Nishino, T., Takahashi, S., & Nakayama, T.（2007）*J. Biol. Chem*., **282**, 23581–23590.
11）Kaivosaari, S., Toivonen, P., Hesse, L.H., Koskinen, M., Court, M.H., & Finel, M.（2007）*Mol. Pharmacol*., **72**, 761–768.
12）He, X.-Z., Wang, X., & Dixon, R.A.（2006）*J. Biol. Chem*., **281**, 34441–34447.
13）Cartwright, A.M., Lim, E.-K., Kleanthous, C., & Bowles, D.（2008）*J. Biol. Chem*., **283**, 15724–15731.
14）Kubo, A., Arai, Y., Nagashima, S., & Yoshikawa, T.（2004）*Arch. Biochem. Biophys*., **429**, 198–203.
15）Masada, S., Kawase, Y., Nagatoshi, M., Oguchi, Y., Terasaka, K., & Mizukami, H.（2007）*FEBS Lett*., **581**, 2562–2566.

國兼 聡，中山 亨
（東北大学大学院工学研究科バイオ工学専攻）

Recent advances in plant secondary product glycosyltransferase research
Satoshi Kunikane and Toru Nakayama（Graduate School of Engineering, Tohoku University, Aoba, Aramaki, Aoba-ku, Sendai, Miyagi 980–8579, Japan）

糖鎖インフォマティクスの概要

1. はじめに

糖鎖はDNAとアミノ酸配列に加え，情報を担う第三の分子と考えられている．情報として糖鎖を扱う研究分野は「グライコミクス」と「グライコームインフォマティクス」に分けられる．「グライコミクス」では糖タンパク質における糖鎖付加部位と糖鎖構造を網羅的に決定することにより，糖鎖の機能解析を行う．一方，「グライコインフォマティクス」（糖鎖インフォマティクス）では，グライコミクスから得られた構造情報とゲノム，プロテオームなどの他のオーム情報を組み合わせて，生物学的に有用な情報を抽出することにより，糖鎖の機能解明を行う．糖鎖インフォマティクスの方法論として，この数年間，新しいアルゴリズムやモデルが次々に開発されてきた．この促進には，以下の三箇所の大規模糖鎖データベースプロジェクトが大きな役割を果たした．

- ドイツがん研究センターのGLYCOSCIENCES. deデータベース[1)]
- 米国Consortium for Functional Glycomics（CFG）の糖鎖データベース[2)]
- 京都大学化学研究所のKEGG GLYCANデータベース[3)]

これらの糖鎖データベースの基となるデータは1990年代に開発された，米国ジョージア大学のCarbBankデータベースに由来する．CarbBankプロジェクトの終了後，新しいデータベースが各々構築され，個別に糖鎖構造情報記述の形式を決め，データ収集を行っていった（**表1**）．このため，現在，これらのデータベース間ではデータ交換が困難である．データ交換を容易にするために，GLYDE-IIと呼ぶ糖鎖構造情報のためのXML（eXtensible Markup Language）標準が提案された[4)]．今後，ユーザーは，GLYDE-IIによって仮想的に統合された糖鎖情報を容易に入手できるようになる．

糖鎖インフォマティクスの研究は，主に次のテーマに関して行われている．

- 糖鎖バイオマーカーの予測
- 糖鎖構造解析
- 糖鎖構造マイニング
- 糖鎖構造予測

各々を以下に簡単に紹介する．

表1 主な糖鎖構造データベースの一覧

データベース名	内容	URL	形式
GLYCOSCIENCES. de	CarbBank及びPDBより糖鎖構造を抽出した．糖鎖構造と質量分析情報が含まれている．	http://www.glycosciences.de	LINUCS
KEGG GLYCAN	KEGGデータベースの一部であり，糖鎖構造がKEGG GENESやPATHWAYの情報にリンクされている．また，糖転移酵素や糖結合タンパク質情報はKEGG BRITEに分類されている．	http://www.genome.jp/kegg/glycan/	KEGG Chemical Function（KCF）
CFG	CarbBankの*N*型と*O*型糖鎖の情報に加えて，GlycoMinds社のシードデータベースが含まれている．また，CFGの組織や細胞情報，糖鎖アレイ情報とCFG独自で合成した糖鎖の情報も蓄積されている．	http://www.functionalglycomics.org/	IUPAC

