

Grid技術を用いた創薬プラットフォームの構築

中馬 寛
徳島大学薬学部

hchuman@ph.tokushima-u.ac.jp

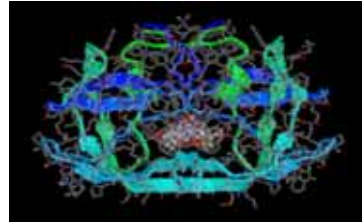
化学とコンピュータ

	カテゴリー	例
化学ファクトリーオートメーション Computer Assisted Instruction	メカトロ技術 教育	
化学実験ロボット		ハイスループットスクリーニング(HTS)
化学データベース 化学構造表記 合成反応予測 化学構造自動推定 化合物物性自動推定	情報化学 情報化学 情報化学・エキスパートシステム 情報化学・エキスパートシステム 情報化学・エキスパートシステム	
ケモトリックス	情報化学	統計学、スペクトル処理、パターン認識、人工知能
構造活性相関 分子モデリング	計算(機)化学・理論化学	Hansch - 藤田法、3次元構造活性相関
分子科学計算 分子シミュレーション	計算(機)化学・理論化学 計算(機)化学・理論化学	分子力場法、分子動力学法、分子軌道法

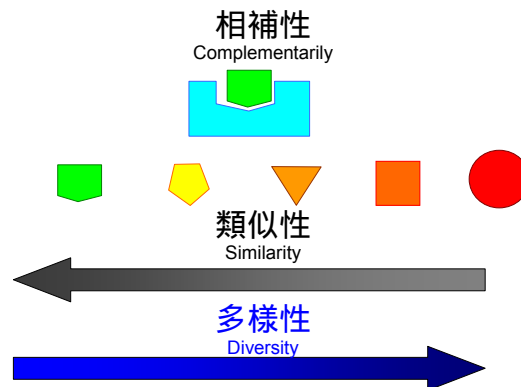
第24回情報化学討論会・第29回構造活性相関シンポジウム(徳島、11月、2001年)

- | | |
|--|---|
| <ol style="list-style-type: none"> 1) 化学情報学(化学論理学、情報検索、DBコンピューターネットワーク、CG、グラフ理論、反応設計など) 2) 理論化学・計算化学(MO, MM, MD, MC, 各種シミュレーションなど) 3) ケモトリック及びそのソフトウェア(ニューラルネットワーク、ファジィ、カオス、遺伝的アルゴリズム、構造物性相関、データマイニングなど) 4) 化学教育・学習システム 5) その他情報化学に関するもの(コンビナトリアル・ケミストリー、遺伝情報、環境ホルモン情報など) | <ol style="list-style-type: none"> 1) QSAR基本パラメーター 2) 医薬への応用 3) 構造生物学と創薬 4) ゲノム創薬と創薬 5) コンビナトリアルケミストリーと創薬およびQSARの情報数理的アプローチ 6) 吸収・分布・代謝・毒性・環境毒性 7) その他 |
|--|---|

薬物-受容体相互作用



10HR
HIV-virus : AG1343



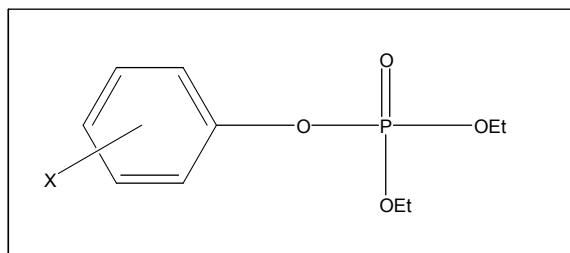
生物活性値 = $f(\text{分子又は置換基のパラメータ})$
= $f(\text{立体構造、電子構造、疎水性})$

定量的構造活性相関 (Quantitative Structure-Activity Relationships)

- ・Free-Wilson 法
- ・Hansch - 藤田法
- ・線型判別分析法
- ・パターン認識法
主成分分析、クラスター分析、SIMCA、人工知能
- ⋮
- ・3次元定量的構造活性相関解析法 (3D QSAR)

Anti acetylcholine esterase activity

(T.R. Fukuto et al, 1956)

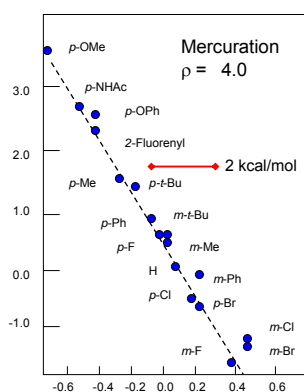


pI_{50} : $p\text{-NO}_2 > p\text{-SO}_2\text{Me} > p\text{-Cl} > p\text{-Bu}$

($\sigma_p = 0.78, \quad 0.72, \quad 0.23, \quad -0.29$)

Linear Free-Energy Principle (1)

Louis P. Hammett, *Physical Organic chemistry*, McGraw-Hill, 1970



Mercuration of Xc1ccc(C(C)(C)Cl)cc1

$$\log k_x/k_0 = \rho \sigma^+ = -4.0 \sigma^+$$

$$\begin{aligned} \sigma^+ &= \log k_x/k_0 / (-4.0) \\ &= \Delta\Delta G / (2.302 \times RT \times 4.0) \\ &= \Delta\Delta G / 5.53 \\ (T = 300 \text{ K}) \end{aligned}$$

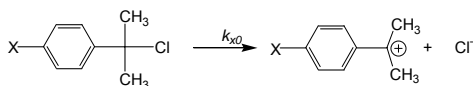


Fig. 11.5 Comparison of specific rates for mercuration Of substituted benzenes with σ^+ constants.

New Azole Fungicides

launched in 1994
from Kureha Chemical



Ipconazole
(Techlead)

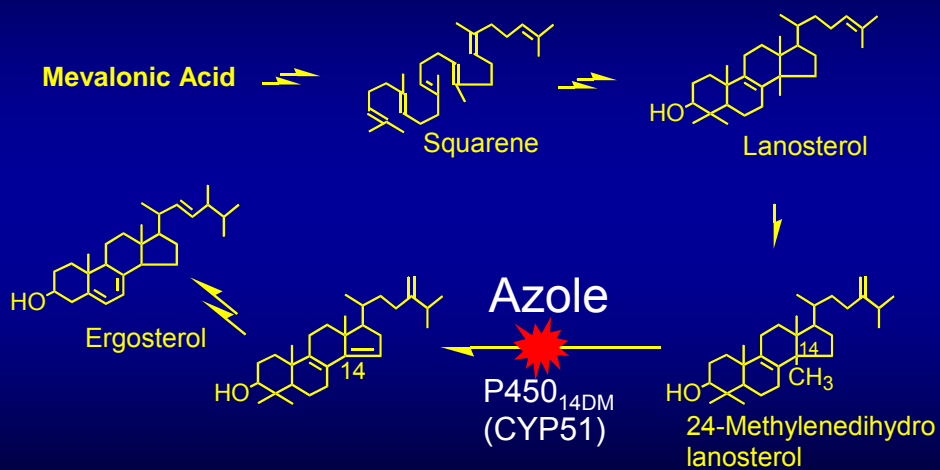
Seed treatment on rice
"Bakanae" disease
Helminthosporium leaf spot
Blast



