平成２１年度厚生労働科学研究費補助金

エイズ対策研究事業

課題番号　H19-エイズ-若手-003

多剤耐性 HIV における将来的な変異・構造予測と新規抗 HIV 薬開発

総括・分担　研究報告書

平成２２年５月

主任研究者　川下　理日人

大阪大学大学院薬学研究科・助教

目　次

厚生労働科学研究費補助金（エイズ対策研究事業）

総括・分担研究報告書

研究課題：多剤耐性 HIV における将来的な変異・構造予測と新規抗 HIV 薬開発

課題番号：H19-エイズ-若手-003

主任研究者：川下　理日人（大阪大学大学院 薬学研究科 微生物動態学分野 助教）

当該年度は３年計画の３年目にあたる。

＜研究要旨＞

　HIV 感染症は種々の抗 HIV 薬の開発やカクテル療法（HAART）の確立により、先進諸国において
は慢性疾患へと変化している。しかしながら、HIV は変異速度がきわめて速く、阻害剤耐性株の生じ
やすいウイルスであり、現実として多剤耐性ウイルスも出現している。それゆえ、既存のカクテル療
法に限界が生じるのは時間の問題である。

　そのような状況に対処すべく、既存の抗 HIV 薬の改変や新しい作用機序を有する抗 HIV 薬の開発が
急務となっている。これらは既存のカクテル療法と組み合わせることにより、さらに大きな効果を上
げるものと期待されている。しかし、それら新しい作用機序を有する抗 HIV 薬においても、耐性ウイ
ルスが既に報告されており、これらに対して何らかの対策を取る必要が生じている。

　そこで、もしこれらの耐性変異が生じる位置、およびそれらの構造変化を前もって予測することが
できれば、耐性ウイルスが生じた際にも速やかに対処できると思われる。よって、本研究ではそのよ
うな背景下、理論面と実験面から研究を行う予定であり、本年度はプロテアーゼの網羅的な配列解析
データを用いた系統樹作成と構造活性相関研究を行うことにより、薬剤耐性プロテアーゼに関する変
異予測に利用する基礎データの蓄積を行った。また、膜融合阻害剤に関してはランダムスクリーニン
グを用いることで、既知の阻害剤と同等程度の活性を有するペプチドを発見した。

分担研究者（１名）
・岡本晃典（大阪大学大学院 薬学研究科 微生
物動態学分野 助教）

A. 研究目的

　HIV は変異速度がきわめて速く、阻害剤耐性株
の生じやすいウイルスであるため、多剤耐性ウイ
ルスが出現し、既存の HAART に限界が生じると
予想される。そのような状況に対処すべく、既存
の抗 HIV 薬の改変や新しい作用機序を有する抗
HIV 薬の開発が急務となっている。

　そこで、これらの耐性変異が生じる位置・構造

変化を前もって予測することができれば、耐性ウ
イルス出現時にも速やかに新規阻害剤を設計する
ことが可能となる。また、その耐性を計算科学的
に速やかに評価することにより、HAART における
薬剤選択にも有用となる。

　このような背景下、本年度は以下の項目を中心
に検討を行った。

　１）多次元尺度構成法（MDS）を利用したプロ
テアーゼ薬剤耐性出現傾向の検討およびプロテア
ーゼの系統解析による選択圧の抽出

　２）T20 耐性ウイルスにも有効な膜融合阻害剤
の設計

３）新規薬剤開発を指向したシクロフィリン A に対する阻害剤のスクリーニング

B. 研究方法
＜膜融合阻害剤＞
　前年度に設計した 10 種のペプチド中、２種に T-20 と同等の活性を有するものが見出された。今回、さらなる活性向上および T20 耐性ウイルスへの適用を目指し、コンピュータによる知見に基づき、いくつかの領域を改変したペプチドをさらに 10 種設計した。これらと対照ペプチド T-20 を用いて、８種の CRF01_AE、B、C 型ウイルスおよび T20 耐性ウイルス（B 型）に対し、実験的な評価を行った。なお本実験は、研究協力者の亀岡正典特任准教授が行ったものである。

＜シクロフィリン A に対する阻害剤のスクリーニング＞
　MOE 2008 を用いて CypA の結晶構造（1cwa.pdb）を読み込み、Site Finder を用いて結合部位を検出した。続いて、検出された二つの結合部位（site 1 および 2）に対して、MOE Dock による MOE leadlike database（775,441 化合物）を用いたスクリーニングを行った。その際、Placement は Alpha Triangle、Refinement は行わず、その他の設定はデフォルトとした。ここで得られた結果から、同一化合物のトップスコアのみを抽出後、スコア順に並び替えを行った。

（倫理面への配慮）
　本研究では、ヒトの遺伝子や個人情報等の利用がないため、考慮する必要はない。もし、研究協力者がウイルス関係の実験を行うに当たって、個人の血液サンプル等を用いる場合は、倫理委員会の規定に則り、担当研究者以外にその個人情報が漏れないよう十分配慮する。

C. 研究結果
＜膜融合阻害剤＞
　前年度に T20 と同様の活性を示した２種のペプ

チドは、いずれも T20 耐性ウイルスに活性を示さなかったが、新規に購入したペプチドのうち、５種は耐性ウイルスにも活性を示し、そのうち３種は耐性株感受性株サブタイプに関わらず、強力な活性を有することがわかった。

＜シクロフィリン A に対する阻害剤のスクリーニング＞
　MOE Dock の S score による結果は以下の通りであった。Site 1 では、-19 以下のものが１個、以下 -18 以下が８個、-17 以下が 42 個、-16 以下が 144 個であった。同様に site 2 では、-19 以下のものが１個、以下 -18 以下が４個、-17 以下が８個、-16 以下が 65 個であった。ここで得られたフラグメントと CypA との結合状態を確認したところ、スコアの良いものに関しては両者とも標的蛋白質中の結合部位を埋めるような構造が得られた。

D. 考察
＜膜融合阻害剤＞
　今回は内部の疎水性相互作用を向上させるため、内部に存在する親水性残基を疎水性残基への変異を行い、さらに元の残基と大きさの類似した残基へと変異させたため、一部で活性が向上したものと考えられる。なお、前回設計したペプチドでは溶解性に問題があったが、今回は溶媒接触面に親水性の残基を配置したため、溶解性も向上した。

＜シクロフィリン A に対する阻害剤のスクリーニング＞
　先の結果より、S score はより結合部位の小さい site 1 で高いスコアを示す傾向がある。これは、フラグメントが各々小さいものであるために、より小さい site 1 で良好なスコアを示したのではないかと考えられる。なお、この結果は各フラグメント中のトップスコアのみの検証であり、その他のスコア分布を検証することによって、より強い親和性を示すフラグメントを選別することが可能であろう。

E. 結論

　今回我々は、系統解析・構造活性相関など、計算機を用いた手法を利用して、薬剤耐性プロテアーゼに関する情報を得た。また、膜融合阻害剤に関しては、分子動力学計算で得た結果をもとに設計したペプチドを用いて実験的な評価を行い、T-20 と同等程度の活性を得ることができた。

　今年度までの検討結果を活用して、薬剤耐性を有するウイルスの情報提供、薬剤耐性を有するウイルスに対抗できる阻害剤の開発を行い、今後もエイズ治療対策への貢献を行っていきたい。

F. 健康危険情報
　特になし

G. 研究発表
１．論文発表
＜欧文＞

1) Hidaka, S., Yamasaki, H., Ohmayu, Y., Matsuura, A., Okamoto, K., **Kawashita, N.**, Takagi, T.
Nonlinear classification of drug-induced arrhythmogenic activity by unsupervised classification method.
J. Toxicol. Sci., in press.

2) Yamashita, A., **Kawashita, N.**, Kubota-Koketsu, R., Inoue, Y., Watanabe, Y., Ibrahim, M. S., Ideno, S., Yunoki, M., Okuno, Y., Takagi, T., Yasunaga, T., Ikuta, K.
Highly conserved sequences for human neutralization epitope on hemagglutinin of influenza A viruses H3N2, H1N1 and H5N1: Implication for human monoclonal antibody recognition
Biochem. Biophys. Res. Commun., **393**, 614–618, (2010)

3) Hidaka, S., Shiraishi, H., Ohmayu, Y., Yamasaki, H., Okamoto, K., **Kawashita, N.**, Yasunaga, T., Takagi, T.
Ecotoxicity Prediction Using 3D Descriptors

J. Comput. Aided Chem., **11**, 11–18, (2010).

4) Sakudo, A., Xue, G., **Kawashita, N.**, Ano, Y., Takagi, T., Shintani, H., Tanaka, Y., Onodera, T., Ikuta, K.
Structure of the prion protein and its gene: an analysis using bioinformatics and computer simulation.
Curr. Protein Pept. Sci., **11**, 166–179, (2010).

