

創薬分野での利活用・成果促進に 関する提案

創薬分野における計算戦略策定のための調査委員会

委員長 姫野龍太郎

(理化学研究所・情報システム本部研究開発部門)

副委員長 北村一泰(元大正製薬、現摂南大学)、

委員 池口(横浜市立大)、追永(ソシオネクスト)、大田(中外製薬)、
奥田(RIST)、小田(大正製薬)、

粕谷(第一三共)、白井(アステラス)、住元(富士通)、泰地(理研)、
高木(阪大)、中田(理研)

創薬分野における計算戦略策定のための調査委員会

- ・ 製薬企業の現実のニーズを把握、今後のコンピュータの速度向上とを考え合わせ、近未来(5,10年後)の創薬の戦略を考え、提言にまとめる
- ・ 活動期間
 - 2015年4月～2016年12月
- ・ 委員
 - 製薬企業4名、コンピュータ関係3名
 - 大学3名
 - 理研3名
- ・ 委員会を20回開催、最後にシンポジウムを開き、報告書をまとめた

委員長 姫野龍太郎

副委員長 北村一泰(元大正製薬・現摂南大学)

委員 池口満徳(横浜市立大)

追永勇次(ソシオネクスト)

大田雅照(中外製薬)

奥田基(RIST)

小田晃司(大正製薬)

粕谷敦(第一三共)

白井弘樹(アステラス)

住元真司(富士通)

泰地真弘人(理研)

高木達也(阪大)

中田真秀(理研)

* 所属は2016/12の時点

創薬の現場からの要件(1)

- 創薬の現場でのコンピュータの利用状況・対象とする物理化学現象・計算手法

創薬ターゲット設定	ヒット・リード化合物の取得	化合物最適化、薬理活性・体内動態・安全性・物性	その他
<ul style="list-style-type: none"> •PPI インターフェイス予測 	<ul style="list-style-type: none"> •化合物ライブラリクラスタリング/検索 •バーチャルスクリーニング •ドッキングスコア関数計算/ポーズの予測検証 •合成化合物の活性予測、 •構造活性相関性解析 •活性部位の立体的・電子的特徴の推定。等 	<ul style="list-style-type: none"> •ADME(薬物の吸収・分布・代謝・排泄) •薬物の代謝予測・軽減予測 •化合物物性改善/最適化/毒性予測 •遺伝毒性予測 •副作用回避 等 	<ul style="list-style-type: none"> •タンパク質の構造予測

物理化学現象	分子会合	<ul style="list-style-type: none"> •結合自由エネルギー •計算結合ポーズの解析 •エンカウンター複合体形成の計算方法
	立体構造予測、構造変化	<ul style="list-style-type: none"> •立体構造予測 •タンパク質複合体構造予測 •立体構造変化
	生化学反応 2018/12/18	<ul style="list-style-type: none"> •化学反応 •光反応 •化合物物性

計算手法	MD
	MO

創薬の現場からの要件(2)

カテゴリ		期待する事
実験の代替による効率化	物理化学計算に基づく結合親和性の推算	<ul style="list-style-type: none"> 化合物のターゲット分子に対する親和性の測定の結合自由エネルギーのMDによる高精度推算 MDでは計算が難しい、化合物の近接効果をMOで計算。FMO法による巨大系の相互作用エネルギー計算。
	物理化学計算に基づく薬物動態の推算	<ul style="list-style-type: none"> 物理化学計算を推し進めれば、薬物動態の分子レベルについての物理化学的シミュレーションが可能 CYP代謝産物：薬物部分のCYPへのアクセシビリティ予測(MD)。その後、反応性予測(MO)。 薬物がCYPのヘムと共有結合し、他の薬物の酸化を阻害する場合の予測 グルタチオンの反応性代謝物との付加反応の予測(MO) 薬物動態に関わる試験を対象タンパク質への結合親和性や構造変化の計算により評価(MD)。 薬物と核酸塩基との化学反応による遺伝毒性を解析(MO)
今まで実現していなかったことの実現 2018/12/18	新しいターゲット分子を理解し、同定する	<ul style="list-style-type: none"> 革新的なターゲット分子の同定の支援のために重要なインシリコ技術 「計算創薬」による活性化合物取得方法のブレークスルー

