

エネルギー表示法による溶解度の計算、 および光毒性予測/VCDによる構造決定

古賀 良太

株式会社クロスアビリティ 代表取締役

rkoga@x-ability.jp

共同研究者:

エネルギー表示法: 松林伸幸、櫻庭俊(京都大学)

光毒性予測/VCD(計算、ソフト開発): 山口徹(TSテクノロジー)
、古川祐貴(クロスアビリティ)、安田耕二(名古屋大学)

- エネルギー表示法(ER)
 - 京都大学・松林准教授が2000年に考案した高速な自由エネルギー計算手法。応用論文あり
 - 2011年10月から弊社と北浦教授によるJST(A-STEP)『フラグメント分子軌道法を中核とした量子創薬計算システムの構築』で共同研究中。
 - GAMESS FMO-Amber-ERのAmber-ERのI/Fを開発
- 光毒性予測／VCD
 - 某社のブラインドテストを契機として、弊社開発のGPGPU QMが活用できる分野であると考えている。応用論文なし
 - 本発表は背景と現状説明

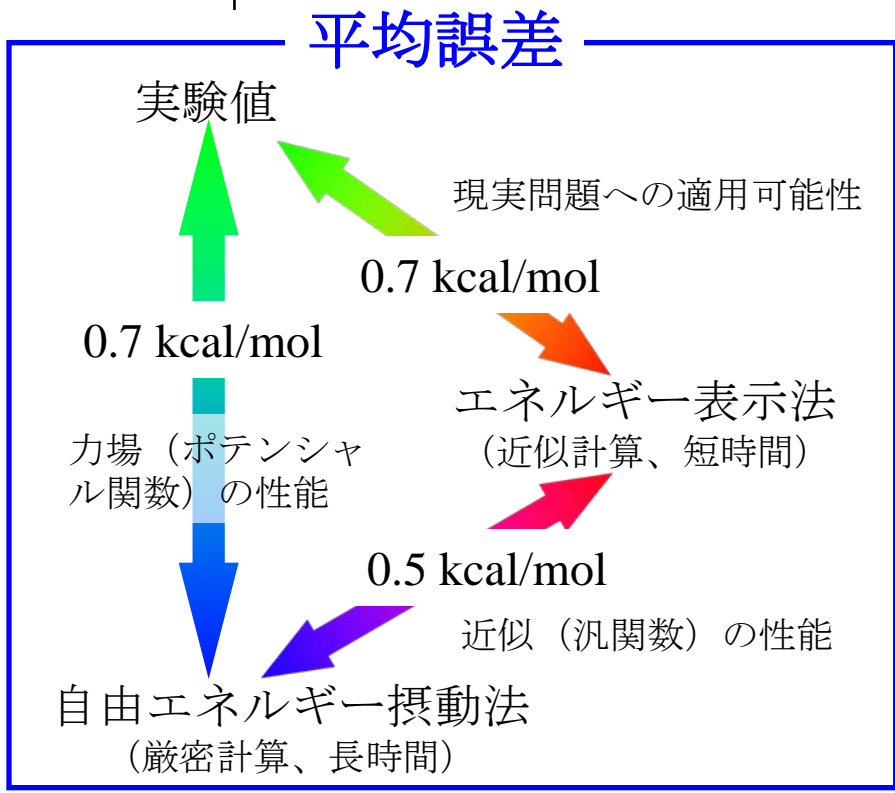
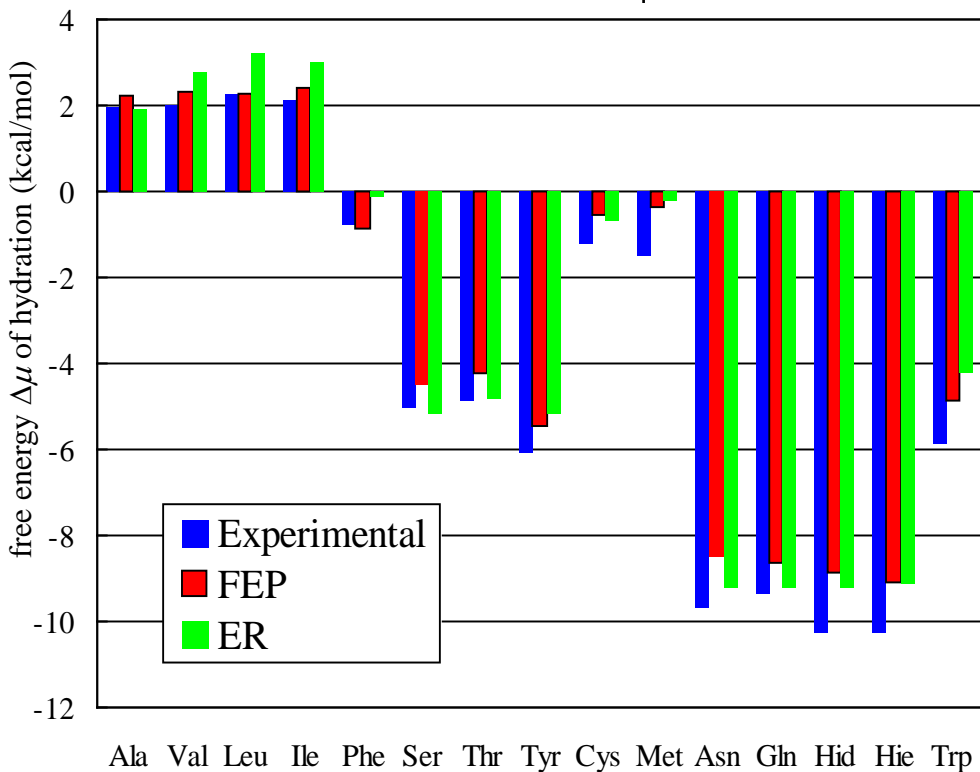
『創薬段階における物性評価の重要性』

日薬理誌 (Folia Pharmacol. Jpn.) 127, 213-216(2006)

- 薬理活性の増強だけに注力して探索研究を進めてしまうと、化合物特性に物性情報を反映させるような展開が困難。
- 迅速な溶解性評価がほしい。計算時間かかると意味がないのでlogP (cheminfoがベター)
 - 迅速な溶解性評価と分配係数logDによる経口吸収性の脂溶性が必要とされる
 - 精査段階で結晶からの溶解度が極めて低くなる場合がある
 - 結晶性や結晶形を収集することが重要で、結晶多形の事前評価が望まれる

※ 面による速度論の違いもあり平衡論だけでは難しい

amino acid	analog solute	amino acid	analog solute	amino acid	analog solute
Ala	methane	Val	propane	Leu	<i>iso</i> -butane
Ile	<i>n</i> -butane	Ser	methanol	Thr	ethanol
Phe	toluene	Tyr	<i>p</i> -cresol	Cys	methanethiol
Met	methyl ethyl sulfide	Asn	acetamide	Gln	propionamide
Trp	3-methylindole	Hid	4-methylimidazole	Hie	4-methylimidazole



厳密法に遜色無い精度を数十分の一の時間で達成する自由エネルギー計算手法

溶液・界面・生体分子系

統計力学に基づく計算

➤ 構造を表す分布関数や
ダイナミクス

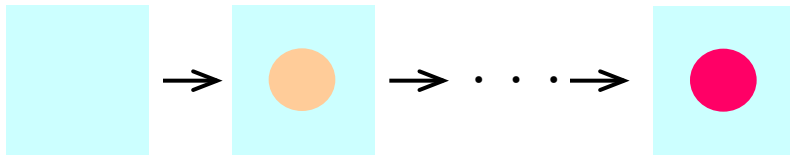
➤ **自由エネルギー**

高速で
正確な
計算

量子計算における
エネルギー

計算が難しい
(時間がかかる)

熱力学積分法 (厳密手法)



純溶媒

溶液系

純溶媒系と溶液系を結ぶ
(仮想的) 中間状態を (多数) 用意

MDシミュレーションと 溶液理論の結合手法

溶媒和自由エネルギー

$$\Delta\mu = \text{〇〇 kcal/mol}$$

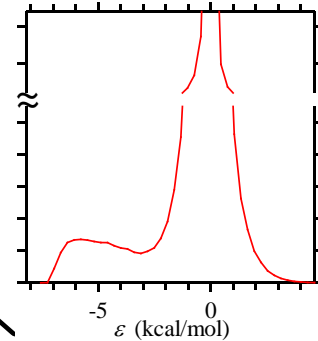
現実の
使用に
耐える
正確さ

近似汎関数

$$\Delta\mu = \int d\varepsilon \varepsilon \rho^\varepsilon(\varepsilon) - F[\rho^\varepsilon(\varepsilon)]$$

エネルギー分布関数

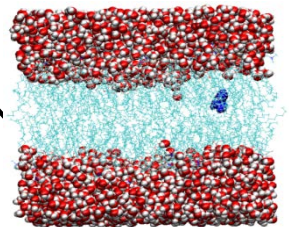
$$\hat{\rho}^\varepsilon(\varepsilon) = \sum_i \delta(\varepsilon - v(\psi, \mathbf{x}_i))$$



容易に
計算可能

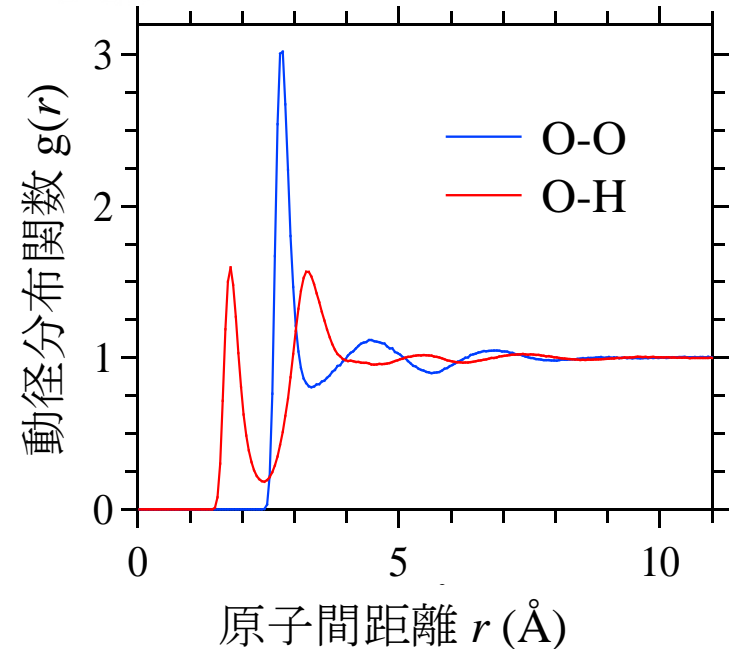
MD

(典型的に)

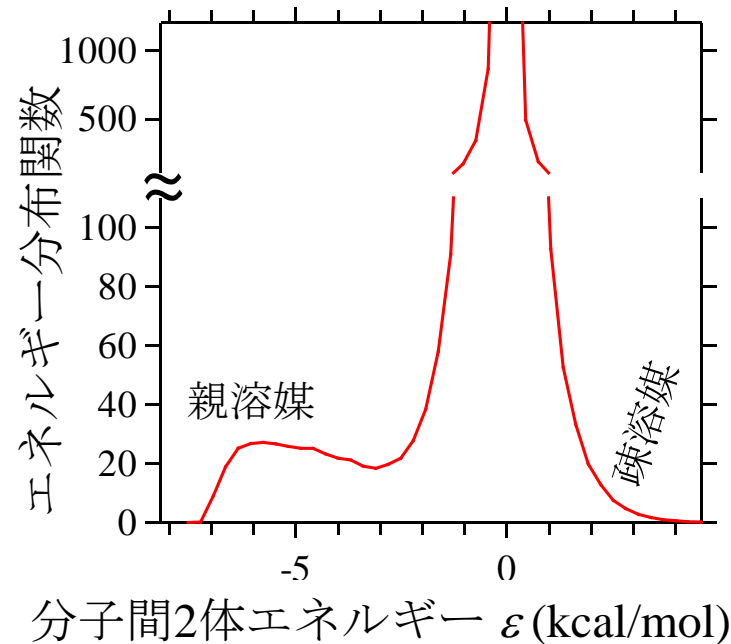
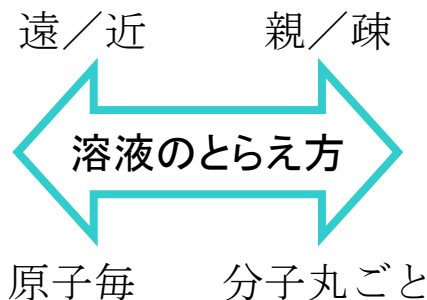


