク質と化合物との結合における熱力学的サイクル



$$\Delta G_{\text{bind}} = \Delta G_{\text{complex}} - \Delta G_{\text{lig-solvate}}$$

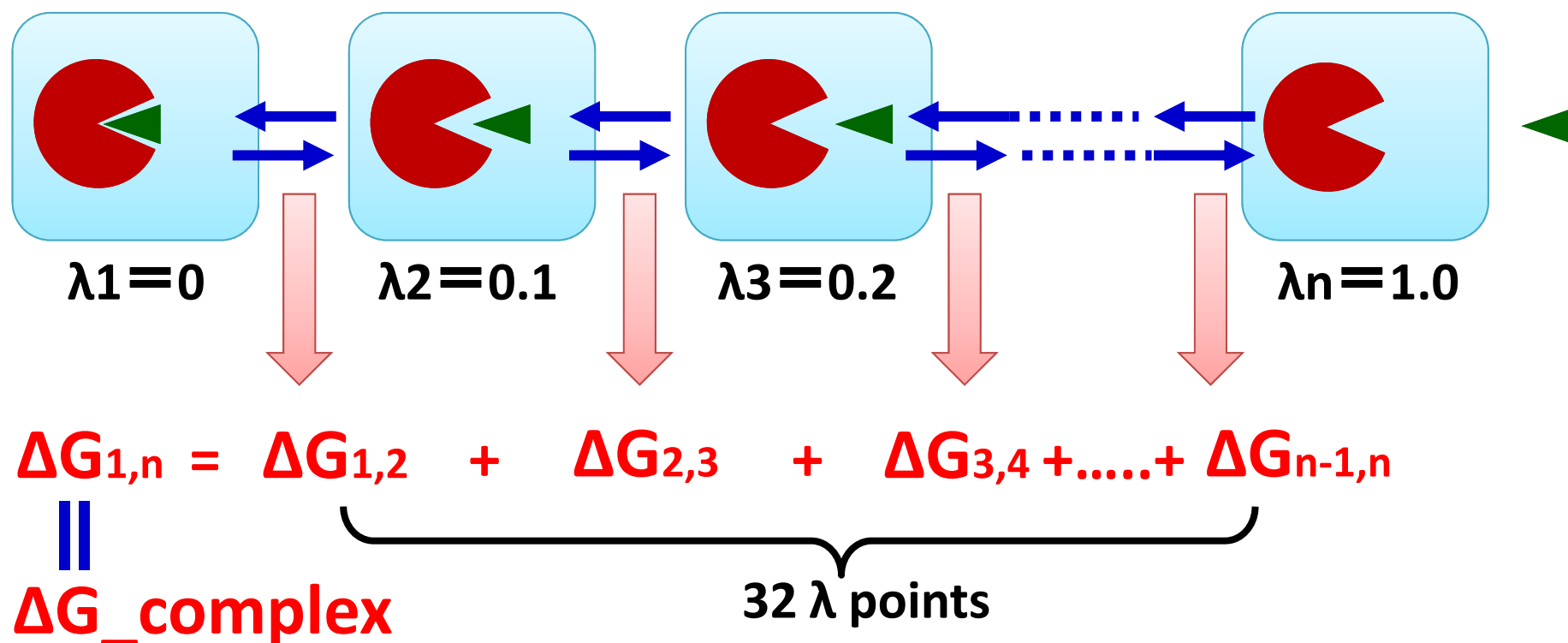
ΔG_{bind} の算出には $\Delta G_{\text{complex}}$ と $\Delta G_{\text{lig-solvate}}$ を計算する必要がある

MP-CAFEE法: ΔG complexの算出:

Soft-core Potential導入によるアンサンブルシミュレーション

Soft-core Potential:
タンパク質と化合物の
分子間相互作用

$$U(\lambda^c, \lambda^L) = \sum_{i,j} \left[(1 - \lambda^c) \frac{q_i q_j}{r_{ij}} + 4(1 - \lambda^L) \varepsilon_{ij} \left\{ \frac{1}{(\alpha^L \lambda^L + (\frac{r_{ij}}{\sigma_{ij}})^6)^2} - \frac{1}{\alpha^L \lambda^L + (\frac{r_{ij}}{\sigma_{ij}})^6} \right\} \right]$$



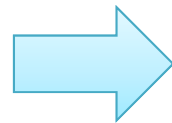
6 sampling velocities $\times$ 32 λ points = 192 ensemble MDs

MP-CAFE法の計算コストの「京」の威力

1組の $\Delta G_{\text{complex}}$ を算出するための計算時間

平衡化MD (50 ns) :

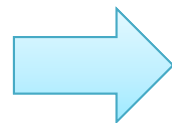
104プロセッサ × 5初速度 × 120時間 = 62,400時間
(約7年)



「京」 520並列で
120時間

非平衡MD (2ns) :