5) Shirakuni, Y., Okamoto, K., **Kawashita, N.**, Yasunaga, T., Takagi, T.
Signal Detection of Drug Complications Applying Association Rule Learning for Stevens-Johnson Syndrome.
J. Comput. Aided Chem., **10**, 118–127 (2009).

２．学会発表
＜国際学会＞

1) Shirakuni Y., Uejima E., Takahara J., Okamoto K., **Kawashita N.**, Yasunaga T., Inui S., Takagi T.
Statistical Study on Chemical Risk Factors for SJS/TEN
The 9th Annual Meeting of the International Society of Pharmacovigilance, Reims(France), 2009.10.

2) Tian Y.-S., **Kawashita N.**, Kashiwada R., Sugimoto H., Yasunaga T., Okamoto K., Takagi T.
A QSAR Study using GA-PLS on anti-HIV-1 Small Molecules Targeting Cyclophilin A
The 6th International Conference on Partial Least Squares and Related Methods (PLS'09), Beijing (China), 2009.09.

＜国内学会＞

1) 川下理日人, 田雨時, 杉本裕昌, 岡本晃典, 安永照雄, 高木達也
シクロフィリンAを標的としたフラグメントスクリーニングによる抗HIV薬の探索研究
日本薬学会第130年会、岡山、2010年3月

2) 田雨時, 川下理日人, 岡本晃典, 安永照雄, 高木達也

HIV1 膜融合阻害剤としての 2-Aryl 5-(4-Oxo-3-phenethyl-2-thioxo thiazolidinylidenemethyl)furans の QSAR 解析
日本薬学会第 130 年会、岡山、2010 年 3 月

3）川下理日人、田雨時、亀岡正典、岡本晃典、安永照雄、生田和良、高木達也
計算機による分子設計を利用した HIV 膜融合阻害剤の開発研究
第 8 回次世代を担う若手のためのフィジカル・ファーマフォーラム、岡山、2010 年 3 月

4）川下理日人，安田匡志，田雨時，岡本晃典，川瀬雅也，安永照雄，生田和良，高木達也
ドッキングスタディを用いた薬剤耐性評価法の開発
第8回感染症沖縄フォーラム、沖縄、2010年2月

5）田雨時，杉本裕昌，岡本晃典，川下理日人，安永照雄，高木達也
シクロフィリン A を標的とする抗 HIV 薬のスクリーニング
第 8 回感染症沖縄フォーラム、沖縄、2010 年 2 月

6）杉本裕昌，田雨時，岡本晃典，川下理日人，安永照雄，高木達也
シクロフィリン A を標的とする抗 HIV 薬の開発を指向した構造活性相関研究
第 37 回構造活性相関シンポジウム、東京、2009 年 11 月

7）安田匡志，川下理日人，柏田理恵，田雨時，岡本晃典，川瀬雅也，安永照雄，高木達也
ドッキングスタディを利用した抗インフルエンザ薬の新規薬剤耐性評価法
第37回構造活性相関シンポジウム、東京、2009 年11月

8）栗花落昇平，山﨑広之，日高伸之介，川下理

日人，岡本晃典，高木達也
多変量解析による化学物質の加水分解反応予測
第37回構造活性相関シンポジウム、東京、2009 年11月

9）奥井朋子，井上 藍，栗花落昇平，白國優子，川下理日人，岡本晃典，高木達也
医薬品副作用データベースからの情報抽出並びに解析手法の検討
第32 回情報化学討論会、山口、2009年10月

10）米倉 聡，岡本晃典，川下理日人，日高伸之介，山﨑広之，川瀬雅也，高木達也
新規ノンパラメトリック回帰手法の応用
第 32 回情報化学討論会、山口、2009 年 10 月

11）山下明史，川下理日人，纐纈律子，井上雄嗣，柚木幹弘，高木達也，安永照雄，奥野良信，生田和良
抗 H3N2 ヒト型単クローン中和抗体が認識するエピトープ領域の保存性
第 57 回日本ウイルス学会学術集会、東京、2009 年 10 月

12）山﨑 広之，岡本 晃典，日高 伸之介，川下理日人，高木 達也
改良型双対尺度法を用いた医薬学データの数量化
第12回情報論的学習理論ワークショップ（IBIS 2009）、福岡、2009年10月

13）田中真奈美，川下理日人，高木達也，安永照雄，山口明人，西野邦彦
大腸菌 RND 型薬剤排出ポンプの進化
第11回日本進化学会大会、札幌、2009年9月

H. 知的所有権の出願・取得状況（予定を含む）
なし。

研究課題：多剤耐性 HIV における将来的な変異・構造予測と新規抗 HIV 薬開発

課題番号：H19-エイズ-若手-003

分担研究者：岡本　晃典（大阪大学大学院 薬学研究科 微生物動態学分野 助教）

分担研究課題：多次元尺度構成法（MDS）を利用したプロテアーゼ薬剤耐性出現傾向の検討およびプロテアーゼの系統解析による選択圧の抽出

＜研究要旨＞

　HIV 感染症は種々の抗 HIV 薬の開発やカクテル療法（HAART）の確立により、先進諸国においては慢性疾患へと変化している。しかしながら、HIV は変異速度がきわめて速く、阻害剤耐性株の生じやすいウイルスであり、現実として多剤耐性ウイルスも出現している。それゆえ、既存のカクテル療法に限界が生じるのは時間の問題である。

　そのような状況に対処すべく、既存の抗 HIV 薬の改変や新しい作用機序を有する抗 HIV 薬の開発が急務となっている。これらは既存のカクテル療法と組み合わせることにより、さらに大きな効果を上げるものと期待されている。しかし、それら新しい作用機序を有する抗 HIV 薬においても、耐性ウイルスが既に報告されており、これらに対して何らかの対策を取る必要が生じている。そこで、もしこれらの耐性変異が生じる位置、およびそれらの構造変化を前もって予測することができれば、耐性ウイルスが生じた際にも速やかに対処できると思われる。

　本分担研究では、多次元尺度構成法（MDS）を利用したプロテアーゼ薬剤耐性出現傾向の検討およびプロテアーゼの系統解析による選択圧の抽出を行い、変異傾向の解析を行った。

A. 研究目的

　HIV は変異速度がきわめて速く、多剤耐性ウイルスが出現しやすいウイルスであるため、既存のHAART に限界が生ずるのは時間の問題である。

　そのような状況に対処すべく、既存の抗 HIV 薬の改変や新しい作用機序を有する抗 HIV 薬の開発が急務となっている。しかし、それらに対しても耐性ウイルスが既に報告されており、さらなる対策が必要となる。

　そこでこれら耐性変異が生じる位置、それらの構造変化を前もって予測することができれば、耐性ウイルス出現時にも速やかに対処可能となる。

　そのような背景下、本年度は前年度までに行った系統解析で得られた情報を元に、多次元尺度構成法（MDS）を利用したプロテアーゼ薬剤耐性出現傾向の検討およびプロテアーゼの系統解析による選択圧の抽出を行い、変異傾向の解析を行った。

B. 研究方法

　LosAlamos の HIV databases より、SIV、B 型のプロテアーゼリファレンス配列および CRF01_AE プロテアーゼの塩基配列を取得した。このうち配列が全く同じ株を除いた 297 株に対して mafft でアライメントを行い、MEGA による距離マトリックスの作成後、距離データを用い R で多次元尺度構成法（MDS）を用いて各株を三次元空間にプロットした。また、Stanford University HIV Drug Rresistance