分子内・分子間相互作用

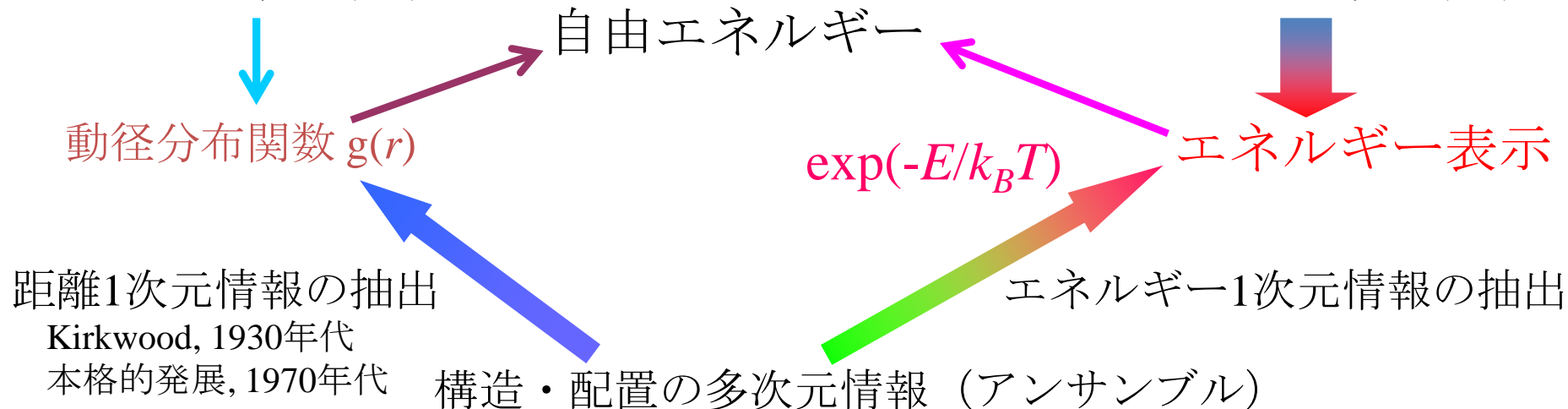
動径分布関数とエネルギー分布関数



1 g/cm³ · 25 °Cにおける
水の動径分布関数

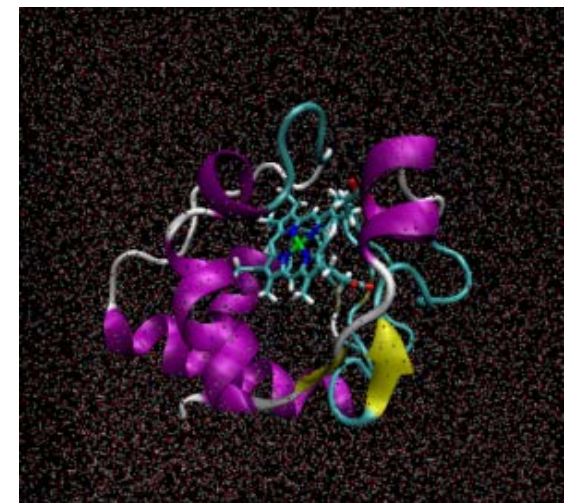


1 g/cm³ · 25 °Cにおける
水のエネルギー分布関数



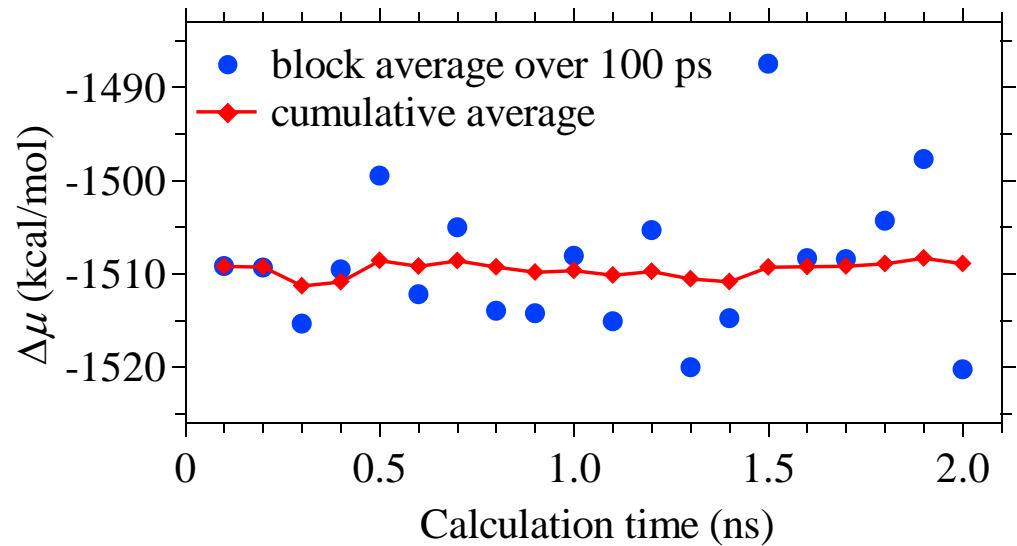
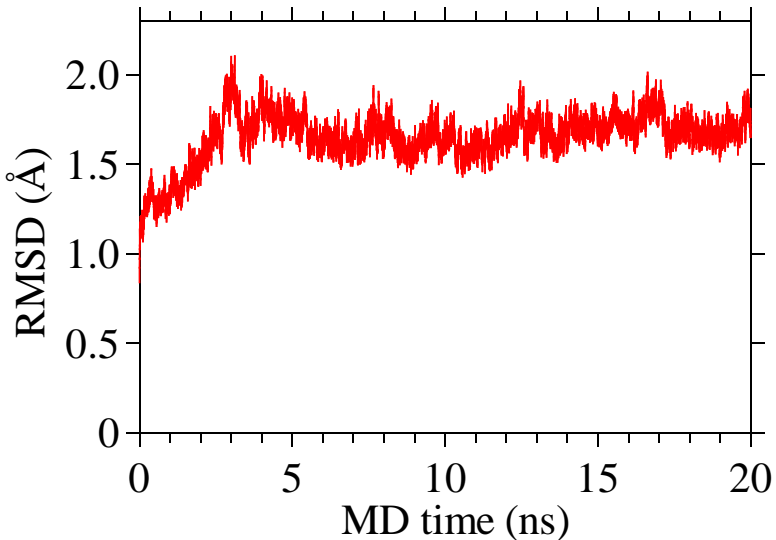


solute: horse heart cytochrome c (1HRC),
104 residues and heme, 1748 atoms,
total charge of +7
solvent: 20000 water molecules (TIP3P)
temperature: 300K



タンパク質の分子構造
ゆらぎ (RMSD)

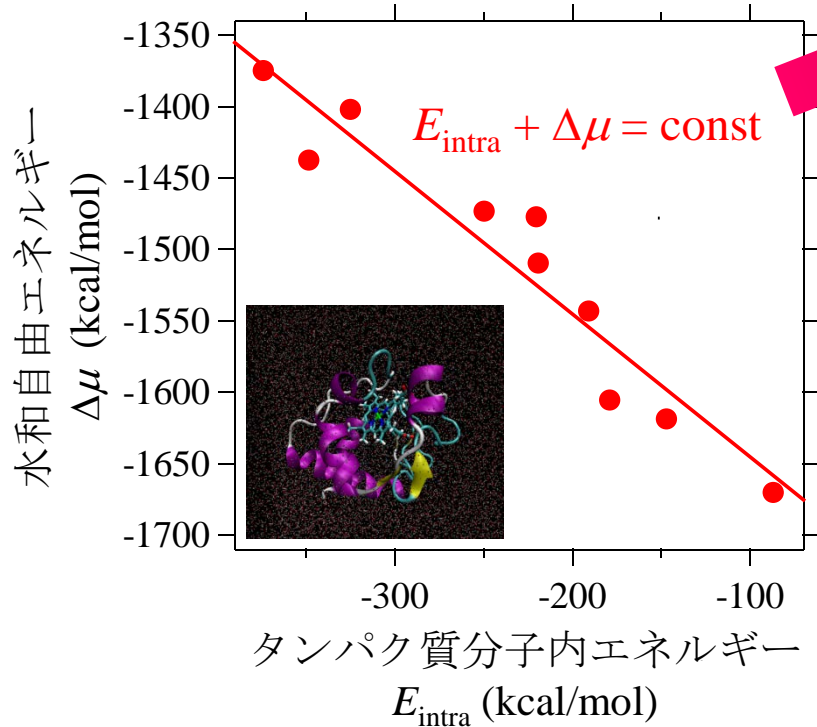
タンパク質固定構造 (配座) における
水和自由エネルギー $\Delta\mu$ の収束挙動



- ✓ 固定構造につき、1 nsで ~ 3 kcal/molの誤差で、水和自由エネルギーを計算
- ✓ タンパク質構造 (配座) と水和効果の関連
- ✓ タンパク質 (溶質) - 水 (溶媒) 相互作用の引力・斥力成分の役割

平衡ゆらぎにおける
タンパク質構造と水和の関係

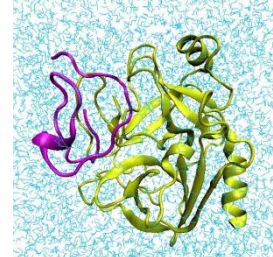
タンパク質分子内エネルギーと
水和自由エネルギーの補償



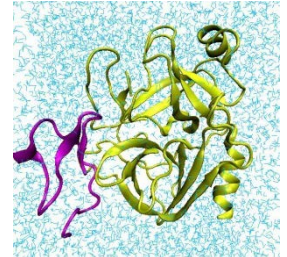
✓ スクリーニング／モデリングでは・・・
serine protease (1PPE)



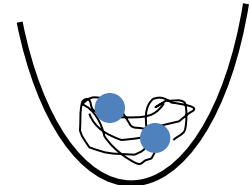
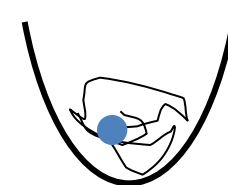
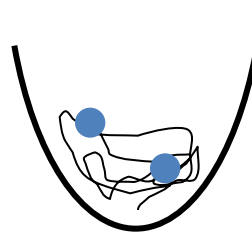
x-ray



rmsd 5.5 Å



rmsd 9.5 Å



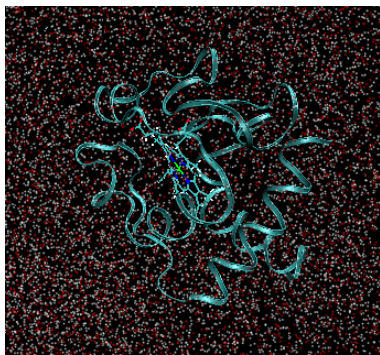
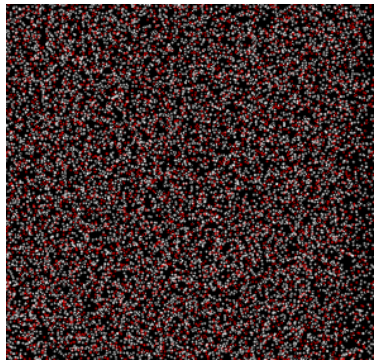
少数のスナップショット配置でスクリーニング

- ✓ 平衡ゆらぎにおける引力効果 (主に水素結合) の支配
- ✓ 大振幅運動における斥力効果 (主に排除体積効果) の効果

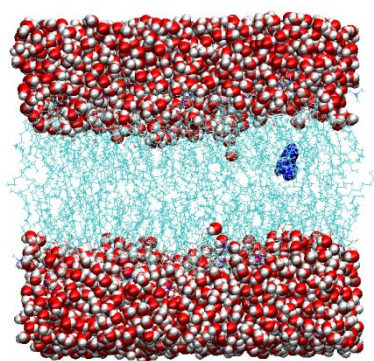
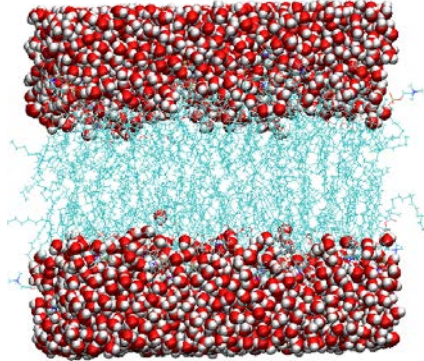
タンパク質の水和自由エネルギー解析が全原子レベルで可能に

「溶質」「溶媒」「溶媒和」概念の拡張 ～統一的な概念構成に向けて

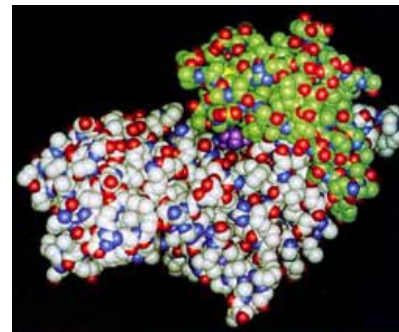
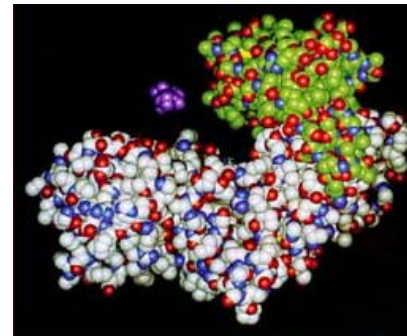
「溶媒」 = 溶液系に最初からあるもの
「溶質」 = 溶液系に後から入ったもの



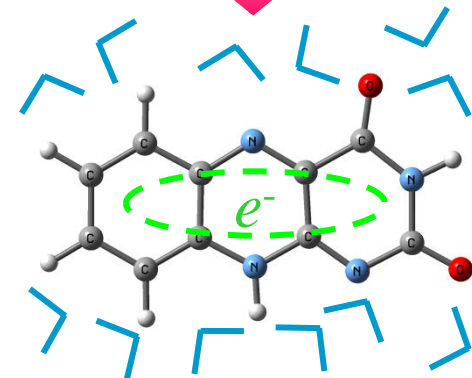
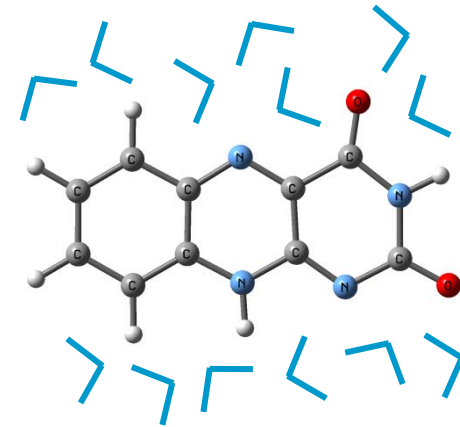
通常の意味での
溶媒和



脂質膜・ミセル
への分子結合



タンパク質への
基質結合



電子の付加
(還元)

溶液理論構成の要件

➤ 内部自由度のある分子

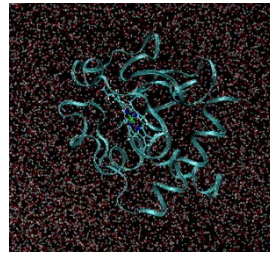
➤ 不均一系

➤ 混合溶媒

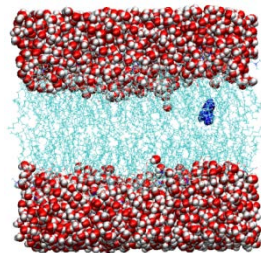
➤ 気体様低密度(1対1の「結合」)から
液体様高密度領域をカバー

➤ 量子論との結合(QM/MM法)

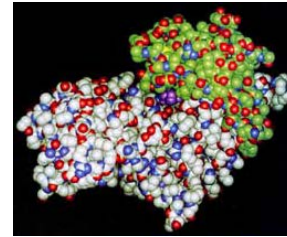
➤ 非物理的挙動を導かないような数学的に堅牢な定式化



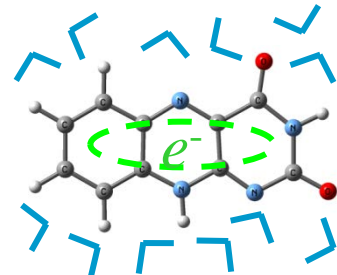
通常の意味での
溶媒和



脂質膜・ミセル
への分子結合



タンパク質への
基質結合



電子の付加
(還元)

スナップショット配置における2体エネルギー分布 (ヒストグラム)

$$\hat{\rho}^e(\varepsilon) = \sum_i \delta(\varepsilon - v(\psi, \mathbf{x}_i))$$

ψ : 溶質の座標

\mathbf{x}_i : i 番目の溶媒分子の座標

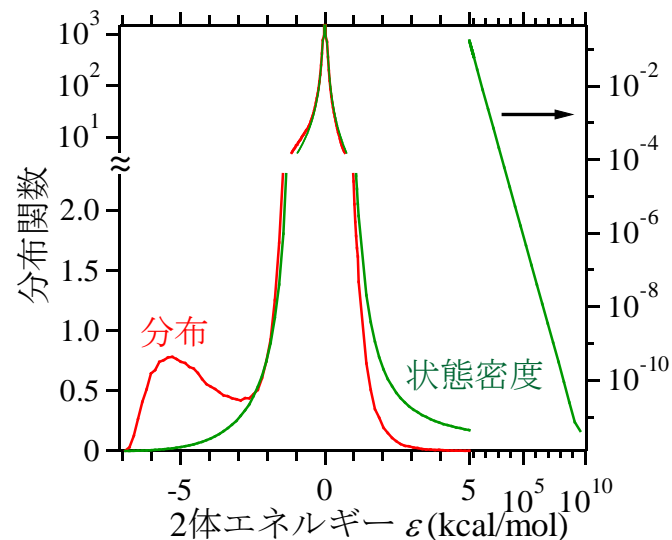
$v(\mathbf{y}, \mathbf{x})$: 溶質-溶媒間の2体相互作用

ε : 2体エネルギー値 (分布関数の横軸)

溶媒和自由エネルギー

$$\Delta\mu = -k_B T \log \left\langle \exp \left(-\beta \int d\varepsilon \varepsilon \hat{\rho}^e(\varepsilon) \right) \right\rangle_{\text{reference solvent}}$$