24プロセッサ × 6初速度 × 32λ points × 16時間 = 73,728時間
(約8年)

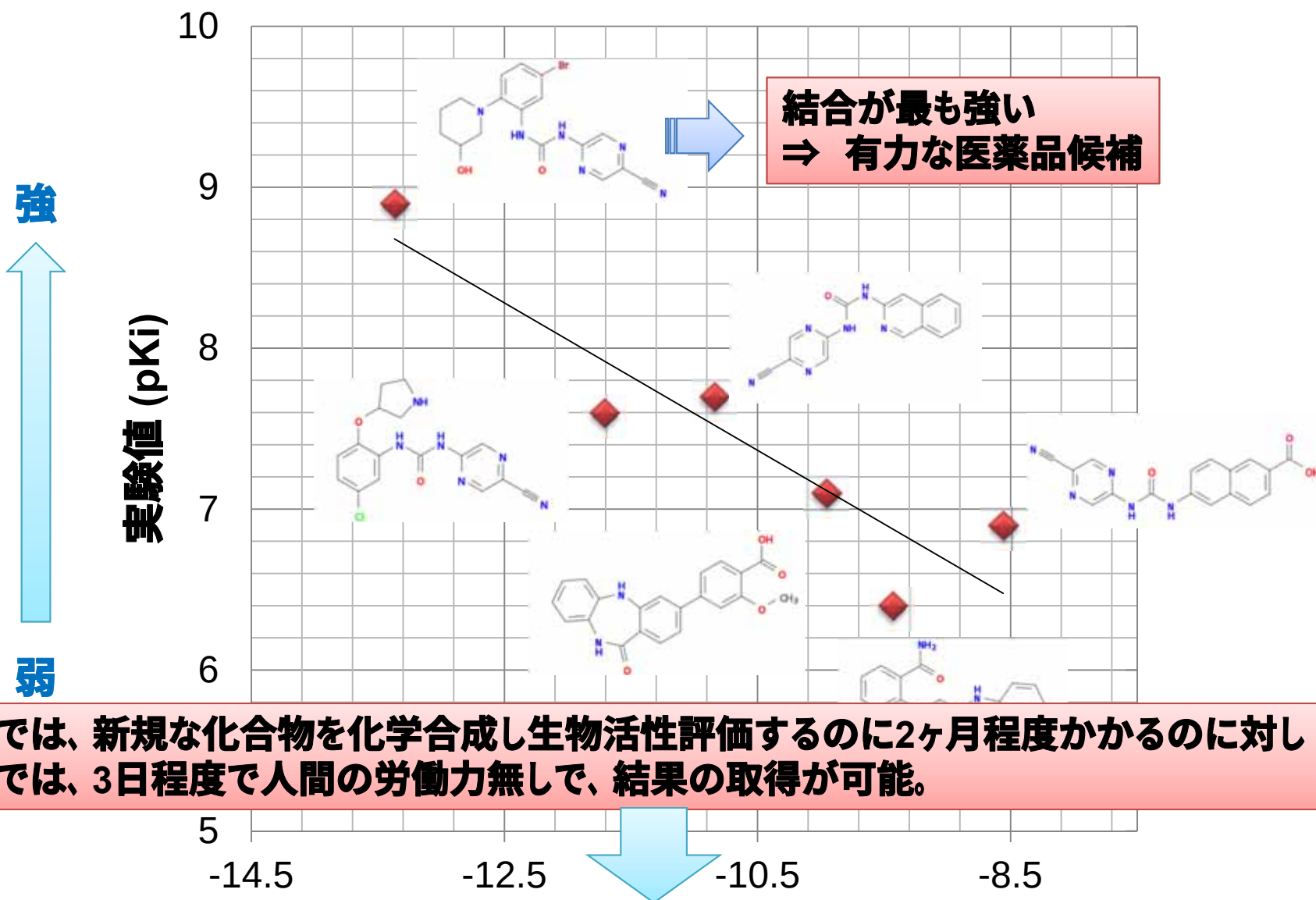


「京」 4,608並列で
16時間

(※ ΔG_{bind} の算出には $\Delta G_{\text{lig-solvate}}$ の算出も必要)