新薬開発を取り巻く状況の変化と取るべき対策

これまでの医薬の大半は低分子化合物、

- 疾患によっては十分効力を発揮してきた

一方、ガンや自己免疫疾患に対しては、半世紀に及ぶ研究開発にもかかわらず、副作用を回避した十分満足いく医薬が開発されなれていなかった

今、有望な創薬手法は

1)スクリーニング創薬の更なる発展

中分子のスクリーニング、

ペプチドリーム(環状ペプチド、一億種)などの方式のスクリーニング

2)抗体に代表されるバイオ医薬の開発

そして、ポスト京で可能になる**新たな手法**として

3)**分子間相互作用のシミュレーション**を元にした**新規薬物分子の設計**

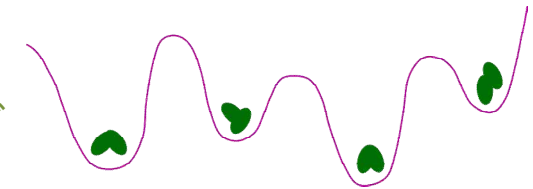
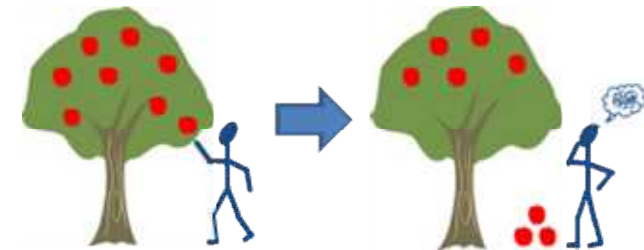
(生体の種々の機能は生体分子間の相互作用で起き、その時構造変化を伴う)

この第3の新たな手法とは、コンピュータシミュレーションで

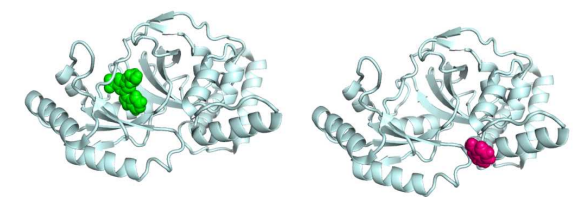
1. 長時間の分子動力学(MD)計算とその遷移過程の解析から、標的タンパク質の構造変化と生物学的機能を探る
2. 標的タンパク質のどこに候補物質を結合させると構造変化を制御できるか、場所を探る
3. 構造変化を制御する低中分子を探し、活性を高める最適化を行う
4. ターゲット分子が他のタンパク質分子などと相互作用するようすをシミュレーション、相互作用ネットワークを知る

こうすることで、特異性の高い(副作用の少ない)低中分子創薬を可能にする

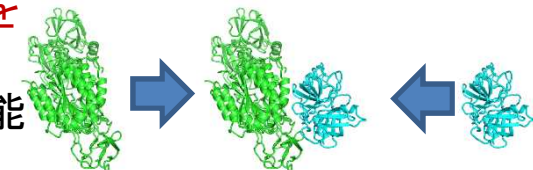
2018/12/18



1.タンパク質の長時間の遷移過程を解析



2.通常の結合部位以外にも結合部位があり、構造変化を起こす



4.タンパク質分子の相互作用

分子間相互作用のシミュレーションによる新規薬物分子の設計が可能か？

- ・ 標的タンパク質との分子間相互作用をMDによりシミュレーション
 - どのような方法で、どの程度の計算が実現でき、どの程度の創薬へのインパクトが見込まれるかを推算
- ・ 分子間相互作用(分子会合過程)のシミュレーションはいつ頃、どの程度可能か、具体的な3つの場合で推計結果を紹介
 1. 結合ポーズ推定(構造変化を引き起こす場所と結合の仕方を同時に推定)
 - ✓ 結合ポケットの大まかな位置を与え、その結合ポーズを推定する
 2. 結合自由エネルギー計算(活性の強さを推定)
 - ✓ 結合ポーズを与え、その親和性の強さ、すなわち結合自由エネルギーを推定する
 - ✓ ごく少量でも効くような特異性の向上を目指す(副作用を防ぐ)
 3. エンカウンター複合体形成の計算(副作用の推定)
 - ✓ 離れた位置にリガンドを置き、その自発的な結合過程を観察する
- ・ 計算機の性能向上の想定(5年後、10年後)
 - 汎用スパコン(保守的&アグレッシヴ性能向上パターン)
 - MDアプリ専用機

