

10-1 NMR 研究会

主題 = 多変量解析と NMR 一埋もれた情報を発掘する

＜趣旨＞ NMR 法は現在、化学、生物学の分野において、構造解析や分子運動解析に欠くことのできない分光法として存在しています。しかしながら、複雑な系になると観測されるピークの帰属に多大なエネルギーが割かれることも事実で、いかに効率的に帰属するかという方法論もここ 10 年の間に大変な進歩を遂げています。一方、数年前から複雑なピークの完全帰属は行わず、一連のスペクトル変化（化学シフト変化や強度変化）を統計的に処理し、問題解決につながる情報を取り出そうという解析手法（多変量解析）が、NMR 法の分野にも浸透してきています。この多変量解析は、統計学の分野では古くから知られている手法であり、早くから近赤外分光法などには応用されていました。ここ 2～3 年で、多変量解析が NMR 法の分野においても、その活躍の場を飛躍的に広げる可能性を秘めていることが明らかにされつつあり、今最もホットな研究分野の一つとなっています。今回、多変量解析の基礎、NMR 法へ応用するときの注意点、代謝系多変量解析の実際、高分子系への応用の注意点や現状等について討論する場を設けました。今回の研究会が、皆様の研究の発展につながる討論と交流の場として大いに活用されることを期待いたします。また、あわせてポスター発表の機会を設けました。様々な関連分野からのポスター発表を期待します。今回から新たに「若手ポスター賞」を選考・授与しますので、奮ってご参加ください。

主 催 高分子学会 NMR 研究会
協 賛 (予定) 日本化学会 日本核磁気共鳴学会 日本分析化学会
日 時 平成 22 年 5 月 21 日 (金) 10:00～17:00
会 場 独立行政法人 産業技術総合研究所 臨海副都心センター 別館 11 階 第一会議室
(東京都江東区青海二丁目 41 番地 6)
交 通 東京臨海新交通ゆりかもめ テレコムセンター駅下車 徒歩約 4 分
(<http://unit.aist.go.jp/waterfront/jp/index.html>)

プログラム

＜10:00～10:45＞

1. 可視・近赤外スペクトルを用いた土壌成分値予測モデルの構築

(東京大学大学院工学系研究科化学システム工学専攻) 船津 公人..... 1

＜10:45～11:30＞

2. 多変量解析の前処理－より正確な定量性を求めて

(旧日本女子大学理学部物質生物科学科) 高橋 征三..... 5

＜12:30～14:30＞ ポスター発表

ポスタープログラムは次ページをご覧ください。

＜14:30～15:15＞

3. 代謝物からバイオマスへの物質創製・情報発現相関解析の試み

(理化学研究所／名古屋大学大学院／横浜市立大学大学院) 菊地 淳..... 7

＜15:15～16:00＞

4. 合成高分子の NMR スペクトルへの多変量解析の応用

(東レリサーチセンター鎌倉 生物科学研究部) 木村 一雄..... 11

＜16:00～16:45＞

5. ^{13}C NMR スペクトルの多変量解析によるアクリル系共重合体のキャラクタリゼーション

(三菱レイヨン中央技術研究所／徳島大学大学院ソシオテクノサイエンス研究部) 百瀬 陽..... 13

＜16:45～17:00＞ 若手ポスター賞表彰式

・ポスタープログラム

- P1.新規の含フッ素ポリマー／多糖ホストゲスト包接化合物の調製と固体 NMR 法による構造解析
(東工大院理工)○小糸 祐介・山田 和彦・安藤 慎治
- P2.¹³C CP/MAS 法によるポリアルキルチオフェンの動的構造の解析
(東工大生命理工)○矢澤 宏次・井上 義夫
(物材機構)清水 禎・丹所 正孝
(群馬大工)浅川 直紀
- P3.固体 NMR 法によるポリアニリン/ β -シクロデキストリン包接複合体の動的構造解析
(東工大生命理工)○長谷川 裕一・井上 義夫
(物材機構)清水 禎・丹所 正孝
(東工大生命理工)矢澤 宏次
- P4.固体 ¹⁹F/¹³C NMR 法を用いた PVDF 系ファイバーの結晶構造解析
(東工大工)○大橋 龍二
(東工大院理工)滝上 義康・山田 和彦・安藤 慎治
- P5. 無機酸化物粉体と高分子の相互作用に関する研究
(群馬大院工) 浜本彰子・上原宏樹・山延 健
(太陽誘電) 都竹浩一郎
- P6.アルキル側鎖を有する全芳香族液晶性ポリエステルが形成する層状構造の磁場配向特性と NMR による構造解析
(名工大院工)○傘 俊人・辻 梨沙・吉水 広明
- P7.温度可変高速 MAS 中の温度分布と ¹H 緩和挙動
(防衛大応化)○北村 成史・浅野 敦志・田中 千香子・黒津 卓三
- P8.系内に存在する気体のスピン - スピン緩和測定を用いたガラス状高分子の高次構造評価
(名工大院工)○岡澤 誠裕・深澤 有美子・吉水 広明
- P9.半結晶性高分子における NMR 法による気体輸送特性の評価
(名工大院)○奥村 祐生・鶴永 実希・吉水 広明
- P10.Gelation mechanism and structure in polysaccharide gels studied by gradient NMR
(Tokyo Univ. of Marine Sci. & Tech.)○Dai Bona・Matsukawa Shingo
- P11.アルカリ金属型イオン液体のダイナミクスと物性評価
(日産アーク)○上口 憲陽・高橋 洋平・若松 寛子・佐藤 秀紀
- P12.固体 NMR と統計力学計算による家蚕絹フィブロイン結晶部モデル化合物のラメラ構造に関する研究
(農工大院工)○鈴木 悠・青木 昭宏・朝倉 哲郎
- P13.延伸に伴うクモ牽引糸の構造変化に関する研究
(農工大院工)○佐藤 佑哉・池永 万希子・亀谷 俊輔・朝倉 哲郎
- P14.アラニンオリゴマー及び天然絹繊維の固体 NMR 構造解析
(農工大院工)○亀谷 俊輔・山崎 静夫・朝倉 哲郎
- P15.細胞接着性配列を導入した絹様タンパク質モデルペプチドの ¹³C CP/MAS NMR 構造解析
(防衛大応化)○田中 千香子・浅野 敦志・黒津 卓三
(農工大工)朝倉 哲郎
- P16.¹³C 標識セルロースを用いたバイオマス分解因子のデータマイニング
(理研)○伊達 康博・飯倉 智弘・菊地 淳
- P17.主成分分析を用いた NMR メタボロミクスの臨床応用ー慢性腎臓病の病態解析
(東北大薬)○藤原 正子・今井 潤
(医療法人・宏人会中央クリニック)関野 宏
(東北大薬/ 医療法人・宏人会中央クリニック)竹内 和久
(産総研) 根本直・安藤一郎