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

MP-CAFEE法による予測の実例



「京」によるコンピュータ創薬の根本課題への挑戦

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

CGBVS法

②結合の強さを予測

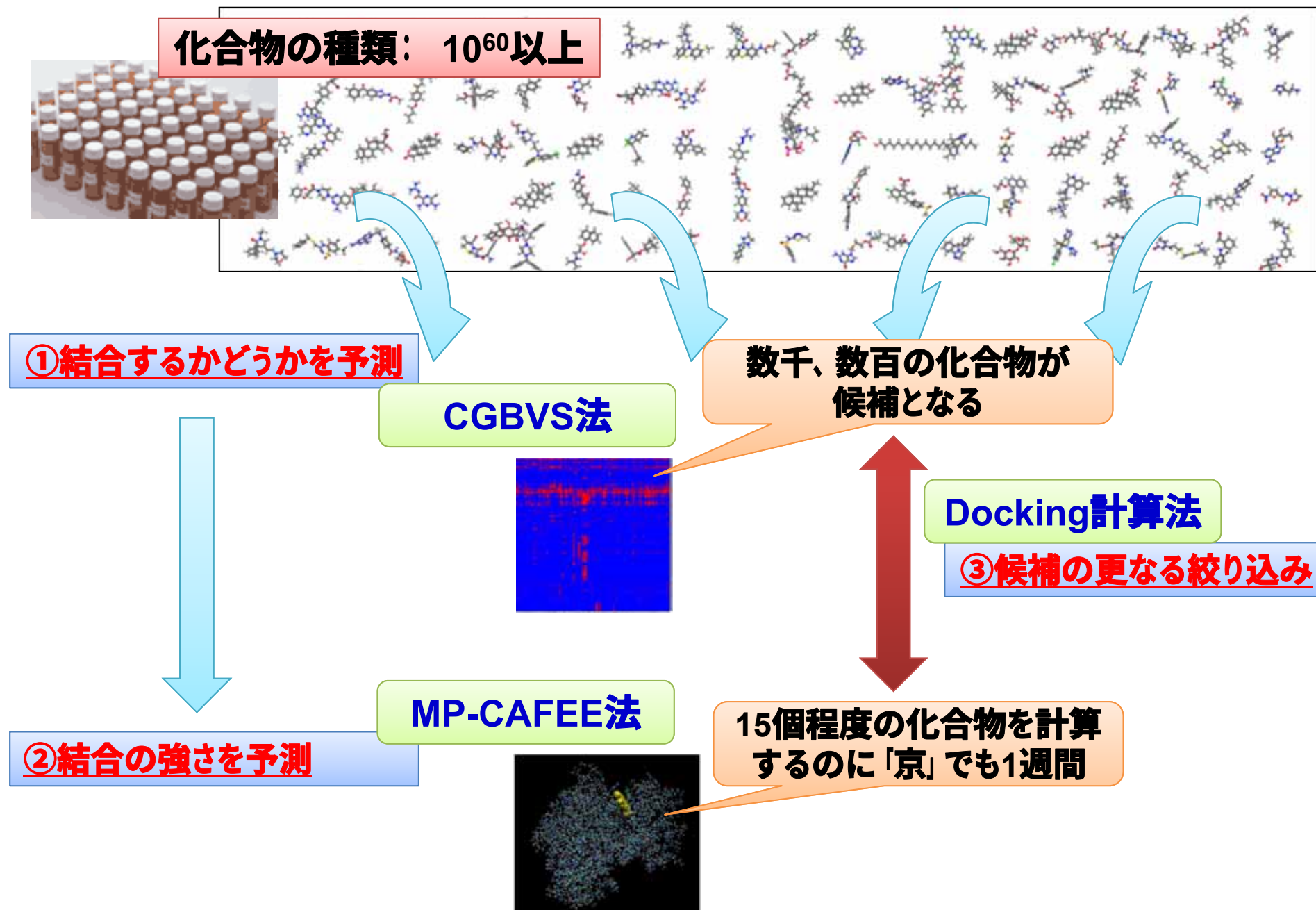
予測精度が悪すぎる: 正答率: 5%程度

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%)