Database の HIVdb program を用いて各プロテアーゼ阻害剤に対する耐性予測を行い、その強度によってさらに4つのレベルに分類した。

また、アライメントを行ったデータセットを用いて、MP 系統樹を作成し、宮田・安永らの方法（sqdif）により Ks/Ka を求め、Ks/Ka>1 の組み合わせを抽出した。さらに、正の選択の有無を、MEGA の codon-based z-test で検定し、どちらの手法でも正の選択が起こっているとされるものがあるかを確認した。

（倫理面への配慮）
　特になし。

C. 研究結果

MDS を行った結果、取得した株は大きく分けて2つの領域にプロットされた。しかし、これらは薬剤耐性の強度では分かれておらず、引き続き検討が必要である。

また、選択圧の抽出に関しては、Sqdif と z-test の両方で正の選択が起こっているとされたものは180通りあった。これらを系統樹上で詳細に確認したところ、これらは系統樹上で遠い位置にある組み合わせばかりであり、正の選択が起こっているといえる組み合わせは得られなかった。

D. 考察

MDS でプロットされた株のうち、B型リファレンス配列および CRF01_AE 9株の合計10株からなるクラスターは、系統樹上でも他の株と離れた位置に分類されているが、MDS の結果の方がよりはっきりと離れている。両者の解析結果から、CRF01_AE の中でもこれら9株とそれ以外の株では変異傾向が異なることが考えられる。

E. 結論

今回、MDS を用いた進化傾向の解析、および正の選択に関する検討を行った。今回の解析結果だけでは説明できない点も多く、また配列情報は随時蓄積されるため、今後も同様の解析を続け、よ

り詳細な変異傾向の解析・進化予測等を行っていきたいと考えている。

F. 健康危険情報
　特になし

G. 研究発表
1. 論文発表
＜欧文＞

1) Kanbayashi, Y., Nomura, K., **Okamoto, K.**, Matsumoto, Y., Horiike, S., Takagi, T., Taniwaki, M. Statistical examination to determine whether only 48-hr value for serum concentration during high dose methotrexate therapy is a predictor for clinical adverse events using ordered logistic regression analysis.
Ann. Hematol., in press.

2) Hidaka, S., Yamasaki, H., Ohmayu, Y., Matsuura, A., **Okamoto, K.**, Kawashita, N., Takagi, T. Nonlinear classification of drug-induced arrhythmogenic activity by unsupervised classification method.
J. Toxicol. Sci., in press.

3) Hidaka, S., Shiraishi, H., Ohmayu, Y., Yamasaki, H., **Okamoto, K.**, Kawashita, N., Yasunaga, T., Takagi, T.
Ecotoxicity Prediction Using 3D Descriptors
J. Comput. Aided Chem., **11**, 11–18, (2010).

4) Shirakuni, Y., **Okamoto, K.**, Kawashita, N., Yasunaga, T., Takagi, T.
Signal Detection of Drug Complications Applying Association Rule Learning for Stevens-Johnson Syndrome.
J. Comput. Aided Chem., **10**, 118–127 (2009).

＜和文＞

1) **岡本晃典**, 山口進康, 馬場貴志, 高木達也, 那須正夫, 細菌数測定法における誤差分布の推定,

医薬品研究, **40**, 1–8, (2009).

2．学会発表

＜国際学会＞

1) Shirakuni Y., Uejima E., Takahara J., **Okamoto K.**, Kawashita N., Yasunaga T., Inui S., Takagi T.
Statistical Study on Chemical Risk Factors for SJS/TEN
The 9th Annual Meeting of the International Society of Pharmacovigilance, Reims(France), 2009.10.

2) Tian Y.-S., Kawashita N., Kashiwada R., Sugimoto H., Yasunaga T., **Okamoto K.**, Takagi T.
A QSAR Study using GA-PLS on anti-HIV-1 Small Molecules Targeting Cyclophilin A
The 6th International Conference on Partial Least Squares and Related Methods (PLS'09), Beijing (China), 2009.09.

＜国内学会＞

1) 神林祐子，山崎大典，古谷和春，**岡本晃典**，高木達也，稲野辺厚，倉智嘉久
抗がん剤の薬物相互作用予測に向けた薬物代謝酵素CYP3A4リガンドのファーマコフォア3D-QSAR解析
日本薬学会第130年会、岡山、2010年3月

2) 川下理日人，田雨時，杉本裕昌，**岡本晃典**，安永照雄，高木達也
シクロフィリンA を標的としたフラグメントスクリーニングによる抗HIV薬の探索研究
日本薬学会第130 年会、岡山、2010 年3月

3) 田雨時，川下理日人，**岡本晃典**，安永照雄，高木達也
HIV1 膜融合阻害剤としての 2-Aryl 5-(4-Oxo-3-phenethyl-2-thioxo thiazolidinylidenemethyl)furans の QSAR 解析
日本薬学会第130 年会、岡山、2010 年3月

4) 川下理日人、田雨時、亀岡正典、**岡本晃典**、安永照雄、生田和良、高木達也
計算機による分子設計を利用した HIV 膜融合阻害剤の開発研究
第8回次世代を担う若手のためのフィジカル・ファーマフォーラム、岡山、2010 年3 月

5) 川下理日人，安田匡志，田雨時，**岡本晃典**，川瀬雅也，安永照雄，生田和良，高木達也
ドッキングスタディを用いた薬剤耐性評価法の開発
第8回感染症沖縄フォーラム、沖縄、2010年2月

6) 田雨時，杉本裕昌，**岡本晃典**，川下理日人，安永照雄，高木達也
シクロフィリンA を標的とする抗HIV 薬のスクリーニング
第8 回感染症沖縄フォーラム、沖縄、2010 年2月

7) 杉本裕昌，田雨時，**岡本晃典**，川下理日人，安永照雄，高木達也
シクロフィリンA を標的とする抗HIV 薬の開発を指向した構造活性相関研究
第37 回構造活性相関シンポジウム、東京、2009年11 月

8) 安田匡志，川下理日人，柏田理恵，田雨時，**岡本晃典**，川瀬雅也，安永照雄，高木達也
ドッキングスタディを利用した抗インフルエンザ薬の新規薬剤耐性評価法
第37回構造活性相関シンポジウム、東京、2009年11月

9) 栗花落昇平，山﨑広之，日高伸之介，川下理日人，**岡本晃典**，高木達也
多変量解析による化学物質の加水分解反応予測
第37回構造活性相関シンポジウム、東京、2009年11月

10) 奥井朋子，井上 藍，栗花落昇平，白國優子，川下理日人，**岡本晃典**，高木達也
医薬品副作用データベースからの情報抽出並びに解析手法の検討
第32 回情報化学討論会、山口、2009年10月

11) 米倉 聡，**岡本晃典**，川下理日人，日高伸之介，山﨑広之，川瀬雅也，高木達也
新規ノンパラメトリック回帰手法の応用
第32 回情報化学討論会、山口、2009 年10 月

7

12）山﨑 広之，**岡本 晃典**，日高 伸之介，川下
理日人，高木 達也
改良型双対尺度法を用いた医薬学データの数量
化
第12回情報論的学習理論ワークショップ（IBIS

2009）、福岡、2009年10月

8

Ⅲ．研究成果の刊行物・一覧

書籍：なし

雑誌

発表者氏名	論文タイトル	発表誌名	巻	ページ	出版年
Kanbayashi, Y., Nomura, K., Okamoto, K., Matsumoto, Y., Horiike, S., Takagi, T., Taniwaki, M.	Statistical examination to determine whether only 48-hr value for serum concentration during high dose methotrexate therapy is a predictor for clinical adverse events using ordered logistic regression analysis.	*Ann. Hematol.*			in press
Hidaka, S., Yamasaki, H., Ohmayu, Y., Matsuura, A., Okamoto, K., Kawashita, N., Takagi, T.	Nonlinear classification of drug-induced arrhythmogenic activity by unsupervised classification method	*J. Toxicol. Sci.*			in press
Yamashita, A., Kawashita, N., Kubota-Koketsu, R., Inoue, Y., Watanabe, Y., Ibrahim, M. S., Ideno, S., Yunoki, M., Okuno, Y., Takagi, T., Yasunaga, T., Ikuta, K.	Highly conserved sequences for human neutralization epitope on hemagglutinin of influenza A viruses H3N2, H1N1 and H5N1: Implication for human monoclonal antibody recognition	*Biochem. Biophys. Res.Commun.*	393	614–618	2010
Hidaka, S., Shiraishi, H., Ohmayu, Y., Yamasaki, H., Okamoto, K., Kawashita, N., Yasunaga, T., Takagi, T.	Ecotoxicity Prediction Using 3D Descriptors	*J. Comp. Aided Chem.*	11	11–18	2010
Sakudo, A., Xue, G., Kawashita, N., Ano, Y., Takagi, T., Shintani, H., Tanaka, Y., Onodera, T., Ikuta, K.	Structure of the prion protein and its gene: an analysis using bioinformatics and computer simulation	*Curr. Protein Pept. Sci.*	11	166–179	2010
Shirakuni, Y., Okamoto, K., Kawashita, N., Yasunaga, T., Takagi, T.	Signal Detection of Drug Complications Applying Association Rule Learning for Stevens-Johnson Syndrome.	*J. Comput. Aided Chem.*	10	118–127	2009