P18. ^{13}C NMR スペクトルの多変量解析による高分岐共重合体のキャラクタリゼーション

(徳島大院 STS)○平野 朋広・上池 亮太

(三菱レイヨン・徳島大院 STS)百瀬 陽

(徳島大院 STS)右手 浩一

P19.DOSY によるアクリル系共重合体のキャラクタリゼーションー組成の分子量依存性

(徳島大院 STS)前田 智也・服部 康佑

(三菱レイヨン・徳島大院 STS)百瀬 陽

(徳島大院 STS)平野 朋広・○右手 浩一

P20.固体 ^{13}C CPMAS NMR 擬似的クロマトグラフ法

(日本電子)○西山 裕介

(JEOL USA Inc.)Frey Michael H.・Mukasa Sseziwa

(日本電子)内海 博明

P21.多核固体 NMR 法を用いたナノファイバーの構造解析

(東工大院理工)○山田 和彦・安藤 慎治

P22.ゴム中に溶解した水素の NMR による解析

(産総研 HYDROGENIUS)○藤原 広匡

(九大院工・産総研 HYDROGENIUS)山辺 純一郎・西村 伸

P23.酸素還元活性を有する熱分解ポリピロールの固体 NMR 解析

(東工大院理工)○黒木 重樹

P24.微小磁性体の磁場勾配を用いた MRI 分光装置の開発と高分子への応用

(群馬大院工)川内 由貴・○浅川 直紀

可視・近赤外スペクトルを用いた土壌成分値予測モデルの構築

(東京大学大学院工学系研究科化学システム工学専攻)、船津 公人

E-mail funatsu@chemsys.t.u-tokyo.ac.jp

電話 03-5841-7751 FAX 03-5841-7771

1. はじめに

近年、近赤外分光法を用いた研究が多く行われている。近赤外領域は 700～2500 nm 付近の波長領域を指す。近赤外領域における吸収は赤外領域と比較し微弱であり、結合音や倍音によるバンドが複雑に重なり合っていることや、バンドのシフトが赤外領域よりも大きいことなどが原因で、その取り扱い是非常に困難であった。そのため、分子の構造推定などに用いられてきた赤外領域と比較し研究が遅れていた。しかし、データを統計的に扱うケモメトリックスの発展により、近赤外領域におけるスペクトル解析が可能となってきた。近赤外光は、迅速な測定や非破壊性などが注目され、医薬品や化学物質、食料品など様々な物質を対象とした幅広い分野の研究に応用されている。

近赤外分光法は精密農業における土壌成分値予測においても有用である。精密農業では、圃場を効率的に管理するため土壌の情報をリアルタイムで入手することが望まれており、可視・近赤外スペクトルを用いてこれが実現可能であると期待されている。圃場内において測定されたスペクトルと、その地点の土壌成分値との間で回帰モデルを構築することで、スペクトルデータからの土壌成分値予測が可能となる。このような背景から、土壌のスペクトルデータを用いた解析について、用いるスペクトルの波長や対象とする土壌の物性値、モデリング手法などに関する検討が多く行われている。

スペクトルデータを用いた回帰分析を行う際には、共線性が大きな問題となる。隣接する波長領域におけるスペクトルの挙動は似ていることが多く、そのため説明変数間の相関は非常に高い。このような相関の高い変数を用いた回帰分析においては、回帰モデルが不安定化することが知られており、モデルの予測精度が大幅に低下してしまう。この問題を回避するための方法として有力と考えられているのが変数選択である。すべての波長領域におけるスペクトルを用いるのではなく、適切な方法によって目的変数の変動を説明するために重要であると考えられる変数を選択することで、安定で予測的な回帰モデルの構築が期待できる。そのため、スペクトルデータの解析を対象とした変数選択手法がいくつか提案されている。

Vasques らは、可視・近赤外スペクトルを用いた土壌中の炭素量予測モデルの構築において、Stepwise 法を用いた変数選択を行った結果を報告している。Stepwise 法は、F 値を評価基準とし、説明変数をモデルへと追加、あるいは削除することで変数の選択を行う手法である。アルゴリズムが簡便であり、直観的な変数選択が可能ではあるが、各変数が独立して選択されるため最適変数セットを求めることは困難であり、多くの場合局所解へ収束してしまうと考えられる。また、変数単位での選択であるため、それらの変数に対して化学的な帰属を行うことが難しいという問題がある。

Bras らは近赤外・中赤外スペクトルの多変量解析においてモデリングに重要なスペクトルの波長領域を選択する iPLS-Bootstrap 法の提案を行っている。iPLS-Bootstrap 法は、

Lazraq らが提案した変数選択手法である PLS-Bootstrap 法を、連続した領域単位で選択を行うよう発展させた手法である。iPLS-Bootstrap 法では、bootstrap の各サンプリングにおいて PLS 法による回帰モデルを構築し、回帰係数の信頼区間に 0 が含まれない変数を有意であると考え、そして、ある一定の波長領域区間において連続的にそのような変数が存在する場合、その領域の変数を選択する。この手法においては、選択された領域の変数が他の領域の変数と組み合わせられたときの影響が考慮されていないという問題が存在する。また、有意な説明変数の数が少なく、それらの間隔が広い場合には、波長領域が 1 箇所も選択されないことが考えられる。