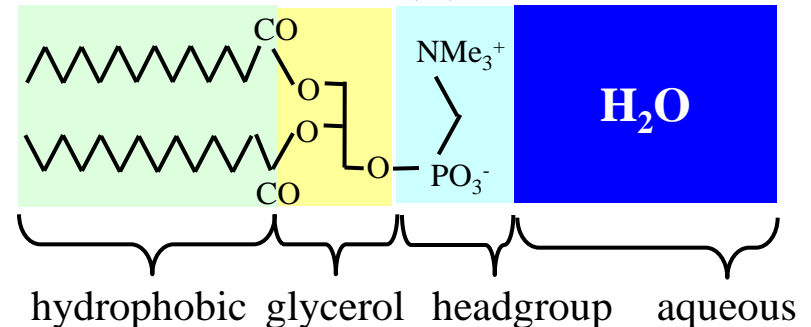
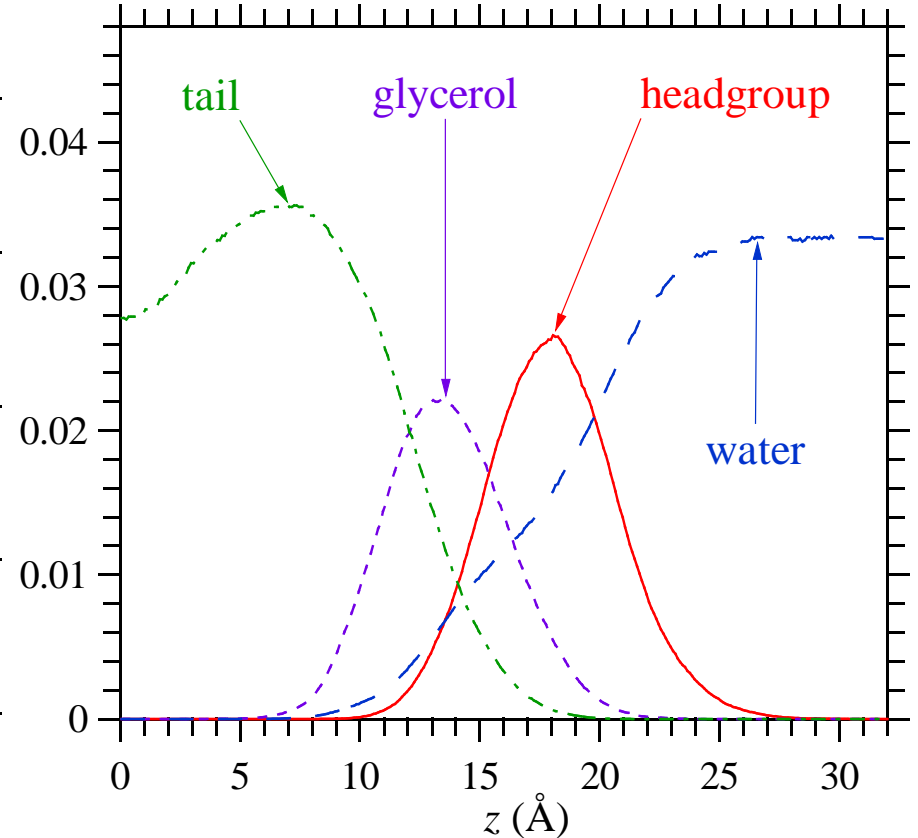
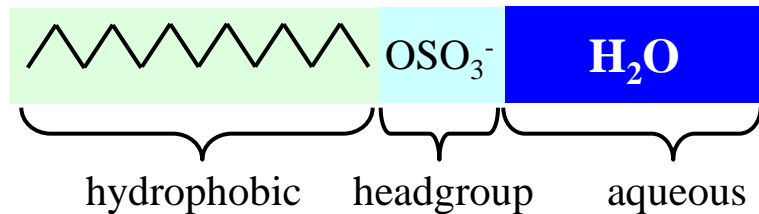
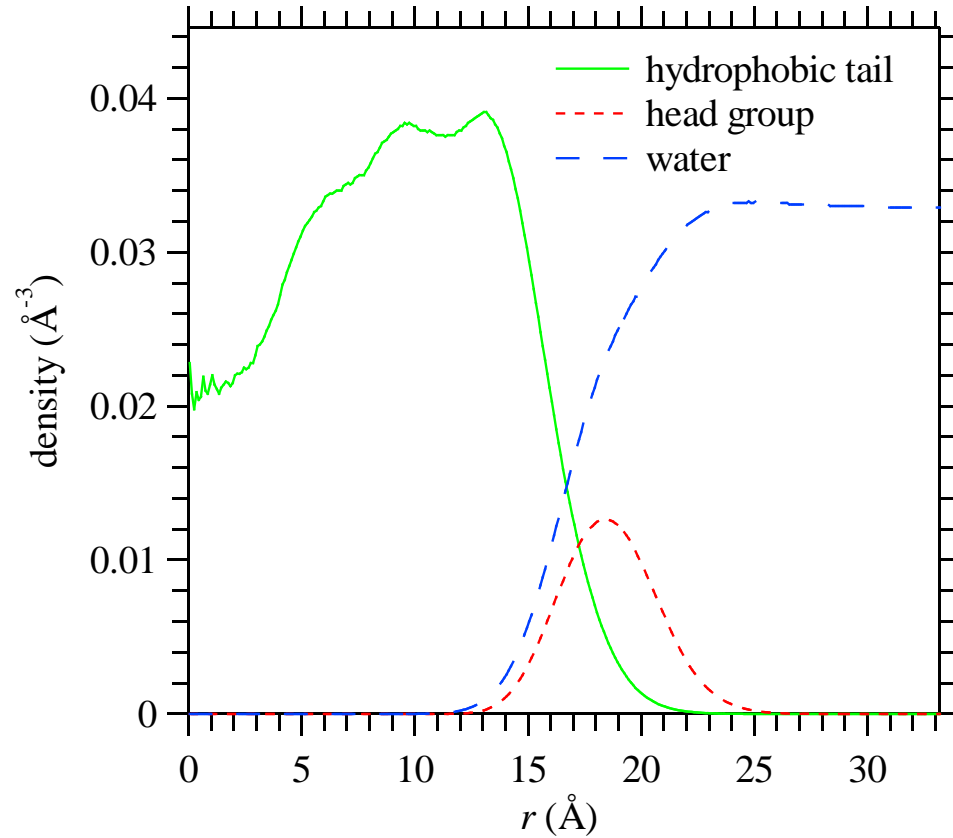
統計平均



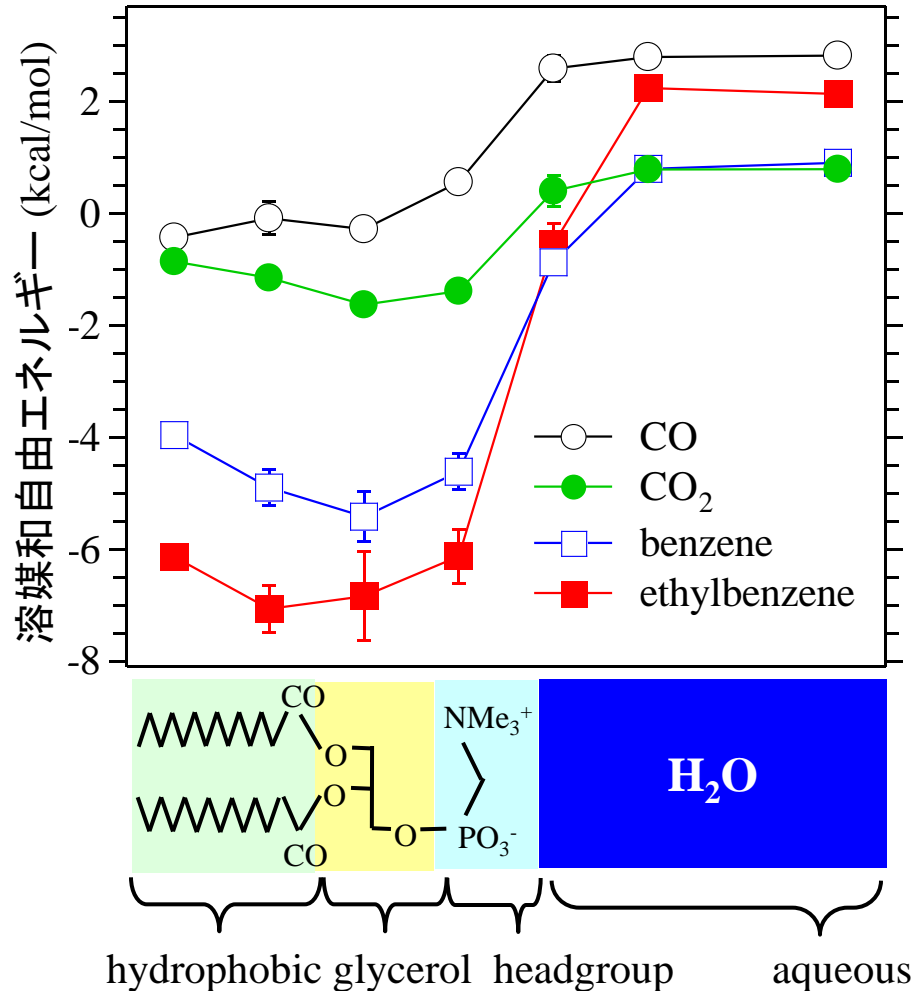
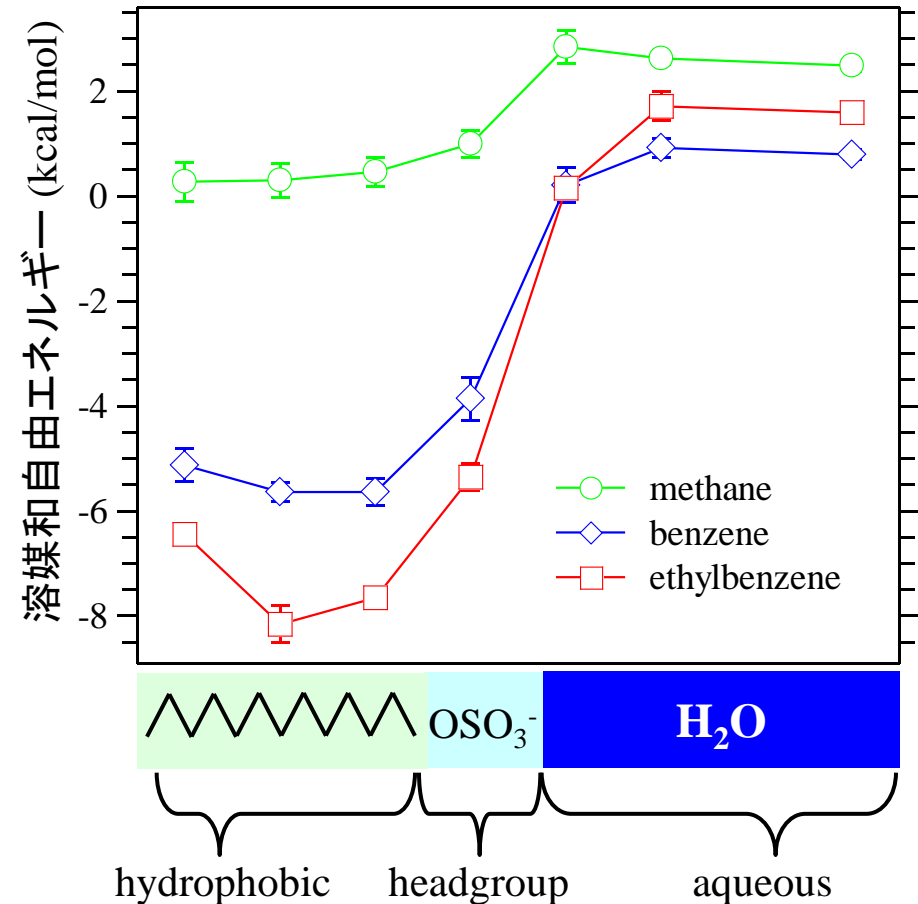
- ✓ 同じ溶質-溶媒相互作用を持つ配置(構造)をグループ化(射影)する
- ✓ どのような近似でも、同じ溶質-溶媒相互作用を持つ配置(構造)が溶媒和自由エネルギーに同じ重みで効くという統計力学の原理を破らない ⇒ 精度の高い計算
- ✓ どのようなポテンシャル関数に対しても1次元の座標 ⇒ 計算スピード向上
- ✓ 溶質および溶媒分子が、共に、全体として1つのものとして扱われる
 - ⇒ 内部自由度 ⇒ 外場 ⇒ 不均一系・QM/MM系
 - ⇒ 低密度/濃度極限 ⇒ 超臨界流体・混合溶媒
- ✓ DFT(密度汎関数理論)的な理論構成 ⇒ 系統的定式化

SDS (sodium dodecyl sulfate) ミセル

DMPC (1,2-dimyristoyl-*sn*-glycero-3-phosphatidylcholine) 二重膜



各領域における溶媒和自由エネルギーの算出

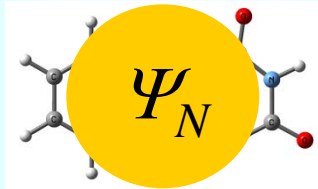


- ✓ 疎水性領域における顕著な安定化
- ✓ 疎水性・親水性領域の違いは、ミセルの場合により顕著
- ✓ 疎水性溶質の分布は、ミセル中でより局所的
- ✓ ミセル・膜内部と水領域での、溶質の安定性の逆転

分子結合の強度(分配係数)とサイトを決定可能
結合寿命も評価可能

還元過程の解析—電子1個を溶質と見る—

溶液内



酸化状態

溶質種: 電子

混合溶媒種: 電子が付加される分子(イオン)

(普通の意味での)溶媒(水など)

$$\varepsilon_{\text{QM}} = \langle \Psi_{N+1} | H_{\text{QM}} | \Psi_{N+1} \rangle - \langle \Psi_N | H_{\text{QM}} | \Psi_N \rangle$$

$$\varepsilon_{\text{MM}} = v(\Psi_{N+1}, \mathbf{x}) - v(\Psi_N, \mathbf{x})$$

$$\exp(-\beta\Delta\mu) = \left\langle \exp\left(-\beta\left\{\varepsilon_{\text{QM}} + \sum_i \varepsilon_{\text{MM},i}\right\}\right) \right\rangle_N$$

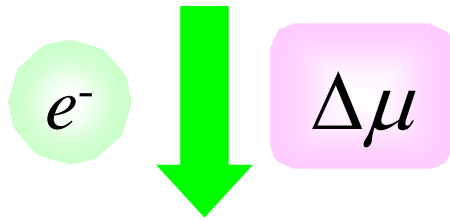
電子分布は、溶媒配置によって揺らぐ
エネルギー座標を構成する相互作用は多体的
 $\rho(\varepsilon_{\text{QM}})$ と $\rho(\varepsilon_{\text{MM}})$ (及び関連相関関数) から、 $\Delta\mu$ を構成

FAD (Flavin Adenine Dinucleotide) の活性部位(イソアロキサジン環) の還元への適用

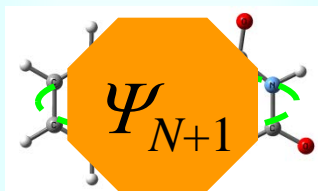
$$\Delta\mu_{\text{QM}} = -39.5 \text{ kcal/mol}$$

$$\Delta\mu_{\text{MM}} = -40.6 \text{ kcal/mol}$$

大きな水の効果(波動関数部分に匹敵)



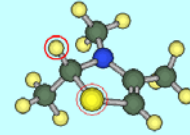
溶液内



還元状態

ERのまとめ

- Gromacs・NAMDに対応、Amberは対応中
 - GUIはフィアラックス社のMF Amberに移植予定
- 利点(積分方程式主体のRISMと比べて)
 - 自由エネルギーが定量的に計算できる。
 - 量子化学計算を用いた場合、電子雲を部分電荷に落とさず、電子雲のまま適用できる。
 - 分子間サイトの関係がexplicitな構造座標として表す必要がないので、サイト間の相互作用の近似が不要で、rigidでも溶質flexibleでもどちらでもよい。また、均一系であるか不均一系であるかを問わないので生体膜や界面活性剤にも適用できる。



J.Chem.Inf.Model.
2011, 51, 69-82

MMPBSA -> ER

C1の溶媒和自由エネルギー
(GAFF、溶質flexible、
Tip4p/amber99sb)

: -22.2422 ± 0.1073 kcal/mol

※計算時間は1化合物4コアで
4~5時間程度(5割がMD)