汎用スパコンのMD性能予測

- ・ 予測条件
 - 計算対象系規模 : 150,000 原子
 - ステップあたりの物理時間 : 2.5 fs/step
 - 演算量 : 25,000 演算量/step
- ・ スケーラビリティ想定
 - 32ノード程度が現行システムの限界。
 - ✓ 報告例 : 32ノード/64cpuソケット (LINPACK性能32TFlops相当) のクラスタシステムで、 $0.2 \mu\text{s/day}$ 程度)
 - 32ノード構成のクラスタ単位に将来の2ケースの性能を推定。
 - ✓ アグレッシブなケース : 汎用スパコンの性能トレンドがアーキテクチャの進化と半導体テクノロジーの向上により達成されるとするケース。ノード性能には性能トレンドの性能向上率を適用
 - ✓ 保守的なケース : ノード性能は主に半導体テクノロジー(プロセスルール)の向上により達成されるとするケース(この場合、性能トレンドはノード性能向上とノード数の増加で達成されると考える)。
 - ✓ MD計算アルゴリズムの改良による性能向上の効果は含まない。

汎用スパコンのMO性能予測

・ 想定条件

– FMO法による高精度ドッキング計算

- ✓ 対象：アミノ酸残基数 500、計算方法はFMO2/MP2/6-31G*
- ✓ 計算回数：1候補分子あたり、リガンドの結合部位10箇所を10ポーズで計算（1候補分子あたり、100回のドッキング計算）。
- ✓ 1回の計算量：1ドッキング計算あたりの計算量を27 TFlops × hr. = LINPACK性能 27TFlopsのシステムで1時間の計算が必要な計算量（地球シミュレータを利用した6-31G法による計算例より、6-31G*への計算量増加を考慮）

– QM/MM + B3LYP-D法(分散力を考慮した密度汎関数)による不可逆的CYP阻害計算

- ✓ 対象：候補物質及び代謝箇所を指定して計算
- ✓ 計算回数：鉄は三通り(II, III,IV)の酸化状態、候補分子は5箇所の阻害箇所での構造最適化計算、計算時間は一点計算の100倍を仮定(候補分子あたり、1,500回の計算)。
- ✓ 1回の計算量：27TFlops × hr. (地球シミュレータによる計算例より)

・ スケーラビリティ想定

- 小規模なジョブを大量に実行するスループットジョブとなる。
- 利用できる計算能力をそのまますべての計算に利用できると考える。(MD計算とは異なる)

(1)結合ポーズ推定(不可逆的CYP阻害の例、MD+MO)

- ・ 例えば、デザイン化合物が、CYPの活性中心であるヘムと共有結合し、他の分子の代謝を長時間阻害するかどうか判定できれば、創薬の現場ではインパクトがある
- ・ 反応部位までたどりつけるかどうかはMDで計算、その後MO計算によって共有結合が形成できるかどうか判定する(QM/MM + B3LYP-D法*)
- ・ このアプローチにより、共有結合性が弱いと考えられるデザイン化合物の優先度を上げたり、共有結合性が強いと考えられるデザイン化合物のCYP阻害アッセイを優先的に実施したりするという積極的な対応が可能となる
- ・ 代表的CYP7種に対し、HTSでヒット化合物の代表20クラスに対して、不可逆的CYP阻害が1週間で判定・予想できれば、創薬に対してインパクトがある
 - CYP7種 × 20化合物 / 7日 = 20構造最適化エネルギー推算 / 1日
- ・ このような計算で、**post京の1/10の使用でインパクトのある結果が出せる**

