

	transcription, DNA-dependent)			
5	転写調節(RNA ポリメラーゼ II プロモータ) (regulation of transcription from RNA polymerase II promoter)	36.071	0.232	0.05
6	転写調節(RNA ポリメラーゼ II プロモータ)(-) (negative regulation of transcription from RNA polymerase II promoter)	21.602	0.101	0.085
7	翻訳制御 (-) (negative regulation of transcription)	20.231	0.13	0.048
8	DNA 依存性転写制御(-) (negative regulation of transcription, DNA-dependent)	17.793	0.101	0.058
9	エネルギー産生 (generation of precursor metabolites and energy)	14.282	0.116	0.033
10	ストレス応答 (response to stress)	11.814	0.217	0.015

Gene Ontology

rank	name	score	score(v)	score(c)
1	transcription (GO:0006350)	83.89	0.565	0.041
2	transcription, DNA-dependent (GO:0006351)	83.32	0.536	0.044
3	nucleobase, nucleoside, nucleotide and nucleic acid metabolism (GO:0006139)	69.196	0.609	0.027
4	transcription from RNA polymerase II promoter (GO:0006366)	66.346	0.406	0.052
5	regulation of transcription (GO:0045449)	57.655	0.42	0.039
6	regulation of transcription, DNA-dependent (GO:0006355)	57.226	0.391	0.043
7	regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolism (GO:0019219)	55.665	0.42	0.037
8	regulation of cellular metabolism (GO:0031323)	51.61	0.435	0.032
9	cellular metabolism (GO:0044237)	51.246	0.725	0.016
10	regulation of metabolism (GO:0019222)	48.887	0.435	0.03

厚生労働科学研究費補助金(化学物質リスク研究事業)
分担研究報告書

超高速分析法の比較解析と評価に関する調査研究

菅野 純 国立医薬品食品衛生研究所・毒性部・部長

研究要旨

本研究班では、内分泌かく乱化学物質問題の解明に向けた厚生労働省の「試験スキーム」に沿って要求される順位付けの為の大規模スクリーニングをすすめるとともに、内分泌かく乱物質を始めとする、核内受容体を標的とする化学物質について分子間相互作用への影響や遺伝子発現への影響など生体作用の科学的根拠に関わる諸要因に関する基礎的研究を進め、そのメカニズムに立脚した受容体標的毒性物質の評価研究への応用を目指している。本年度は、これまでに得られている大規模スクリーニング結果について OECD における内分泌かく乱物質の試験・評価プログラム(EDTA)タスクフォースにおける第4回 VMG-NA(非動物試験検証管理グループ)会議およびこれに関連する会議等において本邦及び本研究班における成果として報告するとともに、諸外国の取り組みについての情報収集及び今後の国際的な枠組みでスクリーニング系の評価の進め方について各国専門家との協議を行った。本研究班において大規模スクリーニングによる検証を進めている HeLa 細胞系エストロゲン受容体レポーター遺伝子スクリーニング系については、これまでの成果からアゴニストアッセイ系のバリデーション結果についてガイドライン化に必要な条件を満たすことが確認され、ガイドライン化に向けて外部の専門家を含めたピアレビューが実施されている。一方、QSAR に関しては OECD では、全ての QSAR 系の検証に関与しないという方針が出されたことから、VMG-NA における今後の QSAR 系検討の進め方について議論した結果、内分泌かく乱物質スクリーニング系としての QSAR の役割は重要であることが確認され、QSAR ワーキンググループの役割や体制について見直したうえで今後も引き続き国際的な協力のもとで議論や検討を行っていくこととなった。

A. 研究目的

本研究に先立つ研究において、これまでに我々は、内分泌かく乱化学物質 HTPS としてエストロゲン受容体 α 及び β ($ER\alpha$) レポーター遺伝子導入 HeLa 細胞系及び *in silico* スクリーニング手法としての $ER\alpha$ 及び β 受容体ドッキングモデル構築を行い、実際に大規模スクリーニングを実施し HTPS 系としての検証を進めてきた。その研究成果は、厚生労働省「内分泌かく乱化学物質の健康影響に関する検討会」における「試

験スキーム」の提案の基礎データ、及び内分泌かく乱化学物質の健康影響に関する検討会・中間報告追補における試験スキーム中のスクリーニング試験前2段階回に関する試験法及びその評価に資する関連データとなっている。

上述の「試験スキーム」におけるスクリーニングの位置付けは、各種のスクリーニング試験をバッテリー試験として実施し、後に続く確定試験に供する物質の優先順位を決めることにある。それぞれ特色の異なる試験を組み合わせで評

価することにより、得られる優先順位リストは、常に最新の科学的知見に基づき更新され、取りこぼしの無いスクリーニングを行うことを可能とする。本研究班では、これまでに構築した各スクリーニング手法間の比較解析による検証と精度の向上をさらに進めるとともに、パスウェイスクリーニング、分子間相互作用解析等によるスクリーニングスキームの拡充を目的として研究を進めて、これらの研究結果により有用な新規スクリーニング手法の構築を進めている。

一方、内分泌かく乱化学物質の問題解決に向けた取り組みは、欧米諸外国においても進められており、その評価基準やガイドライン策定は、国際的な強調のもとに行われるべきである。本邦及び国際的なガイドライン及び評価基準策定における、スクリーニング試験法の評価基準として、自然、あるいは人の生活の環境中に存在していて、既に何らかの内分泌学的生体影響が自然、実験を問わず報告されている化学物質を当面の陽性対象物質群として、それらの影響の機序、或いは、影響の強度が十分な精度で観測可能であることが挙げられる。本研究では、各国のリスクアセスメントにおけるスクリーニング系の開発状況について情報収集や情報交換を行うとともに、本研究班における各スクリーニング手法からの結果の比較解析を行い、我々の研究成果のうち、内分泌かく乱化学物質のスクリーニング系としての評価基準を満たすものを、国際的・国内的なバリデーション(有効性確認)、あるいはその前段階のプレ・バリデーションの対象となる手法としてガイドライン化を念頭に国内外で提案することを目的とする。

B. 研究方法

本年度は、これまでに得られている大規模スクリーニング結果の比較解析を行い、その結果について OECD における内分泌かく乱化学物

質の試験・評価プログラム(EDTA)タスクフォースにおける第4回 VMG-NA(非動物試験検証管理グループ)会議及びこれに関連する会議等において本邦及び本研究班における成果として報告するとともに、諸外国の取り組みについて情報収集を行った。また、専門家との協議により今後の国際的な枠組みでの進め方について情報交換を行った。本年度は、これまでに得られている大規模スクリーニング結果の比較解析を行い、その結果について OECD における内分泌かく乱化学物質の試験・評価プログラム(EDTA)タスクフォースにおける第4回 VMG-NA(非動物試験検証管理グループ)会議及びこれに関連する会議等において、本邦及び本研究班における成果として報告するとともに、諸外国の取り組みについて情報収集するとともに、専門家との協議により今後の国際的な枠組みでの進め方について情報交換を行った。特に本研究班において大規模スクリーニングによる検証を進めている HeLa 細胞系エストロゲン受容体レポーター遺伝子スクリーニング系については、これまでの成果からアゴニストアッセイ系のバリデーション結果についてガイドライン化に必要な条件を満たすことが確認され、ガイドライン化に向けた作業を行った。一方、*in silico*スクリーニング系に関しては OECD では QSAR の検討グループにより全ての QSAR 系の検証に OECD として関与しないという結論が出されたが、VMG-NA においては引き続き国際的な協力のもとでの検討を進める必要があることを提案し、今後の進め方について協議した。

C. 研究結果

OECD では、内分泌かく乱化学物質の問題解決に向けて、内分泌かく乱化学物質の試験・評価プログラム(EDTA)タスクフォースを設置し検討を進めている。本研究班において構築し検

証を進めている各スクリーニング手法は、いずれも EDTA により示されている OECD conceptual framework for the testing and assessment of endocrine disrupting chemicals の Level 2 に相当する。第4回 VMG-NA では、OECD における内分泌かく乱化学物質スクリーニング手法のガイダンス化に向けて、参加各国代表がそれぞれの国における研究開発状況について発表し討議した。わが国では、レポーター遺伝子系の開発及びバリデーション結果、*in silico* スクリーニング系(QSAR)系の開発状況等に関して厚生労働省・経済産業省の共同で、また厚生労働省では独自に開発を進めている cell free スクリーニング系等の報告を行った。

本研究班で開発を進めているスクリーニング系のうちレポーター遺伝子アッセイ系については、もともと米国が提案した化学物質の内分泌かく乱作用の有無の可能性を検討する方法の有用性を確認する事から開始されたが、系の構築は我が国が先行しており、これまでに構築した4受容体(ER α 、ER β 、AR、TR)系のうち国内における多施設バリデーションが終了しているER α 恒常発現Hela細胞を用いた系については、国内バリデーション結果を第2回 VMG-NA において報告し、第3回ミーティングにおいてPVAP(Preliminary Validation Assessment Panels)を設置しガイドライン化に十分なデータが揃っているか検討することとなった。今回のミーティングにおいては、ER α 恒常発現 Hela 細胞を用いたアゴニスト系についてはPVAPによりガイドライン化に向けた最終のピアレビューを実施する条件が整っていると結論されたことが報告された。さらに現在、外部の専門家を入れたピアレビューがすでに始まっており、2007年3月ごろに将来のガイダンス化に向けた WNT において次の作業に向けた最終レポートが出される見通しであると報告された。今回の経験から、