Metconazole
(Caramba)

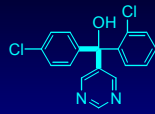
Foliar treatment on cereals
Rust
Powdery mildew
Septoria
Fusarium

Ergosterol Biosynthetic Pathway

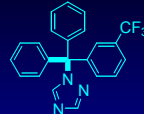


Azole-type Fungicides

N=2

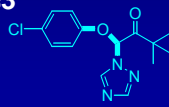


Fenarimol



Fluotrimazole

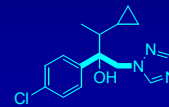
N=3



Triadimefon



Propiconazole



Cyproconazole



Diniconazole

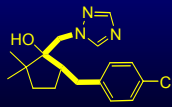


Flusiconazole

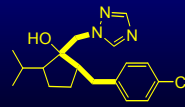


Triflumizole

N=5

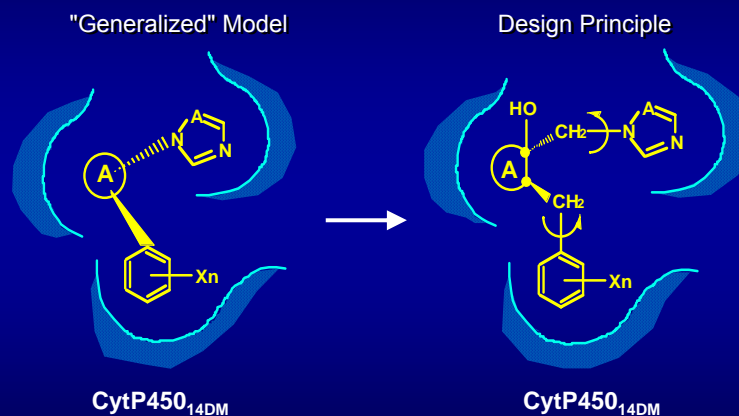


Metoconazole



Ipconazole

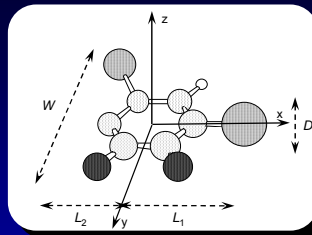
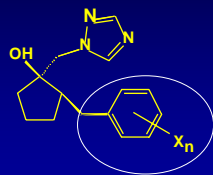
Generalized Substructural Pattern in Azole Fungicides and Design Principle of New Analogs (Kumazawa *et al.* 1984)



A : "Hydrophobic" moiety often with OH.

Conformationally fixed "A", and "flexible" rings.

QSAR (I)



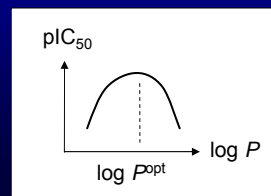
$$pIC_{50} = (\text{steric parameters}) + (\log P) + (\log P^2)$$

$$= -0.833L_2 - 0.711W - 0.435D + 2.896 \log P - (0.356 \log P)^2 + 8.735$$

$$(n = 16, r = 0.938, S = 0.237, \log P^{opt} = 4.07)$$

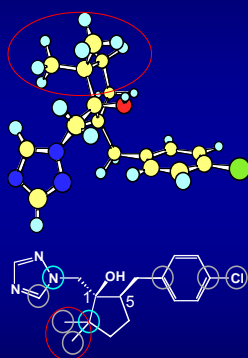
G.f (in vitro)

$$P = \frac{C_{oct}}{C_{H_2O}}$$

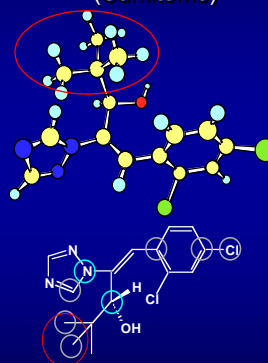


3D Shape Comparison with that of Precedent (Chuman *et al.* 1985)

"Metconazole (1S, 5R)"

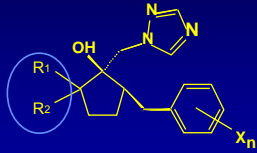


Diniconazole(R)
(Sumitomo)

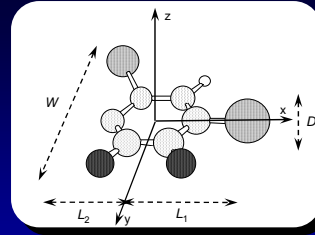


Circled atoms are used in 3D superposition with the flexible fitting method similar to the least squares procedure.

QSAR (II)



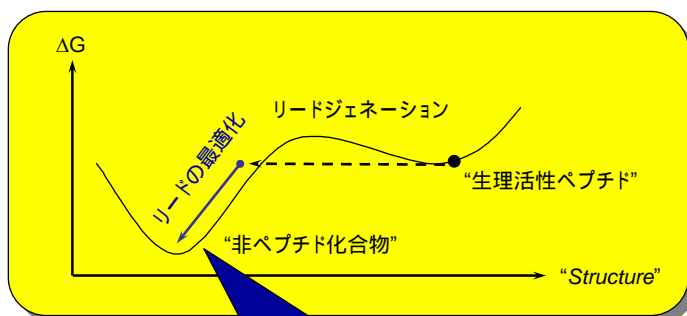
$I = 0$ when $(R_1, R_2) = (H, H)$
 $I = 1$ when $(R_1, R_2) \neq (H, H)$



$$pIC_{50} = (\text{steric parameters}) + (\text{indicator variable}) \\ + (\log P) + (\log P^2)$$

$$= -0.790 L_2 - 0.270 D + 1.012 I + 3.350 \log P - 0.471 (\log P)^2 + 6.145 \\ (n = 30, r = 0.966, s = 0.245, \log P^{opt} = 3.56)$$

G.f (in vitro)



~ 5000 equations

自由エネルギー変化の線形則
 置換基定数 (Hammett σ , π , Taft E_s , ...)
 重回帰分析

Hansch-Fujita 法
 (構造活性相関解析: Quantitative Structure-Activity Relationships)