また、Jiang らは近赤外・中赤外スペクトルデータにおいて波長領域を選択する MWPLS (moving window partial least squares) 法を提案している。MWPLS 法は、ある大きさのウィンドウを波長領域で移動し、ウィンドウ内に含まれる変数のみを用いて作成した PLS モデルを評価することで波長領域を選択する手法である。この選択を繰り返すことによって、複数の波長領域に対応する変数セットの選択が可能である。逐次的な選択であるため、高速に変数選択を行うことが可能であるという利点がある一方、局所解に陥る可能性が高いと考えられる。

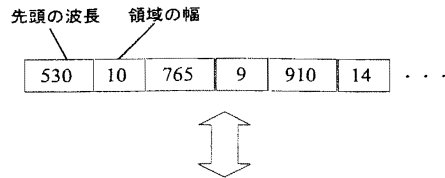
以上のように、スペクトルデータを用いた回帰分析のための変数選択手法が複数提案されているが、決定的な手法はまだ開発されていないのが現状である。そこで我々は、遺伝的アルゴリズム (genetic algorithm : GA) を用いて波長領域を選択するアルゴリズムである GAWLS (genetic algorithm-based wavelength selection) 法を提案した。土壌水分量を目的変数として GAWLS 法による変数選択を行い、従来手法と比較し良好な結果を得ている。

本研究の目的は、土壌水分量に加え、炭素量、窒素量、電気伝導率、pH を目的変数とした解析を行うことで、GAWLS 法の有用性、汎用性を示すことである。1,000 点のトレーニングデータを用いて回帰モデルの構築を行い、1,372 点のデータに対して土壌成分値の予測を行うことで、GAWLS 法が従来手法と比較して優れていることを示した。

2. GAWLS 法

GAWLS 法は GAPLS 法と同様に、GA を用いてモデルの Q^2 を大きくする説明変数の組を選び出す変数選択手法である。Figure 1 に GAWLS 法におけるコーディング方法を示す。GAPLS 法においては染色体の一つのビットが一つの説明変数に対応し、0 と 1 で表現されていたのに対し、GAWLS 法では二つの値を用いて一つの波長領域を表現する。左側の数字は波長領域の先頭の波長を表し、右側の数字はその領域の幅を表している。二つの数字の組を複数用意することにより任意の数の領域を表現することができる。例えば 5 nm 間隔で吸光度を測定したスペクトルデータにおいて、Figure 1 に示された染色体は 530~580 nm、765~810 nm、910~980 nm という三つの波長領域を表している。GAWLS 法における染色体の評価法は、GA 法による染色体で表現された説明変数を用いた PLS によって得られる Q^2 値である。

GAWLS 法は、説明変数をそれぞれ独立に選択する GAPLS 法と異なり、幅を持った領域単位で変数の選択を行う。可視・近赤外スペクトルにおいては、構造の特徴がバンドとして現れるため、変数を領域の形で選択することが適切であると考えられる。



530~580 nm 765~810 nm 910~980 nm . . .

Figure 1 Coding rule of GAWLS.

3. 解析データ

GAWLS 法による波長領域選択の有用性を確認するため、精密農業における土壌の可視・近赤外スペクトルデータへの適用を行った。国内の 13 地域 115 圃場で測定された土壌のスペクトルデータと、土壌水分量(MC) [%]、土壌炭素量(TC) [%]、土壌窒素量(TN) [%]、電気伝導率(EC) [mS/cm]、4 つの土壌成分値との中の回帰モデルを構築した。測定を行った地域は全国に分散しており、また測定時の天候や土壌の性質も様々である。このような多様なデータを用いたモデリングを行うことで、適用範囲の広い汎用的なモデルの構築を目指した。

スペクトルデータは、トラクター後部に取り付けられたリアルタイム土壌センサーによってリアルタイムに測定された。約 300~1700 nm の可視・近赤外領域において測定されているが、測定精度の低いと考えられる部分を取り除いた 600~1550 nm を解析に用いた。また、測定波長間隔が 5 nm になるように補間処理を行っており、変数の数は 190 となっている。測定データ点数は 2,372 点であり、これらをランダムに 1,000 個のトレーニングデータと 1372 個のテストデータに分割し、トレーニングセットで構築したモデルを用いてテストセットに含まれるデータ点の土壌成分値の予測を行った。なお、明らかに外れ値と考えられるデータ点が 6 点存在したため、これらはあらかじめ除外した。

一般に、スペクトルデータには測定機器の特性や環境的要因によるノイズが含まれているため、適切な前処理を行い、ノイズの影響を軽減することが必要である。本論文では前処理として Savitzky-Golay 法によるスムージングと、二次微分、センタリングを行った。Savitzky-Golay 法は前後数点の値を用いて多項式近似を行うスムージング処理の手法である。Savitzky-Golay 法を用いると、一般的なスムージング手法である移動平均法を用いた場合よりもスムージングの効果は小さくなるが、可視・近赤外領域に多く存在する微小なピークがスムージングの際に失われてしまう危険性は低くなる。今回は近似に用いる点を 7 点、近似の次数を 2 とした。

微分処理は一定間隔離れた 2 点の値の差分をとる手法である。微分処理を行うことにより、重なり合った複数のピークの分離やベースライン変動の除去・軽減が可能である。今回は差分をとる 2 点間の間隔を 20 nm とし二次微分処理を行った。

4. GAWLS 解析

GAWLS 法を用いた波長選択により、各土壌成分値の予測モデルの構築を行った。GA のパラメータを世代数 1,000、染色体数 100 として 10 回の計算を行った結果を Table 1 に示す。いずれの目的変数に対しても、すべての変数を用いた PLS モデルと同程度の精度・予測性を示すモデルが得られている。

Table 1 Results of GAWLS

	R^2	Q^2	最適成分数	RMSE	R_{pred}^2
MC	0.855	0.841	10	7.226	0.804
TC	0.849	0.835	10	0.707	0.777
TN	0.821	0.796	10	0.059	0.776
EC	0.555	0.522	10	0.027	0.552