PVAP は義務ではないが、今回のような目的での検証にあたり必要な事項が必ずしも GD34 では明確ではなく、VMG-NA の専門家によるPVAPが有用であったことをWNTに報告すべきだという意見が出された。また、バリデーションデータが一般に入手可能であることも重要だとの意見が出された。TA アッセイに関して、我が国ではさらに、ER α 一過性発現系及びAR恒常的発現系による TA アッセイについてもすでに多施設バリデーションが終了しており、ピアレビュー可能な状態であることを報告した。一方、米国で開発されたER α TA アッセイ系(LUMI cell ER α アッセイ系)の国際バリデーションが、ICCVAM、ECVAM 及び JaCVAM の共同で進められているとの報告があった。また、ECVAM では、MELN (ER アゴニスト)、ER-CALUX (アンタゴニスト)、PALM (AR アゴニスト)及びAR-CALUX (アンタゴニスト)を開発し、全ての系について2007にバリデーションを実施する予定であると報告があった。

ED QSAR の開発については、我が国では、本研究班で構築した ER α 、ER β ドッキングモデルによる *in silico* スクリーニング及び、それを用いた約 15000 化合物のスクリーニング結果について報告を行った。他の QSAR では、化合物構造からのモデルの適用可能な化合物の範囲を限定したモデルバリデーションを進めているが、ドッキングモデルでは構造式を基準にした化合物分類を行っていないことを特徴として強調した。また、これにあわせて本研究班で開発を進めている SPR バイオセンサーを用いた、スクリーニング手法についても報告した。EC/ECB からは、ER による遺伝子活性化の分類モデル構築を行い、現在、モデル構築に用いる化合物数を増やして再構築をしているとの報告があった。分類モデル構築とモデルの評価についてはすでに論文化されているとのことであった。ま

た、ECB では、QSAR そのものの開発とは別に新たな QSAR をレポートする際の統一フォーマットの作成を進めているとのことであった。OECD では、様々な QSAR 手法による各種モデルや予測結果を登録するための QSAR ツールボックスの準備を進めており、2007 年 9 月にベータテストを行う予定をしている。また、EC における ED レギュレーションの QSAR 戦略については、Community Strategy for Endocrine Disruption (1999) で示されている。ECB はそのための QSAR データベース開発を進めており、ECB の QSAR インベントリーにもモデルをアップロード出来るようになるとのことであった。US EPA では、EPA における QSAR 開発と評価の枠組みについて報告があった。EPA では、殺菌剤、殺虫剤、洗剤、残留農薬などを対象物質として、ER 結合予測をエンドポイントした QSAR の開発を進めているとのことであった。モデル構築においては、ER 結合データのある化合物は少ないので、まず全ての化合物について化学構造を確認し、高精度のモデル開発のため実際の試験を実施する化合物を選択するため、最小限の化合物で構造情報を最大化できるようなツールを開発して化合物選択を実施し、rtER 結合試験、競争阻害試験、ピテロゲニン誘導試験を実施したと報告があった。現在は、測定結果から、階層化分類による、意思決定サポートシステムのドラフトが完成し、研究者と EPA とでその評価を行っているとのことであった。

QSAR 開発については、OECD ではアドホック QSAR グループにより、全ての QSAR 系の検証に OECD として関与すべきでないという結論が出されたことをうけ、これまで VMG NA において国際共同により推進してきた内分泌かく乱化学物質スクリーニング手法としての QSAR の検討について、今後、どのように対応すべきかについて議論を行った。その結果、内分泌かく乱化

学物質のスクリーニングにおける QSAR の役割は、その他の QSAR と若干異なるため、VMG-NA における ED QSAR タスクグループにおける検討は継続することとなった。ただし、ED QSAR タスクグループの役割については、個別の QSAR の検証よりも情報交換をメインとして、これまで取りまとめを行ってきた EC/ECB に代わり、今後は US EPA で対応することとなった。まずは ED QSAR タスクグループで検討すべき事項について、US EPA がなるべく早くに取りまとめ、興味のある参加者を求めて ED QSAR タスクグループの再編成を行うこととなった。

上記のほか、本研究班では実施していないが、ER への化合物結合試験のバリデーションが、US EPA, EC/ECVAM, CERI 及び BIAC の共同で進められており、プロトコルの最適化を終了し、バリデーション化合物を決定し、2007 年より多施設バリデーションを開始するとのことであった。また、US EPA よりヒト副腎皮質腫瘍細胞である H295R 細胞を用いた、ステロイド合成系アッセイについて報告があった。この方法は化学物質のステロイド生合成系への影響を見るもので、H295R 細胞はステロイド生合成に関わる全ての本質的な酵素を発現している恐らく唯一のライン化細胞であるとの説明がなされた。2007 年中には、化合物を増やして本格的なバリデーションを始める予定であるが、このアッセイのエンドポイントについていくつかの問題点が挙げられているとのことであった。

D. 考察

本年度は、OECD における内分泌かく乱化学物質の試験・評価プログラム (EDTA) タスクフォースにおける第4回 VMG-NA (非動物試験検証管理グループ) 会議及びこれに関連する会議等において、本研究班で開発を進めている各スクリーニング系について、これまでに得られてい

る大規模スクリーニング結果の比較解析を行い、その結果について報告した。

先立つミーティングで、我が国におけるバリデーション結果をもとに OECD ガイダンス化を提案してきた、ER α 恒常発現 Hela 細胞を用いたアゴニスト系については、PVAP によりガイドライン化に向けた条件が整っていると結論され、最終のピアレビューが実施されており、その結果を待ってガイドライン化に向けた作業を行う。一方、我が国で多施設バリデーションが終了している ER α 一過性発現系及び AR 恒常的発現系による TA アッセイについても、今後、ガイドライン化にむけた積極的な取り組みを行う必要があると考察された。これらの系に関しては、OECD 事務局からも、日本の関係機関はこれらのアッセイが速やかにガイドライン化候補となるよう積極的に取り組んでほしいとのコメントがあった。

本研究班で開発を進めてきた *in silico* スクリーニング系について、これまで我々は OECD における ED QSAR のバリデーション候補として提案してきたが、OECD ではアドホック QSAR グループにより、全ての QSAR 系の検証に OECD として関与すべきでないという結論が出された。そのため、これまで VMG NA において国際共同により推進してきた内分泌かく乱化学物質スクリーニング手法としての QSAR について、今後、どのように対応すべきかについて議論を行った結果、ED QSAR タスクグループの役割については、個別の QSAR の検証よりも情報交換をメインとして、今後は US EPA で対応することとなった。一方、OECD では、様々な QSAR 手法による各種モデルや予測結果を登録するための QSAR ツールボックスの準備を進めており、また、ECB では、QSAR そのものの開発とは別に新たな QSAR をレポートする際の統一フォーマットの作成を進めており、今後、これに対応し、国際的に評価されるような系の構築を進めていくべきであると考え

られる。

個別のスクリーニング系開発とは別に検討すべき課題として、特定の特許技術を用いた系のガイドライン化について現在の OECD では想定いないことから、そうした技術をガイドライン化するための Performance Based Test Guidelines が必要であることが指摘された。これについては、担当者を決めて整理し、次回 VMG-NA で検討することとなった。本研究班で開発を進めている一部のスクリーニング系にも関連する部分でもあり、今後、この問題に関しても積極的に取り組んでいく必要があると考えられた。その他、バリデーションに使用する化合物選択や比較すべきエンドポイントの選択基準、ED スクリーニング系に特有の専門用語の整理などが検討すべき課題として挙げられ、今後の VMG-NA において検討していくこととなった。

E. 結論

本研究班では、内分泌かく乱化学物質問題の解明に向けた厚生労働省の「試験スキーム」に沿って要求される順位付けの為の大規模スクリーニングをすすめるとともに、スクリーニングスキームの拡充を目的として、パスウェイスクリーニング、分子間相互作用解析等の応用による、核内受容体の生体作用メカニズムに立脚した新規スクリーニング手法の構築を進めている。一方、内分泌かく乱化学物質の問題解決に向けた取り組みは、欧米諸外国においても進められており、その評価基準やガイドライン策定は、国際的な強調のもとに行われるべきである。

我々はこれまでの研究成果のうち、スクリーニング手法としての基準を満たすものを、国際的・国内的なバリデーション(有効性確認)、あるいはその前段階のプレ・バリデーションの対象となる手法としてガイドライン化を念頭に国内外で提案し、それぞれ評価を得ている。これまでの取り

組みにより ER α 恒常発現 Hela 細胞を用いたアゴニスト系については、OECD ガイドライン化に向けた条件が整っていると結論され、最終のピアレビューが実施されるに至ったが、スクリーニング系バリデーションに用いる化合物や評価エンドポイントの選択基準などについて国際的な合意は出来ておらず、今後、検討すべき課題として挙げられており、我が国としても対応すべく積極的な取り組みを行うと共に、引き続き本研究班における大規模スクリーニングによる独自検証を進め、それぞれのスクリーニング手法について、適時、国際ガイドライン化に向けた積極的な提案を行う必要があると結論付けられた。

F. 健康危険情報

特になし

G. 研究発表

1. 誌上発表

菅野 純、北嶋 聡、相崎健一、五十嵐勝秀、中津則之、高木篤也、小川幸男、児玉幸夫、Percellome Project による毒性トランスクリプトミクスの新しい試み、細胞工学、2007 年 1 月号、株式会社秀潤社

2. 学会発表

菅野 純、マイクロアレイや定量 PCR から細胞当たりの mRNA コピー数を得る Percellome 法の概略と生物研究への応用、九州大学医学生研セミナー、2006 年 4 月 17 日、福岡

菅野 純、基礎と応用のリンケージ・ツールとしての Percellome System、第 95 回日本病理学会総会、2006 年 4 月 30 日-5 月 2 日、東京

菅野 純、マイクロアレイや定量 PCR から細胞当たりの mRNA コピー数を得る Percellome 法*

の概略と生物研究への応用、第 104 回熊本大学発生研・拠点形成 A セミナー、2006 年 6 月 5 日、熊本

Kanno Jun, Aisaki Ken-ichi,
Igarashi Katsuhide, Nakatsu Noriyuki,
Kitajima Satoshi, Kodama Yukio,
"PERCELLOME" TOXICOGENOMICS
PROJECT FOR THE MECHANISM-BASED
TOXICOLOGY, the SOT 46th Annual
Meeting March 25-29, 2007