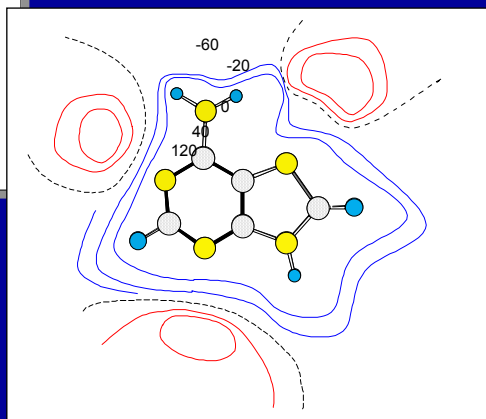
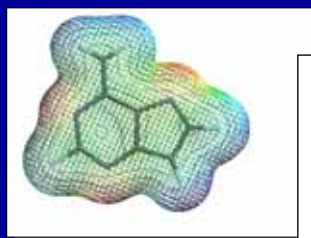
$$\log(1/C) = a \sigma + b \pi + c E_s \dots$$

Three Dimensional QSAR

Ligand - Receptor Interaction
Complementarily
Similarity

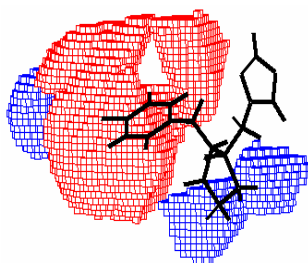


Electrostatic Field
Steric
Field



Voronoi Field Analysis (3D QSAR)

Metconazole derivatives
(*in vitro*, *Gibberella fujikuroi*)



Classical QSAR

$$pIC_{50} = (\text{steric parameters}) + (\text{indicator variable}) + (\log P) + (\log P)^2$$

$$= -0.790L_2 - 0.270D + 1.012I + 3.350 \log P - 0.471(\log P)^2 + 6.145$$

$$n=30, r=0.966, s=0.245, \log P^{opt} = 3.56$$

Voronoi Field Analysis

$$pIC_{50} = (\text{steric field}) + (\log P) + (\log P)^2$$

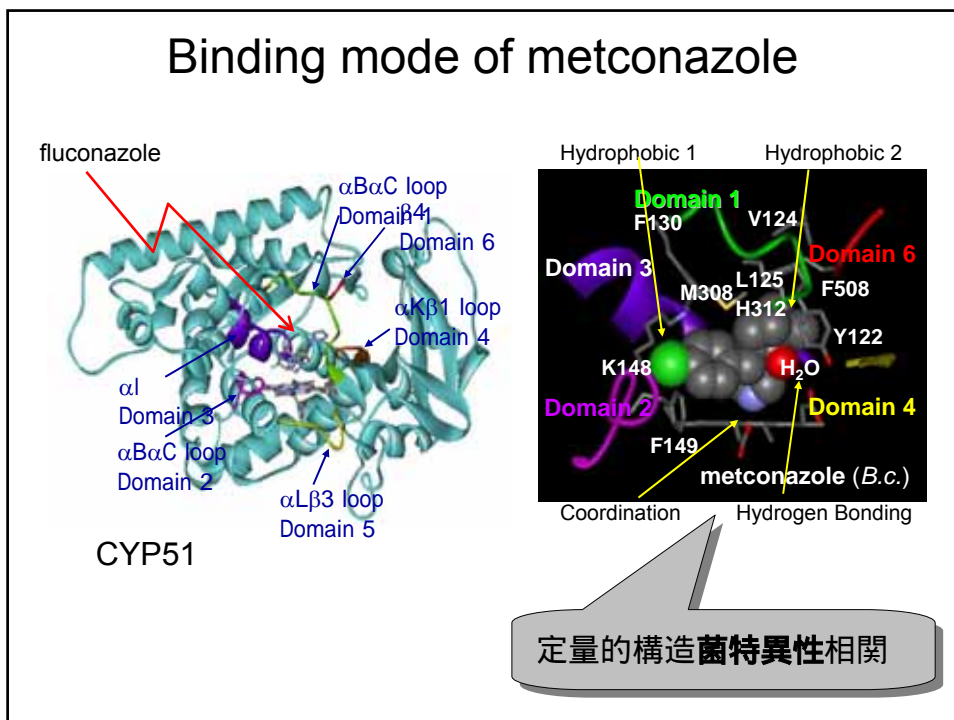
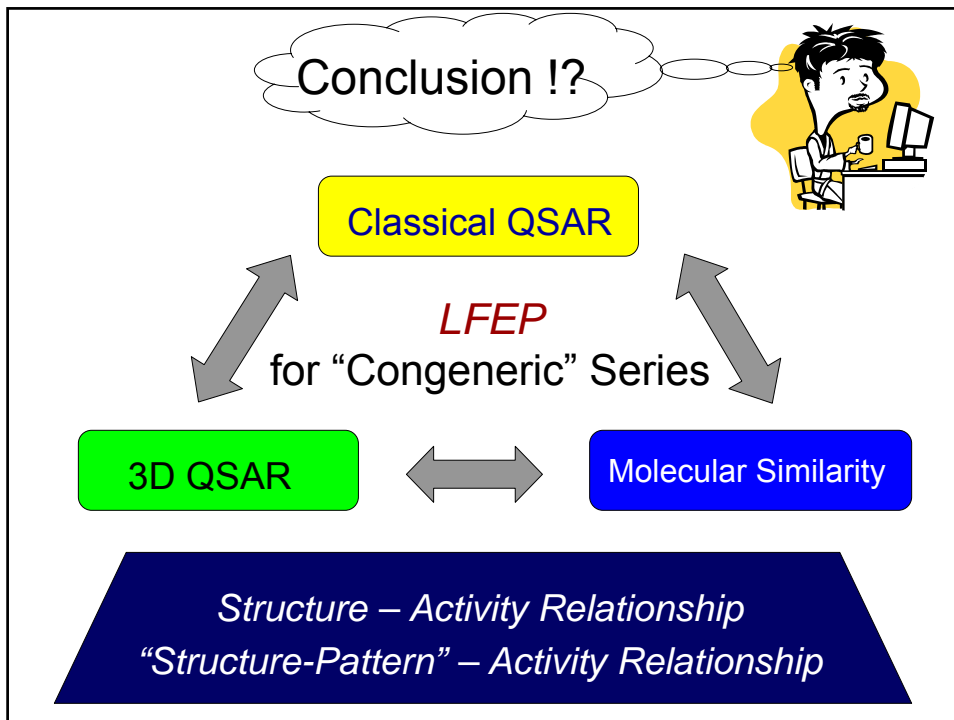
$$= 0.928Z^{ste_3} + 2.505 \log P - 0.337(\log P)^2 + 1.795$$