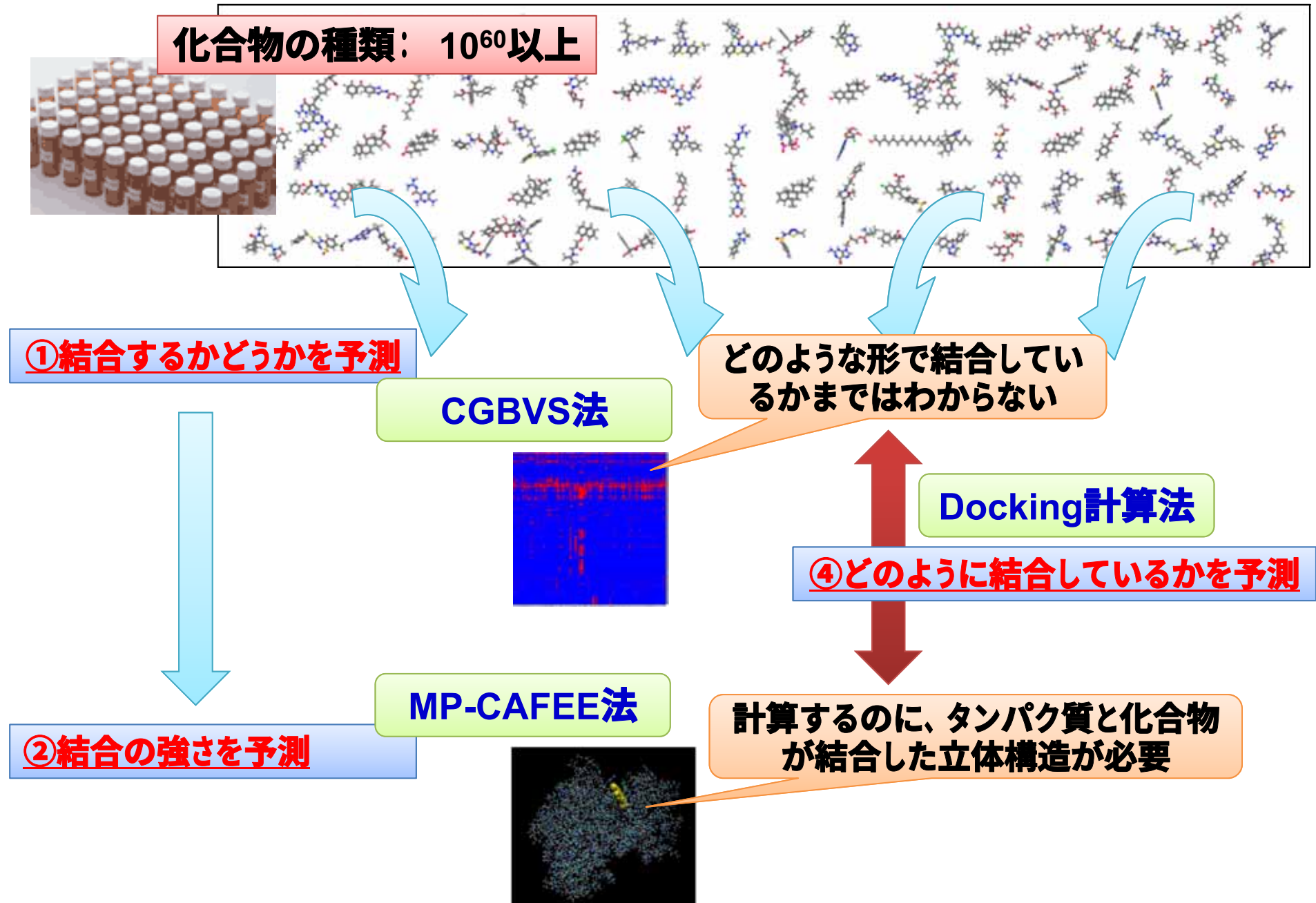
Docking計算法

MP-CAFEE法

候補化合物の更なる絞り込み: Docking計算法



どのような形で結合しているかを予測: Docking計算法



創薬現場でMP-CAFEE法が使いものになるための必須条件

MP-CAFEE法で正確な結合の強さを予測するためには、
正確なタンパク質と化合物の複合体立体構造が必要である



正確な複合体立体構造を得るためには、X線結晶構造解析を行う必要がある



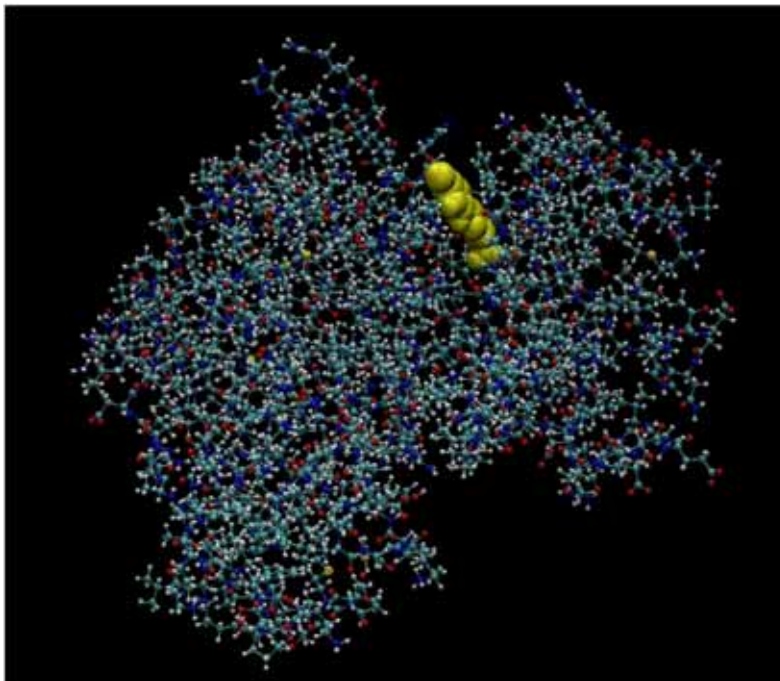
X線結晶構造解析を行うには、化合物の化学合成、タンパクの生成、複合体結晶化などの実験を実施する必要がある。