H. 知的所有権の取得状況

1. 特許取得

なし

2. 実用新案登録

なし

3. その他

なし

研究成果の刊行に関する一覧表レイアウト

雑誌

発表者氏名	論文タイトル名	発表誌名	巻名	ページ	出版年
菅野純、北嶋聡、相崎健一、五十嵐勝秀、中津則之、高木篤也、小川幸男、児玉幸夫	Percellome Project による毒性トランスクリプトミクスの新しい試み	細胞工学	26	71 - 77	2007
Mizutani, M. Y.; Takamatsu, Y.; Ichinose, T.; Nakamura, K.; Itai, A.	Effective handling of induced-fit motion in flexible docking.	PROTEINS	63	878 - 891	2006
Mizutani, M. Y.; Nakamura, K.; Ichinose, T.; Itai, A.	Starting point to molecular design: Efficient automated 3D model builder Key3D.	Chem. Pharm. Bull.	54	1680 - 1685	2006
小原有弘, 水澤博	JCRB 細胞バンクの事業の概要	分子細胞治療	vol. 5 no. 2		2006
Takeuchi M, Takeuchi K, Kohara A, Satoh M, Shioda S, Ozawa Y, Ohtani A, Morita K, Hirano T, Terai M, Umezawa A, Mizusawa M.	Chromosomal Instability in Human Mesenchymal Stem Cells Immortalized with Human Papilloma Virus E6, E7 and hTERT	Genes. • In Vitro			2007 in press

Special Review

Percellome Projectによる毒性トランスクリプトミクスの新しい試み

Percellome Project as a New Approach to Toxicology Transcriptomics

菅野 純 北嶋 聡 相崎 健一 五十嵐 勝秀 中津 則之 高木 篤也 小川 幸男 児玉 幸夫

Jun Kanno, Satoshi Kitajima, Ken-ichi Aisaki, Katsuhide Igarashi, Noriyuki Nakatsu, Atsuya Takagi, Yukio Ogawa, Yukio Kodama

身の回りの物質の毒性(有害性)を予測し、その被害を未然に防ぐのが毒性学の役割である。この精度向上を目指したトキシコゲノミクス研究を実施する際に、マイクロアレイなどから細胞1個当たりのmRNAコピー数を得るPercellome法を開発した。90化合物のマウス肝初期応答データを採取し終え、新たな対象(反復投与、胎児毒性、吸入毒性、多臓器連携)を加えたPercellome Projectを展開している。

key words

トキシコゲノミクス, 分子毒性学, 遺伝子発現カスケード, 標準化, Percellome 法, 3次元多層(Millefeuille) データ

菅野 純 国立医薬品食品衛生研究所 安全性生物試験研究センター 毒性部 E-mail: kanno@nihs.go.jp

1985年東京医科歯科大学大学院医学研究科博士課程修了。人体病理学、実験病理学専攻。国立医薬品食品衛生研究所毒性部室長を経て、2002年より同部長。内分泌かく乱関連などの分子毒性学研究、トキシコゲノミクスプロジェクトなどを厚生労働所掌業務との有機的連携のもとに推進。

北嶋 聡, 相崎 健一, 五十嵐 勝秀, 中津 則之, 高木 篤也, 小川 幸男, 児玉 幸夫 国立医薬品食品衛生研究所 安全性生物試験研究センター 毒性部

はじめに

医薬品, 食品, 化粧品, 生活関連用品など, 身の回りの物質が我々の身体に取り込まれた際に生じる可能性のある毒性(有害性)を予測し, それらの使用に際しての被害を未然に防ぐのが毒性学の役割である^{注1}(図1)。具体的には, 人々の安全を確保するために使用法(用途)や使用量(残留量)を制限したり, 場合によっては禁止したりするための科学的根拠を提供するが, その際, 人の身代わりとして実験動物を用いる場合が多い。このような毒性学の精度向上の一環として, 従来からの毒性研究(毒性症候学, 毒性病理学, など)に加えてのトキシコゲノミクス(Toxicogenomics)研究が進められている。

トキシコゲノミクスでは, 物質が生体に及ぼす影響をトランスクリプトームとして観測・解析する。その際, ①分子毒性学を構築し種差や個体差の問題, 複合暴露の問題などを解決するためには, 遺伝子発現カスケードの全容解明を目指す必要がある, ②形態学的に変化が現れた段階のトランスクリプトームは, 遺伝子発現カスケードの最終段階に過ぎない, ③形態変化の現れないごく初期段階を含む遺伝子発現カスケードを描出するためにはまとまった量のデータの蓄積が必須である, との観点から, 筆者らは, マイクロアレイや定量PCRから細胞1個当たりのmRNAコピー数を得るPercellome手法と, そのデータ解析のための3次元多

層(Millefeuille)システムを開発・実用化した。遺伝子発現量が共通の尺度, すなわち“コピー数/細胞”で表現されることから, 検体間, 実験間, マイクロアレイのバージョン間, 異なったプラットフォーム間, などのデータ比較が直接的に行えるようになり, 数年かけて蓄積したデータの有機的活用が可能となった。現在, 90種類の化学物質によるマウス肝の初期応答データを採取し終えたところである。新たな対象(反復投与, 胎児毒性, 吸入毒性, 多臓器連携)を加えたPercellome Projectの概要を紹介する。

I. Percellome 法: 細胞1個当たりのmRNA絶対量を得る方法

原理は単純である。サンプルの細胞数を計測し, 外部標準mRNA(スパイクRNA)を細胞1個当たり決まった分子数だけそのサンプルに添加し, そしてRNA抽出, 測定に移る。サンプルのRNAの測定値を, スパイクRNAの値を基準に, 細胞1個当たりのコピー数に換算する。実際には細胞数を直接計測するのが困難なことが多いため, その代替指標として細胞核内のゲノムDNA量を用いる^{1), 2)}。定量性・直線性の検証にはLBM標準サンプル(肝[L]と脳[B]を100:0, 75:25, 50:50, 25:75および0:100に混合した5サンプルから成るセット)を用いる。なお, スパイクRNAは, 5種類の枯草菌遺伝子のmRNAを濃度公比3で混合したカクテル(dose-graded spike cocktail; GSC)として用意した。高精度を要求されるDNA定量法は手作業プロトコルおよび自動ロボット(PerkinElmer JANUS)のプロトコルを準備

注1 環境への配慮も含まれる。



図1. 毒性学の対象

毒性学は、身の回りの物質が引き起こす障害を予測し、その発生を未然に防ぐことを目的としている。トキシコゲノミクス (毒性ゲノミクス) は、最先端の網羅的遺伝子発現解析技術を用いて、従来の毒性学の予測の精度を著しく向上、迅速化させることで、国民の健康安全の確保にさらに貢献することを目指している。

中である。カクテルとも共同研究ベースで供給可能である (連絡先: kanno@nihs.go.jp)。また、ERCC (The External RNA Control Consortium) と連絡をとるとともに、国際的標準化への関与を深めるため平成18年度厚労科研費「医薬品などの有効性・安全性評価に資する遺伝子発現解析の国際的標準化に関わる研究 (H18-特別-指定-023)」を立ち上げた。現在、この他にシックハウス症候群を考慮した低用量域での吸入毒性トキシコゲノミクス、1匹のマウスから多臓器を採取しそれらの連携状況をトランスクリプトームから解析する多臓器トキシコゲノミクスを開始し、特徴的な遺伝子について組織内の発現分布を *in situ* ハイブリダイゼーションで確認する作業を並行している。また、下記の3次元データをweb公開するサーバを整備し、一部の化合物から3次元多層 (Millefeuille) データを順次閲覧可能とした (<http://toxicomics.nihs.go.jp/db/>)。

II. 3次元多層 (Millefeuille) データシステム: 生物系研究者に優しいデータ可視化と解析

医薬品を含む毒性既知の90化合物について単回経口投与後のトランスクリプトームデータを取得して、初期応答遺伝子カスケードを解析するための基盤データベースを構築した。現在、第二段階として反復暴露データ集積を開始し

た。データは、用量軸、時間軸、および遺伝子発現軸から成る3次元表示により、遺伝子発現の用量および時間に依存した変化を1枚の曲面として表すことで可視的に変化を判別しやすいように配慮した (図2)。これにより、コンピュータが選び出した遺伝子クラスターの中身を確認する際、特に、mRNAの合成分解のスピードなどの知見から生物学的にありえないパターン (用量軸の方向にも時間軸の方向にもジグザグな変化など) を排除する際に威力を発揮している。

1つの実験から排出されるGeneChip約50枚のデータを一括処理する能力を持ったPercellome自動換算・データ品質管理 (QC) に関わるソフトウェアに加えて、3次元多層 (Millefeuille) データに最適化した、発現パターン類似性による候補遺伝子検索、およびそれを発展させた教師無しクラスタリング³⁾を中心とした解析システム (MF System, MFシリーズ, 開発: 相崎 健一) を独自に実用化し、開発継続中である (図3)。これらにより、データQCはその日のうちに、基本的な発現情報検索から全遺伝子の教師無しクラスタリングまでを3日間で完遂できるものとなっている。

この基本解析を用いて、発現パターンによって分類された候補遺伝子リストが多数生成される。一部の幸運な例ではただちに新規と思われる毒性関連反応を見いだすことができた。またそうでない場合のための1つの補強手段とし