十分なサンプリングをしている

①c1 + 1000water = 1ns, 10fs毎に
snapshot ----> 10^5 frames

②1000water = 10ps, 10fs毎に
snapshot ----> 10^3 frames

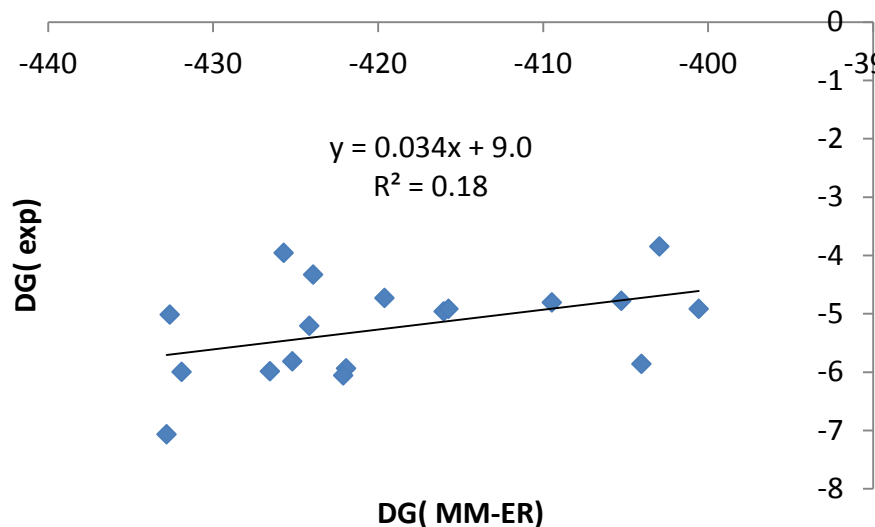
③c1単体 = 10ns, 10fs毎に
snapshot ----> 10^6 frames

**計算時間短縮は溶質rigidにする、
サンプリングを必要最低限に減らす、
MD & ERをGPUで動かす etc**

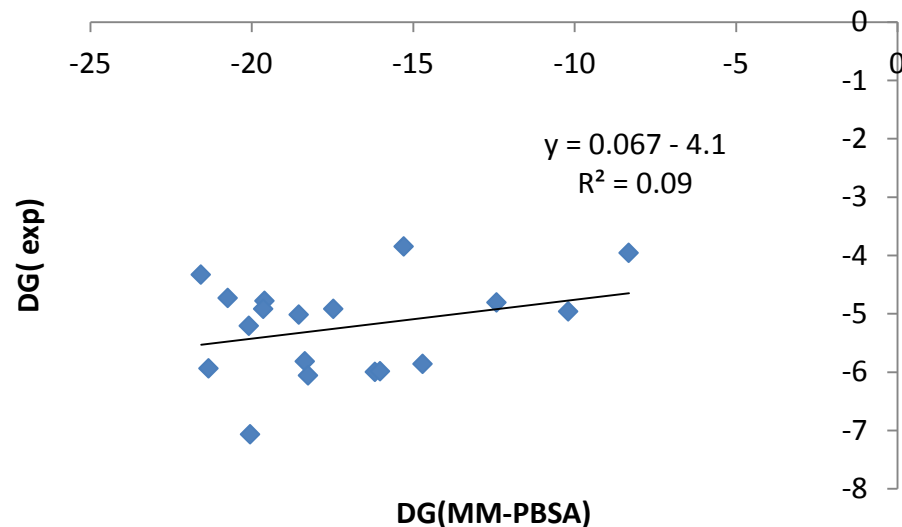
c1/1000water	: -22.2422 ± 0.1073
c2/1000water	: -22.4811 ± 0.1422
c3/1000water	: -27.9471 ± 0.2576
c4/1000water	: -25.2569 ± 0.1451
c5/1000water	: -39.5630 ± 0.3022
c6/1000water	: -33.8582 ± 0.2325
c7/1000water	: -33.5672 ± 0.2735
c8/1000water	: -31.5105 ± 0.1425
c9/1000water	: -32.3265 ± 0.2298
c10/1000water	: -4.1927 ± 0.1029
c11/1000water	: -27.4158 ± 0.1881
c12/1000water	: -32.0990 ± 0.1704
c13/1000water	: -31.8347 ± 0.2275
c14/1000water	: -31.0157 ± 0.1804
c15/1000water	: -37.2129 ± 0.2295
c16/1000water	: -31.3534 ± 0.1880
c17/1000water	: -32.7385 ± 0.1674
c18/1000water	: -34.5272 ± 0.2568

MM-ERの評価(1): MM-ERとMM-PBSAとの比較

検証: MM-PBSAのsolvationをERに変更することで、結果は改善するか?



J. Chem. Inf. Model. 2011, 51, 69–82の
PB部分をERに変更してプロット



J. Chem. Inf. Model. 2011, 51, 69–82

結果: MM-PBSAのsolvationをERに変更することで、結果は改善した

1. R2乗値は0.09から0.18に改善

2. 実験的に結合自由エネルギーが上位3個の化合物の順位は、
MM-PBSAは5位、10位、12位、MM-ERは1位、9位、3位と改善している。

Winmostarでsdfファイル読込

The screenshot displays the Winmostar software interface. On the left, a 3D ball-and-stick model of a molecule is shown. The main window contains a command window with the following text:

```

- Settings¥rkoga¥My Documents¥Gromacs-ER files¥comp.mr
Winmostar 17 120.15 0 -2.2343 -0.9637 1.1366
9-4-1-1 Leng=5.1835 Ang=11.3 Dihed=0 Lper=0 N

%nprow=1
!%chk=temp
# b3lyp/3-21g opt pop=fu

Winmostar
0 1

1 C 0 1 0
2 C 1.39929 1 0
3 C 1.39390 1 121.
4 C 1.39181 1 120.
5 C 1.39194 1 119.
6 C 1.39410 1 120.
7 C 1.40252 1 119.
8 N 1.38449 1 121.
9 N 1.36540 1 126.
10 H 1.10283 1 121.
11 H 1.10326 1 120.
12 H 1.10313 1 120.
13 H 1.10338 1 119.
9 N 1.365402 126.7

Debug 1 1 1
  
```

An "Animation" dialog box is open in the foreground, showing a list of atoms and their corresponding PDB IDs and MID values. The list includes atoms a1 through a11 and b1 through b11. The dialog also features playback controls such as "Rewind", "Last", "Slow", "Fast", "temp", "3D anime", "jpeg", "gif", "autorew", "3D", "Excel", and "Quit".

At the bottom of the screen, there is a footer with the text: "Chem-Suite ナューニク 版による自動Protonatio".

Gromacs-ERのバッチ処理(前半)

```
#!/bin/bash
```

```
mpd
acpype -i $1.pdb -n $2
cd ./${1}.acpype
cp ../c1.acpype/water.xtc ../c1.acpype/water.log .
cp ../c1.acpype/c1.mdp .
mv c1.mdp $1.mdp
cp ../c1.acpype/c1w_*.mdp .
cp ../c1.acpype/c1w.top .
sed s/c1/${1}/g c1w.top > $1w.top
rm -f c1w.top
mv c1w_nvt.mdp $1w_nvt.mdp
mv c1w_npt.mdp $1w_npt.mdp
mv c1w_run.mdp $1w_run.mdp
grompp -maxwarn 10 -f $1.mdp -c $1_GMX.gro -p $1_GMX.top -o $1.tpr
mpiexec -l -n 4 mdrun -s $1.tpr -c $1.gro -x $1.xtc -g $1.log
genbox -cp $1_GMX.gro -box 3.2 3.2 3.2 -cs tip4p.gro -maxsol 1000 -o $1w.gro -p $1w.top
grompp -f $1w_nvt.mdp -c $1w.gro -p $1w.top -o $1w_nvt.tpr
mpiexec -l -n 4 mdrun -s $1w_nvt.tpr -c $1w_nvt.gro
grompp -f $1w_npt.mdp -c $1w_nvt.gro -p $1w.top -o $1w_npt.tpr
mpiexec -l -n 4 mdrun -s $1w_npt.tpr -c $1w_npt.gro
grompp -f $1w_run.mdp -c $1w_npt.gro -p $1w.top -o $1w_run.tpr
mpiexec -l -n 4 mdrun -s $1w_run.tpr -c $1w_run.gro -x $1w_run.xtc -g $1w_run.log
```

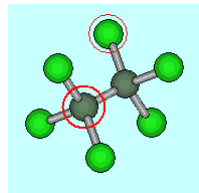
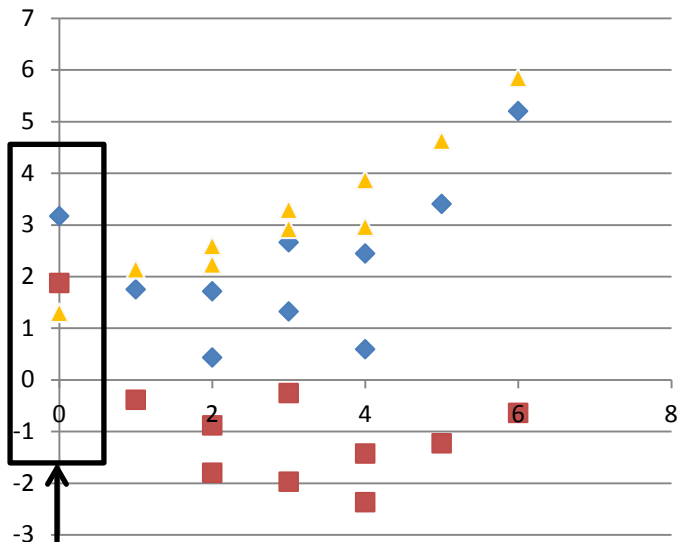
Gromacs-ERのバッチ処理(後半)

(前ページの続き、、、)

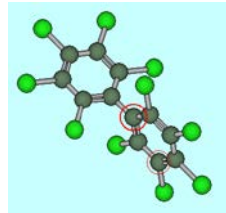
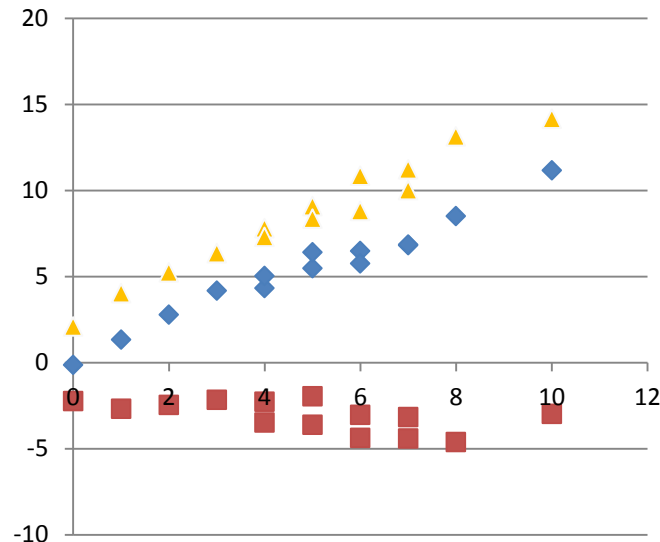
```
mpiexec -l -n 4 mdrun -s $1w_run.tpr -c $1w_run.gro -x $1w_run.xtc -g $1w_run.log
# ../../tools/gromacs/gen_structure -l/opt/gromacs-4.5.3/share/top -t $1w.top
$HOME/enerepr-0.1.214/tools/gromacs/gen_structure_2 -l/opt/gromacs-
4.5.3/share/top -t $1w.top $1
cd soln
$HOME/enerepr-0.1.214/tools/gromacs/gen_input -x ../$1w_run.xtc --
log ../$1w_run.log
$HOME/enerepr-0.1.214/ermod_gromacs
cd ../refs
$HOME/enerepr-0.1.214/tools/gromacs/gen_input --traj ../water.xtc --log ../water.log
--flexible ../$1.xtc
$HOME/enerepr-0.1.214/ermod_gromacs
cd ../
$HOME/enerepr-0.1.214/slvfe > slvfe_$1.log
cd ../
```

※enereprは京大版ERソースコード(GPLで3月末公開予定)

MM-ERの評価(2): SAMPLのchlorideの評価

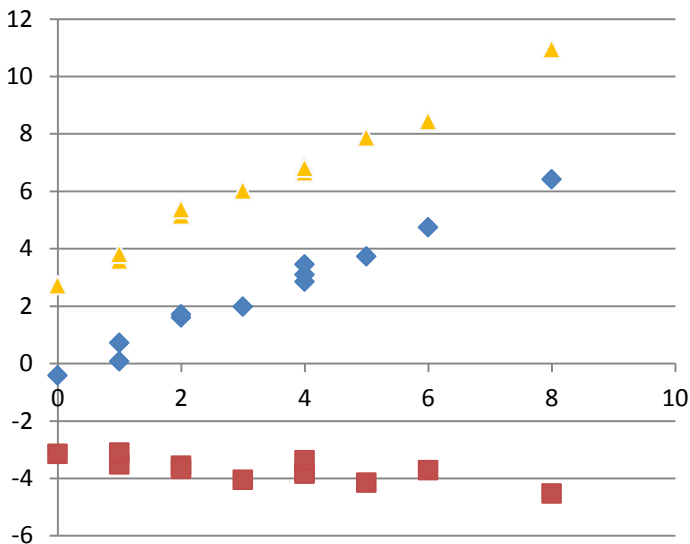


◆ ethane_calc
■ ethane_expt
▲ ethane_calc-expt



◆ biphenyl_calc
■ biphenyl_expt
▲ biphenyl_calc-expt

ethaneのみOPLS-AA



◆ dioxane_calc
■ dioxane_expt
▲ dioxane_calc-expt

横軸: chlorideの数

縦軸: kcal/mol

Gromacs 4.5.3、力場Amber-GAFF
※ethane単体(横軸0)だけOPLS-AA
+ R.E.DとGAMESS-USで電荷を決めることで精度が改善した。

結果は、Clが無い場合に自由エネルギーがプラスに出ており、その後、Clの数にほぼ比例して更に増えていく。これに対して、実測ではClの数への依存性は小さい。

MM-ERを使ってみて

- MDをGromacsで行うことでトラジェクトリを得て、それを使い、複数のリガンドセットやSAMPL(オープンアイ社による溶媒和自由エネルギーコンテスト)の提供化合物群を用いて、ERで溶媒和自由エネルギーを得た。
- $\Delta\Delta G$ の比較のためレンジも違うし全うな評価はできないが、そもそも化合物／蛋白質間の相互作用のところで誤差があり、精度向上には限界があるので、MM-PBSAよりMM-ERの方が改善されたといえる。
- SAMPLで活用したAmber-GAFFでは、Clが、単なる斥力源になっている。原因について。Amber-GAFFのLJパラメータの元になったOPLS-AAのChlorobenzeneのパラメータを見ると、Clに非常に強い-0.180の電荷が付けられている(opls_264)。つまり、OPLS-AAでは、Clに強い負電荷を載せることを前提に設定したためLJが弱くなっており、そのままパラメータを持ち込んだので、電荷決定をRESPで行うAmberと噛み合っていないことが推測される。