	2016	2021	2026
リーディング・スパコン	6	223	927
共用システム	5	23	161
社内中規模	0.5	2.3	16

リーディングスパコン : 最大40MW程度の消費電力のシステム
共用システム(クラウド) : 買取で50億円程度の規模のシステム
社内中規模システム : 買取で5億円程度の規模のシステム

(2)結合自由エネルギー計算(MD)：創薬へのインパクト



- ・ MP-CAFEEによる計算
- ・ 創薬の場面：合成デザイン案 40 個が提案された場合
 - 1日 で結合親和性評価ができれば分子設計上大きなインパクトがある
- ・ 京：産業利用枠への課題申請では1%も使えないので、難しい
- ・ Post京：
 - 1%使うだけで、インパクトのある結果が得られる
 - ✓ 1日計算可能回数：75回(75化合物)

		2016	2021	2026
汎用スパコン	リーディング・スパコン	212	7520	31296
	共用システム	168	785	5420
	社内中規模	17	79	542
	社内小規模	2	8	54

リーディングスパコン：最大40MW程度の消費電力のシステム
 共用システム(クラウド)：買取で50億円程度の規模のシステム
 社内中規模システム：買取で5億円程度の規模のシステム
 社内小規模システム：買取で5千万円程度の規模のシステム

(3)エンカウンター複合体形成計算の実現可能性

離れた位置に候補物質を置き、その自発的な結合過程を観察する副作用の推定に使う副作用の推定に威力を発揮



- ・ 汎用スパコン:
 - 1日 20 μ sec の MD 計算の実現:
 - ✓ 保守的性能向上パターンの場合: 今後10年困難
 - ✓ アグレッシブ的性能向上パターンの場合: 2026年にかろうじて実現
 - 汎用スパコンの能力向上は並列化が主
 - ✓ 長時間のMD計算の実行が難しい

		2016	2021	2026
汎用スパコン	リーディング・スパコン	0	0	0~404
	共用システム	0	0	0~70
	社内中規模	0	0	0~7
	社内小規模	0	0	0~1

リーディングスパコン : 最大40MW程度の消費電力のシステム
 共用システム(クラウド) : 買取で50億円程度の規模のシステム
 社内中規模システム : 買取で5億円程度の規模のシステム
 社内小規模システム : 買取で5千万円程度の規模のシステム

整理すると

分子間相互作用のシミュレーションのうち

1) 結合ポーズ推定 (構造変化を引き起こす場所と結合の仕方を同時に推定)

- Post京で全体の1/10の利用でインパクトのある結果が出せる
- 使える仕組みの構築と人材育成が課題
- 2026年には製薬企業独自でも計算環境を構築可能になる
- PostPost京では広い対象に対して利用が可能になる

2) 結合自由エネルギー計算 (活性の強さを求める)

- Post京の1%でインパクトのある結果が出せる
- 製薬企業での取り組みが始まっている
- 日本で不足しているのは技術者と計算環境(教育とテストベッド構築が鍵)

3) エンカウンター複合体形成計算(副作用の推定)

- Post京でも難しい(小規模な計算のみ)

ただし、これらの計算技術は方法論的に未成熟な部分もあり、計算方法の確立、高精度分子力場の開発等の課題はある。実現に向けてチャレンジし、現実に創薬現場で実施できるようにしていくことが必要である → 研究開発を産官学合同で担う

今後の創薬の革新を加速するために

- 計算のテストベッドの構築と専門家の育成
 - Post京の例えば1/10をある期間占有する枠を設ける
 - その枠を企業とアカデミアに実際に使って貰う
 - ✓ 大学や公的研究機関と製薬企業で計算創薬のコンソーシアムを組織
 - ✓ その中で実際に計算を行い、専門家を育成してゆく
 - 現在の計算方法での問題点を洗い出し、定量的な予測の信頼性を高くする改良を進める
 - ✓ AIの活用
- 更なる高速なコンピュータの開発