このような準備実験ができるのなら、
面倒なX線解析をしなくても、
実験で直接結合親和性を計測できてしまう

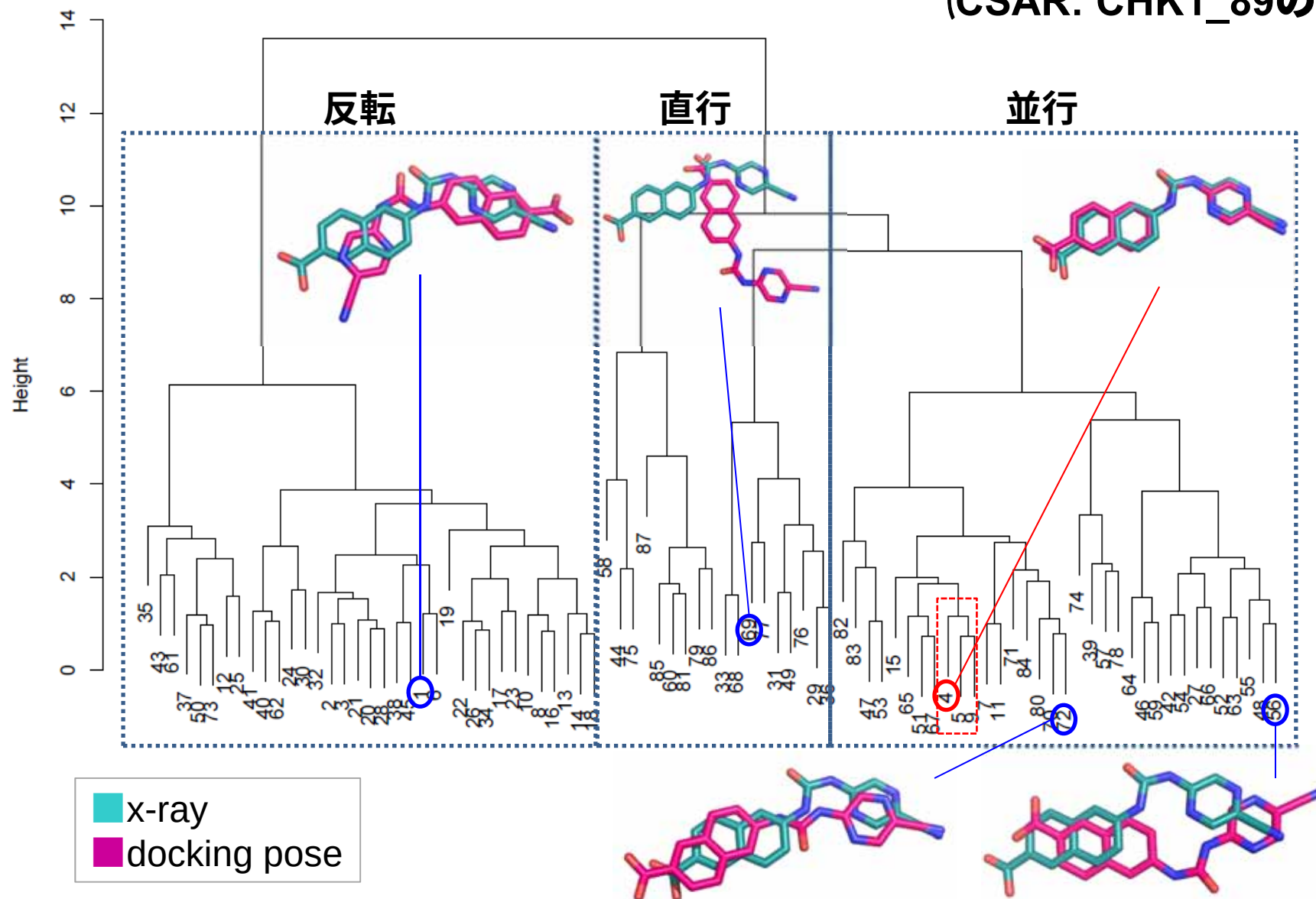


実験をしないで、タンパク質と化合物の
複合体構造を予測することが求められる