力場の問題であり、ERの理論自体の問題によるものではない。

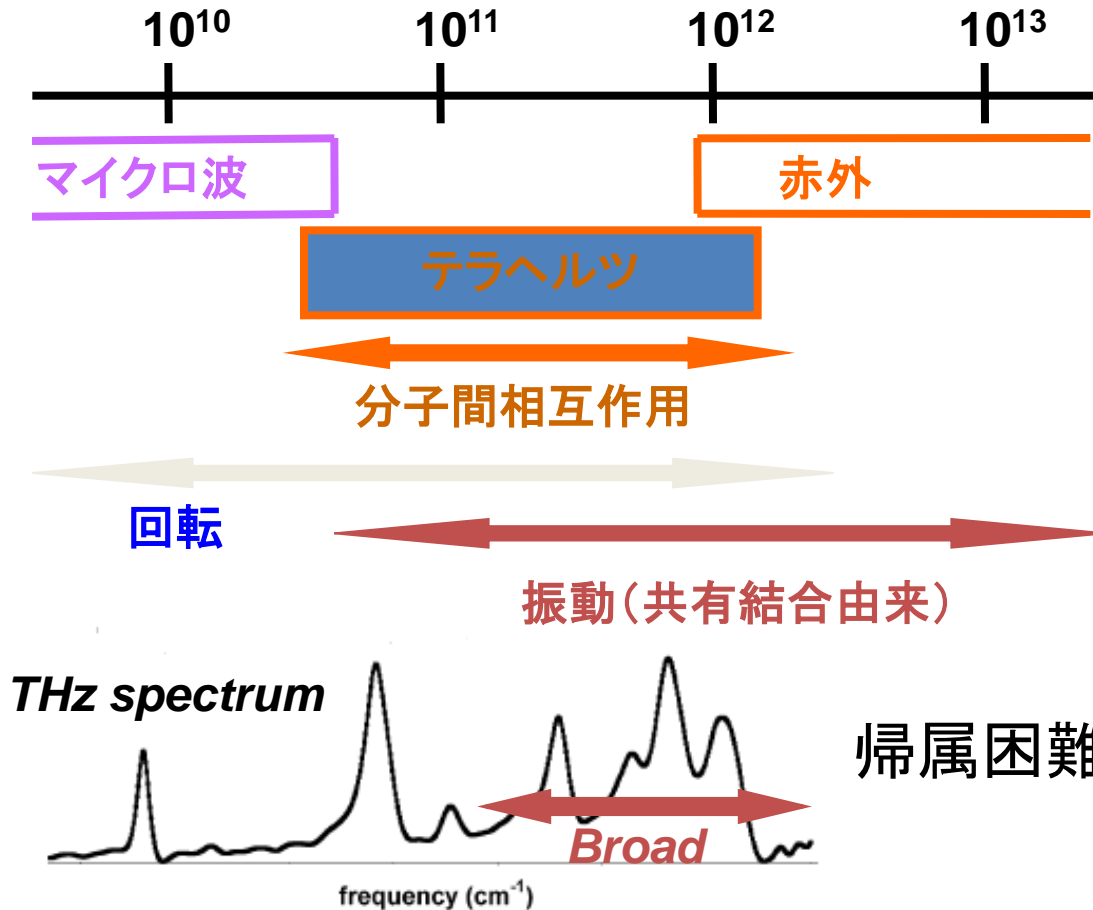
ERの欠点

- FEPに比べて適用範囲が狭い
 - 現時点でリガンドバインディングはうまくいっていない(サンプリングスキームの改善を検討中)
 - 現状、一番良い結果で $R^2=0.4$ 程度
 - ρ と ρ_0 が重なっていないと使えない(粗視化モデルとの融合により解決可能?)
- 溶媒和自由エネルギーしかやらない
- MDの欠点を改善するものではなく、引き継ぐ
 - MDとの組み合わせによる計算手法である
- 理論がちょっと難しい
 - ただ、理論を理解しなくても使える

固体の溶解度

- 結晶多形の溶解をERをつかって計算できるかもしれないが、速度論の扱いが必要かもしれない。研究中。
- 低分子の溶解度よりも実用上では重要かもしれない。
- テラヘルツスペクトルと基準振動解析（量子化学計算）の比較によるペレット（混合結晶）の構造予測が可能。

(以下、参考)テラヘルツ領域の分光 物質の分子間相互作用に由来する未開拓領域



2008.9.18
11 企業 ★1.4版

【第三種郵便物認可】

大半の物質を手り扱わずに、ごく少量の薬物でも正確に種類を特定できる。こんな特性を持つ「テラヘルツ波」と呼ぶ電磁波が活躍し始めた。爆薬や毒物の検査をはじめ、創薬や次世代薬の開発への応用が期待されている。この分野では日本企業が世界の先導を走っている。海外勢の注目を集めている。

大塚製薬が特殊な分子構造にこだわりの詰め込まれ、突然、装置が警報を発した。封書に麻薬が隠されていた。理化学研究所の大谷行リーダールは「薬や毒物の検査をせずに突止められる。封書中の粉末の正体を分析する。テラヘルツ波の非破壊測定が鍵を握った。

テラヘルツ波は大半の物質を手り扱わずに、ごく少量の薬物でも正確に種類を特定できる。こんな特性を持つ「テラヘルツ波」と呼ぶ電磁波が活躍し始めた。爆薬や毒物の検査をはじめ、創薬や次世代薬の開発への応用が期待されている。この分野では日本企業が世界の先導を走っている。海外勢の注目を集めている。

テラヘルツ波、日本勢が先行

理化学研究所はテラヘルツ波が非破壊で薬物検査に利用できることを実証し、世界に先駆けて「テラヘルツ波」による検査装置を開発した。西沢潤一東北大学教授

テラヘルツ波は電磁波の一種で、金属以外の大半の物質を透過する。波長が0.1mmから1mmの範囲にあり、通常の電磁波よりもエネルギーが低く、人体に安全とされている。この波長は、創薬や次世代薬の開発への応用が期待されている。この分野では日本企業が世界の先導を走っている。海外勢の注目を集めている。

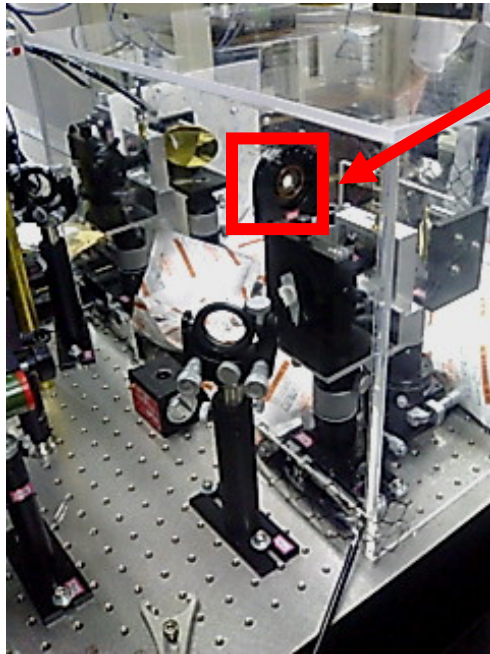
キャノン、微量物質を特定

キャノンは、創薬や次世代薬の開発への応用が期待されている。この分野では日本企業が世界の先導を走っている。海外勢の注目を集めている。

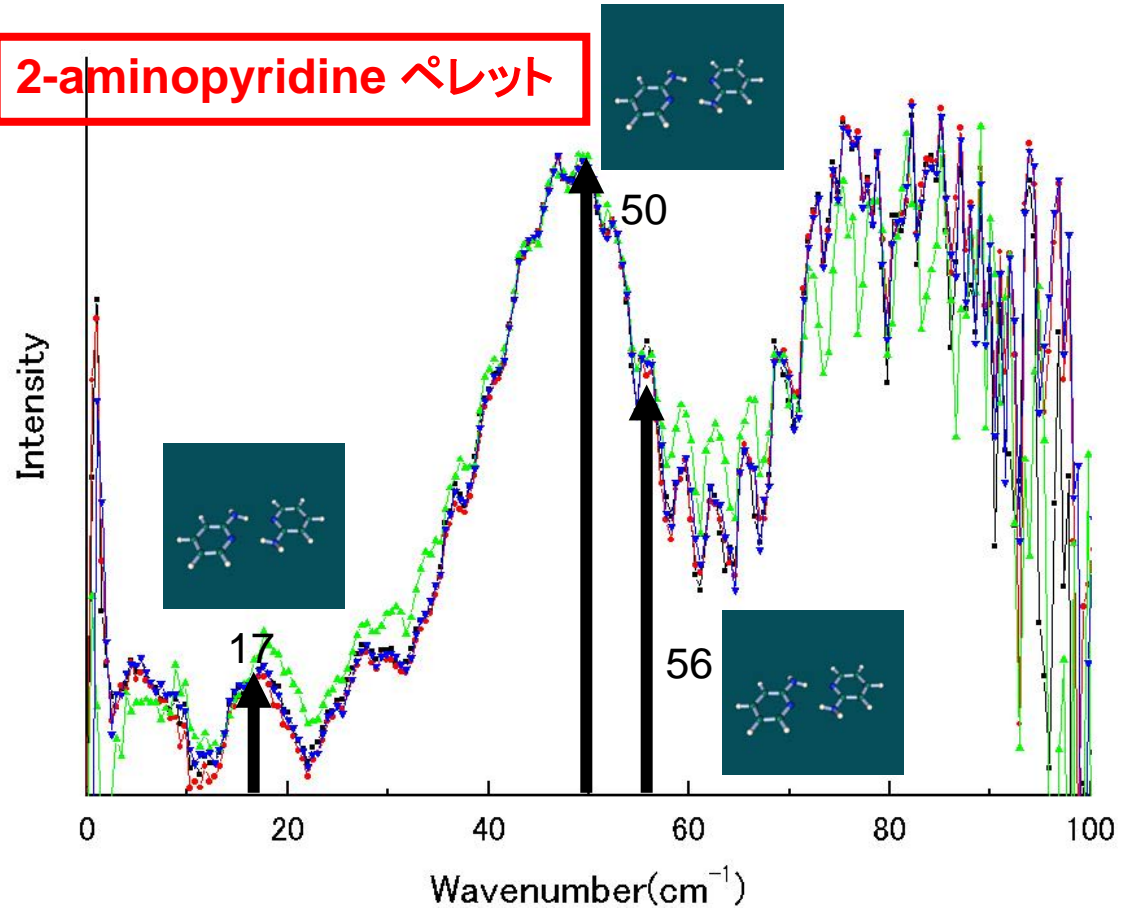
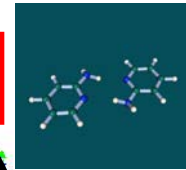
キャノンは、創薬や次世代薬の開発への応用が期待されている。この分野では日本企業が世界の先導を走っている。海外勢の注目を集めている。

キャノンは、創薬や次世代薬の開発への応用が期待されている。この分野では日本企業が世界の先導を走っている。海外勢の注目を集めている。

テラヘルツ領域における基準振動解析



2-aminopyridine ペレット



平面構造の基準振動(cm^{-1})

Twist 19.8

Out-of-plane bend 59.5

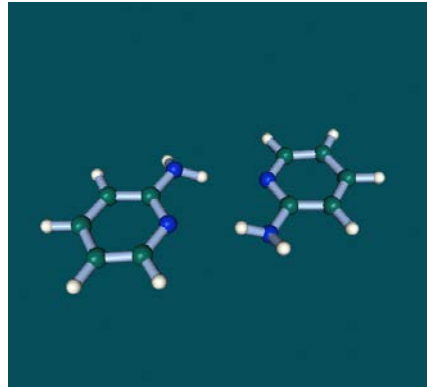
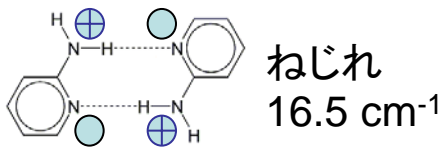
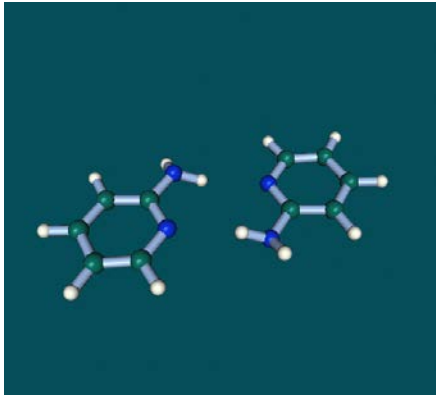
Asymmetric stretch 66.0

-> ($\times 0.847$)

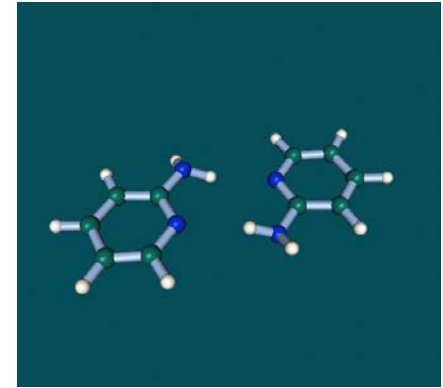
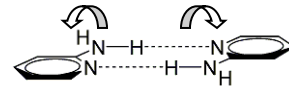
16.8, 50.4, 55.9

量子化学計算による基準振動解析の結果
(孤立系)と、実験値の比較

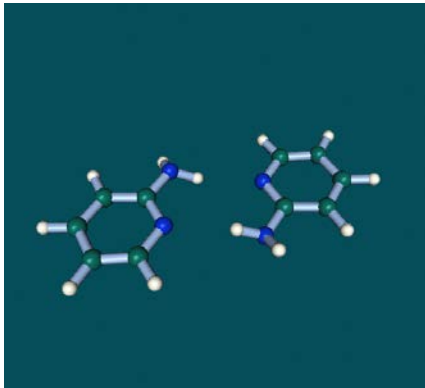
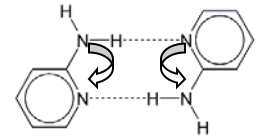
テラヘルツ領域における基準振動解析



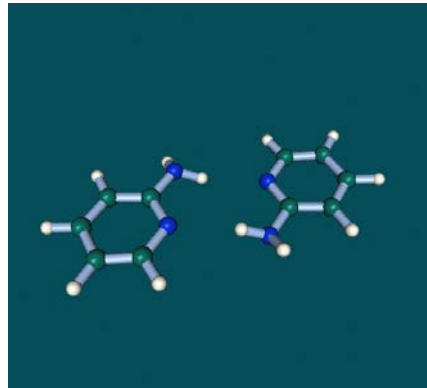
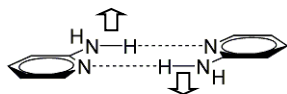
面外変角 28.4



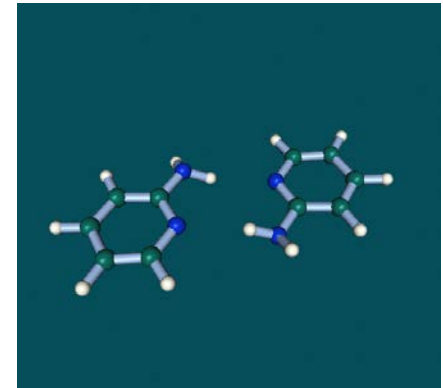
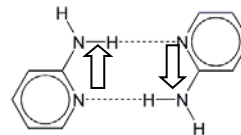
非対象伸縮
63.6



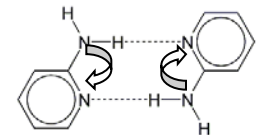
面外ずり 67.7



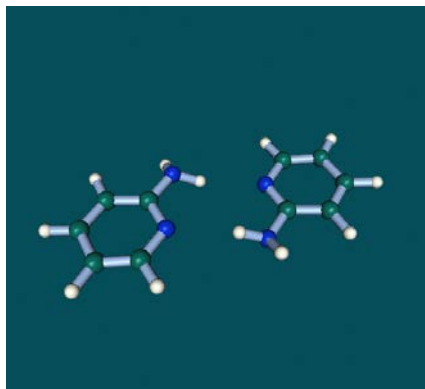
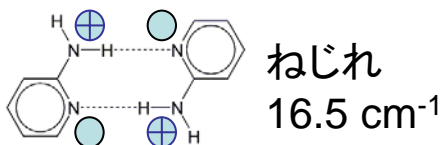
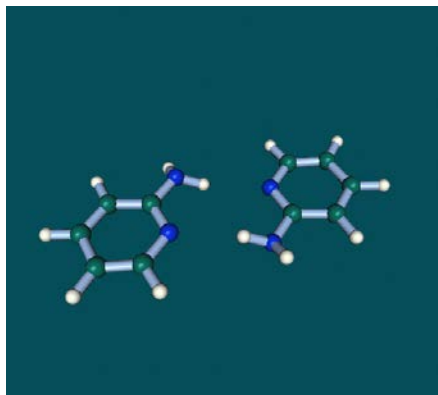
面内変角／
ずり 77.4