2.　糖鎖バイオマーカーの予測

カーネルと呼ぶ学習モデルは大量のデータを分類するために使用される．**図1**に糖鎖カーネルの概念を表す．データのなかから特徴になる情報（図1の糖鎖の場合は結合している「二糖」とその結合様式）をn次元のベクトルとして扱い，カーネル計算よりベクトルのなかで最も特徴を反映する変数を抽出する．この変数は，特徴となる情報（フィーチャ）を示しており，バイオマーカーの候補になる．最初の糖鎖カーネルモデルはLayered Trimer Kernelであり[5]，このモデルを拡張して，我々はq-gram分布カーネルを開発した[6]．一方，Multiple Kernelと呼ぶ別のモデルも開発された[7]．今後，構造のみならず，パスウェイや発現情報をも含めたカーネルモデルを構築することにより，さらに生化学的に妥当な分類ができ，新しいバイオマーカーの発見も期待される．

3.　糖鎖構造の解析

現在まで，計算機科学の分野では二次元情報を扱うアルゴリズムが数多く開発されてきた．しかし，バイオイン

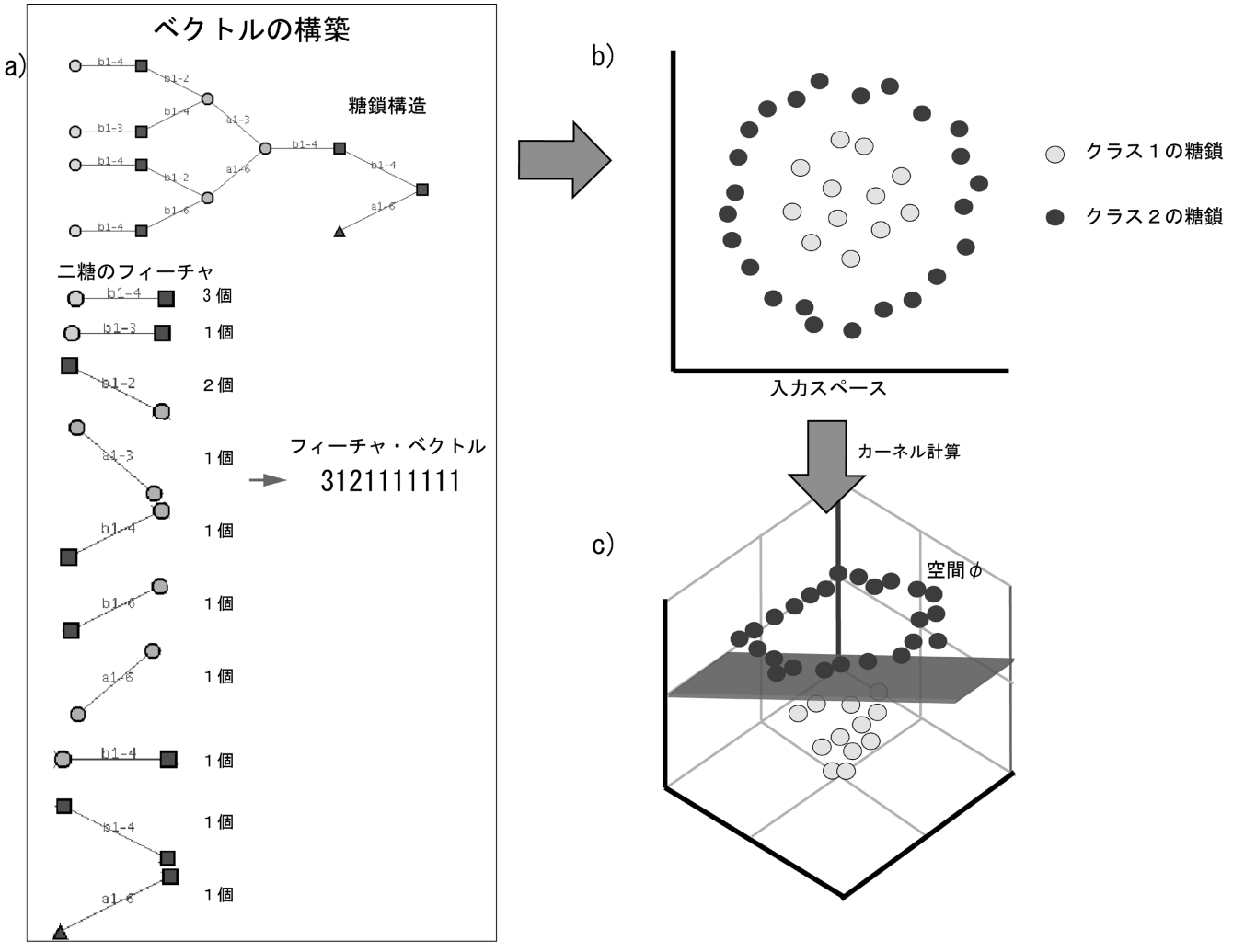

図1　カーネルの計算

a）ベクトルの構築．入力された各糖鎖構造を「二糖とその結合様式」に分解する．この二糖を「フィーチャ」と呼ぶ．それぞれのフィーチャの数を数え，例えばn個のフィーチャが存在する場合，n次元のベクトルとして扱う．これで各糖鎖構造に対して特異的なフィーチャ・ベクトルができあがる．b）二つ以上のクラスに属する糖鎖構造をn次元の空間に置く．c）カーネル計算（内積を使った計算）でクラスを分類しやすい空間に写像し，分類する面に直交するフィーチャを抽出する．このフィーチャがクラス1と2を最も特徴的に分類するので，バイオマーカーの候補となる．

フォマティクス分野におけるその応用はRNAの三次元構造予測や系統解析に限られていた．糖鎖構造が二次元の情報を含むことは，バイオインフォマティクス研究者にとってはとても興味深いものであった．我々は糖鎖構造を表すために，木のアルゴリズムを応用した．始めに，KCaM（KEGG Carbohydrate Matcher）と呼ぶ糖鎖構造アラインメント[8]と糖結合のスコア行列[9]を開発・解析した．