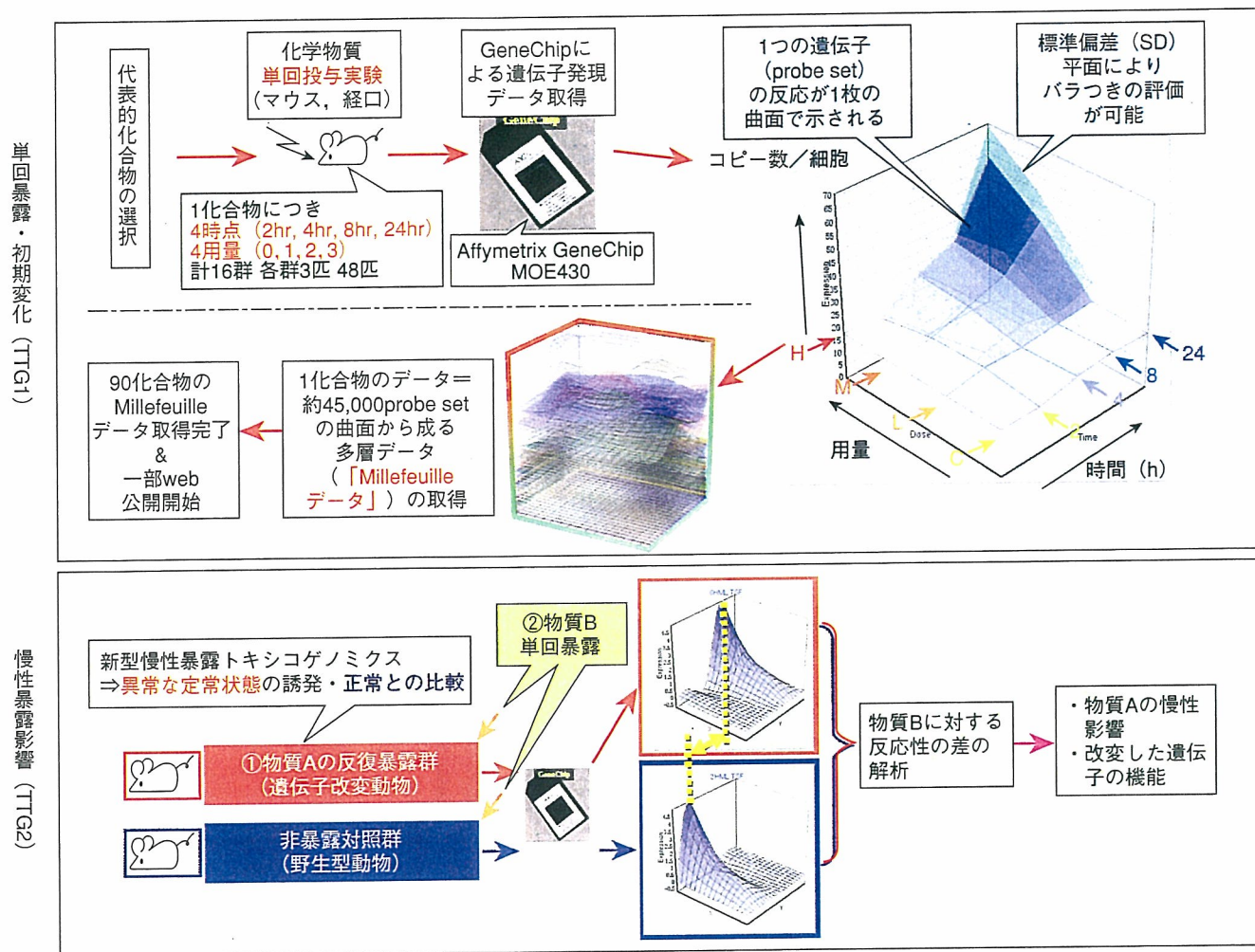


図2. Percellome 法と3次元表示による多層 (Millefeuille) データシステムを用いたプロジェクトの根幹部分の概要
 単回投与による遺伝子発現初期変化データを90化合物について取得 (上段), 現在, 反復投与の影響を検討中 (下段). H; 高用量 (high), M; 中用量 (medium), L; 低用量 (low), C; コントロール (control).

て, Gene Ontology などの既存知識を利用して候補遺伝子リストの理解を支援するソフトウェア (MF GoPlot) を用意した. このツールは一種の化合物クラスタリングとしても利用することができる.

さらに候補遺伝子リストを基に複数化合物間比較を行い, 複数条件下においても同期して発現する遺伝子群を自動抽出するシステムも開発済みである. 本システムで得られた同期遺伝子群はシグナルカスケードの構成単位である可能性があり, データベース化しつつ, その解析を進めている (5TB規模のデータベース部分および, 大量計算アルゴリズム実装は (株) NTT コムウェアおよび (株) 日本NCR/Teradata との共同開発による).

Ⅲ. Percellome 手法のリアルタイムPCR を含む他のプラットフォームへの適用

Percellome 手法は, GSC の受け入れ条件を整えることに

より, 様々なプラットフォームに適用可能である. その1つとして最も定量性が高いとされるリアルタイムPCR (ABI PRISM 7900 HT・96 ウェルプレート) への適用例を示す. 現行のRT-PCR絶対定量法では, 遺伝子ごとに検量線が必要であり, 多数のサンプルについて多数の遺伝子を検討するには不向きである. Percellome RT-PCRでは, マイクロアレイと同様の原理を用いる. すなわち, サンプル破砕液に, その細胞数に比例する量のスパイクカクテル (GSC) を添加し, それらのCt値をPCRプレートごとの検量線とすることにより, 測定したい遺伝子のCt値を細胞1個当たりのmRNAコピー数に換算する. これにより, GAPDHやActinなどのハウスキーピング遺伝子の変動してしまう際の問題, 例えば, 少数の遺伝子を検討する際にGlobal normalization法を適用し難い問題などが解決される. 共通サンプルを測定しデータを比較することにより, Affymetrix GeneChipのPercellome結果と9割程度の整合性が確認され,

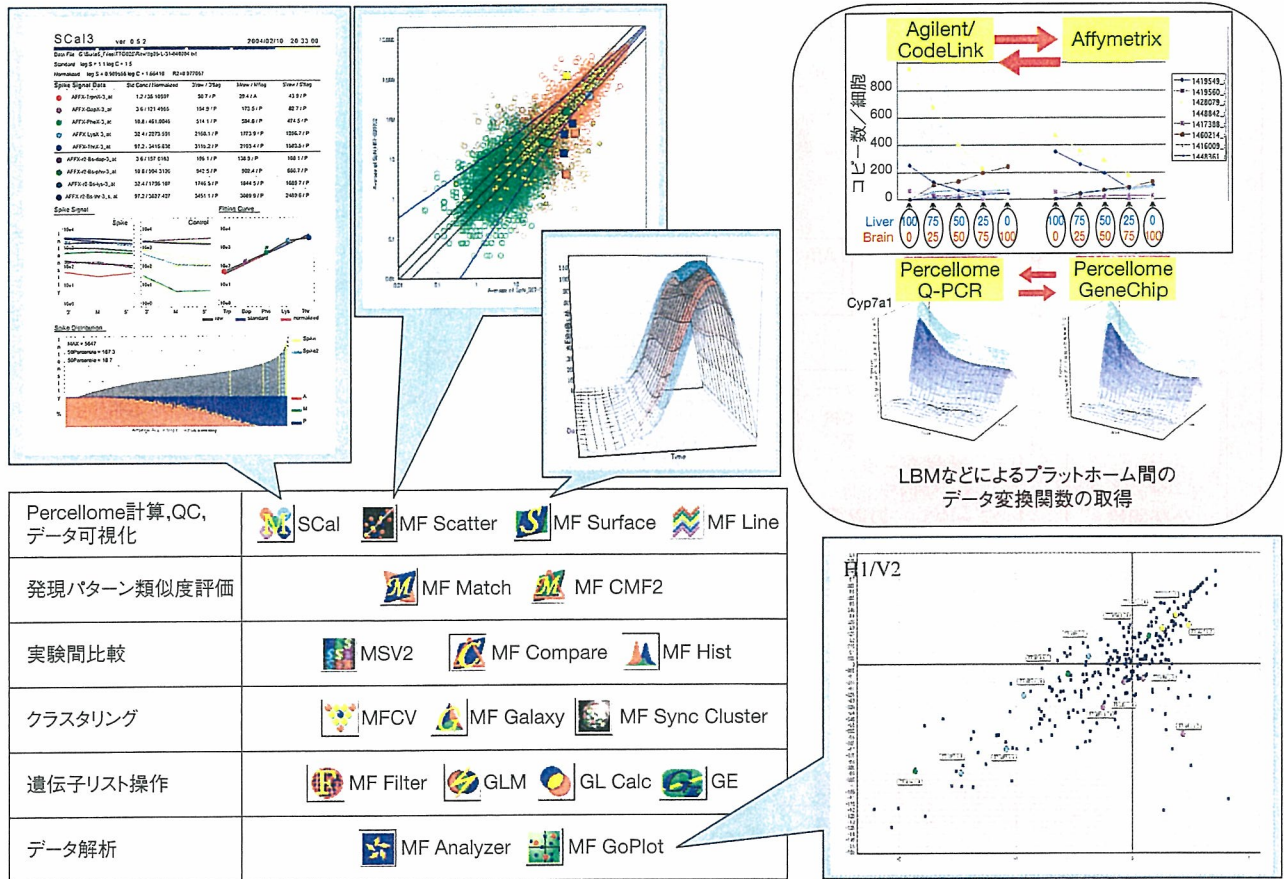


図3. 3次元多層 (Millefeuille) データの解析などに用いる独自開発プログラム群

品質管理とともにPercellome計算を自動的に実施するSCal, Plotソフトウェア, 3次元曲面の描画ソフト (MF Surface), など。右上はプラットフォーム間のデータ変換情報の得方を示す。LBMを用いる方法 (上段) と, 実際の実験サンプルを用いる方法 (下段) がある。いずれも, 一度, 両方のプラットフォームでそれらのサンプルを測定する必要がある。

GeneChipとPercellome RT-PCRとの間でのコピー数の換算式がいくつかの遺伝子について得られている。この他に, Agilent社製の単色マイクロアレイとCodeLinkアレイにGSCを測定可能なカスタムアレイを用意し終え, LBMサンプルのデータなどをもとに, これらとの間の換算式も得つつある (図3右上)。