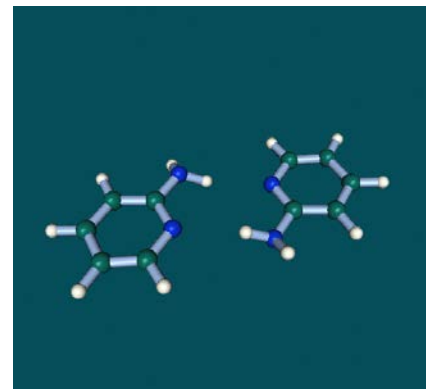
面内伸縮／
変角 109.4



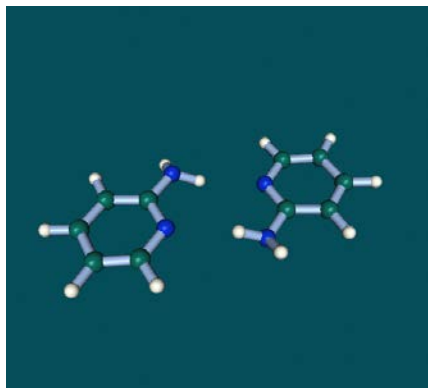
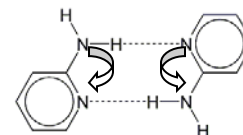
テラヘルツ領域における基準振動解析



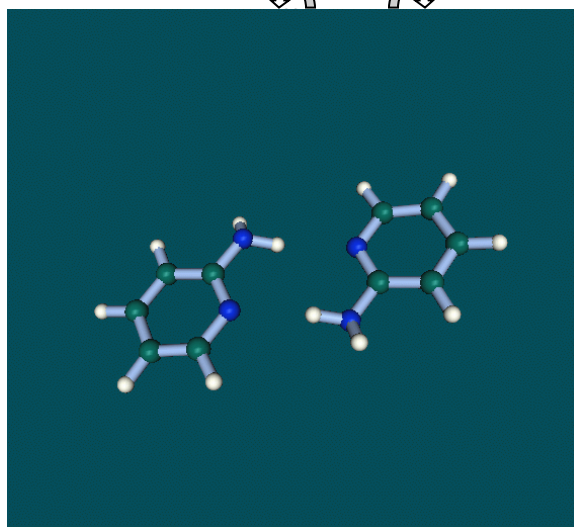
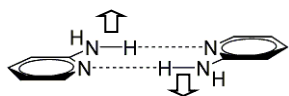
面外変角 28.4



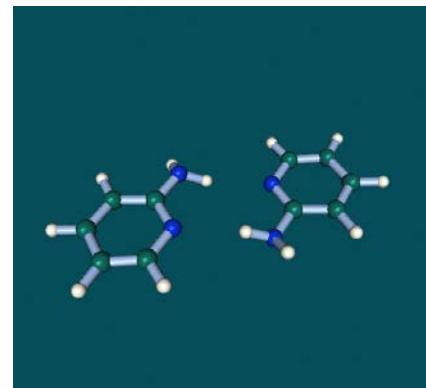
非対象伸縮
63.6



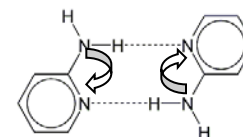
面外ずり 67.7



赤外領域の局所モード
(アミノ基伸縮) 3905.5



面内伸縮/
変角 109.4



既存手法の補完

分子間相互作用および構造解析で使われているサービス

- **Biacore (表面プラズモン共鳴)**
相互作用強度、結合速度 (ノンラベリング)
- **プロテイン・チップ (マイクロアレイ)**
ハイブリダイゼーションの程度を蛍光強度で測定
- **ファーマコフォア・モデリング**
化合物データベースを用いて、リガンドータンパクの構造サンプリング



- **リプレイス・補完例**

既存のデータベース、もしくは計算で構築する化合物情報データベースに基づき、テラヘルツスペクトルを入力として、相互作用部位の挙動、曲率計算 (二次微分) による相互作用強度計算、周辺環境情報を出力する。計算だけで安価。



光毒性予測/VCDによる構造決定

- 光毒性予測は第一励起エネルギー、VCDによる構造決定はVCDスペクトルを計算する。
- 計算結果のコンフォメーション依存性が大きいために、多数のコンフォメーションを発生させ、最安定なコンフォメーションを見つける必要がある
- そのためには配座探索と高速な量子化学計算が不可欠である
- 計算化学の専門家ではなく、安全性・物性の専門家が担当するケースも多く、windowsのUIを含めたユーザビリティが重要になる

The screenshot displays the X-Ability software interface. The main window shows a ball-and-stick model of a fluoroketone molecule. A terminal window in the foreground contains the following input and output:

```

Tencube/WM V3.805q
ファイル(F) 編集(E) 表示(V) 計算(C) 計算2_ その他(Z) ヘルプ(H)
Add Del -CH3 -C2H3 -C6H5 -CH3 Rep H Chngt BS2 1.15 Connect
C:\Documents and Settings\rkoga\My Documents\fluoroketo
Winmostar 33 22029 0 -0.084 1.9823 -1.6077
33-33-33-2 Leng=0 Ang=0 Dihed=0 Lper=-999 H
Number All Atoms Mark
Zoom 1
Atom 0.25
Bond 10
undo
%nproc=4
#p b3lyp/6-31g* opt=(maxcyc=100) pop=full gprint
freq=vcd
Winmostar
0 1
21 H 1.09662 1 110.6344 1 -58.2401 1 20 3
22 H 1.09488 1 111.391 1 61.7927 1 20 3
23 H 1.095
24 H 1.097
25 H 1.099
26 C 1.515
27 H 1.099
28 H 1.094
29 H 1.092
30 H 1.097
31 H 1.097
32 H 1.096
33 H 1.095
33 H 1.095
Debug
47 0.3892
56 1.0298
96 3.4724
115 0.5296
176 2.5726
200 0.1490
207 0.4932
246 0.4127
256 1.8638
274 2.1170
277 0.7762
304 0.1925
333 2.7952
385 3.4816
378 0.5112
414 0.8586
421 0.5584
458 13.2041
  
```

The IR/VCD Spectrum window shows a plot of intensity versus wavenumber (cm⁻¹) from 4000 to 500. The plot includes a red trace for the IR spectrum and a blue trace for the VCD signal. The X-axis is labeled 'X-min 0 X-max 4200' and the Y-axis is labeled 'Y-Level 0.5 0.0 1/cm 0.0000'.

The terminal window also shows a list of atoms and their coordinates:

```

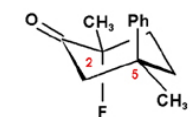
19 ! A2 A(2,1,12)
20 ! A3 A(10,1,12)
21 ! A4 A(1,2,3)
22 ! A5 A(1,2,4)
23 ! A6 A(3,2,4)
24 ! A7 A(2,4,5)
25 ! A8 A(2,4,6)
26 ! A9 A(5,4,6)
27 ! A10 A(4,6,7)
28 ! A11 A(4,6,8)
29 ! A12 A(7,6,8)
30 ! A13 A(6,8,9)
31 ! A14 A(6,8,10)
32 ! A15 A(9,8,10)
33 ! A16 A(1,10,8)
34 ! A17 A(1,10,11)
35 ! A18 A(8,10,11)
36 ! A19 A(1,12,13)
37 ! A20 A(1,12,15)
38 ! A21 A(13,12,15)
39 ! A22 A(12,13,14)
40 ! A23 A(12,15,16)
41 ! A24 A(12,15,17)
42 ! A25 A(16,15,17)
  
```

At the bottom, there is a control panel with buttons for 'stat-a', 'stat-s', 'stat', 'stat-Q', 'psall', 'ps', 'top', 'nodes', 'sendsub', 'send', 'Delete *chk', 'sendrun', 'ls', 'qsub', 'run', 'get', '*fchk', 'fluoroketo', 'cat', 'Edit log', 'import', 'qdel', 'kill', 'Quit', and 'clear'.

computed for all relevant conformations. Finally, these best tradeoff between accuracy and computational cost B3LYP or B3PW91—with the 6-31G(d) basis set (a

EXAMPLE: FLUOROKETONES

As a first example, we'll look at the IR and VCD spectrum.



光毒性もVCDもQMで同時計算可

- HOMO-LUMO Gapでdescriptorを扱う報告あり
 - 光毒性には第一励起エネルギーが関係している
 - 計算コストの観点から第一励起エネルギーと相関値の出るHOMO-LUMO Gapを採用
 - ただ、圧倒的に第一励起状態はHOMO-LUMOから構成される化合物のみ有効

Journal of Molecular Structure: THEOCHEM 719 (2005) 191-200

- 第一励起エネルギーが最小コストで計算できるTDDFT計算が必要と考えているが、芳香族系炭化水素ならAM1-ZINDOがベター。

半経験的による励起エネルギー計算

- 最も高速でそれなりなのはoptをAM1、励起エネルギーをZINDOにする。Gaussianの入力例

```
%chk=temp.chk
```

```
#n AM1 Opt
```

```
--Link1--
```

```
%chk=temp.chk
```

```
#n ZINDO Geom=Check td=(nstates=5)
```

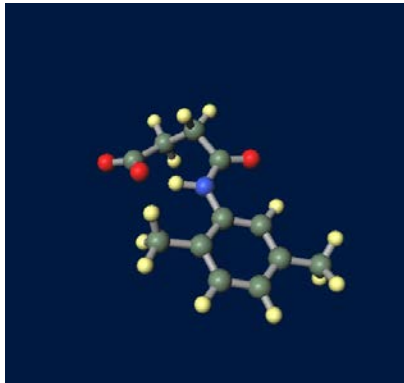
※おそらくMOPACの方が高速

- 半経験的方法だとパラメータがない場合があるし、あってもNa以降の原子を含む系では構造最適化の精度もガクンと落ちる。

QMによる励起エネルギー計算

TDDFT計算 (光毒性スクリーニング用ナミキライブラリの一部)

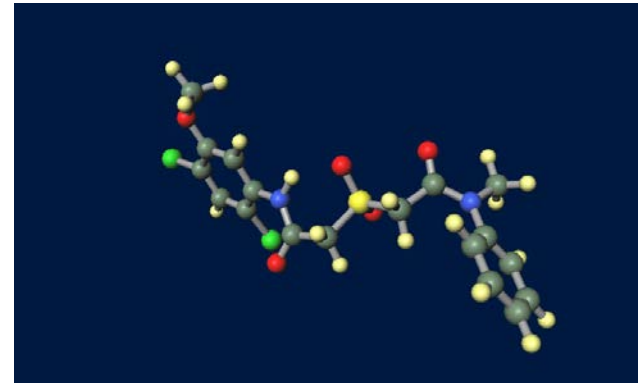
B3LYP/6-31G* td=(nstates=10)の第一励起エネルギー



4.5634(eV)



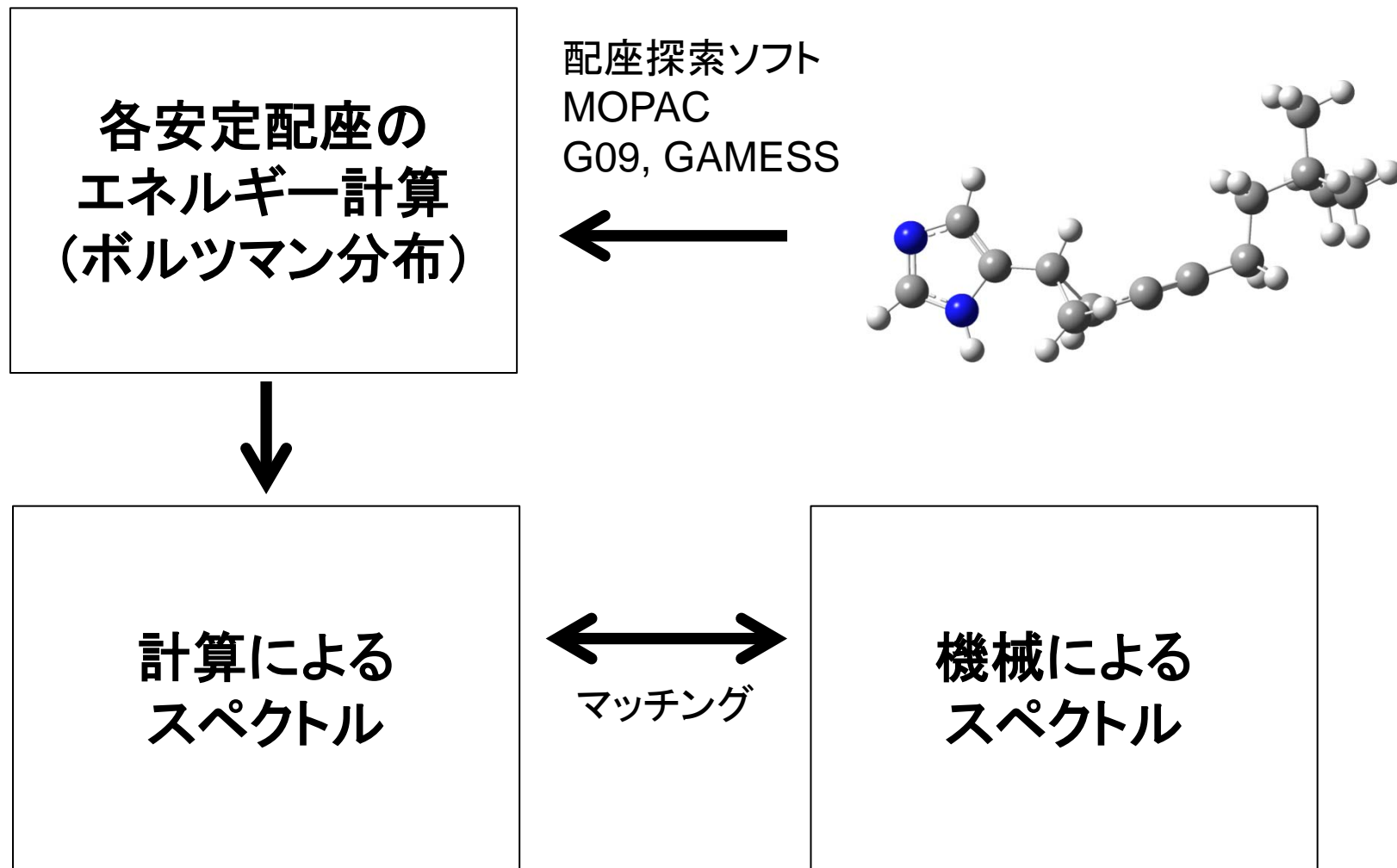
2.7711(eV)



3.7267(eV)

Journal of Molecular Structure: THEOCHEM 719 (2005) 191-200では、 7.2 ± 0.7 (eV)でHOMO-LUMO gapのdescriptorを作ったとあるので、励起エネルギーとの絶対値とはダイブ違う値。

VCDによる絶対構造推定の流れ



VCDによる絶対構造推定

- X線で絶対構造はわかるが、X線解析しづらいサンプルに対してVCDは有効
- Gaussian社のウェブサイトに記述あり
http://www.gaussian.com/g_whitepap/vcd.htm
 - 標準物質からの相対エナンチオマー純度の決定
 - 絶対構造決定
 - 溶液中の配座決定