補足資料

計算するコンピュータの導入規模予測の前提と利用形態

- リーディングスパコン：消費電力で制限
- 企業導入スパコン：導入費用で制限

リーディング・スパコン	国プロで整備。特別な研究開発に公募等の選抜プロセスを通じて利用。超大規模、長時間、大量の処理が実施可能。 規模は運用電力により制約されるとして、 「京」x ~2 = 40MW を想定
企業が業務で利用する スパコン	価格で導入されるシステムの規模が決定されると想定(1\$=100円で価格動向から計算)
共用システム	企業が社外で共用で利用できる大規模システム。商用サービス、クラウドサービス等を想定。必要に応じてある程度タイムリーに利用できる。 規模は価格50億円相当
社内中規模システム	企業内に導入できる中規模システム。企業内で共用で利用する。ある程度タイムリーに利用できる。 規模は価格5億円相当。
社内小規模システム	企業内の部門で導入で小規模システム。目的別に専用で利用することも可能。 規模は価格5千万円相当。

汎用スパコン動向：3つのトレンドまとめ

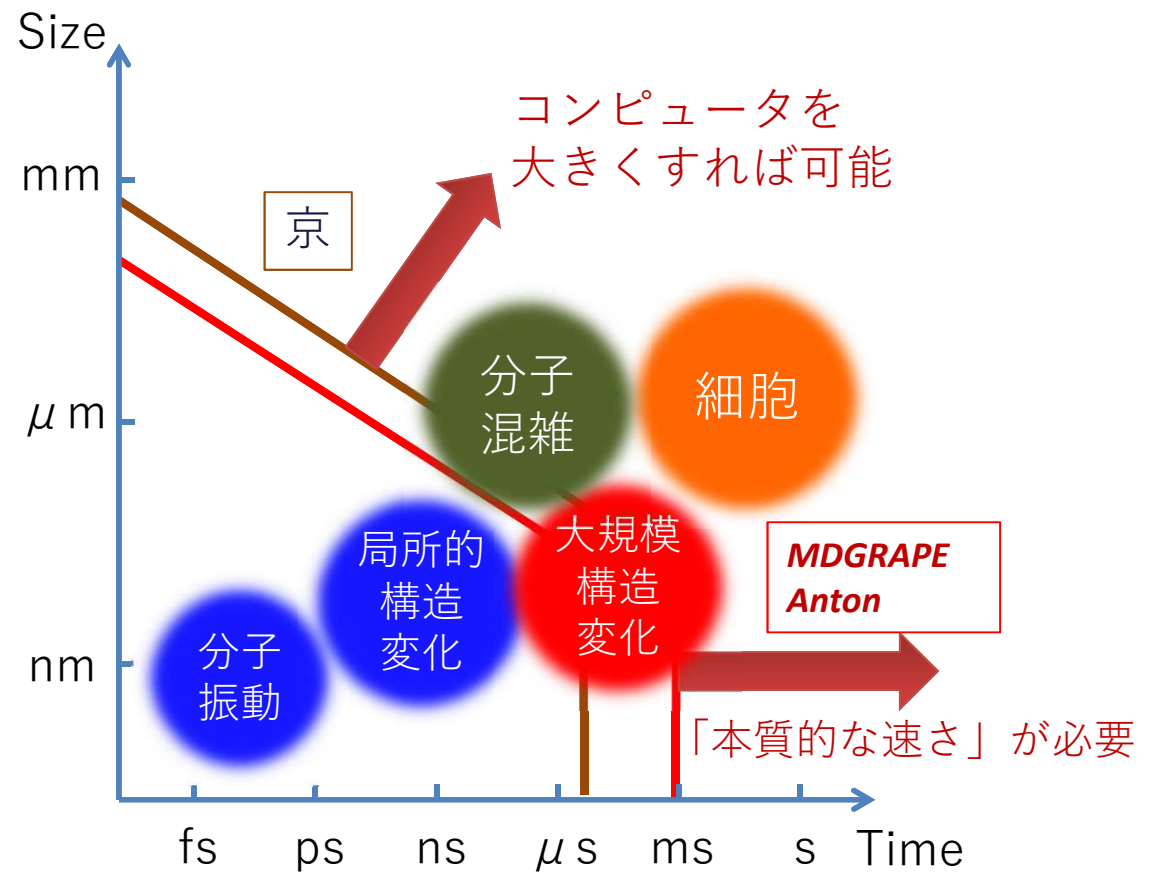
トレンド	改善
LINPACK性能トレンド(TOP1)	年率 1.82倍
電力対性能トレンド	年率 1.33倍
価格対性能トレンド	年率 0.68倍 (32% down)