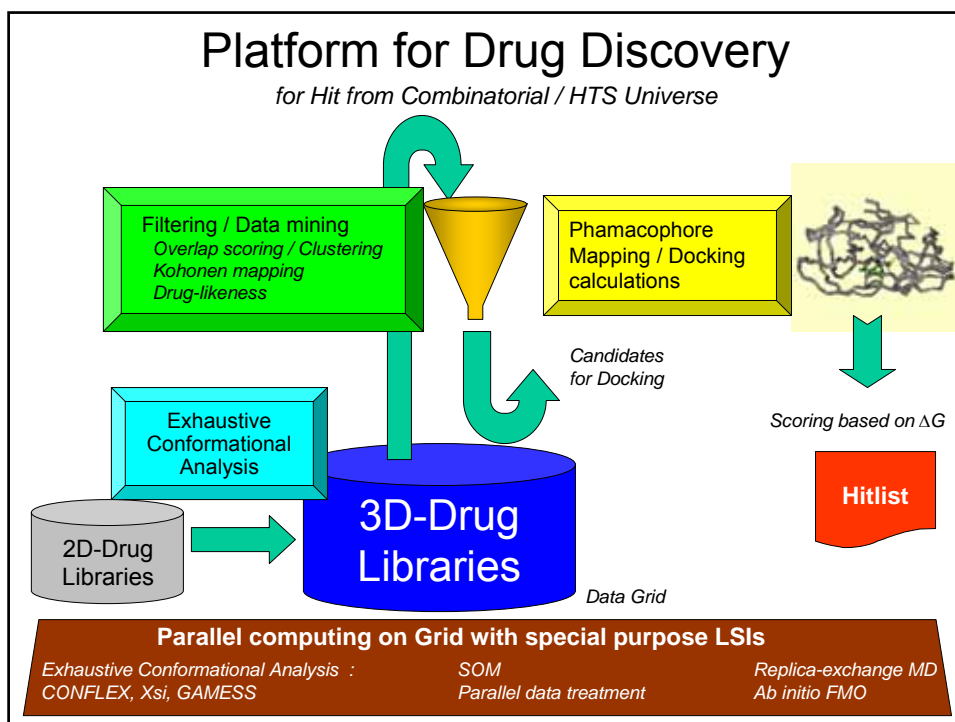
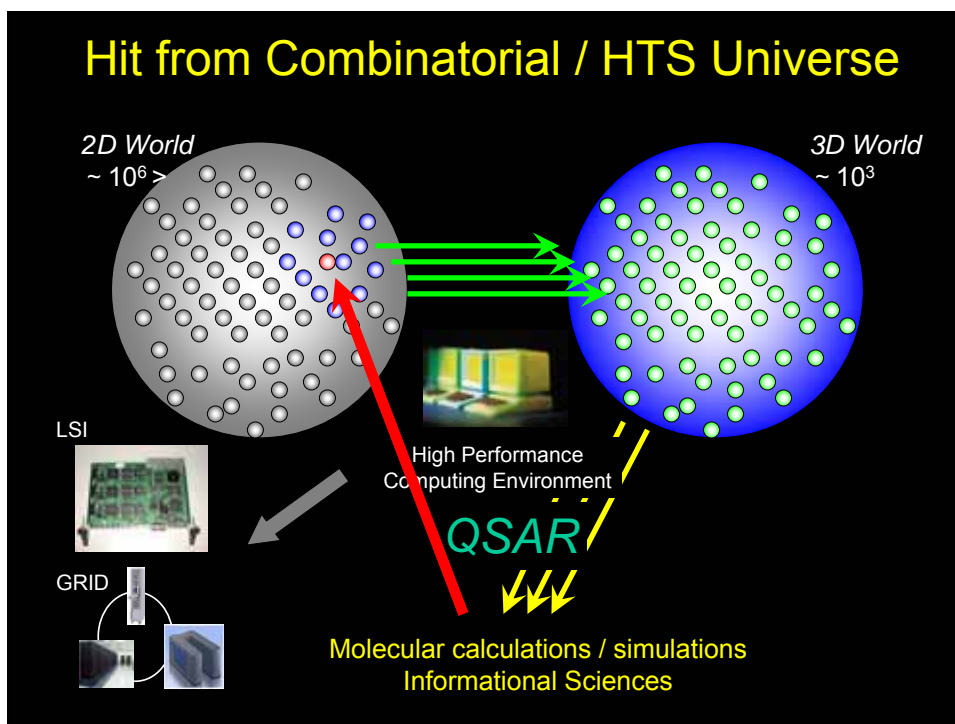
$$n=30, r=0.977, s=0.198, S_{cv}=0.398, \log P^{opt} = 3.72$$

$$Z^{ste_3} = (\text{steric parameters}) + (I)$$

$$= 0.714L_2 - 0.270 D + 1.271I + 5.052$$

$$n=30, r=0.966, s=0.221$$





High Throughput Computing for Drug Discovery

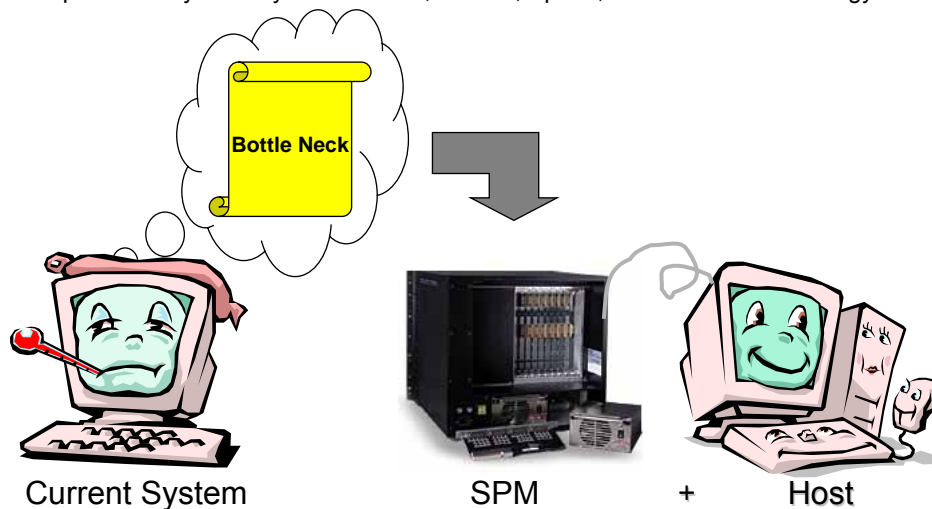
- A. Parallel Computing
- B. Special Purpose Machine
- C. Combinatorial Computing Using GRID



1. Large Scale Molecular Calculations
2. Utilize QSAR / Informational Chemistry
3. Large Scale DB for Drug Design
4. Application for HIV-virus Inhibitors

Platform Architectures for Logic Embedded High-Performance Computing (1999-)

Sponsored by Ministry of Education, Culture, Sports, Science and Technology



Special Purpose Computer



- Molecular Mechanics
- Molecular Orbital
- Density functional theory
- Fluid dynamics

Grid Technology for Drug Discovery

Combinatorial
Computing



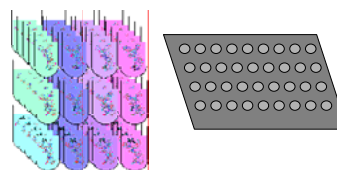
Combinatorial
and HTS

Peta Scale Data Processing !