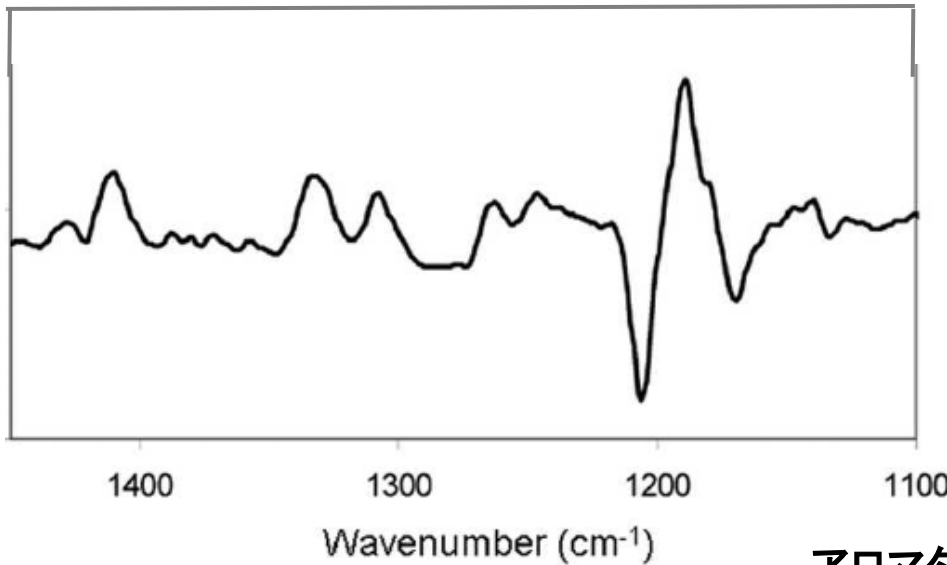
Biotoools社、VCD計測機

エナンチオマーとジアステレオマー

- エナンチオマー間でエネルギー差はない、VCDで区別するしかない。
- ジアステレオマーと幾何異性体(cis/trans)はエネルギー差がある。
- IRでジアステレオマー、VCDでエナンチオマーを区別するため、QMは有効だとGaussian社は言っている。

http://www.gaussian.com/g_whitepap/vcd.htm

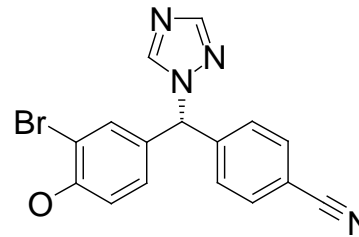
某社のブラインドテスト(1)



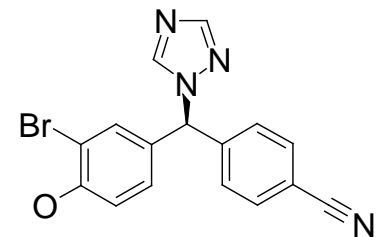
構造リジッド

アロマターゼ阻害剤

光学異性体毎にMOEのlow mode MDで発生した8種類の配座を提供
上記のVCDスペクトルがどちらの光学異性体のものを予想



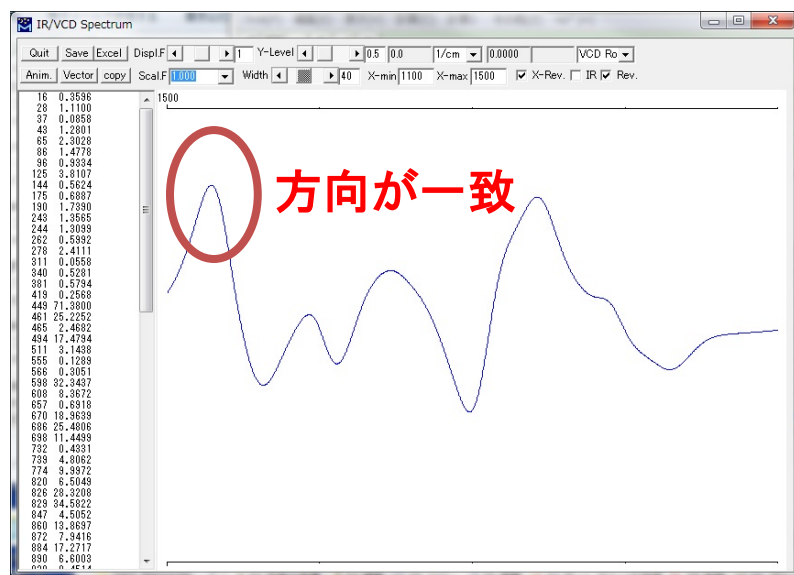
(R)-1-[(4-cyanophenyl)
(3-bromo-4-hydroxyphenyl)
-methyl]-1H-[1,2,4]triazole



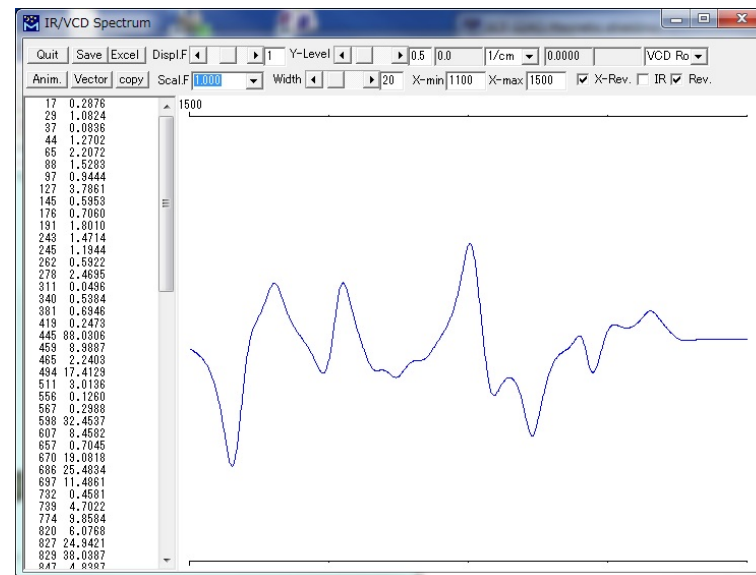
(S)-1-[(4-cyanophenyl)
(3-bromo-4-hydroxyphenyl)
-methyl]-1H-[1,2,4]triazole

下記計算でテスト(1)に合格

aromataseはrigidなので同じconfigurationの配座のvcd傾向は一致しているとみて、sとrのsdfの中の最安定な配座のみ b3lyp/6-31g* freq=vcd一点計算したところ、1400cm⁻¹付近のピークが明らかにr体とs体で逆。方向一致しているr体と判断。
J. Med. Chem., 2008, 51, 4226-4238



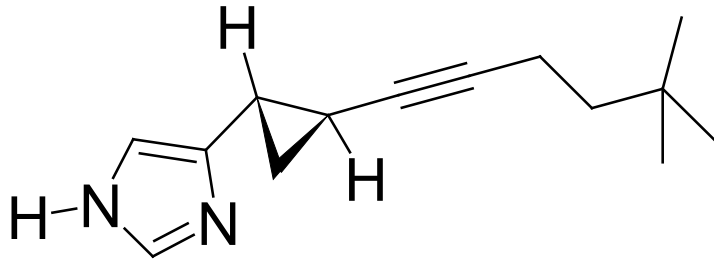
R体



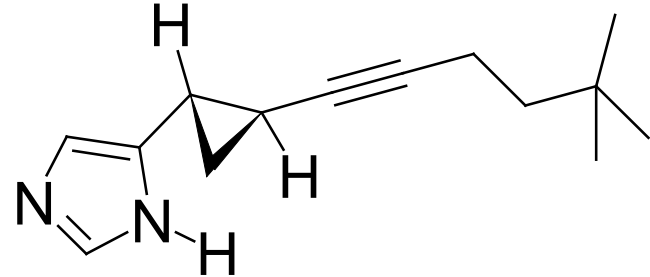
S体

某社のブラインドテスト(2)

構造フレキシブル+tautomer



GT-2331
N3-H type
209 conformers



GT-2331
N1-H type
670 conformers

2種類のTautomerについてMOEのlow mode MDで配座を発生(合計879配座)
それぞれのTautomerの存在確率は？

テスト(2)の問題

	Values	備考
Ei [kcal/mol]		0.8 相対エネルギー
E0 [kcal/mol]		0.0 相対エネルギー
T [K]	298.15	温度
g [-]		1 状態数(singlet:g=1, triplet:g=3)
N(Ei)/N(E0)	0.258986383	存在比
kB [J/K]	1.381E-23	ボルツマン定数
Avog	6.02E+23	アボガドロ定数
ToJou	4.185851821	ジュール変換
	存在比	存在割合
N(E0)	1.000	79.4%
N(Ei)	0.259	20.6%

- 真空中の計算で構造決定して良いか？ $\frac{N_i/g_i}{N_j/g_j} = \exp\left\{-\frac{e_i - e_j}{k_B T}\right\}$
- 量子化学計算の精度で十分か？
 - エネルギー差・温度・状態数が分かれば存在比が計算できる。0.8kcal/molの差で20%変わる。
 - よって、存在比をエネルギーから正確に議論する場合は、1kcal/mol以下の精度が必要(hartreeで点6桁以上の精度)。doubleの計算が必要。

計算時間はかけられない

- 溶媒効果の計算は時間かかるので省きたい
- 構造最適化に時間がかかるから、できる限りエネルギー一点計算で済ませたい
- 励起エネルギーもCCSD/6-31G**くらいでちゃんと計算したいが無理
- ab initioコンフォメーションサーチするにしても、B3LYP/6-311G*くらいが限界と考えるべき

報告によると

- 下記論文では、実験値はN1:N3=65:35といっている。

Minick et al., Chirality, 19:731-740 (2007)

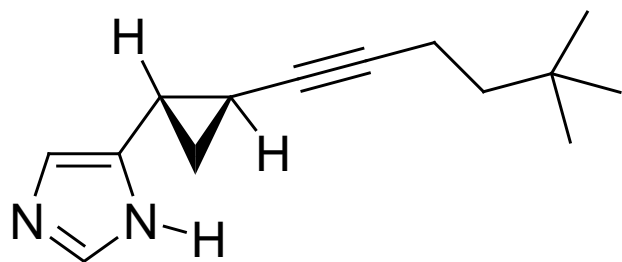
- 真空中の計算では90-95:5-10という報告あり
de Vries AH, van Duijnen PT. Theoretica calculation of tautomer equilibria of 4-substituted imidazoles in the gas phase and in solution. Biophys Chem 1992;43:139 147.

だが、ダイブ古い計算の文献なので信用できない

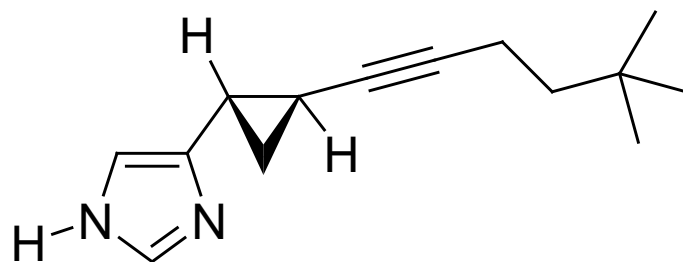
1. 課題

GT-2331 N1-H typeとN3-H typeの存在比を推算

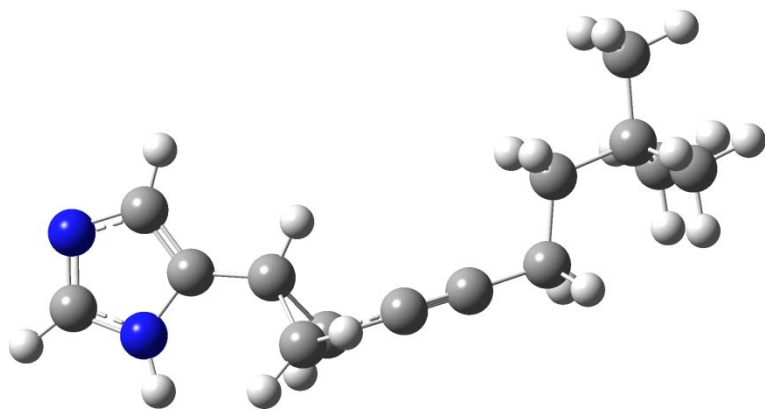
※ 計算はTSテクノロジー社協力



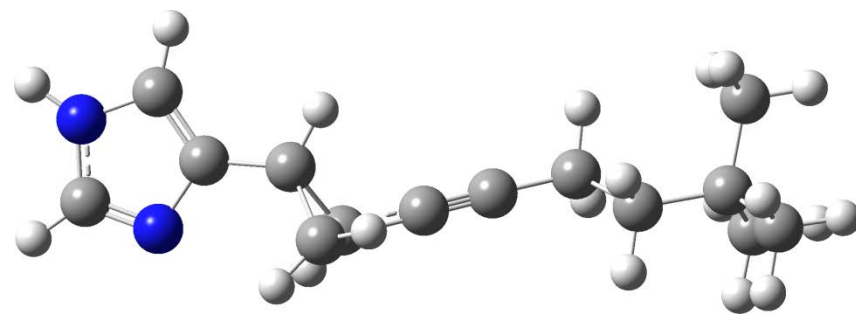
GT-2331
N1-H type



GT-2331
N3-H type



Optimized Structure at B2PLYP/6-31G*

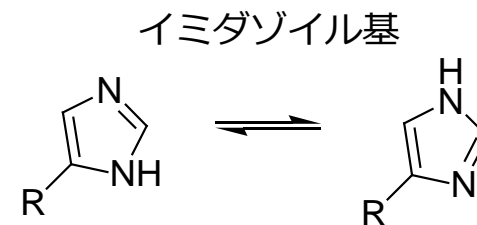


Optimized Structure at B2PLYP/6-31G*

2. 準備

tautomer(互変異性体)の存在比計算で考えられる可能性

- 物質の平衡状態として互変異性体が存在している場合
⇒ 今回のGT-2331はこのタイプ
tautomer間のエネルギー差を計算し、ボルツマン分布を計算することで、存在比を推算できるかもしれない。
- 互変異性を起こさせるために何か化学的な操作（塩基を加える等）が必要な場合
⇒ ケト-エノールなど
化学的な操作の過程（操作=反応過程の遷移状態）を計算、もしくは、原料からの合成過程でtautomerが発生する過程の遷移状態を計算し、その活性化エネルギー差を議論に加えなければならない。
（遷移状態の計算は非常に大変）



ボルツマン分布(分子iとjの比率)

$$\frac{N_i/g_i}{N_j/g_j} = \exp\left\{-\frac{e_i - e_j}{k_B T}\right\}$$

(e: 全エネルギー、g: 状態数、k_B: ボルツマン定数、T: 温度)

3. 最もオーソドックスな手法(手法1)

手順

1. 2D構造を参考に、3D構造を作成
2. **MM2でフルコンフォメーションサーチ。最安定構造を抽出**
3. QM(PM6 or B3LYP or B2PTLYP)で構造最適化 (真空中)
4. ボルツマン分布より存在比を計算

Level	Type	Energy [a.u.]	ΔE [kcal/mol]	N(N1)/N(N3) [-]	Ratio [-]
PM6	N1-H1	0.10369	0.01	0.9812	49.5%
	N3-H1	0.10367	0.00	-	50.5%
B3LYP/6-31G*	N1-H1	-654.95029	2.77	0.0094	0.9%
	N3-H1	-654.95470	0.00	-	99.1%
MP2/6-31G*//B2PLYP/6-31G*	N1-H1	-654.18211	2.99	0.0064	0.6%
	N3-H1	-654.18687	0.00	-	99.4%
B2PLYP/6-31G*	N1-H1	-653.47183	2.84	0.0083	0.8%
	N3-H1	-653.47635	0.00	-	99.2%

結果

- ・計算レベルによって存在比が異なるが、全レベルでN3-H1 typeが安定であると計算されている。
- ・PM6はN3,N1でエネルギー差が殆どなく1:1で存在と推算。
- ・ab initio計算ではいずれもN1-H1は殆ど存在しないと推算。
- ・エネルギー差が3 kcal/mol 離れてしまうだけで、存在比に大きく影響してしまう。
- ・量子化学計算の精度の問題 (3 kcal/mol以下の精度を議論できるか)、測定溶媒等の影響 (計算結果は真空中でのもの。溶媒を考慮すればさらに精度が悪くなると考えられる。)

4. コンフォメーションを意識した手法（最適化なし）（手法2）

手順

1. MOEのlow mode MDで配座を発生(N1:670配座、N3:209配座)
2. MM(MMFF94s)とQM(PM6)で全ての配座に対してエネルギーを計算
3. 最安定配座のエネルギーより存在比を計算

結果

- MM計算でのエネルギーはN1とN3間で23kcal/mol程度開きがあり、信頼できないと考えられる。
- PM6は80%がN1と推算。
- MMFF94sでは、エネルギーの開きが大きいため存在比の計算ができないが、
総じてN3が安定であると推算。