まず，データベースから効率的に正確な構造を検索するために，糖鎖構造のアラインメントアルゴリズムを開発した．さらに，糖結合のスコア行列（置換行列）を作成し，糖鎖構造の特徴の抽出に応用した．アミノ酸のスコア行列はアミノ酸置換の許容範囲を示すものであり，その単位はアミノ酸そのものである．一方，糖鎖のスコア行列の単位は結合している二糖とその結合様式である．我々は，スコア行列でその二糖の結合の関連性を表すことができ，糖転移酵素やレクチンなどの認識部位の類似性の比較に利用できると考えた．既に，*N* 結合型糖鎖，*O* 結合型糖鎖などの各々のクラスに対応したスコア行列を作成し，糖鎖構造アラインメントの向上に成功している．今後，タンパク質，病原体，細胞の糖鎖アレイを用いた結合親和性データに応用し，結合に関わる重要な糖鎖構造を予測することにより糖鎖機能解析への足がかりにしたいと考えている．

4. 糖鎖構造マイニング

レクチンの糖鎖認識機構は予想外に複雑な場合がある．例えば，Siglecは，末端の単糖だけではなく，さらに内部の構造も認識すると考えられている．このような複雑な糖鎖認識は，実際に結合している糖鎖構造のみならず，空間的に近くにある糖鎖構造をも認識している可能性があるため，probabilistic sibling-dependent tree Markov model（PSTMM）が開発された．このモデルは親子関係（結合している単糖．例えば図2の場合，GlcNAcβ1-4GlcNAcやManβ1-4GlcNAc）のみならず，兄弟関係（近傍にある単糖．例えば図2の場合，Manα1-3とManα1-6）も考慮するモデルである[10]．このモデルを学習させるために，効率的な学習アルゴリズムを開発した．さらに，このモデルを改良し，ordered tree Markov model（OTMM）も開発した[11]．OTMMは，PSTMMと同じ予測制度を保ちながら計算量を減少させるために単純化したモデルである．

これらのモデルから認識パターンを読み出すことができなかったため，認識プロファイルの抽出を目的としてOTMMを拡張し，ProfilePSTMMモデルを開発した．この