- LINPACK性能トレンドがもっとも改善率が高い
 - リーディングスパコンが目標とするところ
- さらなる電力対性能、価格対性能向上が鍵
 - アーキテクチャを含めた改善が期待される
 - 改善がない場合、規模拡大による消費電力とシステム単価増加につながる
 - 一般企業向けスパコンはコストに敏感なためこれらのトレンドで制限される

長時間MDシミュレーションの課題

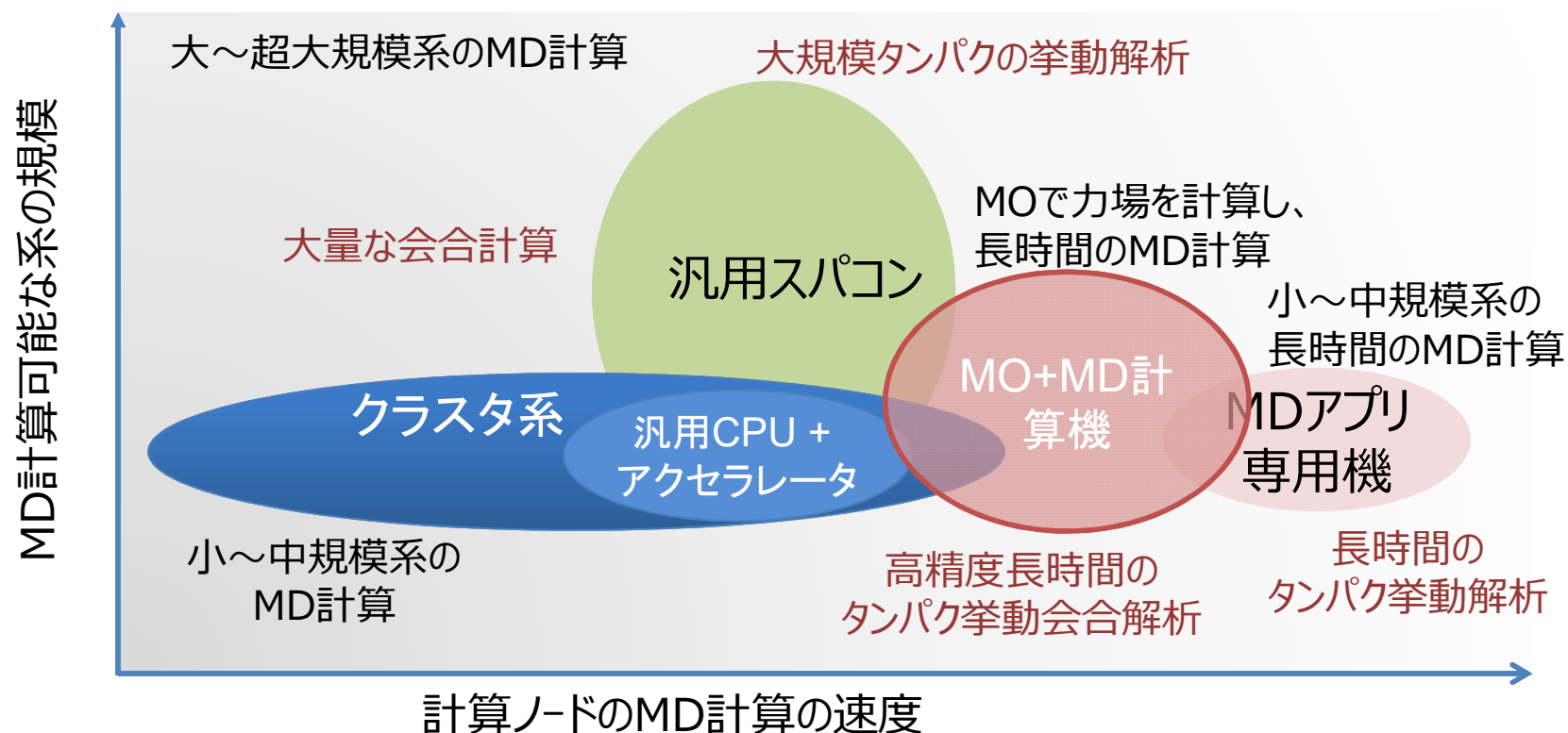
最も重要なタンパク質の大規模構造変化を予測するには

- 時間刻み ~フェムト秒
 - 目標到達時間
マイクロ秒~秒
- 10億~1京のギャップを埋める必要性
- 但し、京のようにコンピュータを大きくするだけでは実現できない
=「**本質的な速さ**」が必要
 - MDに関しては、スパコンのピーク性能は意味を持たなくなっている



利用されている計算機システムと計算手法、そして将来の方向

- MD計算で利用されている計算機システム



- MO計算で利用されている計算機システム

- クラスタ系で中小規模系の計算
- 汎用スパコンで中小規模系の大量計算、FMO法等による大規模系計算

2018/12/18
- アプリ専用機の例は無い

汎用スパコンのMD性能予測(2)

- 汎用スパコンで一日に計算可能な物理時間と一日にその計算を実行できる回数を予測。

		2016年	2021年		2026年		
			ケース1	ケース2	ケース1	ケース2	
1日に計算可能な最長物理時間 (μ s/day)		0.2*	2.2	0.80	24.2	2.5	
最長物理時間を一日に実行できる回数	リーディングスパコン	313	1,068	2,938	404	3912	
	企業が業務で利用できる汎用スパコン	共用システム	262	112	307	70	68
		社内中規模システム	26	11	31	7	7
		社内小規模システム	3	1	3	1	1

汎用スパコンのMO性能想定(2)

- FMO法による高精度ドッキング計算(候補分子数/day)

		2016年	2021年	2026年