9

1．Highly conserved sequences for human neutralization epitope on hemagglutinin of influenza A viruses H3N2, H1N1 and H5N1: Implication for human monoclonal antibody recognition

Yamashita, A., Kawashita, N., Kubota-Koketsu, R., Inoue, Y., Watanabe, Y., Ibrahim, M. S., Ideno, S., Yunoki, M., Okuno, Y., Takagi, T., Yasunaga, T., Ikuta, K.

Biochem. Biophys. Res.Commun., **393**, 614–618, (2010).

2．Ecotoxicity Prediction Using 3D Descriptors

Hidaka, S., Yamasaki, H., Ohmayu, Y., Matsuura, A., Okamoto, K., Kawashita, N., Takagi, T.

J. Comp. Aided Chem., **11**, 11–18, (2010).

Highly conserved sequences for human neutralization epitope on hemagglutinin of influenza A viruses H3N2, H1N1 and H5N1: Implication for human monoclonal antibody recognition

Akifumi Yamashita [a], Norihito Kawashita [b], Ritsuko Kubota-Koketsu [c], Yuji Inoue [c], Yohei Watanabe [c], Madiha S. Ibrahim [c], Shoji Ideno [d], Mikihiro Yunoki [c,d], Yoshinobu Okuno [e], Tatsuya Takagi [a,b], Teruo Yasunaga [a], Kazuyoshi Ikuta [c,*]

[a] Department of Genome Informatics, Genome Information Research Center, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565-0871, Japan
[b] Graduate School of Pharmaceutical Sciences, Osaka University, Suita, Osaka 565-0871, Japan
[c] Department of Virology, Research Center for Infectious Disease Control, Research Institute for Microbial Diseases, Osaka University, Suita, Osaka 565-0871, Japan
[d] Osaka Research Laboratory, Benesis Corporation, Yodogawa-ku, Osaka 532-6505, Japan
[e] Kanonji Institute, The Research Foundation for Microbial Diseases of Osaka University, Kanonji, Kagawa 768-0061, Japan

ARTICLE INFO

ABSTRACT

The epitope sequences within the hemagglutinin (HA) of influenza A virus H3N2 at amino acid residues 173–181 and 227–239 that forms anti-parallel β-sheet structure are similarly recognized by human monoclonal antibodies (HuMAbs), B-1 and D-1 that we recently obtained using the peripheral blood lymphocytes from two influenza-vaccinated volunteers. Both HuMAbs showed strong global neutralization of H3N2 strains. Here we show the significant conservation of the β-sheet region consisting of the above-mentioned two epitope regions in H3N2. In addition, we also identified the corresponding regions with similar structure in other subtypes such as H1N1 and H5N1. These two regions are similarly located underneath the receptor-binding sites of individual subtypes. Analysis of those regions using sequences available from the Influenza Virus Resource at the National Center for Biotechnology Information revealed that compared with those in the known neutralizing epitopes A–E, those sequences were fairly conserved in human H3N2 ($n = 7955$), swine H1N1 ($n = 360$) and swine H3N2 ($n = 235$); and highly conserved in human H1N1 ($n = 2722$), swine-origin pandemic H1N1 ($n = 1474$), human H5N1 ($n = 319$) and avian H5N1 ($n = 2349$). Phylogenetic tree for these regions formed clearly separable clusters for H1N1, H3N2 and H5N1, irrespective of different host origin. These data may suggest a possible significance of those regions for development of alternative vaccine that could induce neutralizing antibodies reactive against wide-range of influenza virus strains.

© 2010 Elsevier Inc. All rights reserved.

Introduction

Influenza pandemics are rarely occurring but recurrent events. Such pandemics are associated with the emergence of new types of influenza virus for which the human population has no immunity. The highly pathogenic avian influenza A virus H5N1 was widely believed to be the most likely causative candidate for the next pandemic [1,2]. However, novel swine-origin influenza A (H1N1) virus, named pandemic (H1N1) 2009 virus (H1N1 (pdm)) thereafter, has emerged in April, 2009 with a human pandemic potential [3,4]. Since it is not known beforehand which strain of influenza A virus could give rise to a next pandemic, prepandemic vaccines that can induce broadly cross-reactive immune responses are mandatory.

Influenza virus undergoes a high rate of antigenic changes giving rise to a new type of influenza strain(s). Influenza A viruses are classified into subtypes based on the antigenicity of their hemagglutinin (HA) and neuraminidase (NA) molecules; i.e., 16 HA (H1–H16) and 9 NA (N1–N9) subtypes [5,6]. Human influenza A viruses of at least three HA subtypes, H1, H2 and H3 have emerged as important pathogens. Currently, H1N1 and H3N2 influenza A viruses are seasonally circulating and causing human infections. In 2003, influenza virus subtype H5 of avian origin emerged as a human pathogen and it is much more lethal than earlier strains [7]. Recently, the H1N1 (pdm) virus emerged in April, 2009 as a novel type of influenza virus and is spreading rapidly in the human population.

Seed viruses for the production of inactivated influenza vaccine, influenza A virus H1N1 and H3N2 as well as influenza B virus, are naturally occurring virus strains that replicate well in the allantoic cavity of embryonated chicken eggs. Further, similar vaccine against

the highly pathogenic H5N1 virus could be prepared by reverse genetics of H5N1-derived HA and NA genes, with the internal genes of a high-yield strain in order to allow well growth of the H5 virus in chicken embryos [1,8].

We recently succeeded to establish two hybridoma cell clones producing human monoclonal antibodies (HuMAbs), by using the peripheral blood mononuclear cells from influenza-vaccinated volunteers, one from each. Both HuMAbs showed high neutralizing activity against wide-range of H3N2 virus strains [9]. Epitope mapping with synthetic peptides identified highly conserved amino acids (aa) 178–181 and 227–239 in the H3N2 HA as the epitope regions recognized similarly by both HuMAbs [9]. These amino acids are located underneath the receptor-binding site of the HA globular region. Herein, we expanded our study to computer-based characterization of HA sequences of human H3N2 that are available from the National Center for Biotechnology Information [10]. Further, we identified conserved sequences in other influenza A subtypes such as H1N1 including H1N1 (pdm), and H5N1 derived from human, swine and/or avian that correspond to the above two epitope regions of human H3N2. The two regions formed independent clusters for individual subtypes and further, the viruses belonging to the same subtypes but derived from different hosts fell into the same clusters by phylogenetic analysis. Furthermore, the sequences of the two epitope regions were highly conserved in all the viruses belonging to the same subtypes. These data may suggest that the two epitope regions of not only H3N2 but also other subtypes such as H1N1 and H5N1 could function as a significant human epitope to induce global neutralizing antibodies.

Materials and methods

Collection of the influenza A HA sequences and extraction of HA1 region as well as the epitope regions recognized by two independent HuMAbs: B-1 and D-1. Full-length protein sequences of HA of human-, swine- and avian-derived influenza A viruses H1N1 (including pandemic 2009), H3N2 and H5N1 were obtained from the Influenza Virus Resource at the National Center for Biotechnology Information (http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html) [10]. The HA sequences were then aligned using mafft v6.705b [11]. To determine HA1 regions, we searched for palindrome sequence identified just after the cleavage point (n′-GLFGAIAGFI-c′, the palindrome region is underlined) [12].