Percellome法は, Affymetrixの新しいエクソンアレイの定量性・直線性の検討にも適応可能である。Affymetrix社のHuman Exon 1.0 ST Arrayと従来型の発現アレイHuman Genome U133 plus 2について, 性質の異なるヒト癌細胞株2株から調製したLBM様標準サンプル (100:0, 75:25, 50:50, 25:75および0:100混合5サンプル) による比較を行い, 両アレイ間の相関性の高いprobe setを多数検出することができた。また, 既知のエクソンに対して設計されたprobe setでは発現が見られ, イントロンに対して設計されたprobe setでは発現が見られない, あるいは, 既知のsplicing variantに対応したprobe setの発現が検出された,

などの基本性能が確認された。しかし, Percellome法を適用して未知のsplicing variantの検出力を向上させるためには, 現状では各エクソン間の定量性に問題があることが示唆された。定量値を算出する補正アルゴリズムの開発など, 何らかの対策が必要であることが考えられ, 現在, Affymetrix社に確認を行っている。

IV. 核内受容体原性毒性のPercellomeトキシコゲノミクス解析

受容体原性毒性とは, 化学物質が受容体 (リガンド依存的転写因子を含む) に選択的に結合してシグナルをかく乱し, その結果生じる有害性を指す。代表例としてはダイオキシンが挙げられる。AhR (Arylhydrocarbon receptor) ノックアウトマウスでは, ダイオキシンを大量に投与しても毒性がほとんど観察されない。すなわち, 野生型マウスがダイオキシンで死ぬメカニズムには, AhRが必須であり, AhRからの異常なシグナルがマウスを死に至らせていることに

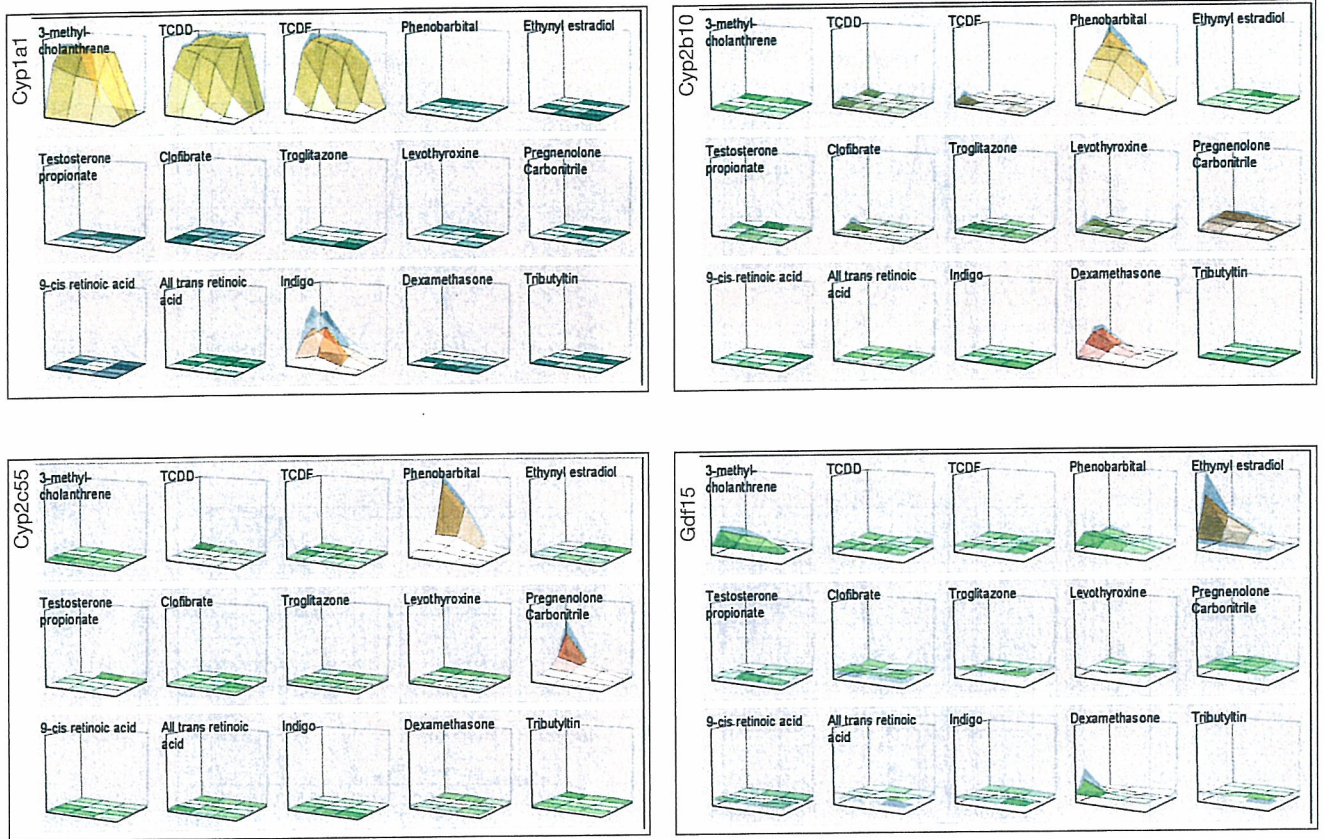


図4. 化合物間の発現比較

15種類の核内受容体リガンド化合物(各3次元グラフ内に表示)によるCyp1a1(左上), Cyp2c55(左下), Cyp2b10(右上)および, Gdf15(右下)の遺伝子発現を3次元表示したもの. 各軸は, 図2のとおり. 縦軸のスケールは遺伝子ごとに共通. リガンドに選択的な遺伝子の発現が確認される.

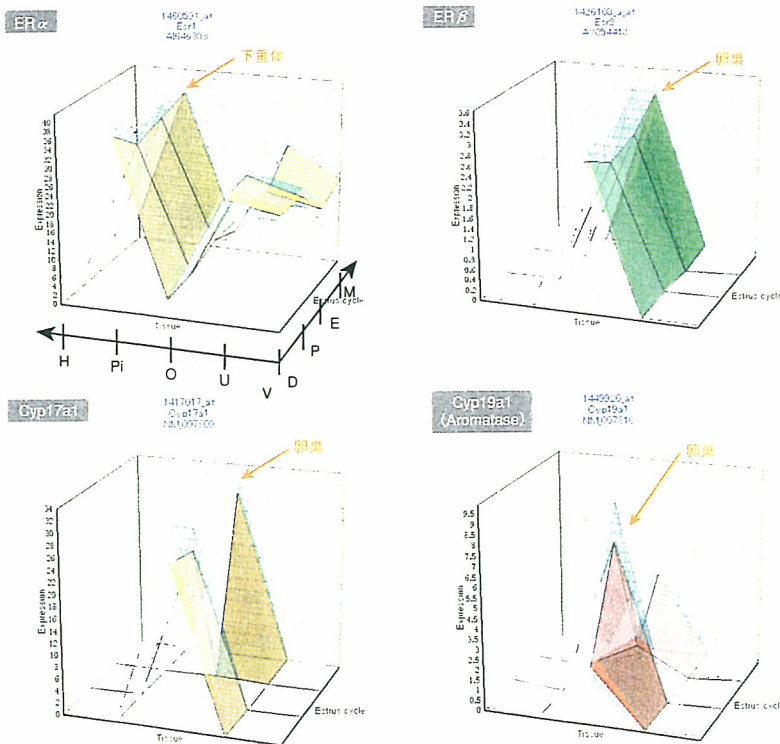


図5. 臓器間の発現比較

マウスの性周期 (Diestrus, Proestrus, Estrus, Metestrus) の4日間で1周期) ごとの視床下部 (H), 下垂体 (Pi), 卵巢 (O), 子宮 (U) および膣 (V) における, ER α , ER β , Cyp17a1 (steroid-17 α -hydroxylase), およびCyp19a1 (Aromatase) の遺伝子発現変動を3次元表示したもの. 後二者の酵素は卵巢において周期性を持って発現している.

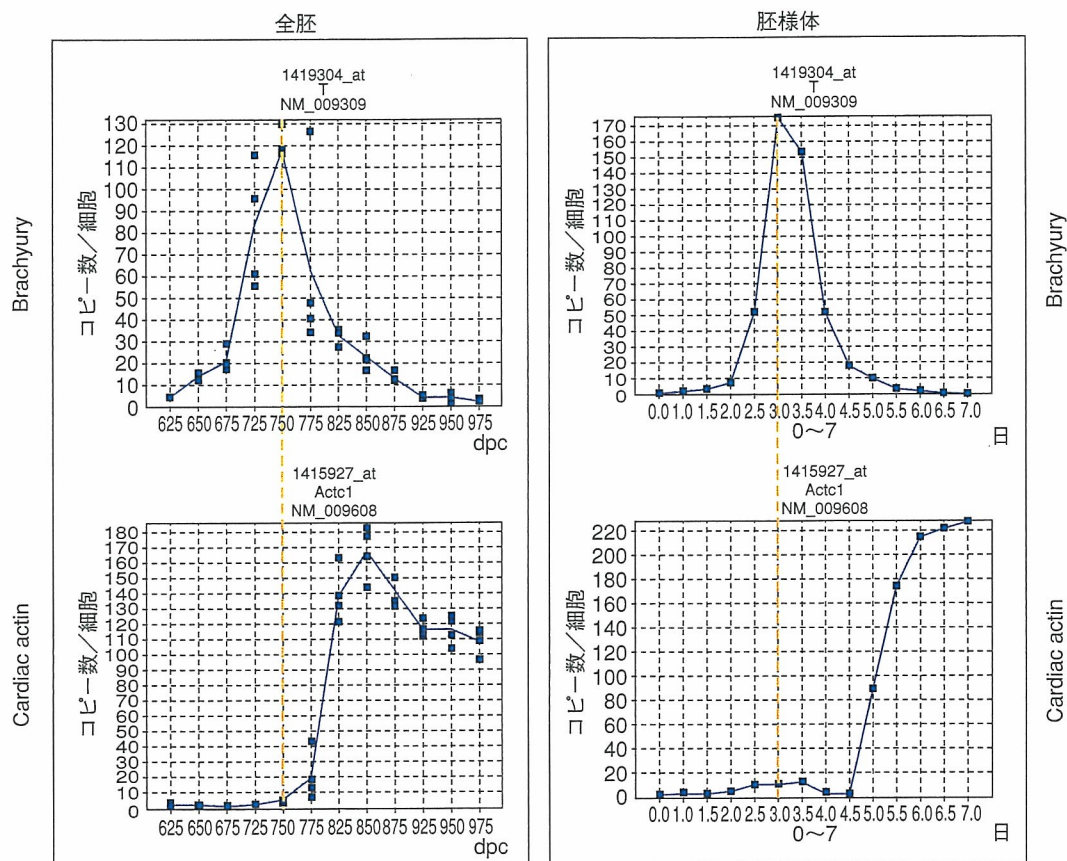


図6. マウス胎児（全胚）と胚様体の発現比較

マウス全胚の胎生6.25日～9.75日までの遺伝子発現と、胚様体の1日～7日目までの遺伝子発現の網羅的データベースから、初期中胚葉分化マーカーであるBrachyury遺伝子と、Cardiac actin遺伝子の経時変化を示す。