Exhaustive computations for enormous cases
mol × conformation × configuration × temperature × pH,,,

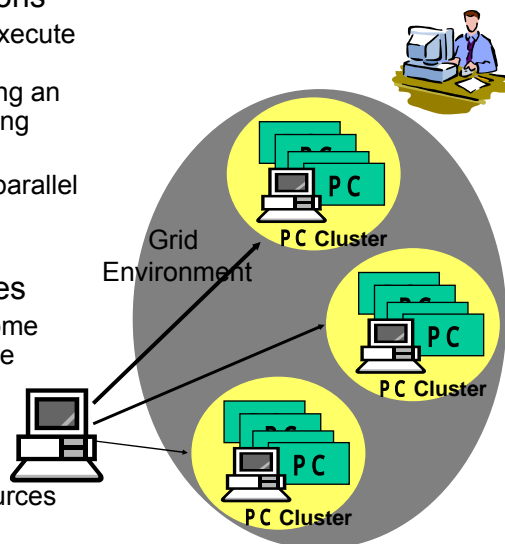


Cluster & Grid

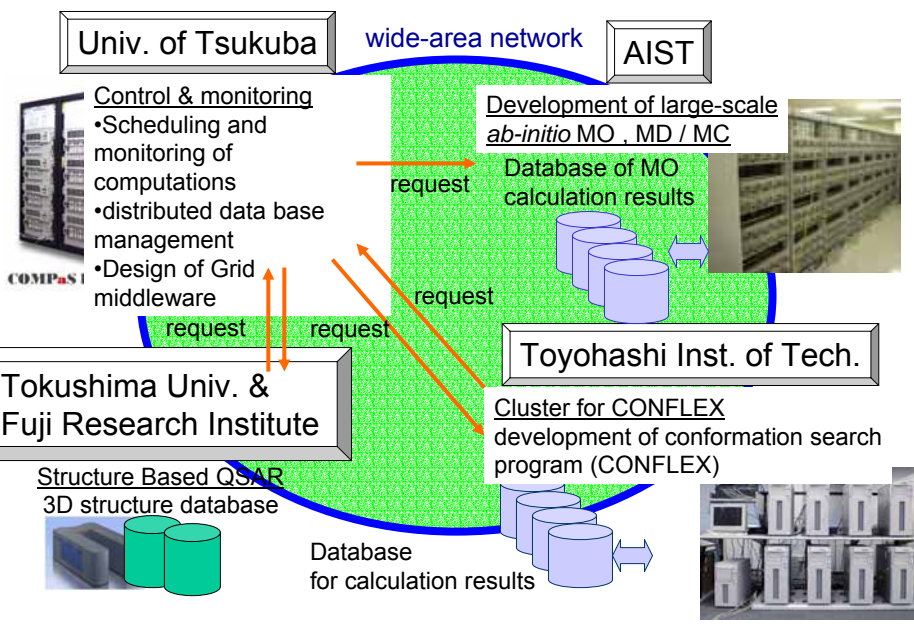


Grid and Parallel Applications

- “Typical” Grid Applications
 - Parametric execution: Execute the same program with different parameters using an large amount of computing resources
 - Master-workers type of parallel program
- “Typical” Grid Resources
 - A Cluster of Clusters: some PC Clusters are available
 - Dynamic resources: load and status are changed time-to-time.
 - Special computing resources like GRAPE6



Grid Network for Drug Discovery Project



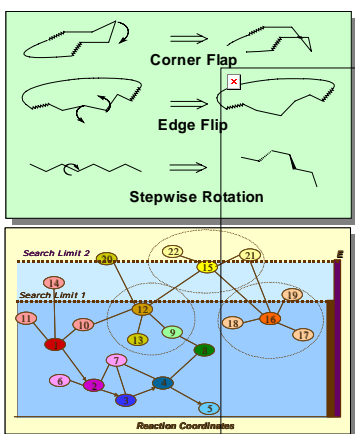
High Throughput Computing for Drug Discovery

1. Large Scale Molecular Calculations

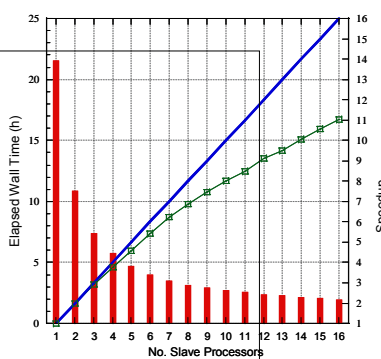
Exhaustive Conformational Search (Conflex / Xsi)
 Replica Exchange Molecular Dynamics (REMD)
Ab initio Fragment MO (FMO)

- Utilize QSAR / Informational Chemistry
- Large Scale DB for Drug DB
- Application for HIV-virus Inhibitors

Conflex



Reservoir filling algorithm



No. Slave Processors	1	2	3	4	6	8	12	16
Elapsed Wall Time (h)	21.51	10.81	7.36	5.73	3.97	3.13	2.36	1.95
Efficiency (%) ^a	100.00	99.48	97.38	93.87	90.29	85.89	75.93	69.09
Speedup ^b	1.00	1.99	3.75	3.75	5.42	6.87	9.11	11.05

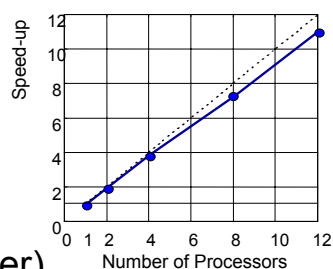
^a Efficiency = (Search Time for single CPU used) / (No. CPUs) * (Search Time for each No. CPUs used).
^b Speedup = (Search Time for single CPU) / (Search Time for each No. CPUs used).