In our previous paper, we prepared two HuMAbs (B-1 and D-1), independently from two vaccinated volunteers, which showed strong neutralizing activities against global strains of influenza A virus H3N2. The epitope regions were identified using a total of 158 sets of 15-residue peptides (overlapping by 13 amino acids) spanning aa 1–329 of the HA1 region of human H3N2: 173–181, named "upper region" according to the positive reactions with four peptides (aa 167–181, 169–183, 171–185 and 173–187) and 227–239, named "lower region" according to the positive reactions with two peptides (aa 225–239 and 227–241) [9]. Sequences corresponding to these regions were also extracted from virus strains of different host origin or from those belonging to other subtypes from the Influenza Virus Resource at the National Center for Biotechnology Information [10]: H1N1 derived from human and swine, and H1N1 (pdm) from human; H3N2 derived from human and swine; and H5N1 derived from human and avian.

As control sequences, we similarly extracted the human H3N2 sequences of known neutralizing epitopes A to E [13,14] from the Influenza Virus Resource at the National Center for Biotechnology Information [10].

Sequence comparison. Sequence logos for the epitope sequences recognized by B-1 and D-1 and the corresponding sequences in the

human-, swine- and avian-derived viruses were constructed using weblogo [15].

The evolutionary relationship was inferred using the minimum evolution (ME) method [16]. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Poisson correction method [17] and are in the units of the number of amino acid substitutions per site. The ME tree was searched using the close-neighbor-interchange algorithm [18] at a search level of 1. The neighbor-joining algorithm [19] was used to generate the initial tree. All positions containing gaps and missing data were eliminated from the dataset (complete deletion option). Phylogenetic analyses were conducted in MEGA4 [20].

Results

Previously we identified human epitope regions to form antiparallel β-sheet structure within the HA of human-derived H3N2 (upper and lower regions) that are recognized by B-1 and D-1 HuMAbs showing neutralizing activities to wide-range of strains in this subtype, but not cross-reactive to other subtypes [9]. To examine the sequence conservation rate of the two regions, we obtained 7955 complete HA sequences derived from human H3N2 that were available at December 28, 2009 from the Influenza Virus Resource at the National Center for Biotechnology Information [10]. Among these, 35 and 48 variants for the upper and lower regions, respectively, were detected. This indicates that they produced one variant per 227.3 and 165.7 strains in these regions, respectively (Supplementary Table 1). In contrast, sequences of the known neutralizing epitopes A to E [13,14] that were extracted from the same 7955 sequences were much more variable, especially in epitopes A and B2, i.e., one variant production per 19.2 and 42.1 strains, respectively (Supplementary Table 1). Epitope C2 and the lower region showed similar frequency for variant production. However, this epitope C2 consisted of only five amino acids, while the lower region consisted of 13 amino acids, indicating much higher conservation in the sequences of this lower region. In Supplementary Table 2, the sequences of top 10 variants and the percentages of each variant in the known neutralizing epitopes A to E are shown. The sequence logo analysis using all these variants (Supplementary Table 1) in individual epitope regions also supported the high heterogeneity of these known neutralizing epitopes, especially in epitopes A and B2, as well as B1 (Fig. 1).

Next, we examined the possible detection of the corresponding β-sheet regions in HA molecules in other subtypes of influenza A virus. A total of 12470, 595, and 2349 full-length HA sequences derived from human (H1N1, H1N1 (pdm), H3N2 and H5N1), swine (H1N1 and H3N2), and avian (H5N1) influenza A virus were obtained from the Influenza Virus Resource at the National Center for Biotechnology Information [10], respectively, at December 28, 2009. Sequences corresponding to the upper and lower regions of H3N2 were also extracted from full-length HA sequences of H1N1 and H5N1 derived from different hosts. The efficiency to produce variants was much higher in swine- than human-derived influenza viruses, H1N1 and H3N2 (Supplementary Table 1). Especially, H1N1 (pdm) showed nearly perfect conservation among the 1474 isolates (Supplementary Table 1).

The top 10 variants and percentages of individual sequences in these upper and lower regions are shown in Supplementary Table 3. These data were summarized and illustratively shown for the individual variant ratios in Fig. 2. The sequences in human H1N1 and H1N1 (pdm) were highly conserved in both upper and lower regions. In case of human H3N2 as well as human and avian H5N1, the sequences of both regions were relatively highly conserved. In contrast, conservation of both regions in H1N1 and

Fig. 1. Sequence logo of the known neutralizing epitopes A–E of human influenza A virus H3N2. The sequences of the known neutralizing epitopes A–E in human H3N2 that were extracted from the Influenza Virus Resource at the National Center for Biotechnology Information were subjected to sequence logo for analysis of variation sites and their ratios.

H3N2 derived from swine were significantly lower. Sequence logo analysis using all the variants (Supplementary Table 3) supported this conclusion (Fig. 3). Appearance of variants was mainly due to the variations in the restricted amino acid residues in most cases: residues 173 and 227 in human H3N2; residues 173, 174, 179, 229 and 230 in swine H3N2; residues 173, 179, 181, 227, 237, 238 and 239 in swine H1N1; and residues 178, 179, 231, 238 and 239 in human and avian H5N1. Consequently, the upper and lower regions were highly conserved in human H1N1 and H1N1 (pdm), while less conserved in the other viruses derived from human, swine and avian.

Although the phylogenetic tree of HA1 amino acid sequences that are the original ones for the extraction of sequences at the upper and lower regions clustered depending not only on subtypes but also on the hosts, those of the upper and lower regions did not clustered depending on the hosts (Fig. 4). This result shows that diversities of these regions are independent of the hosts of the viruses.

Discussion

The HA molecule of influenza viruses plays pivotal roles not only for the receptor-binding and fusion process but also for bud-
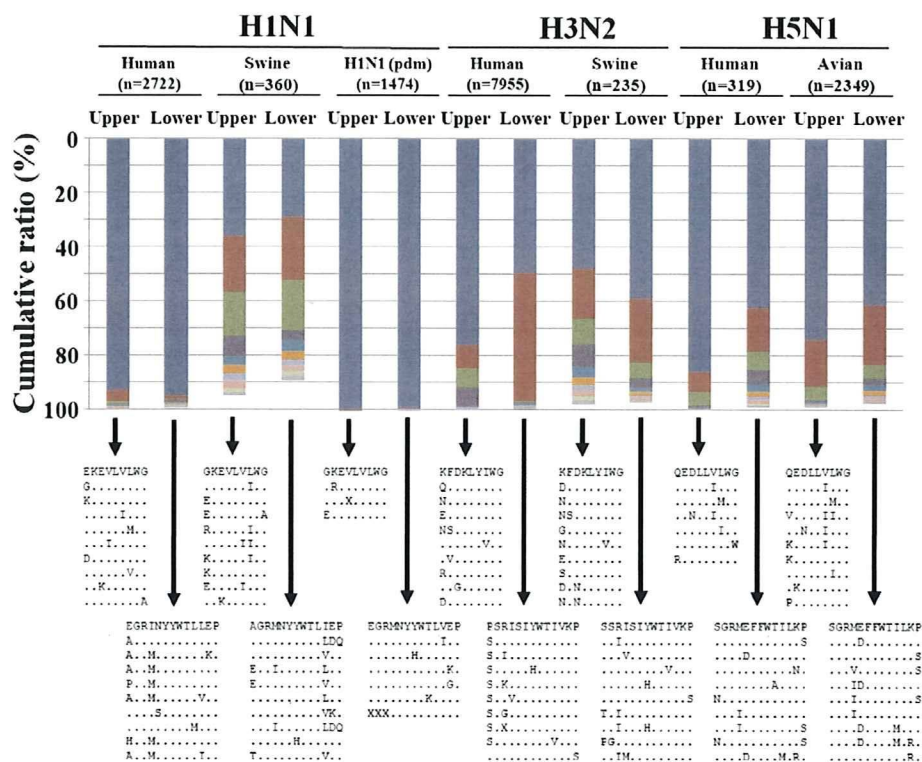