なる。エストロゲン活性化学物質による有害影響（内分泌かく乱化学物質問題）も同様にER（estrogen receptor）を介する受容体原性毒性と考えられ、胎生期にERを発現する組織が、低用量シグナルかく乱影響の重要標的であると考えられている。

ここでは、受容体原性毒性研究の基盤として、Percellome手法を適用して、①核内受容体作動性物質によるマウス雄肝臓の遺伝子発現変動、②性周期に伴うマウス雌生殖器遺伝子発現変動、③生後の発達過程におけるマウス雌生殖器遺伝子発現変動、の3種類のデータベースを構築した。例えば、①では10種類の核内受容体に作用する典型物質について、単回経口投与後、2, 4, 8, 24時間目の変動を解析し、Ethinyl-estradiolがGDF15, TCDDがCyp1a1, 9-cis Retinoic AcidがCyp26a1, DexamethasoneがCyp2b10, ClofibrateがCyp4a14, PCNがCyp2c55など、各々の受容体に特徴的な遺伝子発現を誘導するところがとらえられている（図4）。②の性周期データベースは視床下部、下垂体、卵巣、子宮、膈を対象としており、性周期との関連が網羅的にとらえら

れている（図5）。これらのデータベースは、今後、各種の候補物質が引き起こす変化を詳細に解析する際の基準として利用される。

V. 発生トキシコゲノミクスへの応用

発生毒性学は、個体発生過程におけるダイナミックな遺伝子発現調節の分子機構を把握することにより、さらに正確なものに補強されると考える。現在、C57BL/6マウス胚の器官形成初期初期にあたる胎生6.5～9.5日（プラグ確認日：0.5日）の、①全胚の遺伝子発現変動解析、②遺伝子欠失マウス全胚との比較、および③標的が明らかな既知発生毒性物質投与による本データベースの具体的な適用、を実施している。①についてはすでに0.25日間隔（Time point 計12点）の遺伝子発現データベースを得て、②遺伝子欠失胚のデータといくつかの注目すべき遺伝子についてはwhole mount ISHを用いた発現の検証を加えた。これと並行して、ES細胞からhanging drop法で得た胚様体の0.5日間隔の遺伝子発現データとの比較を実施している。個体発生に関与

する遺伝子群の多くは経時的に激しく変化しており、既知発生毒性物質投与実験については標的遺伝子シグナルカスケードを解析中である(図6)。

おわりに

ノーザンブロットでは実験サンプルにだけバンドが見られ、対照サンプルには遺伝子発現がないという結果を得ても、細胞1個当たりで定量してみると、対照が10コピーに対して実験サンプルが20コピーである場合がある。“無”が“有”になったのではなく、“10”が“20”になったのである。

さて、筆者らの属する毒性学でも、医学の分野でも、疾患概念や毒性概念が整理され、患者や実験動物を診断するには、まず、そのどれに当てはまるかを検討する。すなわち、どの“典型”に近い症例であるかを検討することから始まることが多い。

しかし、最近の医学・生物学には多因子疾患・多因子形質発現制御の概念が導入され、今から何年かの後には、“21世紀初頭までは、患者の遺伝子多型を調べずして治療を行っていた時代”として、“血液型を調べずに輸血していた時代”と並び称されるようになる可能性がある。このような多因子概念が定着すると、その多くは、“有(100%)”“無(0%)”の組み合わせではなく、“70%”“50%”“90%”といった半端な数の組み合わせであることが考えられる。すなわち、今までの離散値的な“典型”例を基準とするアプローチから、

連続値的な病態“スペクトラム”を直接扱うアプローチに変革していく可能性が考えられる。その際の網羅的データの解析とその蓄積の必要性を考えると、遺伝子発現データの定量化・標準化という問題は、今まで以上に重みを増すと考えられる。生命現象の網羅的解析にはトランスクリプトームだけでは不十分であることは自明であるが、この定量性を確保することは、これから実現されるであろう網羅的プロテオミクスなどの基盤としても重要ではないかと考える。

マイクロアレイなどから得られるトランスクリプトーム情報が、今後の医薬品審査や化学物質の安全性評価の際に必須なものとなる時代がすぐそこまで来ていることを念頭に、筆者らはPercellome法をさらに展開し、Percellome Projectデータベースを可能な限り高精度に保ちつつ毒性学的内容を充実させるべく最大限の活動を継続して行く所存であるが、この技術、あるいは研究内容が毒性学以外の研究分野にもお役に立つことができれば幸甚である。

謝辞 本システムの開発とプロジェクトの遂行に当たっては、当毒性部の全メンバー、特に松田菜恵、辻昌貴、森田紘一、今井あや子、安東朋子、安部麻紀、森山紀子、近藤優子、青柳千百合、相原妃佐子、渡辺忍の各氏の卓越した働きに深謝する。本研究は厚生労働科学研究費補助金H13-生活-012、H13-生活-013、H14-トキシコ-001、H15-化学-002、H18-化学-一般-001などによる。

文献

- 1) Kanno J, et al: BMC Genomics (2006) 7: 64
- 2) 菅野 純ら: 細胞工学 (2004) 23: 685-693
- 3) Matsumoto S, et al: Genome Informatics (2005) 16: 183-194

Effective Handling of Induced-Fit Motion in Flexible Docking

Miho Yamada Mizutani, Yoshihiro Takamatsu, Tazuko Ichinose, Kensuke Nakamura, and Akiko Itai*
Institute of Medicinal Molecular Design, Inc., Tokyo, Japan

ABSTRACT For structure-based drug design, where various ligand structures need to be docked to a target protein structure, a docking method that can handle conformational flexibility of not only the ligand, but also the protein, is indispensable. We have developed a simple and effective approach for dealing with the local induced-fit motion of the target protein, and implemented it in our docking tool, ADAM. Our approach efficiently combines the following two strategies: a vdW-offset grid in which the protein cavity is enlarged uniformly, and structure optimization allowing the motion of ligand and protein atoms. To examine the effectiveness of our approach, we performed docking validation studies, including redocking in 18 test cases and foreign-docking, in which various ligands from foreign crystal structures of complexes are docked into a target protein structure, in 22 cases (on five target proteins). With the original ADAM, the correct docking modes (RMSD < 2.0 Å) were not present among the top 20 models in one case of redocking and four cases of foreign-docking. When the handling of induced-fit motion was implemented, the correct solutions were acquired in all 40 test cases. In foreign-docking on thymidine kinase, the correct docking modes were obtained as the top-ranked solutions for all 10 test ligands by our combinatorial approach, and this appears to be the best result ever reported with any docking tool. The results of docking validation have thus confirmed the effectiveness of our approach, which can provide reliable docking models even in the case of foreign-docking, where conformational change of the target protein cannot be ignored. We expect that this approach will contribute substantially to actual drug design, including virtual screening. *Proteins* 2006;63:878–891.

© 2006 Wiley-Liss, Inc.

Key words: structure-based drug design; docking validation; docking accuracy; foreign-docking test; redocking test; ligand conformation; protein movement; ADAM; structure optimization; virtual screening

INTRODUCTION

The starting point of successful structure-based drug design is to obtain accurate docking models of arbitrary ligands to the protein of interest. For the purpose of

accurate docking, one of the most important problems is how to explore the whole solution space of docking exhaustively to find the correct docking solution for every protein–ligand. Because the ligand conformation of the bound form often varies greatly from the most stable conformation in the unbound state, an effective docking algorithm needs to consider the degree of freedom for not only relative rotation and translation between protein and ligand, but also ligand conformation. Furthermore, “induced-fit” motion of the target protein accompanying ligand binding has also been widely recognized.^{1–3} The induced-fit motion depends on the structure of the bound ligand. Accordingly, the flexibility of the target should be taken into account in practical drug design, including virtual screening, where it is required to generate correct docking models of various small molecules with different sizes and shapes to one three-dimensional (3D) structure of the target protein. However, the number of possible docking modes to be searched increases explosively with increasing degree of freedom of ligand conformation, let alone protein flexibility. A docking method that can fully handle flexibility of both the ligand and the target protein has not yet been reported, and this remains a very difficult problem.

Regarding ligand flexibility alone, many automated docking tools that treat it (and even small movements of protein side chains, in some cases) have been developed. Pioneering work includes the method of DesJarlais et al.⁴ and that of Leach and Kuntz,⁵ both of which were implemented in the well-known docking tool DOCK.⁶ We subsequently developed a unique automated docking method, ADAM, for flexible ligands and rigid proteins.^{7–9} Since then, many flexible docking tools, which are now widely used, have been reported, for example, AUTODOCK,¹⁰ FlexX,¹¹ GOLD,¹² and Glide.¹³ Most of these methods do not exhaustively explore the solution space of docking, even as regards ligand flexibility, owing to the use of discrete sampling of ligand conformation in some methods and application of a stochastic algorithm in other methods (see Discussion for details). Nevertheless, these methods

K. Nakamura's present address is the Quantum Bioinformatics Group, Center for Promotion of Computational Science and Engineering, Japan Atomic Energy Research Institute, Kyoto, Japan.