Xsi (ku-su-shi: 薬師 : *Pharmacist*)

- MMFF Force Field
- Exhaustive Conformational Search
- Docking Calculations
- Normal Mode Analysis / MD / Monte Carlo

- Generate 3D Descriptors
- Superimposing Facilities
- Similarity / Cluster Analyses

- ◆ Configurable by Scripts
- ◆ Highly Paralleled (on PC cluster)

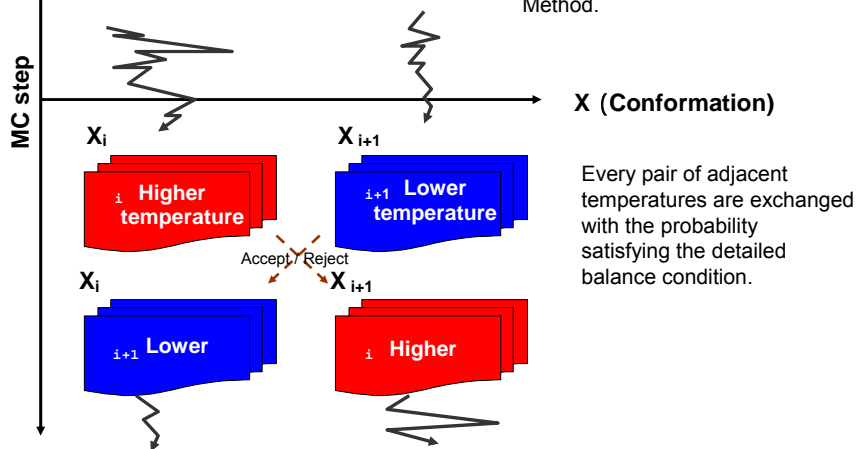


REMD (replica exchange molecular dynamics)

Accelerating conformational sampling and Parallelizing Molecular Simulations

A set of MC/MD simulations are performed simultaneously and independently for a given steps

Sampling efficiency of the system with subverted ergodicity can be enhanced by Replica Exchange Method.



REMD (2)

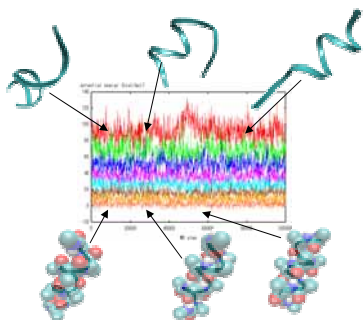


Fig.1 Time series of the potential energy and structures

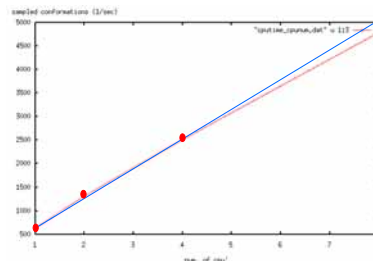


Fig.2 The sampling rate as a function of the num. of cpu's

- Object-oriented framework
 - Extensible through combination with various force field implementations and Grid middle-wares
- Combined with NAMD package (CHARMM force field)
 - Efficient secondary structure sampling and evaluation of thermodynamic properties through WHAM
 - Parallel efficiency 94% (MPI)

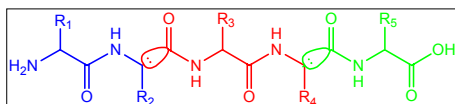
Ab initio Fragment MO for Biological Polymers (FMO)

Problems in current MM and *ab initio* MO methods

- MM: Unreliability in estimating interaction energy between drug and receptor
- *ab initio* MO : Enormous computational time

Fragment Molecular Orbital (FMO) as a New Approach

- Divide a system into fragments



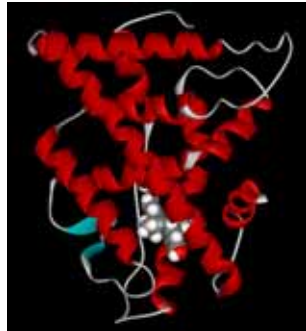
Advantages of FMO

- Compute a whole system by adding contributions from fragments and fragment pairs
 - Dramatic reduction of computational time
 - 4 ~ 5 hrs (256 cpu) with HF/STO-3G basis set for 250 residues (~ 4000 atoms)
 - Without loss of accuracy
 - Error = ~ 0.4 kcal/mol for ~ 50 residues
- Obtain binding energy between drug and receptor
- Analyze interaction energy among fragments

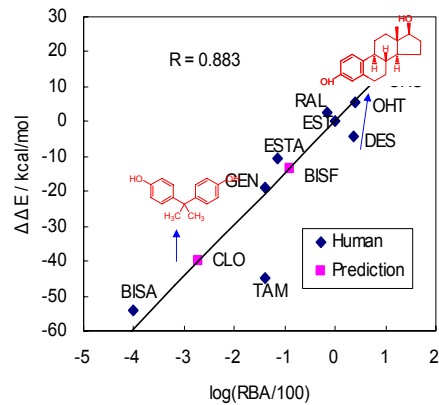
T.Nakano, T.Kaminuma, T.Sato, Y.Akiyama, M.Uebayasi and K.Kitaura, *Chemical Physics Letters*, 318, 614-618, 2000.

FMO (2)

Estrogen Receptor – Ligand Interaction



250 residues (4000 atoms)
FMO-HF
STO-3G (12000 basis sets)
4–5 hrs
Hitac SR8000(256CPU) at TACC



High Throughput Computing for Drug Discovery

2. Utilize QSAR / Informational Chemistry

Beyond LFEP ?

WHIM-like Descriptors

Self Organization Maps (Kohonen Maps)

1. Large Scale Molecular Calculations
3. Large Scale DB for Drug Design
4. Application for HIV-virus Inhibitors

Alignment Independent 3D Descriptors

WHIM (Weighted Holistic Invariant Molecular)

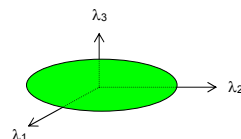
$$\sigma_{jk} = \sum w_i (r_{ij} - \langle r_j \rangle)(r_{ik} - \langle r_k \rangle) / \sum w_i$$

where n is the number of atoms, w_i is the weight of the i -th atom, $r_{ij(k)}$ represents the $j(k)$ -th coordinate [$j(k)=1,2,3$ (x,y,z)] of the i -th atom, and $\langle r_{j(k)} \rangle$ is the average of the $j(k)$ -th coordinates ($\langle r_{j(k)} \rangle = \sum w_i r_{ij(k)} / \sum w_i$).

$$\vartheta_m = \lambda_m / \sum \lambda_m \quad (m=1,2,3)$$

where λ_1, λ_2 and λ_3 are the eigenvalues of the covariance matrix (σ_{jk}).

R. Todeschini and P. Gramatica, in *3D QSAR in Drug Design*, Volume 2, pp.355-380, KLUWER/ESCOM, 1997.



Comparative Moment Molecular Analysis

Multipole Moment Expansion (monopole, dipole, quadrupole, ...)

$$V(r_1, r_2, \dots, r_n) = \sum \sum c_{lm} Q_l^m \quad (l=0, 1, \dots, m=-l, \dots, +l)$$

e_i represents the net atomic charge on the i -th atom.