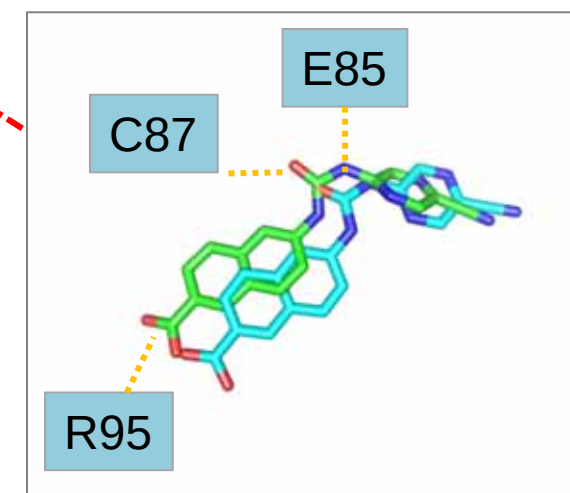
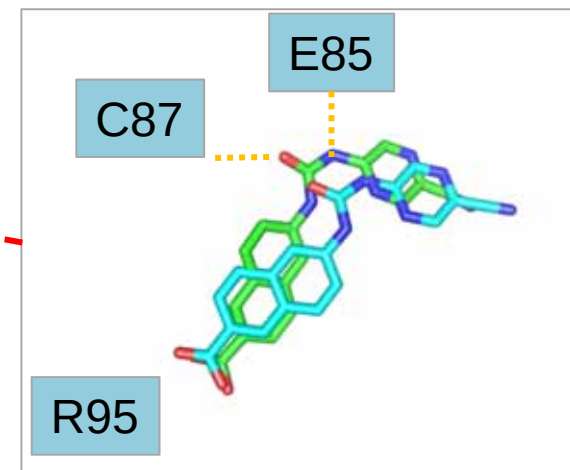
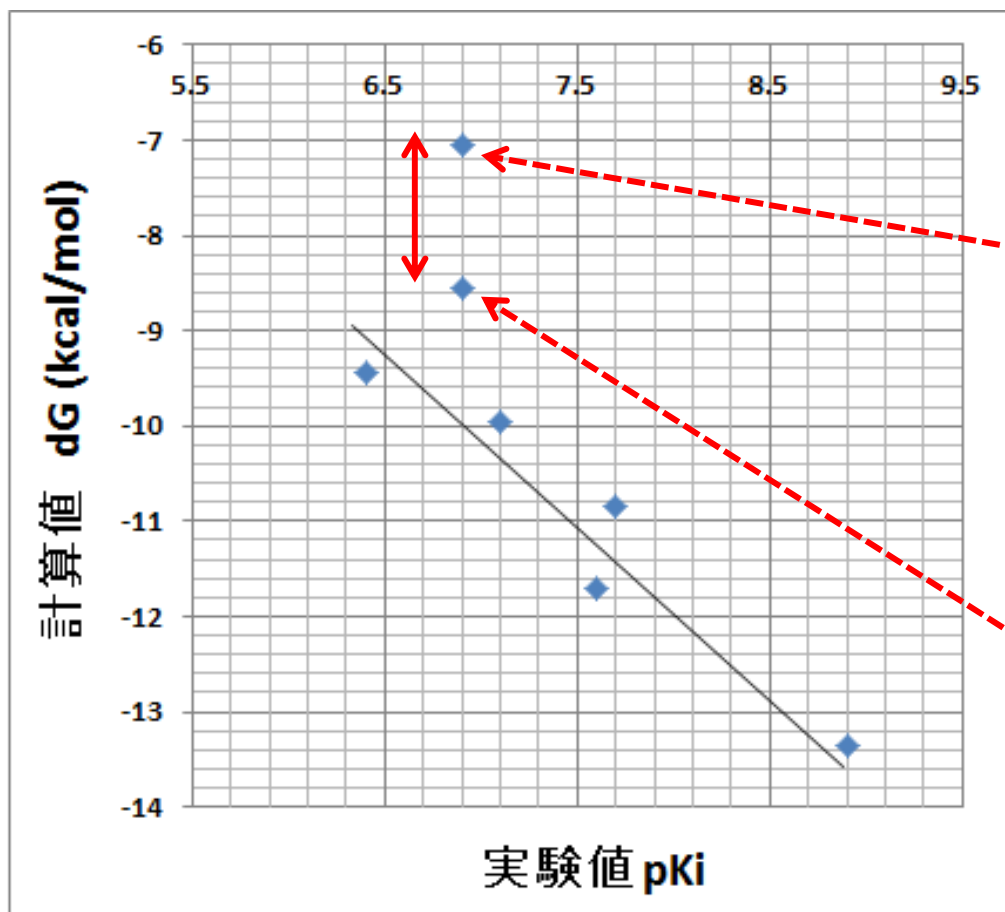


Docking計算法で予測された結合ポーズとX線構造との比較

(CSAR: CHK1_89の例)