MCについて、GAWLS法によって選択された波長領域をFigure 2に示す。GAPLS法を用いた場合と異なり、10回の試行において選択される波長領域には共通した傾向が見られた。繰り返し選択されている波長領域から土壌成分値と関係性の高い重要な波長を見つけ出すことが可能であると考えられる。

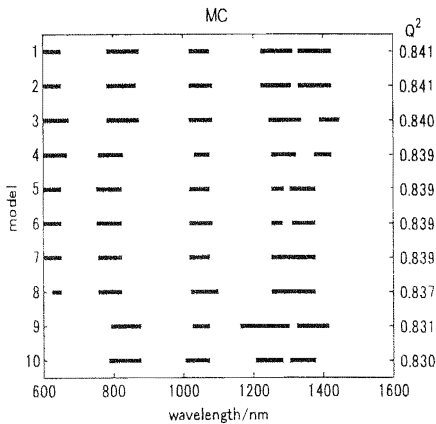


Figure 2 Wavelengths selected by GAWLS.

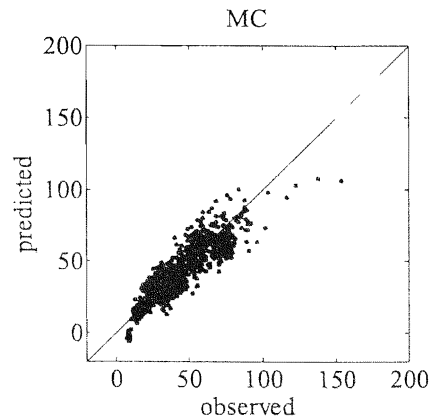


Figure 3 Plot of observed and predicted values.

Figure 3にMCを目的変数とした場合の実測値とGAWLS法による予測値との散布図を示す。ほぼ全ての点が対角線付近に位置しており、特異な外れ値はほとんど見られなかった。

5. まとめ

精密農業におけるリアルタイムな圃場情報の獲得を目的とし、土壌成分値を予測する回帰モデルの構築を行った。全国の広い範囲で測定された2,000点を超える土壌データにおいて、土壌の可視・近赤外スペクトルを説明変数とし、MC、TC、TN、ECの4つの目的変数に対する予測モデルを構築した。その結果、GAWLS法を用いて波長領域選択を適切に行うことで、MC、TC、TNに関しては実用的な回帰モデルが構築可能であることが示された。また、ECについても十分なモデルが構築された。この結果から、多様な土壌をまとめた汎用的なモデルで実用的な予測性を得ることができたと言える。

GAWLS法は、可視・近赤外スペクトルを用いた土壌成分値の予測のみでなく、その他の解析においても有用であると考えられ、今後様々な対象への適用が期待される。

謝辞: スペクトルデータおよび土壌データを提供していただいたエスアイ精工株式会社様に感謝いたします。

多変量解析の前処理 より正確な定量性を求めて

(旧 日本女子大学理学部) 高橋 征三

((株) 実行データサイエンス) 山口 行治

E-mail rs_takahasi@s7.dion.ne.jp

電話 03-3642-5020

はじめに

今日ではプログラムさえあれば高度な統計処理が誰でもいとも簡単にできるが、結果が、正しい答か否かは別問題である。NMR の世界に限っていえば、実感として、従来のスペクトル解析法だけでほとんど不便を感じない。そのためか多変量解析の導入は遅々として進んでいないという印象を受ける。私も、できれば避けて通りたいが、必要に迫られてしぶしぶ勉強しはじめた段階である。ここでは統計解析にまったく無知な初心者が、未知の世界である、多変量解析を使うにはどうしたらよいか分らず、膨大なデータを抱えて悪戦苦闘している状況をお伝えしたい。

発端は、共同研究者とつぎの議論をしたことにはじまる。新薬開発の最大のハードルは、ヒトを対象とする臨床試験の段階になると、薬物応答に個体差があるため、有効性を証明できなくてストップしてしまうケースが大部分である。この現状を打破する方法として、NMR を使って、多数の指標を同時に使えば、多次元空間のなかで個体差を表現できないだろうか。うまくすれば代謝物の変動に有意差が見えるかもしれない。ここで Metabonomics という分野の研究がすでに存在することを初めて知り、勉強をはじめた。研究報告を比較すると、しばしばデータ数に依存して有意差を示す領域も変動する例が多い。どうやら本質は、有意差を出すことではなく、データ数で変らない安定な解を得ることにあるらしい。

実験対象には若年女性の唾液を選んだ。唾液は容易に採取できる分泌液で、その 90% 以上が体内に再吸収される。これは血液と尿の中間的性質をなす。いっぽう唾液は交感神経および副交感神経の支配を受けるので、環境の変化でその成分は大きく変動する。これを個体差変動に見立てた。また若年女性の周期的なホルモン変動を薬物とみなせば、薬物応答を NMR でモニターすることは容易であろうと考えた。さらに唾液分泌のメカニズムは、薬物が血液から組織に輸送されるメカニズムに類似し、モデル系として有効であると考えた。ちなみにヒト組織での薬物動態については、モニターする方法がないために、組織中の濃度は血中濃度に比例するという仮定の下に行われるのが、そうならない例は多数報告されている。

実験

20歳以上の若年女性のボランティアを最終的に30名選び、起床直後の体温と唾液を約30日分採取した。さらに高温期と低温期の2回を1日に5回、唾液を採取直後に NMR を測定した。唾液は脱脂綿に吸収させたものを遠心して分離し、380uL を NMR チューブにとり重水 20uL と混ぜた。測定は Bruker AMX-400 を使い室温で行った。

結 果

最初に成功例を紹介する。水を照射後に測定するために、消え残りの水の信号のパターンは FID ごとに変動する。装置は古く劣悪な環境のもとに設置されているために、かなり不安定である。あるとき FID の積算結果があまりにひどいので、データを2次元に並べて主値解析を行った。その結果、主値の大きいほうから4項を使えば、磁場変動をほぼ完全に補正できた。