*Correspondence to: Akiko Itai, Institute of Medicinal Molecular Design, Inc., 5-24-5 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. E-mail: itai@immd.co.jp

Received 25 July 2005; Revised 23 November 2005; Accepted 6 December 2005

Published online 10 March 2006 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/prot.20931

are extensively used for drug design and virtual screening, and successful ligand discoveries have been reported.^{14–16}

On the other hand, a variety of approaches have been developed to deal with protein flexibility, although this is a more challenging task than a consideration of only the ligand flexibility. Some of the reported approaches have restricted the protein flexibility to be considered to changes of side-chain conformation in the ligand-binding region, by rotating the side chains at minimum,^{17,18} by including the side-chain movements in the Monte Carlo step,¹⁹ or by sampling discrete side-chain conformations based on rotamer libraries.^{20–22} Other approaches have tried to deal with the flexibility of protein backbones in addition to side chains, by using ensembles of experimentally determined protein structures^{23–26} or representative structures sampled through molecular dynamics simulations.²⁷ These methods should be capable of wider application than the approaches considering only side-chain motions. However, the variety of available experimental structures or the quality of sampled structures has a great influence on the docking performance with these methods. There have been other approaches to treat larger protein movements.^{28–30} In addition, soft potentials, where van der Waals (vdW) repulsion is reduced, have been used for implicitly handling the small protein motions.^{31,32} This approach is simple and inexpensive. However, Shoichet and coworkers³³ reported that the identification of true ligands is difficult in the case of using soft potential, because the false positives can get as good scores as the true actives.

In this article, we present a simple, practical, and effective approach for docking simulation with handling of both ligand flexibility and protein induced-fit motion. Furthermore, we have validated the performance and usefulness of our approach by applying it to several test systems of target proteins where induced-fit motion has been experimentally observed. The results are also compared with published results obtained with other docking tools.

From a very early version, our docking tool ADAM has been able to explore the solution space extensively and continuously, taking account of all possible conformational flexibility of the ligand, by means of an effective combination of systematic conformational search and several structure-optimization steps.^{7–9} The successful results of application of ADAM for virtual screening include identification of many novel inhibitors of aldose reductase (AR) and acetylcholinesterase (AChE).^{34,35} The unique algorithm of ADAM, which makes efficient use of the H-bonding pattern between ligand and protein as an initial clue to the rapid finding of promising docking modes, has had a great impact on the subsequent development of other flexible docking methods, for example, FlexX,¹¹ SLIDE,^{17,18} Ph4Dock,³⁶ and the method of Makino et al.³⁷ ADAM meets the requirements for extensive exploration of the solution space of docking involving the ligand conformational flexibility, but it has not so far coped with the induced-fit problem. To improve ADAM in this respect, we have developed an approach in which the protein flexibility around the ligand-binding region is implicitly consid-

ered during the docking process, and then is converted to explicit movements of protein atoms through the postdocking optimization. Generally, it is not feasible to exhaustively explore the solution space of docking including protein conformational flexibility, unlike the case of treating only the ligand flexibility. In this regard, our approach is very practical and has the advantage of being easily applicable to virtual screening. In-house use of our approach has already yielded several novel lead compounds as ligands of attractive drug targets (unpublished results).

To confirm the effectiveness of the induced-fit handling and the performance of the current version of ADAM, we have also carried out a series of validation tests. Two methods were used for this docking validation, that is, redocking and foreign-docking. In redocking, the protein and ligand structures are extracted from the crystal structure of their complex and docked together again. Because the protein structure in the complex obtained by cocrystallization with the test ligand is used as the docking input, consideration of induced-fit motion is not necessary. So, the redocking test should be rather easy. Although redocking tests have been performed extensively for evaluation of various docking tools,^{38–41} actual drug design does not require separation and redocking of a protein and ligand from a complex structure, so good results in the redocking test do not necessarily guarantee usefulness in actual drug design. On the other hand, another docking test is to dock various ligands from foreign crystal structures of complexes into a target protein structure. We call this test foreign-docking. Generally, it is hard to obtain the correct solution without taking protein motion at least partly into account, so the foreign-docking test is far more difficult than the redocking test.⁴² However, the nature of foreign-docking test is very similar to that of actual drug design, so the foreign-docking test is likely to be a better touchstone of value for structure-based drug design. Also, as a matter of course, the foreign-docking test, in which protein flexibility cannot be ignored, is indispensable for verifying the effectiveness of our approach to treat induced-fit motion. Therefore, we have focused on the foreign-docking test for evaluation of our method, although the redocking test was also performed for basic checking of the method and to compare the docking accuracy with that of other docking tools.

MATERIALS AND METHODS

Handling Local Movement of Protein

In the docking process of ADAM, the ligand-binding region of the protein is represented by a 3D grid.^{43,44} On each 3D grid point, the potential values including vdW and electrostatic potentials are stored, and the intermolecular energies of intermediate (and final) docking models can be estimated very rapidly by using 3D grid data. The 3D grid is prepared in advance of docking calculation, and the protein structure is assumed to be rigid throughout the ADAM process. In this article, we present an implementation to treat the protein induced-fit motion by means of a combination of the following two strategies.

One is the use in the ADAM docking process of the "vdW-offset" grid, the application of which in virtual screening has already been reported.³⁵ In this approach, the vdW energy curve for each atom pair is shifted in calculating the 3D grid data, so that the interatomic contact distances become smaller as the user indicates. As a result, the vdW potential surface is moved to widen the protein cavity uniformly. By means of this simple idea, the generation of docking models has become possible for ligands with large sizes and various shapes, even in cases where some accommodation of protein atoms is necessary for ligand binding. This approach has contributed our successes in virtual screening, including the discovery of novel AChE ligands. However, several problems arise with the use of the vdW-offset grid. The interaction energy would not be estimated properly by using the potential values on the vdW-offset grid, because both the shifted vdW energy curves and the uniform extension of the protein cavity are contrary to reality. Moreover, in ADAM docking models obtained by using the vdW-offset grid, severe clashes are occasionally detected between protein and ligand. This arises because the atomic coordinates of the protein structure are not changed when the potential surface is modified in the vdW-offset grid. To avoid such intermolecular clashes and to properly estimate the interaction energy, structure optimization including the protein atoms around the ligand-binding region and all ligand atoms is needed after the ADAM docking.

For this reason, we have developed a new energy-minimization program, named BLUTO. The BLUTO program is able to rapidly and sequentially converge the total energies of given docking models to the local-minimum states, considering the local motion of protein atoms. The L-BFGS method,⁴⁵ which has the merits of high speed and small memory use, is adopted as the optimization algorithm. Robust convergence can be achieved by a combination of the steepest decent method and L-BFGS optimization, avoiding pseudoconvergence attributable to saddle points on the energy surface. The total energy optimized by BLUTO consists of intermolecular vdW and electrostatic energies and intramolecular energies of ligand and protein. When BLUTO is used for postprocessing of ADAM docking calculations, all energy terms are estimated by using the AMBER94 all-atom force field,⁴⁶ in accordance with ADAM. The protein atoms to be moved in the energy-minimization process can be easily defined by the user, with cutoff distances from the ligand-binding site, and various settings for the movable region are possible by Boolean operation. The BLUTO optimization yields stable docking models without intermolecular clashes.

With the implementation of the above strategies, a reexamination of parameters in the ADAM process was required. It is sufficient that qualitatively correct docking modes are included in the high-ranked solutions generated by ADAM, because the ADAM docking models are optimized to quantitatively more stable ones through the energy-minimization process considering the local protein movements. Accordingly, for reduced computational cost, we have used a 10 times looser criterion than that in the

former version, for convergence in the optimization steps of ADAM docking; specifically, the fractional tolerance for energy-difference of the Powell optimization step was changed from 1.0e-2 to 1.0e-1.

Preparation and Execution of Docking Validation

Figure 1 shows a flow chart of our total process of docking with induced-fit handling. The docking validation in this paper was performed according to this procedure.

Preparation of target protein structure

The target proteins used for docking validation are listed in Table I (for details, see Results).

The target protein structures were taken from the Protein Data Bank (PDB).⁴⁷ Ligand and water molecules were removed from the PDB structure, except for an important water molecule in the active site of HIV-1 protease (4phv), according to the conditions of the FlexX test.¹¹ All hydrogen atoms were added to the protein structure, because the all-atom model is currently used in our docking. The AMBER94 force field type, AMBER94 atomic charge,⁴⁶ and ADAM H-bonding type⁹ were assigned for each atom. For a metal ion, the charge value of +1.0 was assumed. The above-mentioned preparation of protein structures was automatically performed by the program PDBFIL.^{43,44}

For each target protein structure, the ligand-binding region was indicated so as to fully include the binding site of test ligands. Then, a 3D grid with a regular interval of 0.4 Å was generated inside the ligand-binding region by using the CALGRID program.^{43,44} Two sorts of 3D grid were generated, that is, a normal grid and a vdW-offset grid, for comparison. For the vdW-offset grid in most of the test cases, the offset value of 0.3 Å was applied to all protein atoms inside the ligand-binding region. This setting means that the vdW energy curve for each atom pair is shifted by 0.3 Å to enlarge the protein cavity. The only exception was TK, for which the larger offset value of 0.5 Å was used because of the greater extent of protein conformational change around the ligand-binding cavity. Furthermore, CALGRID also generated the H-bonding dummy atoms, which are necessary for predicting the possible H-bonding schemes and for accelerating the docking process by making use of the distance relation of H-bonding pattern. From among the H-bonding dummy atoms automatically generated, two to six were selected for each protein system. These dummy atoms were not chosen only for the docking of particular ligands (i.e., the ligands used in this validation), but can also be used for the docking of various other ligands or for virtual screening. The criteria for selecting dummy atoms were as follows: (1) select dummy atoms at the bottom of hydrophilic pockets, especially those at sites where strong H-bonds or salt bridges are expected, (2) select dummy atoms at coordination sites to metal ions, (3) in the case of serine protease, select dummy atoms at sites where β -sheet-like interaction with the protein backbone is expected, (4) exclude dummy atoms at the edge of the grid region and in regions exposed to the solvent. Based on these criteria, we selected dummy