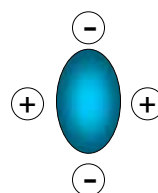
< dipole moment >

$$d_j = \sum e_i r_{ij} \quad (j=1,2,3)$$

< quadrupole moment >

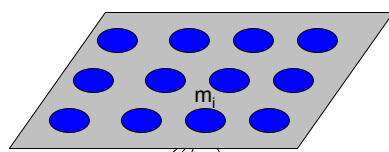
$$q_{jj} = \sum e_i [r_{ij}^2 - r_i^2 / 3] \quad (j=1,2,3)$$

$$q_{jk} = \sum e_i r_{ij} r_{ik} \quad (j,k=1,2,3, j \neq k)$$

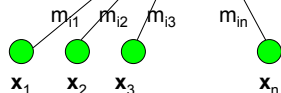


B. D. Silverman, *Quant.-Act. Relat.*, 19, 237-246 (2000)

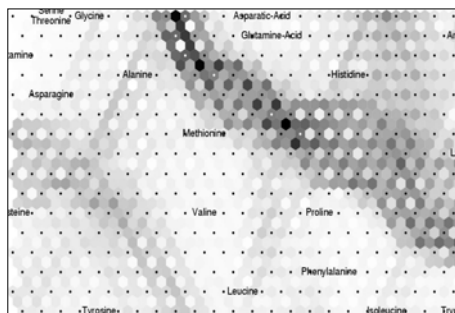
Self-Organization Map (Kohonen Map)



$$m_i(t+1) = m_i(t) + [\alpha(t)x(t) - m_i(t)]$$



< Classification of 20 amino acids >



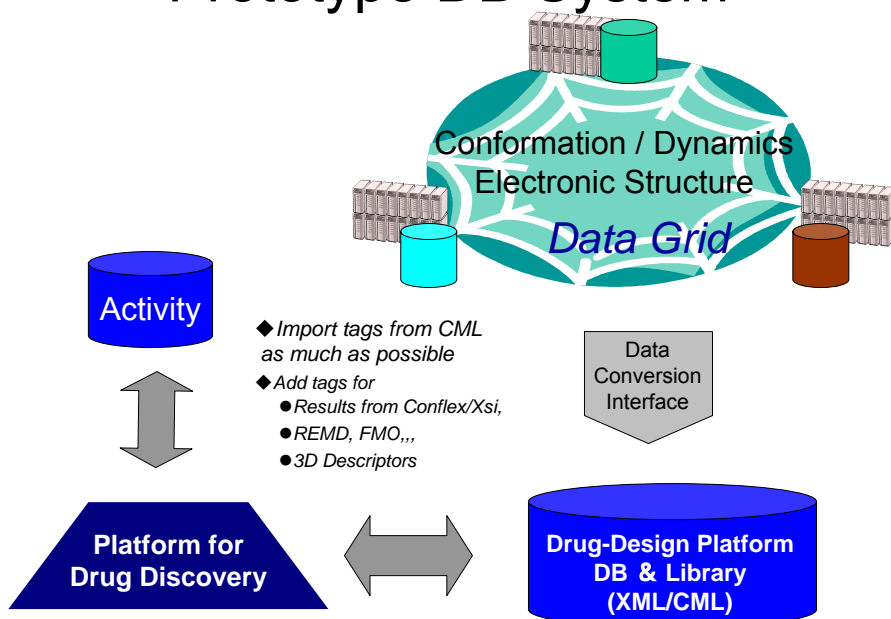
High Throughput Computing for Drug Discovery

3. Large Scale DB

Drug-ML as XML Schema for drug design

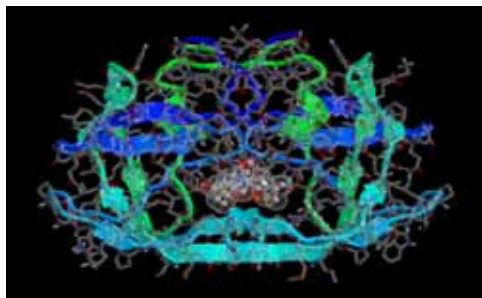
1. Large Scale Molecular Calculations
2. Utilize QSAR / Informational Chemistry
4. Application for HIV-virus Inhibitors

Prototype DB System



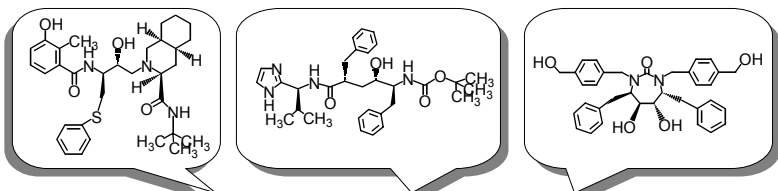
High Throughput Computing for Drug Discovery

4. Application for HIV-virus inhibitors



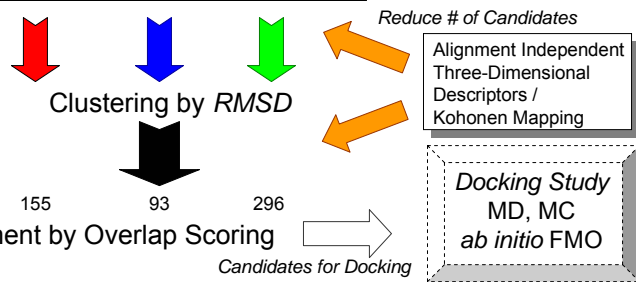
1. Large Scale Molecular Calculations
2. Utilize QSAR / Informational Chemistry
3. Large Scale DB for Drug Design

Analyses of HIV virus Protease Inhibitors (1)



	Viracept* (AG-1343)	SB-203386	DMP-323
# of Initial conformation	8,748	39,366	26,244
# Stable conformation	1,188	683	2,398

* Calculation time(sec) / Conformation = 0.97 <Kusushi>
(total elapsed time = 2111 sec on 1+4 Pentium933MHz Processors)

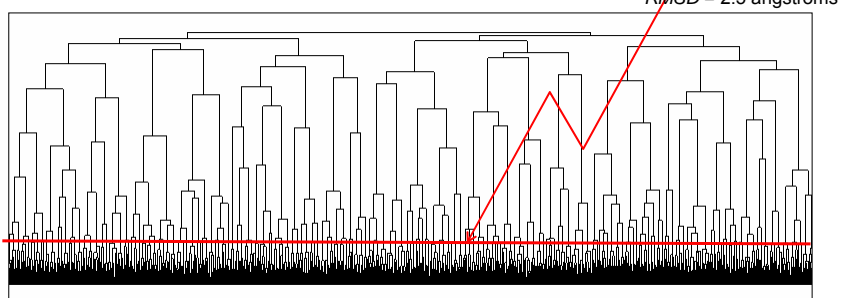
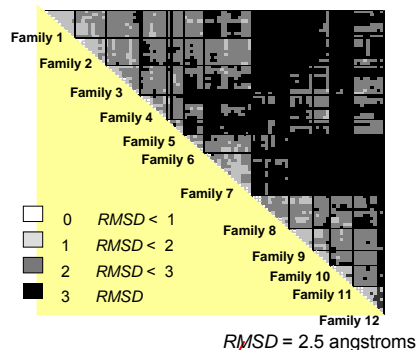