リーディングスパコン		94	3,342	13,909
企業で業務 で利用可能 なスパコン	共用システム	75	349	2,409
	社内中規模システム	7	35	241
	社内小規模システム	1	3	24

- QM/MM + B3LYP-D法(分散力を考慮した密度汎関数)による不可逆的CYP阻害計算(候補分子数/day, カッコ内は一回の計算に必要な日数)

		2016年	2021年	2026年
リーディングスパコン		6	223	927
企業で業務 で利用可能 なスパコン	共用システム	5	23	161
	社内中規模システム	0.5(2)	2	16
	社内小規模システム	0.05(20)	0.2(4)	2

結合自由エネルギー計算：計算条件

- MP-CAFEE法による結合自由エネルギー計算の計算コスト



- 1000結合の結合自由エネルギー計算（結合活性予測）
 - 1結合 / 1化合物ならば 1000 化合物
 - 10 Pflops の計算機で 5 日で計算が終了
- 数nsecから数十nsecレベルの短時間MD計算を多数行う
- 並列性能が極めてよい
 - 多数のCPUコアを使えば、それだけ早く計算できる
 - ほぼLINPACKの値どおりに計算がスケールする
 - ✓ LINPACK：コンピュータの数値計算スピードの指標
- 将来の計算機で想定されるLINPACKの性能から、1日で実施可能な「結合自由エネルギー計算」の計算可能数を見積もった

エンカウンター複合体形成の計算

- リガンドを離しておいた状態でMDを開始し、シミュレーションにより自発的結合を観察する



- 想定していなかった結合ポケットの発見など、今まで得られなかったような新しい情報も提供することが可能な、より実験的な計算手法
- 「エンカウンター複合体形成」の計算は、「結合ポーズ推定」や「結合自由エネルギー」に比べて長時間のMDシミュレーションが必要
- 今後の計算機能力増大によって、大きく発展する可能性
- 「エンカウンター複合体形成」計算において典型的と考えられる 20 μ sec の MD 計算の回数をその指標とした