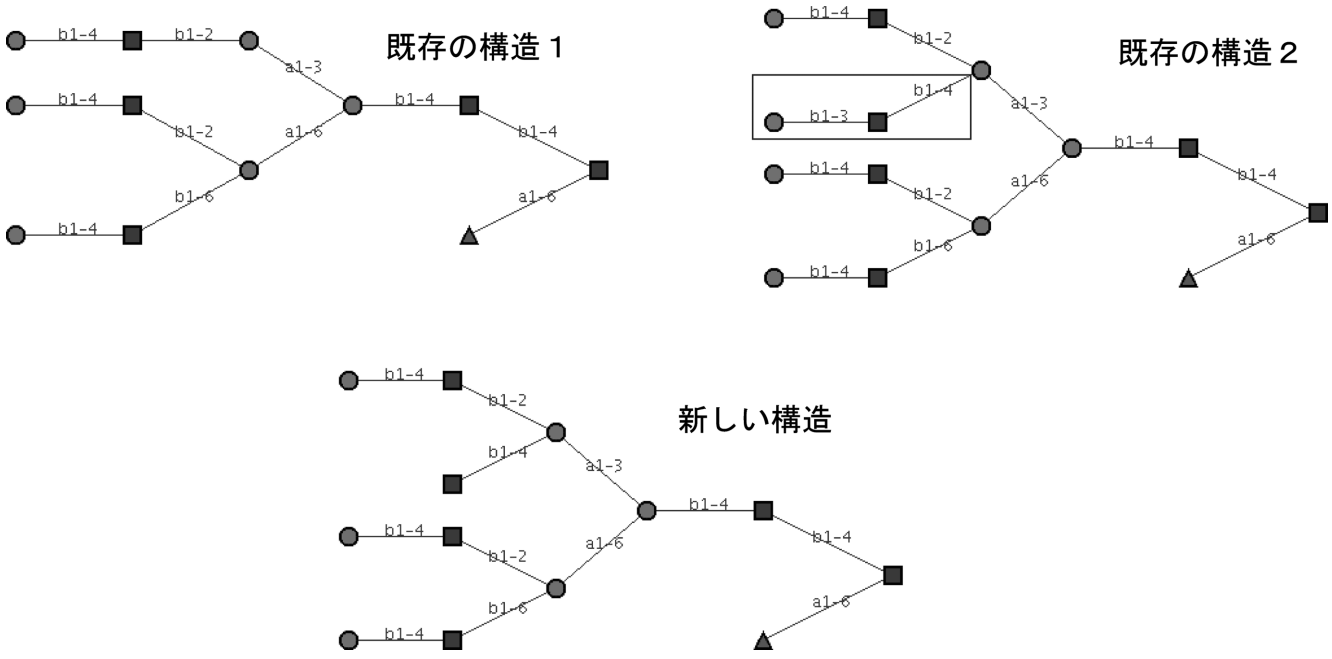

図2 糖鎖構造データベースに存在すべき構造の決定
KCaMを用いて全ての糖鎖構造を比較し，合成過程で認められるはずの二つの単糖のみが変化した全ての合成中間体に対応する糖鎖構造を抽出した．「既存の構造1・2」の間に二つの単糖の違いが認められるため，「新しい構造」も「既存の構造1・2」の間の中間体として存在するはずであり，これをデータベースに追加する．

モデルを用いて，公開されているレクチンと糖鎖の結合親和性情報からパターン抽出を試みたところ，既に推論されていた構造を抽出することができ，本ProfilePSTMMモデルが実用に耐えることが明らかとなった[12]．

5. 糖鎖構造予測

糖転移酵素の発現情報から糖鎖構造を予測する方法が開発されている[13]．この方法では，糖転移酵素が合成する結合およびアクセプターの単糖を一つの単位とし，糖鎖構造データベースに含まれている全ての糖鎖構造をまずこの単位に分解する．次に，各糖鎖構造に対して，二つの単位（リアクションペア，RPと呼ぶ）が同時に出現する頻度を計算し（共起性を計り），共起性行列にまとめる．この行列は糖転移酵素の基質特異性を反映していると考えられる．糖転移酵素の発現情報が与えられた場合，糖鎖構造データベース中の各糖鎖構造に対して，糖鎖構造情報と共起性行列を用いて，スコアを計算することができる．全ての糖鎖構造の中で最も高いスコアを得た構造が発現している糖鎖構造の候補となる．