Analyses of HIV virus Protease Inhibitors (2)

Viracept (AG-1343)

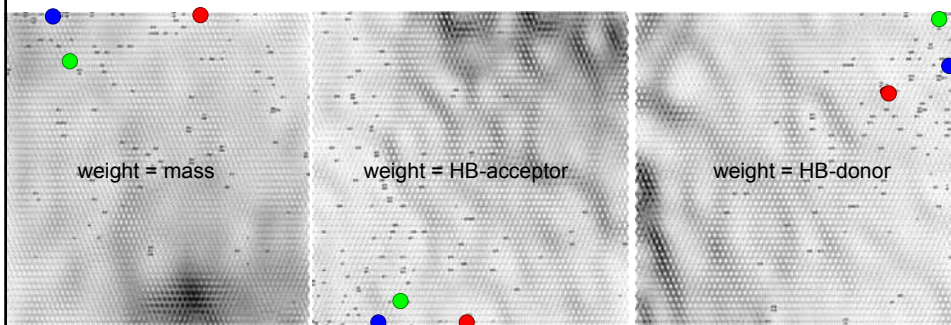
of Families = 155

of Stable Conformers = 1188



Analyses of HIV virus Protease Inhibitors (4)

Kohonen Maps Based on WHIM Descriptors of Stable Conformers



Points which satisfy *RMSD* from X-ray conformation < 2.5 Å are marked

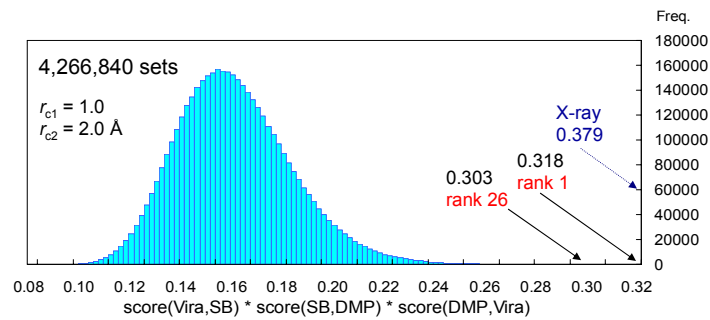
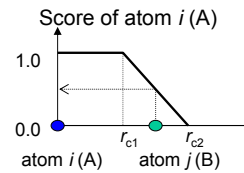
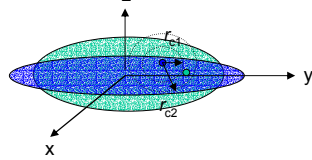
- AG-1343 - X-ray conformation with HIV-protease
- DMP-323 - X-ray conformation with HIV-protease
- SB-203386 - X-ray conformation with HIV-protease

Analyses of HIV virus Protease Inhibitors (3)

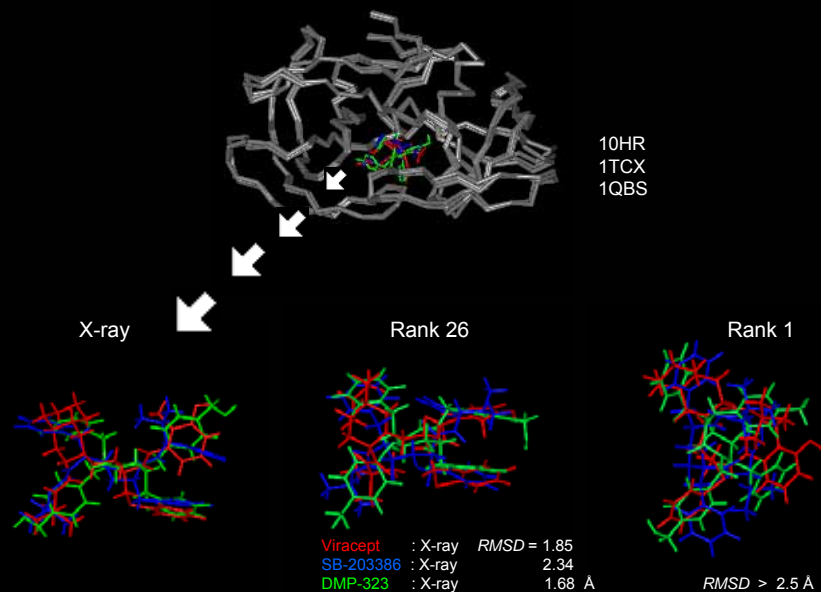
Overlap Score between A and B mols

$$\text{Total Score} = \sum \omega_k [(\sum \text{Score of atom } i \text{ (A)} \times \sum \text{Score of atom } j \text{ (B)})]^{1/2}$$

ω_k for C = 1, ω_k for O, N, ... = 2

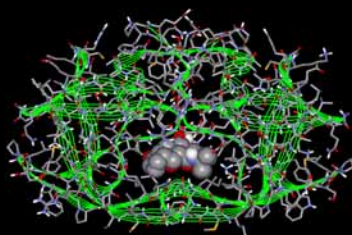


Analyses of HIV virus Protease Inhibitors (5)



Analyses of HIV virus Protease Inhibitors (6)

プロテアーゼ - 阻害剤複合体系への分子軌道法適用



198 残基 + 阻害剤 ~ 3200 原子

FMO: single point : ~ 8000 (sec) 80 CPU (AIST)

AM1-LocalSCF: optimization : ~ 4000 sec (Pent IV)

Development of Platform for Drug-Design By Grid Technology

2001-2004

Sponsored by Japanese Science and Technology Corporation, Japan

Hiroshi Chuman (University of Tokushima), *Project leader*

Umpei Nagashima (National Institute of Advanced Industrial Science and Technology)

Takeshi Nishikawa (National Institute of Advanced Industrial Science and Technology)

Masakatsu Ito (National Institute of Advanced Industrial Science and Technology)

Hitoshi Goto (Toyohashi University of Technology)

Hisashi Nakayama (Toyohashi University of Technology)

Taisuke Boku (University of Tsukuba)

Mitsuhisa Sato (University of Tsukuba)

Cheng Feng (University of Tokushima / Fuji Research Institute Corporation)

Yuichiro Inagaki (Fuji Research Institute Corporation)

March, 2003