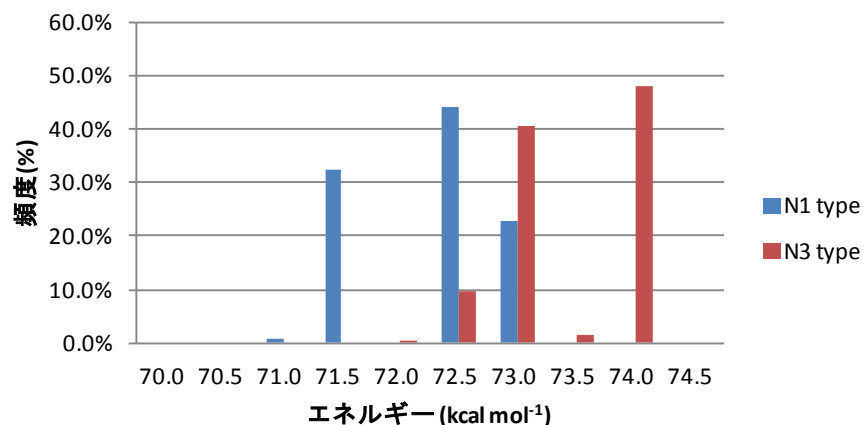
4. コンフォメーションを意識した統計処理的な手法（最適化なし）（手法1）

Energyとして異性体中の最安定構造を採用した場合の結果

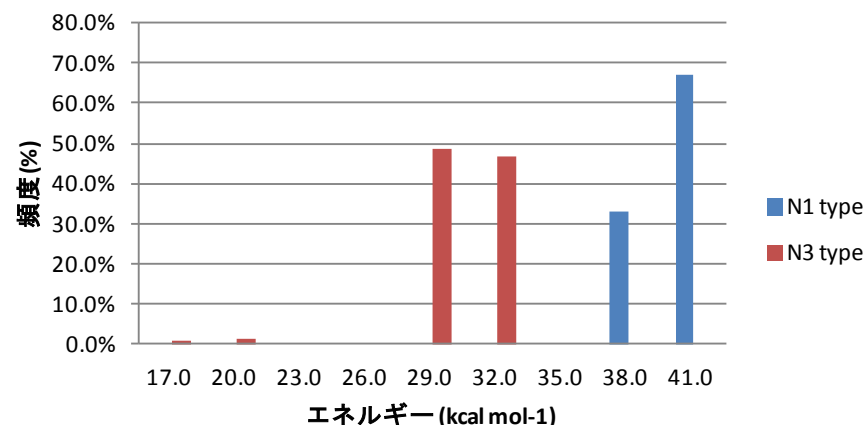
Level	Type	Energy [kcal/mol]	ΔE [kcal/mol]	N(N1)/N(N3) [-]	Ratio [-]
MMFF94s 1SCF	N1-H1	40.58867	23.05	0.0000	0.0%
	N3-H1	17.53611	0.00	-	100.0%
PM6 1SCF (H.O.F.)	N1-H1	71.47418	-0.81	3.9477	79.8%
	N3-H1	72.2873	0.00	-	20.2%

N1、N3各タイプでのコンフォメーションにおけるエネルギー値の分散(縦軸は%で標準化)

GT2331 PM6 1SCF (H.O.F.)



GT2331 MMFF94s 1SCF



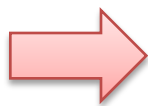
5. コンフォメーションを意識した手法（最適化あり）（手法3）

手順

1. MOEのlow mode MDで配座を発生(N1:670配座、N3:209配座)
2. MM(MMFF94s)とQM(PM6)で全ての配座に対して構造最適化計算を実施
3. 最安定配座のエネルギーより存在比を計算

結果

- PM6での異性体の構造最適化計算の結果は、**N1 : N3 = 76 : 24** となり、**実測結果を支持している**と考えられる。
(実験系（溶液中）のIRでの結果は、N1 : N3 = 65 : 35)
最適化なしの結果(N1 : N3 = 80 : 10)より実測値へ近づいた。
- MM計算でのエネルギーは依然としてN1とN3間で平均10kcal/mol程度の開きがありあまり信頼できないと考えられる。また、N3-H typeの方が安定と計算され、実測とは逆の結果。
- これらはいくまで真空の結果であるので、正確に計算するためには、測定溶媒等の影響も考慮に入れる必要があると考えられるが、ラフなスクリーニングとして使うには充分だと考えている。



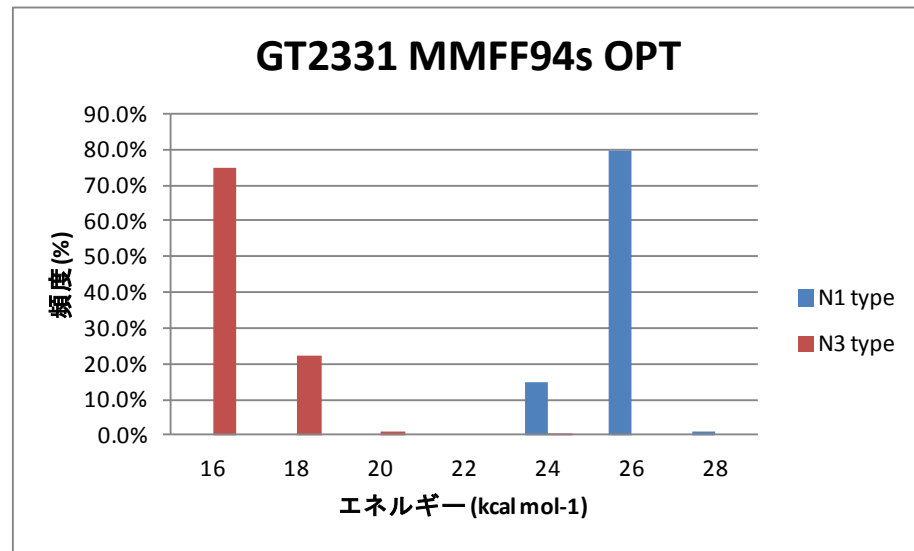
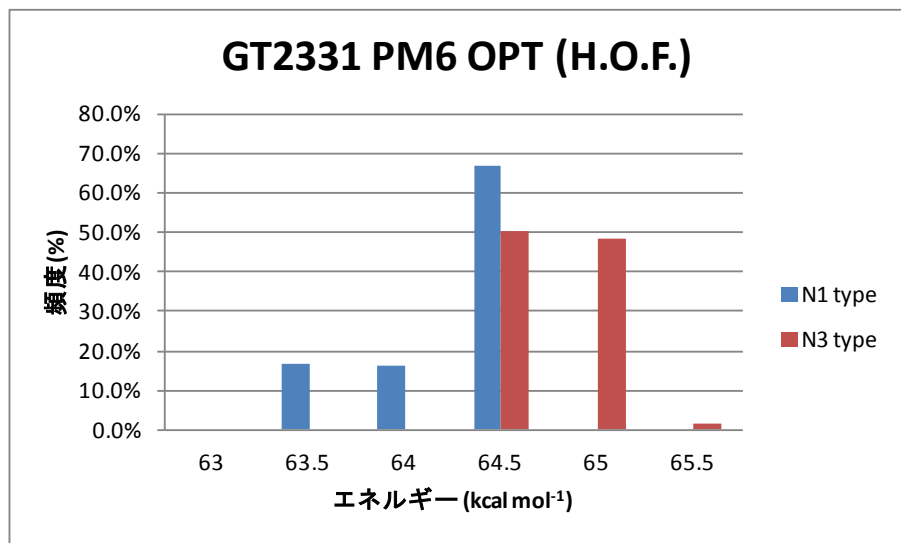
N1、N3の全配座の最安定構造(PM6)を初期構造として、
ab initio計算を行えば、結果は改善される！？
(云わば、**PM6でのフルコンフォメーションサーチ**)

5. コンフォメーションを意識した統計処理的な手法 (最適化あり) (手法3)

Energyとして異性体中の最安定構造を採用した場合の結果

Level	Type	Energy [kcal/mol]	ΔE [kcal/mol]	N(N1)/N(N3) [-]	Ratio [-]
MMFF94s OPT	N1-H1	25.843	8.97	0.0000	0.0%
	N3-H1	16.871	0.00	-	100.0%
PM6 OPT (H.O.F.)	N1-H1	63.97674	-0.69	3.2313	76.4%
	N3-H1	64.67128	0.00	-	23.6%

N1、N3各タイプでのコンフォメーションにおけるエネルギー値の分散(縦軸は%で標準化)



6. PM6でフルコンフォメーションサーチ→ab initio (手法3アドバンス)

手順

1. 手法3によりN1、N3の最安定構造(PM6)を計算(= **PM6でフルコンフォメーションサーチ**)
(N1については211番目、N3については57番目の配座が最安定であった。)
2. QM(B3LYP or B2PTLYP)で構造最適化 (真空中)
3. ボルツマン分布より存在比を計算

Level	Type	Energy [a.u.]	ΔE [kcal/mol]	N(N1)/N(N3) [-]	Ratio [-]
PM6(手法3の結果)	N1-H1	0.10195	-0.69	3.2313	76.4%
	N3-H1	0.10306	0.00	-	23.6%
B3LYP/6-31G*	N1-H1	-654.95215	1.56	0.0713	6.7%
	N3-H1	-654.95464	0.00	-	93.3%
MP2/6-31G*//B2PLYP/6-31G*	N1-H1	-654.18428	1.60	0.0669	6.3%
	N3-H1	-654.18684	0.00	-	93.7%
B2PLYP/6-31G*	N1-H1	-653.47418	1.34	0.1045	9.5%
	N3-H1	-653.47631	0.00	-	90.5%

結果

- ・ PM6での最安定配座を初期構造として、DFT計算を行った結果、N1-HとN3-H間でのエネルギー差は手法1より小さくなったが、依然N3が安定と計算され、実測とは逆の結果となっている。
- ・ MP2、B2PLYP(B3LYP + MP2)での計算結果も概ね、B3LYPの結果と同じであった。
- ・ 本手法では実験結果を支持する結果が得られなかった。

7. 計算時間 (PM6構造最適化計算)

PM6 構造最適化(EF PRECISE)

ソルバ: MOPAC2009 (アカデミックフリー)

<http://openmopac.net/MOPAC2009.html>

※PM6 はGaussian09でも計算できます。

マシン : Intel(R) Core(TM)2 Quad CPU Q6600 @ 2.40GHz

計算時間 :

	N1-H1	N3-H1	合計
合計(1コア使用時)	8386 秒 (2.3 時間)	3622 秒 (1.0 時間)	12008 秒 (3.3 時間)
平均(1コア使用時)	12.52 秒	17.33 秒	

8. まとめ

- ・オーソドックスな方法 (N1, N3の構造最適化によるエネルギーを比較) では、ab initio の結果でも、実測値を再現できなかった。
- ・それぞれの配座(N1:670配座、N3:209配座)に対して、構造最適化を実施しながらエネルギーの平均値を計算した結果、PM6法で、実測値を再現する結果となった。(ただし、これは真空中での構造最適化計算であり、測定溶媒等の影響は考慮されていない。)
- ・PM6でラフスクリーニングができる可能性を示していると考えられる。
- ・ab initio計算では、1000~1万倍の計算時間が必要であると考えられるため、計算コスト的に不利であると考えられる。
- ・PM6での最安定配座→ab initioの計算では、実験結果が支持できなかった。

トータル計算時間

普通のデスクトップ1台で担当者が計算した概算時間

※第四段階は最安定エネルギーから1kcal/molの配座を選択

- 第一段階(構造構築, ISIS Draw)
- 第二段階(配座探索, MOE)
 - まず、水素付加、MMLレベルの構造最適化 - 30sec
 - 次に配座探索(LowModeMD)を実施- <5min
- 第三段階(semi empirical計算)
 - semi empiricalの構造最適化を実施 - <5min
- 第四段階(DFT計算, Gaussian09)
 - 全41配座のDFTレベルの構造最適化計算(低精度:B3LYP/6-31G*)
 - 1配座にかかる計算は1hour
 - 全41配座のDFTレベルのエネルギー計算(高精度:B3LYP/6-311+G**)
 - 1配座にかかる計算は1hour

テスト(2)の現状

- 今回の計算
 - MM2フルコンフォメーションサーチ→最安定選択→abinitio(手法1) NG
 - PM6フルコンフォメーションサーチ→最安定で比較(手法3) OK
 - PM6フルコンフォメーションサーチ→最安定選択→ab initio(手法3AD) NG
- 議論点
 - 測定溶媒の効果
 - 手法1と手法3ADで、ab initio時のエネルギー差が縮まっているので、ab initioは初期構造にかなり依存するので、難しい。けれども、全配座ab initio計算するのは時間的に大変なので、GPGPU QMが使えるかも
 - 全配座のPM6は意外と使えるかもしれない。
- 将来に向けて
 - 配座発生の方法を検討中
 - 全配座ab initioを実施中

某社担当者のコメント

ブラインドテストを実施した某社からのコメント

- 配座探索アルゴリズムだけで最安定配座を絞りこんだり、半経験的分子軌道法では最安定配座を探索すると失敗する。むしろ、配座探索では可能性のある多数の配座を発生させるに留め、ab initioで最安定(+ α)の配座を探すことが大切である。
- 化合物構造の柔軟性が高かったり、Tautomerがある場合には、可能性のある配座が多数になり、自社の計算リソースに与える負荷が高くなる。XA社のGPGPU QMソリューションは費用対効果に優れており、VCDによる構造決定には有効であると考えている。

GPGPU QMソリューション

- 弊社と名大の共同開発のプログラム

XA-CHEM-SUITE : XA-CUDA-QM

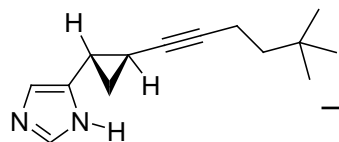
- テラヘルツの基準振動解析、光毒性予測のためのTDDFT(HOMO-LUMO Gap)、VCDいずれも多量の量子化学計算を伴う。
- HF, DFTがマルチコアCPUに比して3~4倍高速
- 現状、GAMESSまたはGaussianにアドオン可能
- 価格対効果は3倍

balloonによる配座探索

- 高速というよりもコストがかからない方法
- 有償のものを含めていろいろあるが、今後はMOEがなくても使えるように、フリーで使えるballoonとの組み合わせを検討
 - MOPACのPM6やGAMESS/Fireflyなどを組み合わせれば完全にフリーになる(vcdはGaussian09のみ)
 - 100個の配座を生成するコマンド例

```
# balloon -c 100 --randomSeed 51277 --output-file 3D.sdf -o sdf --  
input-file 2D.smi --noGA
```

光毒性予測/VCD評価システム



GT-2331
N1-H type

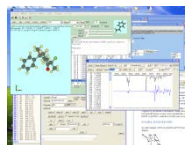
配座発生
(balloon)

各配座について漏れなく
semi-empiricalエネルギー一点計算(MOPAC)
低精度ab initioエネルギー一点計算
(GAMESS on GPGPU計算)

再安定な配座から1kcal/mol
の構造を抽出し
HOMO-LUMO gapと
VCDスペクトルを計算
(e.g. TDDFT + g09 freq=vcd)

安定な配座について
高精度量子化学計算エネルギー&最適化計算

同一ネットワーク内



Windows操作マシン
(ユーザーはこちらのみ
の操作)
配座探索ソフトballoon
Winmostar + MF Amber

Linux演算サーバ
(ユーザーは意識する
必要がない)
ジョブスケジューラ
Gaussian09 B.01

MD/ER + QM(freq/vcd, tddft)

- 京大版ERソースコードはGPLで3月末公開予定
- プリビルド計算サーバ販売予定
 - GUIから溶質、溶媒、力場を決めれば計算可能
 - 弊社が一次サポート
 - 研究の議論はermod_user(アカデミックML)に代理投稿、もしくは松林研の許可を得てermod_userに参加
 - GPUでQM + MD(Amber / Gromacs)が加速