Fig. 2. Sequence variation and accumulation rates at the upper and lower sequences in H1N1, H3N2 and H5N1 derived from different hosts. The sequences and the accumulated rates of top 10 sequences at the upper and lower epitope regions in H1N1 derived from human and swine as well as H1N1 (pdm); H3N2 derived from human and swine; and H5N1 derived from human and avian, as shown in Supplementary Table 3, are shown in this graph. The most abundant sequence was used as consensus and serially laid out the next rate sequences. (.) Indicates the same amino acid residue.
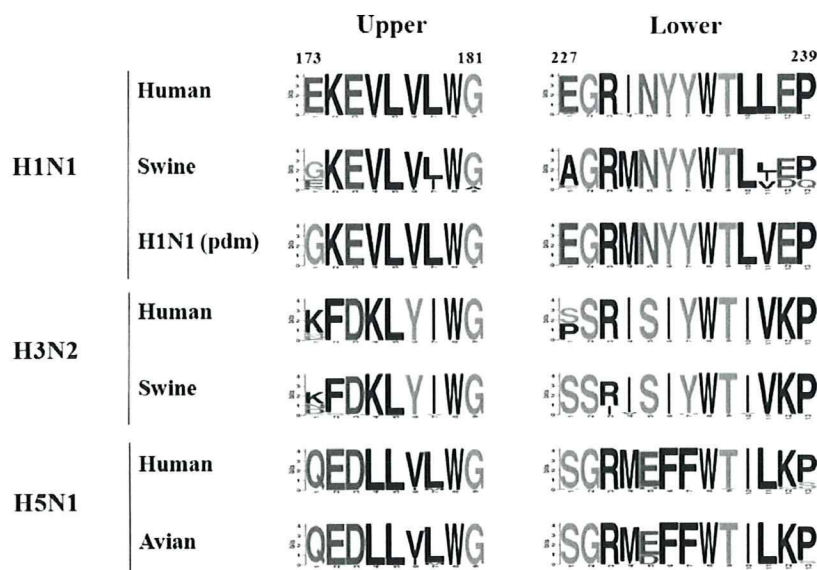
Fig. 3. Sequence logo of the upper and lower regions of various subtypes of influenza A virus. The upper and lower regions were extracted from the Influenza Virus Resource at the National Center for Biotechnology Information: human H1N1 ($n = 2722$), swine H1N1 ($n = 360$), and H1N1 (pdm) ($n = 1474$); human H3N2 ($n = 7955$) and swine H3N2 ($n = 235$); and human H5N1 ($n = 319$) and avian H5N1 ($n = 2349$). All of the variants in individual subtypes with different host origin were used for this sequence logo analysis. Amino acid residues by H3 numbering are shown at the top of the sequence.
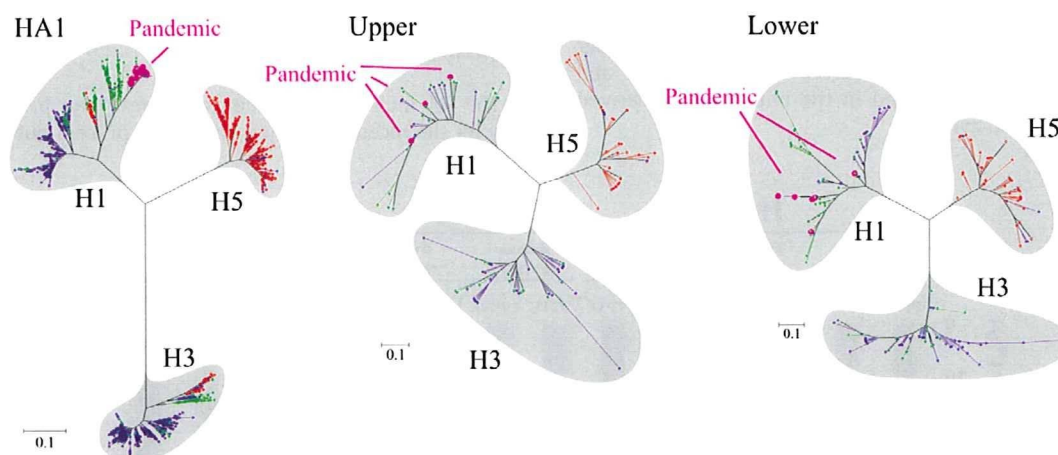

Fig. 4. Evolutionary relationships of the upper and lower epitope regions of various subtype of influenza A virus by ME method. The variable sequences in the upper and lower regions were used for phylogenetic tree construction. As a reference phylogenetic tree, we used the whole HA1 sequences in H1, H3, and H5 that are the original ones for the extraction of the upper and lower regions in this study. Blue means human origin; red means avian origin; green means swine origin; and purple means H1N1 (pdm) from human.

ding and particle formation. In addition, this molecule plays the major role for induction of neutralizing antibodies in the infected hosts. In our previous study, we isolated HuMAbs that can neutralize wide-range strains of human H3N2 subtype [9]. In this study, we identified the β-sheet region that could be the epitope recognized by the HuMAbs and further the corresponding regions from individual subtypes of influenza A virus, such as H1N1 including swine-origin pandemic H1N1 as well as highly pathogenic avian influenza virus H5N1. Sequences in these regions could be the major human epitopes to induce antibodies to neutralize wide-range of strains within individual subtypes.

The sequences of the upper and lower epitope regions in human H3N2 were significantly less conserved than the sequences of the corresponding regions in human H1N1, probably because of their different histories in the human population. In contrast, there was no apparent difference in the conservation of both regions in

H5N1 of human- and avian-origins, indicating the transmission of several variants produced in avian to human (Figs. 3 and 4). On the other hand, although these regions in swine H1N1 and H3N2 were very poorly conserved compared to those of human origin, the 1474 sequences of the upper and lower regions in H1N1 (pdm) were virtually composed of single sequences. This result is reasonable if only a single strain of swine-origin influenza A H1N1 has acquired the ability to transmit to human.

The first X-ray crystallographic structure of the HA was derived from H3N2 [21]. Now, the structures of numerous HAs have been resolved, including that of the H1N1 subtype of the 1918 pandemic influenza virus [22,23] and H5N1 [24]. In this study, we focused on the upper and lower regions that form anti-parallel β-sheet structure. Interestingly, although the overall amino acid sequence is poorly conserved among subtypes, the structure and functions of

such HAs are highly conserved in individual subtypes. This indicates that although a case of evolution and sequence variation proceed to an extreme level, structure and functions have remained conserved [25]. Most of the major neutralization epitopes were shown to be related to the former. Although the anti-parallel β-sheet structure underneath the receptor-binding region is one of the human neutralizing epitope, the regions is also likely to fulfill the latter category, as we found for the regions to be relatively highly conserved, suggesting that this structure could play a role for viral infection.

The level of serum antibody to HA and NA was shown to be correlated with resistance to illness and with restriction of the influenza virus replication in infected individuals [26]. Especially, the HA antibodies can prevent infection by neutralizing the infectivity of the virus [27]. Interestingly, it is known that after infection with a new type of influenza virus, antibodies that react with only a limited number of antigenic sites on HA are generated, whereas after several infections, antibodies show a broad range of specificities [28]. In agreement, we obtained two anti-HA neutralizing HuMAbs from the peripheral blood mononuclear cells from two influenza-vaccinated adult volunteers few weeks after vaccination and both react with the same antigenic sites on the HA [9]. Generally, humans have often been exposed during their lives to influenza virus field strains and also vaccine strains with various mutations. Such infections as well as vaccinations possibly induce selective proliferation of a certain population of memory cells that recognize the common epitope(s) among the influenza virus HAs in strains belonging to the same subtype, resulting in the increase of the levels of antibodies that can react with wide-range of strains belonging to each subtype.

In this study, computer-based characterization of the H3N2 viruses revealed relatively high conservation of two epitope regions. Further, we identified corresponding sequences in influenza viruses of other subtypes, such as H1N1 and H5N1 including those from different hosts. Some of them showed much higher conservation in their sequences than those in human H3N2. These data strongly suggest the possibility for the development of alternative vaccine against influenza A virus that could neutralize a wide-range of virus strains.

Acknowledgment

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.bbrc.2010.02.031.

References

[1] N.J. Carter, G.L. Plosker, Prepandemic influenza vaccine H5N1 (split virion, inactivated, adjuvanted) [Prepandrix]: a review of its use as an active immunization against influenza A subtype H5N1 virus, BioDrugs 22 (2008) 279–292.

[2] M. Pappaioanou, Highly pathogenic H5N1 avian influenza virus: cause of the next pandemic? Comp. Immunol. Microbiol. Infect. Dis. 32 (2009) 287–300.