成功した理由は、磁場変動がスペクトル全体に一様に影響し、変動の原因が限られ、しかも残差が正規分布することはかなり確かだったからである。右図に示すように、PC 1 は単なる平均スペクトルを示し、PC 1 成分を差し引いた残りは平均値からのずれ、つまり磁場強度の変動を意味する。

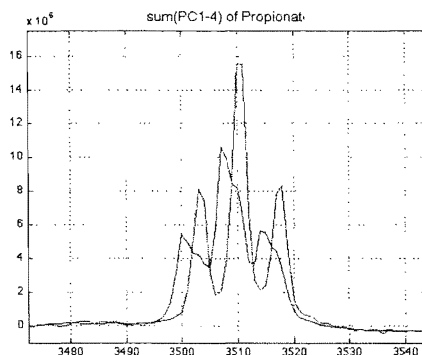


図1 平均スペクトルと FID の主値解析をもとに補正した FT 後のスペクトル。

2番目の例は、私に苦い教訓を与えた。HPLC と NMR の信号強度の間に相関が認められなかったので、

HPLC のデータを因子分析したところ、(1) 吸光度の観測値は吸収係数と成分濃度の積で表され、(2) 濃度はサンプルごとに変動するが、(3) HPLC からの溶出パターンは共通する、とする仮定のどれかに問題があることが分った。どう扱っていいか、いまだに悩んでいる。

唾液の NMR は奇妙な挙動を示し、酢酸の化学シフトは pH 依存性を示さず、強度のヒストグラムはどれも正規分布を示さない。対数正規分布に非常に近いが、対数正規分布から有意に外れるものが多数見つかった。これは今まで誰も予想しなかった事実である。

多変量解析が正規分布を前提としている以上、強度データの正規化が必須である。いろいろ試みた挙句、化学的意味づけは別として、強度データを 0.2 乗するという、古典的な Box-Cox 変換でしのいでいる。そして経日変化の多変量解析は、ランダム変動が系統変化より大きいために成功していない。今までに達成した確実な成果は、基礎体温や唾液成分は外気温度に左右されることである。経日変動が個体差変動の10分の1以下のとき、どうすればその変動を捕らえることができるか、皆様の知恵を拝借したい。

<はじめに> 高分子科学で扱う多くのポリマーは、原油から蒸留・精製されたモノマーを重合させたホモポリマー・コポリマーであろう。筆者も工業化学科時は石油化学漬だったが、後に蛋白質や多糖類といった複雑かつ選択的な機能を発現する生体高分子の世界を知り、この複雑性や生命の神秘に魅了されるようになった。折しも、その20年前時は石油代替への期待が高まった頃でもあり、バイオマスの生産と利活用に関する問題意識をもつことができた。

一方で世間の期待とは裏腹に、バイオマスは我々の生活にさほど浸透しなかった。こうしたトラウマがあったか、暫くは日本の重化学工業界はバイオ分野という反応が鈍かった。しかしここ1,2年は、こうした産業界との意見交換から、急激に前向きになっていると実感する[1]。20年前は、複雑系、多様性が前提の生命科学の世界にアプローチする際、表現型解析の手法として計測科学と情報科学の両方が未発達であった。特に、分子レベルのアプローチは対象が見えないために、

遺伝子をクローニングし、蛋白質や生理活性物質を単離精製して研究するという物質還元主義が横行していた(図1左)。このアプローチは科学としての厳密さとは裏腹に、バイオマス増大といった表現型へのリンクは乏しいものであった。しかし一方で、DNAシーケンサーや計算機の数・容量、さらにはインターネット情報網の急激な進展により、オミクス科学に代表されるDB蓄積・利用の情報科学に重きを置く新しいアプローチが開拓されている(図1右)。これまで、物質還元主義では2体的相関で表現していたグラフが、オミクス科学では多変量を扱うことが前提になっているため、ヒートマップ、ネットワーク図や多変量解析(MVA)で結果を示す(図1下)。ここでは、代謝表現型解析や、それをバイオマス生産・利活用研究に展開する我々のアプローチを紹介する。

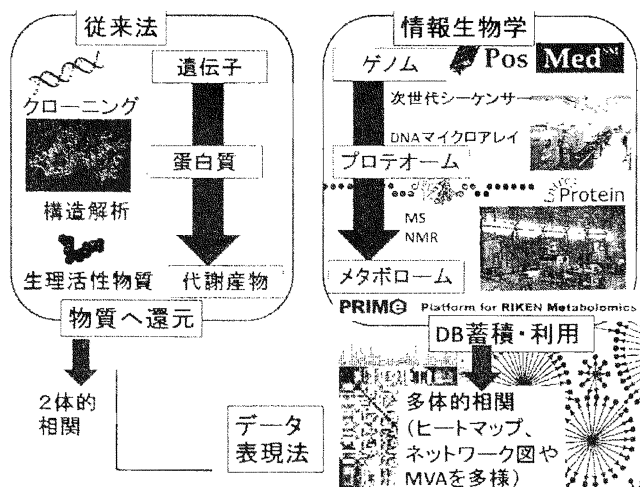


図1 DBの蓄積・利用に重心をおくオミクス科学的アプローチ(右)と従来の物質還元主義的アプローチ(左)の比較

<代謝表現型解析> 今一度前世紀からの生命科学の進展を見直してみると、見た目の表現型に基づき分類していた博物学的な世界から、細胞に共通な現象に沿って体系的に生命を捉えようとする分野が拓けた。分子生物学がそれであり、DNA から蛋白質に至るセントラルドグマがまさに、体系的理解を動機付けた共通原理と言えよう。実際に web 辞書でセントラルドグマを調べると、「生物の表現型を決める遺伝子の働き」と記載されており、あたかも全ての生物が利己的な DNA に情報支配されているかのように、DNA 情報依存での生命科学が推進されている。しかし、温度・乾燥等の物理的、栄養・飢餓等の化学的、さらには常在菌・寄生等の生物的環境要因を受けて、一生が左右されるからこそ、生命活動という表現型の解析に興味がつきないのであろう。これらの環境要因を受けて変化する表現型の解析手段がメタボノミクスである[2]。