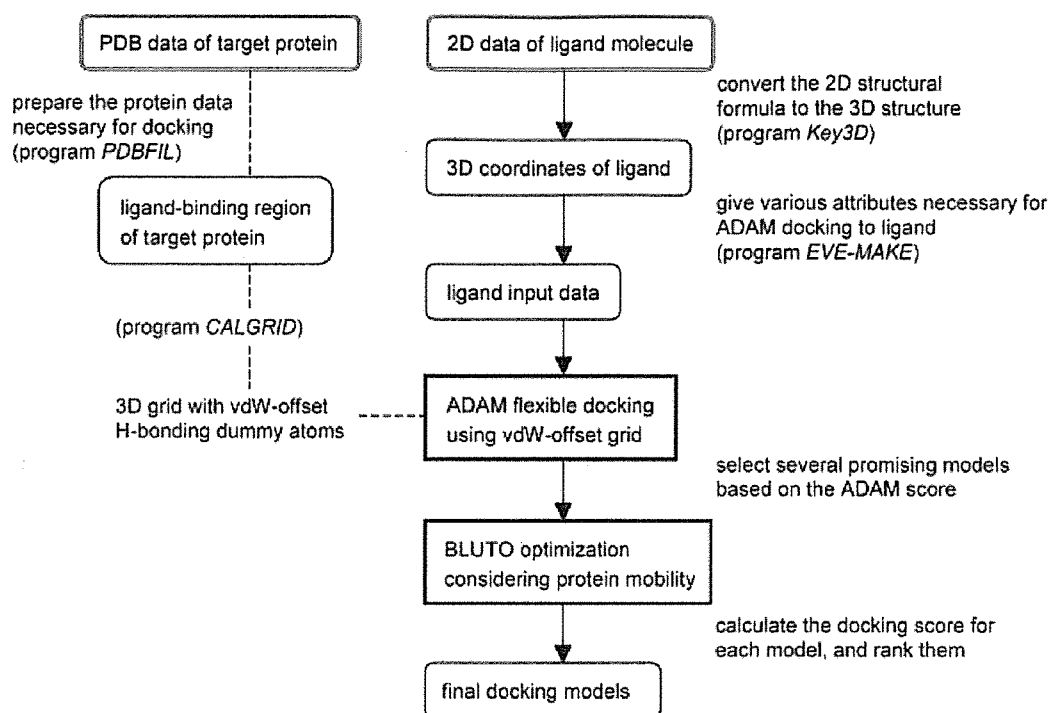


Fig. 1. Flow chart of our docking approach with handling of protein induced-fit motion.

atoms that were distributed sparsely, centering on the bottom of the protein cavity apart from the solvent-accessible region. We considered that a quite small number of dummy atoms was sufficient, because the matching of the H-bonding pattern by means of dummy atoms only provides an initial clue to search intermediate docking modes at the early stage of the calculation, and the solution space of docking is extensively explored in the subsequent steps of systematic conformational search and structure optimization.

Preparation of ligand structure

The 2D chemical structures of all ligands were drawn by ISIS-Draw.⁴⁸ Figure 2 shows the chemical structures of ligands used in the foreign-docking test; for the ligands in the redocking test, see ref. 11. The 2D structures of ligands were automatically converted to 3D structures, by using our program Key3D. Key3D is a useful tool for converting the 2D information (i.e., chemical structural formula) of small molecules to 3D structures with high accuracy of bond lengths, bond angles, and ring conformations, by means of the techniques of distance geometry and structure optimization. In the conversion of 2D ligand structures to 3D ones, we assumed the same protonation states as in the literature reporting the corresponding docking validation.^{11,49} As for the ligands not included in this literature, that is, the foreign-docking test cases other than the thymidine kinase (TK) system, the protonation states were assumed to be as shown in Figure 2. The MNDO-MULLIKEN atomic charge was calculated for each atom in the ligand, by using Key3D as an interface to

the MOPAC93 program.⁵⁰ Furthermore, the KMF¹ atomic type, which is our original force field type, was assigned to each atom through Key3D. Key3D also provided the KMF parameter set that included all force field parameters necessary for calculating the energies of the ligands in test cases based on the AMBER force field function. Details of Key3D and KMF will be reported elsewhere. The other attributes necessary for ADAM docking, that is, H-bonding atom types and settings of bond rotation, were automatically generated by our EVE-MAKE program.³⁵

Execution of ADAM docking and BLUTO optimization

The ADAM calculation was executed with the current default parameters: that is, the threshold of total potential energy for the intermediate docking model (E_{thres}) is 10,000 kcal/mol, the weighting on the H-bonding term in the total potential energy (W_{hb}) is 1.0, and the other parameters are the same as in ref. 9. The top 20 docking models generated by ADAM were subjected to structure optimization by BLUTO. In this step, all the ligand atoms and the protein side-chain atoms within 7 Å around the ligand were allowed to move, and a distance-dependent dielectric constant of $2.0r$ (r is an interatomic distance in Å units) was used. The resulting final docking models were reranked according to the conventional score, which is the summation of the intermolecular vdW and electrostatic energy terms and ligand intramolecular energy term calculated by the AMBER94 all-atom force field. This conventional score value for reranking, together with the

TABLE I. Redocking and Foreign-Docking Test Cases

Code	Target protein		Ligand	
	PDB ^a	Name	PDB ^b	Name
Redocking test				
5tim	5tim	triosephosphate isomerase	5tim	sulfate
1ldm	1ldm	lactate dehydrogenase	1ldm	oxamate
2phh	2phh	<i>p</i> -hydroxybenzoate hydroxylase	2phh	<i>p</i> -hydroxybenzoate
3ptb	3ptb	β -trypsin	3ptb	benzamidine
1ulb	1ulb	purine nucleoside phosphorylase	1ulb	guanine
3tpi	3tpi	trypsinogen	3tpi	Ile-Val
4tsl	4tsl	tyrosyl-transfer-RNA synthetase	4tsl	tyrosine
4dfr	4dfr	dihydrofolate reductase	4dfr	methotrexate
1stp	1stp	streptavidin	1stp	biotin
1dwd	1dwd	α -thrombin	1dwd	NAPAP
1dwc	1dwc	α -thrombin	1dwc	argatroban
1rnt	1rnt	ribonuclease T ₁	1rnt	2'-GMP
1tmn	1tmn	thermolysin	1tmn	N-(1-carboxy-3-phenylpropyl)-L-leucyl-L-tryptophan
4tln	4tln	thermolysin	4tln	L-leucyl-hydroxylamine
3cpa	3cpa	carboxypeptidase A	3cpa	glycyl-L-tyrosine
2ctc	2ctc	carboxypeptidase A	2ctc	L-phenyl lactate
4phv	4phv	HIV-1 protease	4phv	<i>N,N</i> -bis-(2(R)-hydroxy-1(S)-indanyl-2,6-(R,R)-diphenylmethyl-4-hydroxy-1,7-heptanediamide
121p	121p	H- <i>ras</i> p21 protein	121p	guanosine 5'-B,G-methylene-triphosphate
Foreign-docking test				
COX01	4cox	cyclooxygenase-2 (COX-2)	4cox	indomethacin
COX02			6cox	SC-558
COX03			3pgh	flurbiprofen
PKA01	1ydt	cAMP-dependent protein kinase (PKA)	1ydr	H-7
PKA02			1yds	H-8
PKA03			1ydt	H-89
AR01	1ah3	aldose reductase (AR)	1ah3	tolrestat
AR02			1ah0	sorbinil
ER01	3ert	estrogen receptor α (ER α)	3ert	4-hydroxytamoxifen (antagonist)
ER02			1ere	17 β -estradiol (agonist)
ER03			3erd	diethylstilbestrol (agonist)
ER04			1err	raloxifene (antagonist)
TK01	1kim	thymidine kinase (TK)	1kim	deoxythymidine (dT)
TK02			1ki7	5-iododeoxyuridine (idu)
TK03			1ki6	5-iodouracil anhydrohexitol nucleoside (ahiu)
TK04			1e2p	6-(dihydroxyisobutyl)thymine (dhbt)
TK05			1e2m	6-(3-hydroxypropyl)thymine (hpt)
TK06			1e2n	(6-[6-hydroxymethyl-5-methyl-2,4-dioxo-hexahydro-pyrimidin-5-yl-methyl]-5-methyl-1 <i>H</i> -pyrimidine-2,4-dione (hmtt)
TK07			1e2k	(North)-methanocarbathymidine (mct)
TK08			2ki5	aciclovir (acv)
TK09			1ki2	ganciclovir (gcv)
TK10			1ki3	penciclovir (pcv)

^aThe protein structure taken from the PDB data in this column was used in the docking test. For each system in the foreign-docking test, the common PDB structure of the target protein was used for docking of various test ligands.

^bNo 3D information about the ligand was taken from the PDB data in this column. Instead, the 3D coordinates of all ligands were generated from the 2D drawing, by using our Key3D program.

decomposition of each energy term, were output directly from BLUTO when the optimization process finished.

The accuracy of the docking model was estimated on the basis of root-mean-squares (RMS) deviation for all the

nonhydrogen atoms of the ligand from the reference coordinates, that is, the corresponding ligand coordinates in the crystal structure of the complex. In the foreign-docking test, the reference coordinates of the foreign ligand were