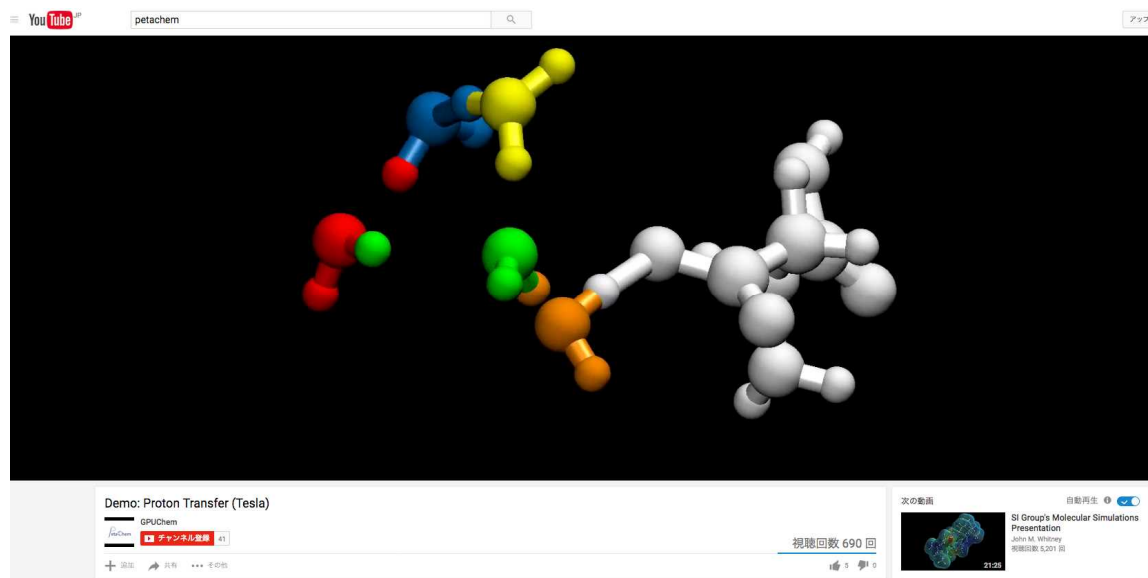
エンカウンター複合体計算による新しい可能性

- 現状では入手・推定不可能な情報が得られるようになる →
 - 創薬の質を変える可能性があるのでインパクトは大きい
 - 「エンカウンター複合体計算」には、その可能性がある
- 以下のようなケースは、「エンカウンター複合体計算」により実現可能なことであり、今までにない分子デザインのアプローチを可能とし、分子デザインの質を向上させる意味からもインパクトが大きい
 - X線結晶解析で複合体構造が得られない場合に、化合物の結合ポーズと結合親和性が推定できるようになる
 - 標的タンパク質の立体構造が解っていれば、HTSヒット化合物について、X線結晶構造解析を行うことなく、結合箇所と結合ポーズが網羅的に推定できるようになる
 - X線結晶解析などの実験手法では困難な、化合物の準安定的な結合部位の同定と、その間の遷移確率の把握を含む、結合過程の詳細な描像が、デザイン化合物も含めて、化合物ごとに把握できるようになる
 - デザイン化合物の結合・解離速度 (k_{on} , k_{off}) の情報が得られるようになる

アプリ専用機のMO性能想定

- 想定条件

- 1000原子 10000軌道の10ns程度のMO計算によるMD計算を実現
 - ✓ MO計算を用い、力を計算し、時間発展を追う。
 - ✓ MDの力場の計算を第一原理のMO計算から行うため膨大な数のMO計算が必要
 - ✓ Born-Oppenheimer MDと呼ばれる。
 - ✓ 生体内での、**プロトン移動や化学反応**を追うことができる。
 - ✓ **これまで全く想定してなかった化学反応を発見**できる可能性もある



2018/12/18 PetaChem 475原子、2006基底関数、B3LYP、106s/step C2070x8
<https://www.youtube.com/watch?v=EXt5l5bYY3w>

「計算創薬」アプローチにおけるAI/機械学習

- MD計算における各種精度改良

- コンピュータが自分でデータを生成しつつ、結果の分かっている問題で、力場や計算領域の取り方、初期位置など、各種のパラメータの改良
- コンピュータ上で無数の組み合わせが実現でき、コンピュータが自己完結的に学習可能なので、有利
- 例えば、結合自由エネルギーが分かっている、予測精度を改良していく
- MO計算を事前に多数行い、力場の改良を行う等