メタボノミクスでは、データ処理に対する考え方が、通常の分析と大きく異なる。メタボノミクスでは分析ターゲットをあらかじめ特定せず、計測データ中に含まれるすべてのシグナルを解析対象として、データから得られる情報の最大化を目指す(図2上)。いったんデータ・マトリクスを得ることができれば、さまざまなデータマイニングが可能となる利点があり、生物学研究への幅広い応用が拓けてくる。例えば代謝物間の協同的な変動等[3]や遺伝子多型との相関[4]等、従来のターゲット分析では予期せぬ情報が得られる。

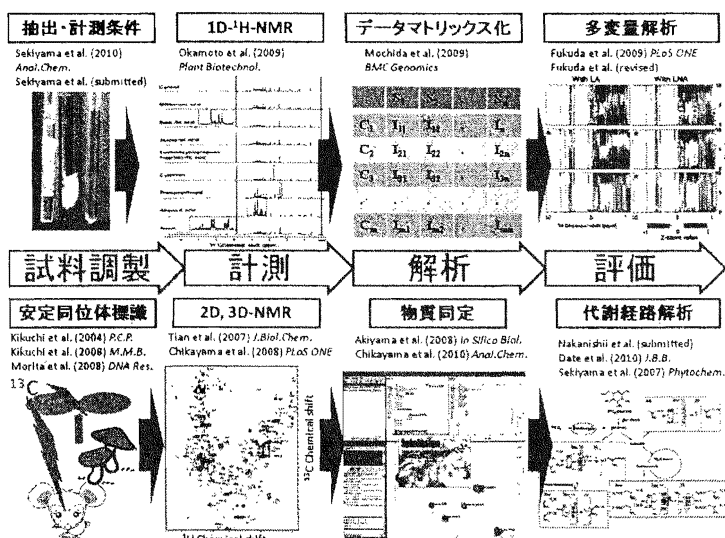


図2 筆者らの推進するメタボノミクス技術開発

これに加えて、筆者らは安定同位体標識を用いることで、従来のメタボノミクスとは異なる動的な解析技術構築を目指している[5]。例えば、 $^{13}\text{C}_6$ -グルコースや $^{13}\text{CO}_2$ など一次代謝の起点あるいはハブとなる化合物の標識体を生体に投与し、安定同位体を広範な代謝物に分布させ、各代謝物における標識の分布やバランスの変化を網羅的に観測する事である(図2下)。各代謝物について、同位体組成の異なる異性体をアイソトポマーと呼び、この中で特に ^{13}C - ^{13}C 結合の位置が異なる異性体をボンドマーと呼ぶ。NMR 法を用いる強みの一つは、分子内の各原子をそれぞれ区別して観測できる事、また、 ^{13}C - ^{13}C 結合定数が分子内の各

結合に特徴的な値を示す事であり、従ってアイソトポマーおよびボンドマーの種類を明確に区別する事が可能である[6]。

一方で NMR 法は、抽出された可溶物質だけでなく、残渣の難溶性物質や生体組織を無傷なまま計測できる[7]。そこで、筆者らは斬新な代謝物アノテーション法の解決策を考案し、植物資源の有効資源化を目指した(図 3 上)。植物の全成分を有効利用するためには、抽出する成分、抽出操作に伴う成分組成の変化、各抽出残渣に含まれる成分を、網羅的で体系的な評価方法が望まれる。筆者らは ^{13}C で均一安定同位体標識化した植物をメタボノミクス手法を用いて、抽出操作に伴う植物成分の組成変化と残渣成分の評価を行った。この過程で代謝物のアノテーションをつけ、抽出操作に伴うそれぞれの挙動について調べたところ、脂質は有機溶媒、糖は水、といった単純な溶解性で分布しないことが分かった。

さらに、我々はメタボノミクスの世界で物質資源のみならず、情報資源も捨てられていることに着目した。例えば、ゲノム解析で頻繁に用いられる BLAST の E-value のような統計数学的手法が、ゲノム科学発展に極めて重

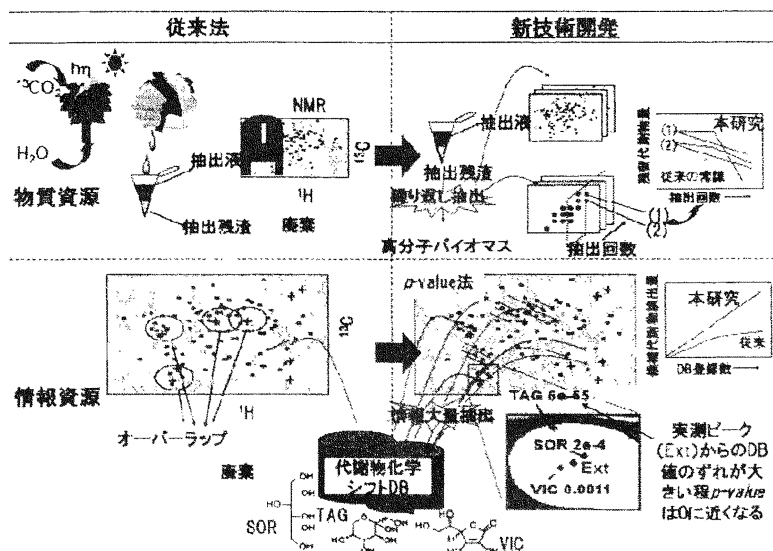


図 3 従来は廃棄されていた物質・情報資源を有効活用する試み
かかわらず、同じオミッ

クス解析分野で比較的歴史の浅いメタボノミクス分野では、まだそのような統計数学的手法が存在していない。そこで筆者らは、BLAST の E-value に相当する p -value という指標をメタボノミクス用に考案した(図 3 下)。この p -value 法を実装した SpinAssign プログラムを用いて ^{13}C 標識化植物細胞の網羅的代謝物アノテーションを遂行したところ、メタボノミクス分野では、世界新記録となる 211 候補代謝物の同時検出を記録した[8]。この手法は、参照代謝物化学シフトデータベースの登録数が増大するほど能力を発揮するため、筆者らはその情報蓄積も続けている。近い将来に登録数を 500 代謝物以上に増大させる予定で、定められた物理化学条件で計測した生物代謝物データベースとしては、世界最大規模となっており、筆者らが公開している PRIME 上の SpinAssign[8]で、