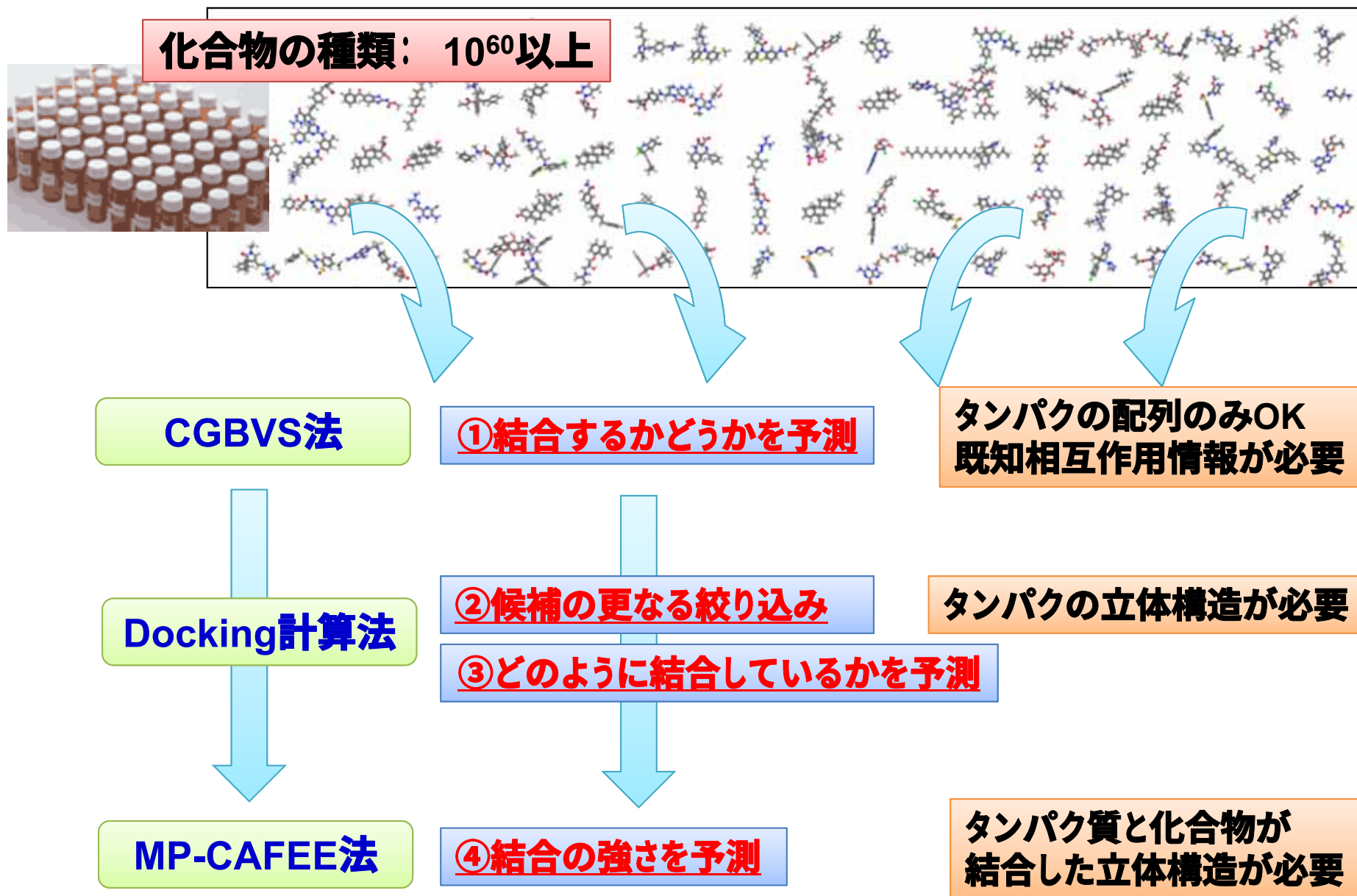


複合体の結合ポーズの違いによる MP-CAFEE法の結合自由エネルギー値への影響



**X線構造の無い状況で、Docking計算法で
最適な結合ポーズを決定することが今後の重要課題**

創薬利用にはCGBVS-Docking-MPCAFEEの統合化が重要



「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

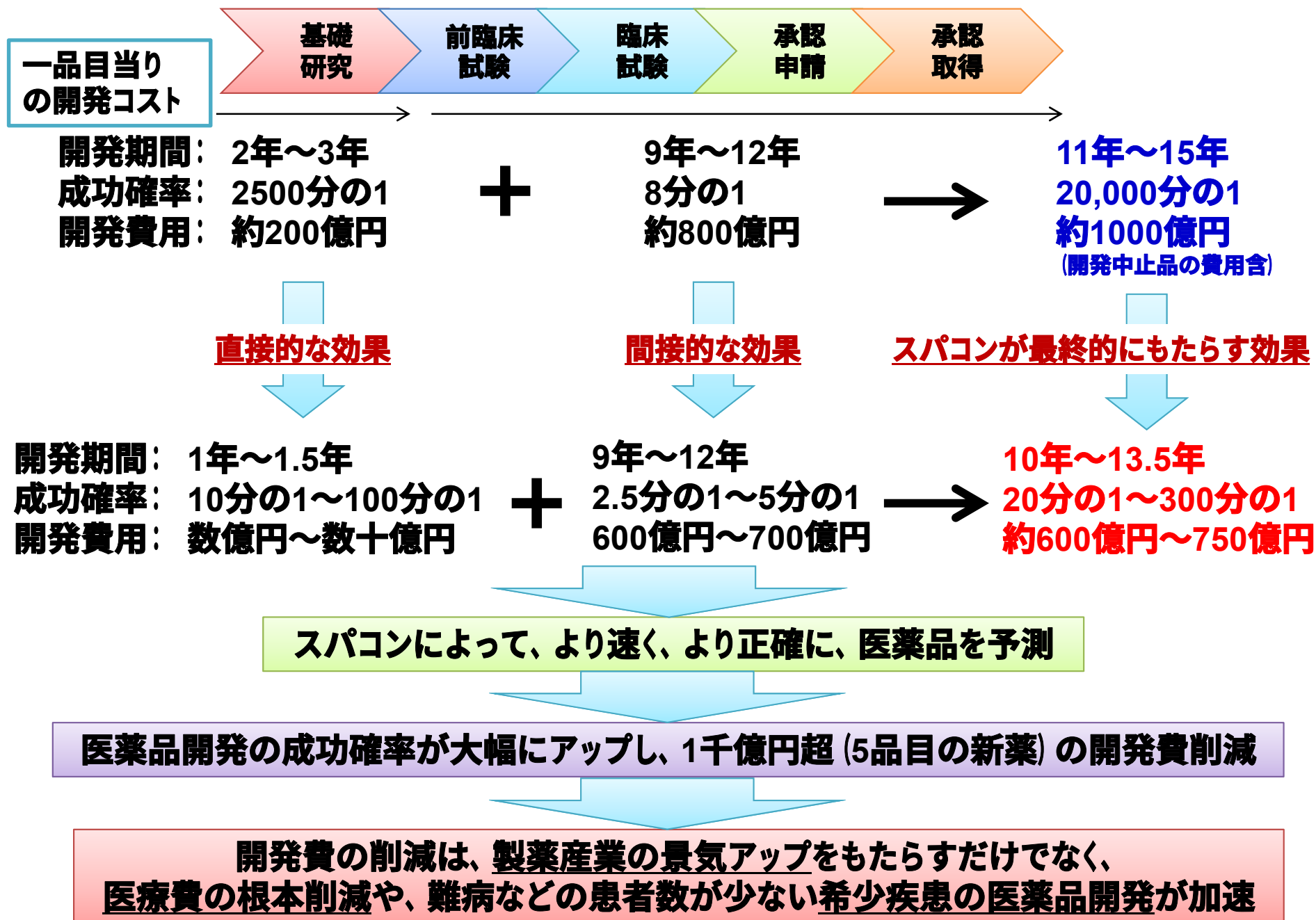
技術的ノウハウの指導・提供
勉強会・講習会の実施



「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年先の創業計算
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

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
- 小野薬品工業株式会社

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's a header with the JST logo and the text 'サイエンスチャンネル (SCIENCE CHANNEL)'. Below this, there's a navigation bar with links like HOME, サイエンスチャンネルとは, ダウンロード・DVD・放送用メディアご希望の方へ, 利用規約, and アンケート. A search bar is also present. The main content area features a video player for a 'Science News' segment. 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 (2013年3月13日配信)'. The video player includes a progress bar and a '再生時間: 3分' (Playback time: 3 minutes) indicator. To the right of the video player, there's 