この方法を拡張し，さらによい結果を得た．改良に当たっては，**図2**で示したような糖鎖構造データベースに存在すべき構造を以下の方法で追加した．まず，「糖鎖構造の解析」で紹介したKCaMを用いて全ての糖鎖構造を比較し，合成過程で認められるはずの二つの単糖のみが変化した全ての合成中間体に対応する糖鎖構造を抽出した．さらに，予測スコアの計算に糖転移酵素の発現の定量値を含めた[14]．実際の定量値を利用して，発現の高い酵素が合成する糖鎖構造に重みをつける方法である．Consortium for Functional Glycomics（CFG）のデータベースに含まれているヒトのがん細胞における糖転移酵素の発現情報に適用したところ，高いスコアを得た糖鎖構造の多くがsialyl-Lewis Xやsialyl-Lewis Aを含む糖鎖構造であった．両糖鎖構造はがんにおける発現が既に報告されており，この方法の有効性が示された．

6. おわりに

糖鎖インフォマティクスが誕生してから5年が経過した．種々のアルゴリズムが開発され，その有効性が示されたが，生化学的検証が今後の課題となっている．また，カーネル法に細胞内局在やパスウェイ情報を含めることにより，糖鎖機能を反映したバイオマーカーの獲得が期待できる．CFGデータベースにおける親和性情報の増加により，確率モデルをレクチンの糖鎖認識部位予測に用いることも可能となる．糖鎖の複雑な構造や機能を効率よく解析するためには，上述のような高度な学習モデルが必須であり，糖鎖インフォマティクスのますますの発展が期待される．

1）Lütteke, T., Bohne-Lang, A., Loss, A., Goetz, T., Frank, M., & von der Lieth, C.-W.（2006）*Glycobiology*, **16**, 71R–81R.
2）Raman, R., Venkataraman, M., Ramakrishnan, S., Lang, W., Raguram, S., & Sasisekharan, R.（2006）*Glycobiology*, **16**, 82R–90R.
3）Hashimoto, K., Goto, S., Kawano, S., Aoki-Kinoshita, K.F., Ueda, N., Hamajima, M., Kawasaki, T., & Kanehisa, M.（2006）*Glycobiology*, **16**, 63R–70R.
4）Packer, N.H., von der Lieth, C.-W., Aoki-Kinoshita, K.F., Lebrilla, C.B., Paulson, J.C., Raman, R., Rudd, P., Sasisekharan, R., Taniguchi, N., & York, W.S.（2008）*Proteomics*, **8**, 8–20.
5）Hizukuri, Y., Yamanishi, Y., Nakamura, O., Yagi, F., Goto, S., & Kanehisa, M.（2005）*Carbohydr. Res*., **340**, 2270–2278.
6）Kuboyama, T., Hirata, K., Aoki-Kinoshita, K.F., Kashima, H., & Yasuda, H.（2006）*Genome Inform*., **17**, 25–34.
7）Yamanishi, Y., Bach, F., & Vert, J.-P.（2007）*Bioinformatics*, **23**, 1211–1216.
8）Aoki, K.F., Yamaguchi, A., Ueda, N., Akutsu, T., Mamitsuka, H., Goto, S., & Kanehisa, M.（2004）*Nucleic Acids Res*., **32**, W267–W272.
9）Aoki, K.F., Mamitsuka, H., Akutsu, T., & Kanehisa, M.（2005）*Bioinformatics*, **21**, 1457–1463.
10）Aoki, K.F., Ueda, N., Yamaguchi, A., Kanehisa, M., Akutsu, T., & Mamitsuka, H.（2004）*Bioinformatics*, **20**, i6–i14.
11）Hashimoto, K., Aoki-Kinoshita, K.F., Ueda, N., Kanehisa, M., & Mamitsuka, H.（2008）*ACM Trans. on Knowledge Discovery from Data*（*TKDD*）, **2**（1）, Article No. 6.
12）Aoki-Kinoshita, K.F., Ueda, N., Mamitsuka, H., & Kanehisa, M.（2006）*Bioinformatics*, **22**, e25–e34.
13）Kawano, S., Hashimoto, K., Miyama, T., Goto, S., & Kanehisa, M.（2005）*Bioinformatics*, **21**, 3976–3982.
14）Suga, A., Yamanishi, Y., Hashimoto, K., Goto, S., & Kanehisa, M.（2007）*Genome Inform*., **18**, 237–246.

木下 聖子
（創価大学工学部生命情報工学科）

An introduction to glycome informatics

Kiyoko F. Aoki-Kinoshita（Department of Bioinformatics, Faculty of Engineering, Soka University, 1–236 Tangi-cho, Hachioji, Tokyo 192–8577, Japan）