公共利用が可能である。

＜低分子代謝物と高分子バイオマス＞ 冒頭で述べたようにオミクス解析の急激な進展に伴い、バイオマスの生産・利活用に関して産業界からも熱い視点が注がれている。既に欧米の化学系メジャーはスイッチグラスや藻類といったバイオマス生産植物のゲノム育種、さらにはシロアリ腸内細菌叢や熱水鉱床のメタゲノム解析といった、バイオマス分解・代謝に関わる遺伝子・酵素資源の探索に巨額の研究資金を投入している。こうした潮流を筆者らはいち早く予見し、バイオマス分解系に関しては、複合生物系から分解微生物や酵素を探索する手法を開発してきた[9]。今後の課題は、リグノセルロースのような複雑系高分子を解析する高分子科学の手法開発と、如何にしてオミクス科学のような最新情報生物学とのリンクを密接にするかであろう。講演ではこのような将来展開について、生産的な議論ができればと願っている。

＜参考文献＞

- [1] 菊地淳 (2008) “統合バイオリファイナリへの NMR 技術開発 ―石油資源が遺伝資源に代替される時代を見据えて―” *植物の生長調節*, **43**, 144-145.
- [2] Fukuda et al. *Nature* (revised); 菊地淳, (2010) *遺伝子医学* **16**, 81-85; Okamoto et al. (2009) *Plant Biotechnol.* **26**, 551-560; 吉田欣史ら, (2009) *ぶんせき* **7**, 371-378; 菊地淳ら, (2007) *メタボロミクスの解析と応用技術*, CMC 出版, 86-99
- [3] Chikayama et al. (2008) *PLoS ONE* **3**, e3805; Fukuda et al. (2009) *PLoS ONE* **4**, e4893.
- [4] Mochida et al. *BMC Genomics*. **10**, e563.
- [5] Kikuchi et al. (2010) *Radioisotopes* (in press). Morita et al. (2008) *DNA Res.* **15**, 151-161; 菊地淳ら, (2008) *バイオサイエンスとインダストリー* **66**, 143-144; Kikuchi et al., (2007) *Method Mol. Biol.* **358**, 273-286; Tian et al., (2007) *J. Biol. Chem.* **282**, 18532-18541; Kikuchi et al., (2006) *Biotech. Agri. Forest.* **57**, 93-101; Kikuchi et al., (2004) *Plant Cell Physiol.*, **45**, 1099-1104.
- [6] Ohyama et al. (2009) *Proc. Natl. Acad. Sci. USA* **106**, 725-730; Sekiyama et al., (2007) *Phytochemistry* **68**, 2320-2329.
- [7] Sekiyama et al. (2010) *Anal. Chem.* **82**, 1643-1652; 菊地淳ら, (2007) *ブレインテクノニュース*, **124**, 16-21; 菊地淳, (2006) *分光研究*, **55**, 320-330.
- [8] Chikayama et al. (2010) *Anal. Chem.* **82**, 1653-1658; 近山英輔ら, (2009) *遺伝子医学* **16**, 126-130; Akiyama et al. (2008) *In Silico Biol.* **8**, e27.
- [9] 菊地淳ら, (2010) *難培養性微生物研究の最新技術 2*, CMC 出版, 147-155. Date et al. (2010) *J. Biosci. Bioeng* (in press); 守屋繁春ら, (2008) *原核生物、特に真正細菌 cDNA の簡易な調製法*, 特願 2008-52503; 菊地淳ら, (2007) *蛋白質の探索方法*, 特願 2007-232792.

合成高分子の NMR スペクトルへの多変量解析の応用

(東レリサーチセンター 生物科学研究部) 木村一雄

〔川口謙、高崎万里〕

E-mail Kazuo_Kimura@trc.toray.co.jp

電話 0467-32-9977 FAX 0467-32-0414

メタボノミクスの発展により、 ^1H NMR スペクトルが本格的に多変量解析の分野で応用され始めた。 ^1H NMR スペクトルによるメタボノミクスは、実際には ^1H NMR スペクトルを用いたケモメトリックス (Chemometrix) 分野への応用であり、メタボノミクス以外にも NMR を用いた分野での応用が考えられる。本発表では、まず ^1H NMR によるメタボノミクスがどのような目的で発展してきたかを簡単に説明し、次に ^1H NMR スペクトルによる多変量解析をナイロン 6 とナイロン 66 の混合物の定量分析に応用した例、及びポリマーの添加剤の定量分析に応用した例について説明する。

1. ^1H NMR によるメタボノミクス

メタボノミクスは代謝物 (主に尿) の全体像を捉え、代謝物プロファイルの統計解析から毒性予測や毒性バイオマーカーの同定などを目的としている。

メタボノミクスでは、まず薬物を投与し経時的に採取した尿 (または血漿などの末梢体液) の ^1H NMR を測定する。NMR を用いたメタボノミクスの手法が通常の NMR 分析と大きく異なる点は、NMR スペクトルのシグナル変化をパターン認識で解析することである。

多変量解析は、「シグナル変化をパターン認識して解析」する過程で、NMR スペクトルを比較する際に必要となる。メタボノミクスの目的は変化を捉えることである。このため 1 セットで 50 個ぐらい尿のスペクトルについて変化を評価することが必要になる。尿のような混合物の多検体の NMR スペクトルを、人が目で見比べて変化を定性的・定量的に評価することは困難である。そこで、NMR スペクトルをパターン化してコンピューターを用いて多変量統計解析し、スペクトルパターンの変化を捉えよう、というのが ^1H NMR メタボノミクスの手法の画期的な点である (と考えている)。図 1 に 600MHz で測定した尿の ^1H NMR スペクトルと、そのデータを縮約してパターン化したスペクトルを示した。NMR スペクトルをパターン化 (短冊のように積分して棒グラフ化 (データ縮約)) することによってデータ容量を約 1/100 に下げて、かつ pH 等による化学シフト値の変化の影響を軽減させて多変量統計計算を実行可能にしている。