[3] Novel Swine-Origin Influenza A (H1N1) Virus Investigation Team, Emergence of a novel swine-origin influenza A (H1N1) virus in humans, N. Engl. J. Med. 360 (2009) 2605–2615.

[4] V. Shinde, C.B. Bridges, T.M. Uyeki, B. Shu, A. Balish, X. Xu, S. Lindstrom, L.V. Gubareva, V. Deyde, R.J. Garten, M. Harris, S. Gerber, S. Vagoski, F. Smith, N. Pascoe, K. Martin, D. Dufficy, K. Ritger, C. Conover, P. Quinlisk, A. Klimov, J.S. Bresee, L. Finelli, Triple-reassortant swine influenza A (H1) in humans in the United States, 2005–2009, N. Engl. J. Med. 360 (2009) 2616–2625.

[5] R.A. Fouchier, V. Munster, A. Wallenstein, T.M. Bestebroer, S. Herfst, D. Smith, G.F. Rimmelzwaan, B. Olsen, A.D. Osterhaus, Characterization of a novel influenza A virus hemagglutinin subtype (H16) obtained from black-headed gulls, J. Virol. 79 (2005) 2814–2822.

[6] R.G. Webster, W.J. Bean, O.T. Gorman, T.M. Chambers, Y. Kawaoka, Evolution and ecology of influenza A viruses, Microbiol. Rev. 56 (1992) 152–179.

[7] R.J. Webby, R.G. Webster, Are we ready for pandemic influenza? Science 302 (2003) 1519–1522.

[8] Y. Takahashi, H. Hasegawa, Y. Hara, M. Ato, A. Nonomiya, H. Takagi, T. Odagiri, T. Sata, M. Tashiro, K. Kobayashi, Protective immunity afforded by inactivated H5N1 (NIBRG-14) vaccine requires antibodies against both hemagglutinin and neuraminidase in mice, J. Infect. Dis. 199 (2009) 1629–1637.

[9] R. Kubota-Koketsu, H. Mizuta, M. Ohshita, S. Ideno, M. Yunoki, M. Kuhara, N. Yamamoto, Y. Okuno, K. Ikuta, Broad neutralizing human monoclonal antibodies against influenza virus from vaccinated healthy donors, Biochem. Biophys. Res. Commun. 387 (2009) 180–185.

[10] Y. Bao, P. Bolotov, D. Dernovoy, B. Kiryutin, L. Zaslavsky, T. Tatusova, J. Ostell, D. Lipman, The Influenza Virus Resource at the National Center for Biotechnology Information, J. Virol. 82 (2008) 596–601.

[11] K. Katoh, H. Toh, Recent developments in the MAFFT multiple sequence alignment program, Brief. Bioinform. 9 (2008) 286–298.

[12] J.J. Skehel, M.D. Waterfield, Studies on the primary structure of the influenza virus hemagglutinin, Proc. Natl. Acad. Sci. USA 72 (1975) 93–97.

[13] R.G. Webster, W.G. Laver, Determination of the number of nonoverlapping antigenic areas on Hong Kong (H3N2) influenza virus hemagglutinin with monoclonal antibodies and the selection of variants with potential epidemiological significance, Virology 104 (1980) 139–148.

[14] D.C. Wiley, I.A. Wilson, J.J. Skehel, Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation, Nature 289 (1981) 373–378.

[15] G.E. Crooks, G. Hon, J.M. Chandonia, S.E. Brenner, WebLogo: a sequence logo generator, Genome Res. 14 (2004) 1188–1190.

[16] A. Rzhetsky, M. Nei, A simple method for estimating and testing minimum evolution trees, Mol. Biol. Evol. 9 (1992) 945–967.

[17] E. Zuckerkandl, L. Pauling, Evolutionary divergence and convergence in proteins, in: V. Bryson, H.J. Vogel (Eds.), Evolving Genes and Proteins, Academic Press, New York, 1965, pp. 97–166.

[18] M. Nei, S. Kumar, Molecular Evolution and Phylogenetics, Oxford University Press, New York, 2000.

[19] N. Saitou, M. Nei, The neighbor-joining method: a new method for reconstructing phylogenetic trees, Mol. Biol. Evol. 4 (1987) 406–425.

[20] K. Tamura, J. Dudley, M. Nei, S. Kumar, MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0, Mol. Biol. Evol. 24 (2007) 1596–1599.

[21] I.A. Wilson, J.J. Skehel, D.C. Wiley, Structure of the haemagglutinin membrane glycoprotein of influenza virus at 3 Å resolution, Nature 289 (1981) 366–373.

[22] S.J. Gamblin, L.F. Haire, R.J. Russell, D.J. Stevens, B. Xiao, N. Vasisht, D.A. Steinhauer, R.S. Daniels, A. Elliot, D.C. Wiley, J.J. Skehel, The structure and receptor binding properties of the 1918 influenza hemagglutinin, Science 303 (2004) 1838–1842.

[23] J. Stevens, A.L. Corper, C.F. Basler, J.K. Taubenberger, P. Palese, I.A. Wilson, Structure of the uncleaved human H1 hemagglutinin from the extinct 1918 influenza virus, Science 303 (2004) 1866–1870.

[24] J. Stevens, O. Blixt, T.M. Tumpey, J.K. Taubenberger, J.C. Paulson, I.A. Wilson, Structure and receptor specificity of the hemagglutinin from an H5N1 influenza virus, Science 312 (2006) 404–410.

[25] P. Palese, M.L. Shaw, Orthomyxoviridae: the viruses and their replication, in: D.M. Knipe, P.M. Howley, D.E. Griffin, R.A. Lamb, M.A. Martin, B. Roizman, S.E. Straus (Eds.), Fields Virology, fifth ed., Lippincott Williams & Wilkins, 2007, pp. 1647–1689.

[26] M.I. Clements, R.F. Betts, E.L. Tierney, B.R. Murphy, Serum and nasal wash antibodies associated with resistance to experimental challenge with influenza A wild-type virus, J. Clin. Microbiol. 24 (1986) 157–160.

[27] J.L. Virelizier, Host defenses against influenza virus: the role of anti-hemagglutinin antibody, J. Immunol. 115 (1975) 434–439.

[28] M.L. Wang, J.J. Skehel, D.C. Wiley, Comparative analyses of the specificities of anti-influenza hemagglutinin antibodies in human sera, J. Virol. 57 (1986) 124–128.

Ecotoxicity Prediction Using 3D Descriptors

Shinnosuke Hidaka[a], Hiroaki Shiraishi[b], Yoshihiro Ohmayu[a], Hiroyuki Yamasaki[a], Kousuke Okamoto[a], Norihito Kawashita[a,c], Teruo Yasunaga[c] and Tatsuya Takagi[a, c, d] *

[a] Graduate School of Pharmaceutical Sciences, Osaka University, 1-6 Yamadaoka, Suita, Osaka 565-0871, JAPAN
[b] National Institute for Environmental Studies, 16-2 Onogawa, Tsukuba, Ibaragi 305-8506, JAPAN
[c] Research Institute for Microbial Diseases, Osaka University, 3-1 Yamadaoka, Suita, Osaka 565-0871 JAPAN
[d] Research Collaboration Center on Emerging and Re-emerging Infections, 3-1 Yamadaoka, Suita, Osaka 565-0871 JAPAN

A vast quantity of chemicals are present in our environment and are considered indispensable to our high technological society. However, there are some chemicals that are hazardous and that can extensively impact both human health and the global environment. In Japan, ecotoxicity tests of chemical substances have been conducted with the goal of contributing to the Organization for Economic Cooperation and Development (OECD) evaluation program for harmful high-production volume (HPV) chemicals since 1995. To date, only about 500 compounds have been tested. There is a possibility that quantitative structure-activity relationships (QSARs) may enable us to predict environmental toxicities and fates as well as the physical and chemical properties of compounds; therefore, the toxicity prediction by QSARs is a possible alternative to the measurements of their ecotoxicities. In this study, we generated QSAR models from toxicity tests of Daphnids using only 3D descriptors to examine the availability of particular 3D descriptors for predicting of the ecotoxicity of compounds with various structures. Predicton accuracy of the model generated in this study was adequate and improved compared to that of the model using only the n-octanol/water partition coefficient, logP(o/w).