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, Affiliation, and Collaborating Institutions) section. The summary mentions that the project started in autumn 2012, aiming to improve efficiency in drug development using the supercomputer 'K'. The cast list includes 奥野希史 (Osano Nozomu) from the University of Tokyo, 江口至洋 (Eguchi Shiyoh) from the RIKEN Advanced Science Institute, and others.

サイエンスチャンネル
(SCIENCE CHANNEL)

知的好奇心を刺激する動画ライブラリ

356シリーズ3827番組を無料公開中!

HOME サイエンスチャンネルとは ダウンロード・DVD・放送用メディアご希望の方へ 利用規約 アンケート

シリーズ検索 サイエンスニュースアーカイブ 受賞作品 ドキュメンタリー ドラマ/アニメ 暮らし 自然 人物 科学館/研究所 イベント 学習/教育

番組検索

検索 ヒットポイント検索 外国語番組 キーワード一覧 ランキング ログイン | 新規ユーザー登録

サイエンスニュース2012(新着情報)

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 (2013年3月13日配信)

再生時間: 3分 制作年度: 2012年

概要

スーパーコンピュータ「京」を使って、医薬品開発を大幅に効率化しようという研究プロジェクトが2012年秋から始まりました。コンピュータを使って医薬品を開発する、いわゆる「IT創業」は、計算機の進歩とともに大きな注目が集まっています。

出演者名・所属機関名および協力機関名

奥野希史(京都大学大学院 薬学研究科 教授) 江口至洋(理化学研究所 HPC計算生命科学推進プログラム 副プログラムディレクター) 岡本敦之(アスビオファーマ株式会社 採薬第一ファカルティ 主幹研究員) 村上喜子(RADIO HAAFU)

キーワード