* ttakagi@phs.osaka-u.ac.jp

Introduction

Reportedly, as of 2008, there were tens of millions of chemicals that have been synthesized or isolated, and most of these are considered indispensable to our high technological society. However, some of these chemicals are hazardous and can extensively impact both human health and the global environment. In Japan, "in order to prevent environmental pollution caused by chemical substances that are persistent and pose a risk of impairing human health or interfering with the inhabitation and/or growth of flora and fauna" the "Act on the Evaluation of Chemical Substances and Regulation of Their Manufacture, etc.", was established in October 1973, around the time of the environmental pollution with polychlorinated biphenyls [1]. However, the influence on the environment and the entire ecosystem, especially the producer and the primary and secondary consumers in the food chain, had not been thoroughly investigated. In the face of increasing environmental problems, restrictions on the production and use of chemical products were implemented during the 1960s in Europe, and today chemical manufacturers in the European Union and the four other countries among the Organization for Economic Co-operation and Development (OECD) signatories are obligated to perform three types of ecotoxicological tests.

Since 1995, the Ministry of the Environment in Japan has been collecting information regarding the ecological effects of chemical products to assess the risk these products pose to the ecosystem with the goal of contributing to the OECD evaluation program for harmful high-production volume (HPV) chemicals [2]. However, as only about 500 compounds have been tested to date, it is realistically impossible to examine all the possible compounds. Accordingly, an alternative to the ecotoxicity tests of these chemicals is desired. There is a possibility that quantitative structure-activity relationships (QSARs) may enable us to predict environmental toxicities and fates as well as the physical and chemical properties of compounds. From this perspective, toxicity prediction by QSARs is one possible alternative [3-6]. The predictive capabilities os QSAR models are necessary to decreasing the number of ecotoxicity tests. It would reduce both the time and expense associated with these tests. Moreover, simplifying the examination model will reduce the number of animals required.

Some applications for predicting ecotoxicity are already being used, including the Ecological Structure Activity Relationship (ECOSAR) program and the TIssue MEtabolism Simulator (TIMES) algorithm [7, 8]. In the application of ECOSAR, firstly compounds in the dataset are classified based on their partial structures and their ecotoxicities are subsequently predicted by a single regression model based on the n-octanol/water partition coefficient, logP(o/w), of their classes. In TIMES, a classification that considers the same mechanism of action and a single or multiple regression models based on one descriptor or two or more descriptors are used.

Recently, research on three-dimensional (3D) structures has flourished due to the increase in computational power associated with the advancement of computer technology. The highest occupied molecular orbital (HOMO) energy and lowest unoccupied molecular orbital (LUMO) energy are 3D descriptors that are used in QSAR models for the prediction of ecotoxicity [6, 9]. However, some 3D descriptors have been used in QSAR analysis, and other descriptors have yet to be considered. It is necessary to research whether other 3D descriptors are applicable to build a QSAR model for the prediction of ecotoxicity.

An evaluation of the ecotoxicity of compounds will be a requirement for both drugs as well as HPV compounds in the near future [10]. In the development of new drugs, classification of drugs based on their structures or mechanisms of action may be ineffective because they have more various structures or mechanisms than pesticides or HPV compounds. A QSAR model that enables the ecotoxicities of various compounds to be predicted simultaneously is essential, in addition to QSAR models for classification of compounds.

The ecotoxicity of a compound is evaluated based on results of toxicity tests using three types of living things. The accuracy of toxicity prediction worsens in order of Fish, Daphnids, and Algae, and it is poor for Daphnids and Algae [5, 11]. In the present study, we expected that the poor predictions depend on the involvement of several toxicity mechanisms and 3D descriptors can be available to predict such toxicity prediction. Therefore, we generated QSAR models from toxicity tests of Daphnids using 3D descriptors to examine the availability of particular 3D descriptors for predicting of ecotoxicities. We used multiple regression analysis (MLR) and partial least squares regression (PLS) analyses to generate QSAR models without the classification based on a structure or toxicity mechanism.

Materials and Methods

Dataset

The results from the 387 compounds used in the *Daphnia* sp. Acute Immobilization Test, one of the ecological effect examinations that the Ministry of the Environment conducted, were used as a dataset [2].

Firstly, compounds were excluded from the dataset if they exhibited low or no toxicity ($EC_{50} > 1000$ mg/L), or if their toxicity values were qualitatively evaluated. Next, compounds that formed salts and whose structures were impossible to specify based on the CAS number were excluded from the dataset. The resultant dataset consisted of 265 compounds. Since the various structures of the compounds were merged in the dataset, two dummy variables that represent whether or not a compound is an aromatic amine and whether or not it is an aliphatic amine were added to the dataset. The number of amines in the resultant dataset are shown in Table 1. The units of acute toxicity were converted from mg/L to mol/L and transformed to log EC_{50}.

Table 1. Number of amines in dataset used in the present study

	Number of compounds
Aromatic amine	35
Aliphatic amine	20
Others	210

Molecular modeling

Molecular modeling and descriptors were calculated by the Molecular Operating Environment (MOE) software (version 2006.08) by Chemical Computing Group Inc. (Quebec, Canada). MOL Files downloaded from the Japan Chemical Substance Dictionary Web service, Nikkaji Web, were used to obtain structural formula information for compounds with a CAS number. If a compound was not registered or the structural formula of a compound could not be identified, the conformational information of the compound was generated using MOE. Structural optimization of compounds and calculation of the descriptors were conducted according to the following procedure. 1) "Wash Molecules"; hydrogenation and desalination are preprocessed. 2) "Partial Charges"; a partial charge is given to each atom. 3) "Energy Minimize"; energy minimization is calculated. PM3 was used for optimization in the present study. 4) "QSAR descriptors"; descriptors are calculated. Typical descriptors are shown in Table 2.

Analysis

Since the dataset consists of 45 explanatory variables [42 3D descriptors, 2 dummy variables and logP(o/w)], variable selection and data compression procedures are important to generate a favorable model for the prediction. In the present study, MLR and PLS analyses were applied to the dataset. We used R (version 2.5.1)

Table 2. Descriptors calculated by MOE

descriptor	
dipole	Dipole moment
E	The total energy [kcal/mol]
Eele	The electronic energy [kcal/mol]
HF	The heat of formation [kcal/mol]
HOMO	The energy of the HOMO [eV]
LUMO	The energy of the LUMO [eV]
E	Value of potential energy
E_ang	Angle bend potential energy
E_ele	Electrostatic component
E_nb	Potential energy with all bonded terms disabled
E_stb	Bond stretch-bend cross-term potential energy
E_str	Bond stretch potential energy
E_strain	Local strain energy
E_tor	Torsion potential energy
E_vdw	Van der Waals component of the potential energy
ASA	Water accessible surface area
ASA[+/-]	ASA of all atoms with [positive/negative] partial charge
ASA[H/P]	ASA of all [hydrophobic/polar] atoms
CASA[+/-]	Charge weighted surface area
DASA	= |(ASA+) - (ASA-)|
DCASA	= |(CASA+) - (CASA-)|
FASA[+/-, H/P]	= ASA[+/-, H/P] / ASA
FCASA[+/-]	= CASA[+/-] / ASA
VSA	van der Waals surface area
pmi	Principal moment of inertia
dens	Mass density: [Weight / volume]
glob	Globularity: 1
std_dim[1/2/3]	Standard dimension
vol	van der Waals volume
logP(o/w)*	Log of octanol/water partition coefficient
aromatic**	Aromatic amines [0 or 1]
aliphatic **	Aliphatic amines [0 or 1]

*2D descriptor
**dummy variable

for the analysis, and package 'pls' (version 2.1-0) was used for PLS [12,13].

Multiple Linear Regression; MLR

The variable selection procedures for the MLR analysis were conducted using two steps in the present study.

Firstly, the model with the lowest Akaike Information Criterion (AIC) was generated by the backward selection procedure using all of the 45 descriptors [14]. Subsequently, variables were selected referring to p values of statistical tests for the coefficients of the explanatory variables in the first model. Significance levels of the tests were set to 0.1, 0.05 and 0.01. Additionally, simple linear regression analysis of logP(o/w) was calculated. Leave-One-Out Cross-Validation (LOO-CV) was used for the evaluations of the predictive performance of these models.