パターン化したスペクトルの多変量解析の手法はいろいろ考えられるが、基本となるのは主成分分析 (Principal Component analysis) と呼ばれるものである。主成分分析は、多変量データを統合し、新たな総合指標を作り出すための手法といえる。新たな総合指標を用いて、データ構造 (サンプルデータの分布) を説明していくことになる。その際、データ構造を最もよく説明するものから順に PC1、PC2 と縮約された空間でデータ構造を解析していくことになる (主成分分析については、多くの教科書があるので詳細はそれらを参照したい)。

主成分分析は、ローディング (固有ベクトル) からスペクトルの情報、スコアから量的

な情報が得られる。これらを利用し、合成高分子を含む混合物の定量分析に応用した例を次に紹介する。

2. ナイロン 6 とナイロン 66 の混合物の定量分析

ナイロン 6 とナイロン 66 の ^1H NMR スペクトルを図 2 に示す。これらは非常に良く似たスペクトルパターンを示し、混合比を議論することは難しい。しかし、数種の混合比の異なるスペクトルを準備し主成分分析（回帰分析）を行うことで検量線を作成し、定量することが可能であることがわかった。

3. ポリマーの添加剤の定量分析

ポリスチレンに添加した添加剤の定量分析例を示す。本分析例では、添加剤の添加量がポリマーに対して少ないため、そのままでは添加剤の定量は難しい。しかし、主成分分析を行い、ローディングスペクトルから添加剤に相当する主成分軸を特定することで、添加剤の定量が可能であることがわかった（これについて発表で示す）。

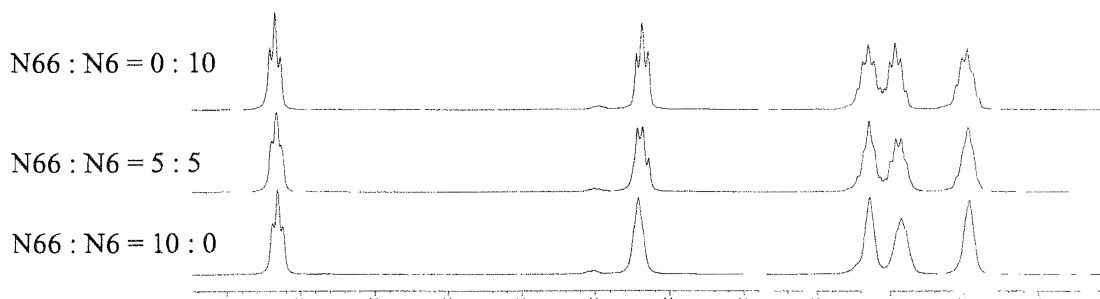
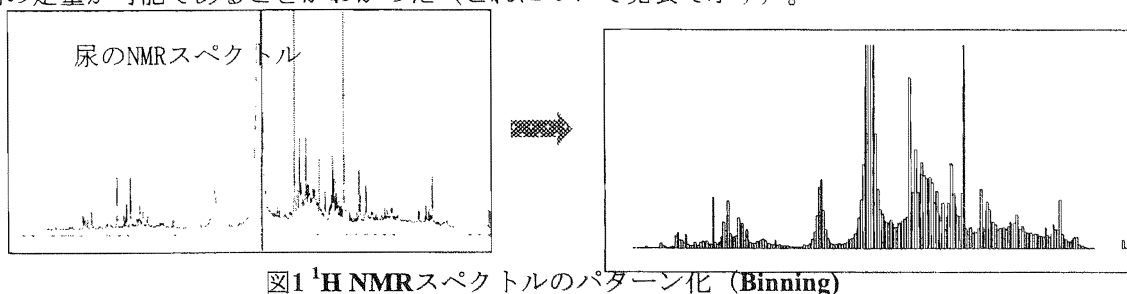


図 2 ナイロン 6、及びナイロン 66 の ^1H NMR スペクトル

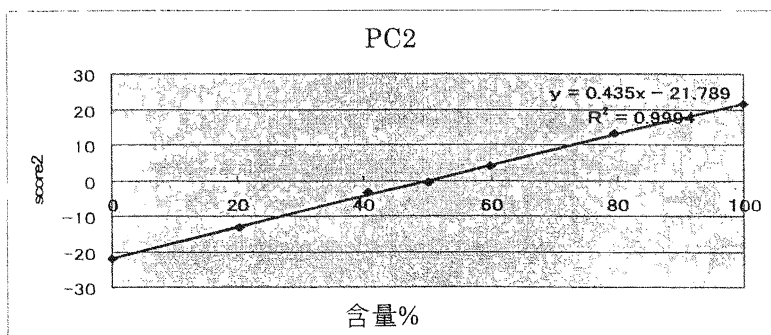


図 3 スコアから作成した検量線 (PC2 がナイロン 6 とナイロン 66 の差に相当する)

^{13}C NMR スペクトルの多変量解析による アクリル系共重合体のキャラクタリゼーション

(徳島大院 STS/三菱レイヨン) 百瀬 陽

E-mail momose@poly.chem.tokushima-u.ac.jp

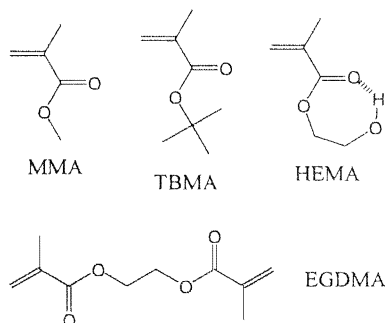
TEL/FAX 088-656-7404

はじめに

機能性を有する高分子材料の多くは多成分共重合体である。そのため、材料の性能向上には多成分共重合体の構造因子を詳細に解析することが不可欠となる。NMR は共重合体の 1 次構造解析に用いられる最も有効な手法の 1 つであるが、その NMR スペクトルは成分数の増加とともに、組成、モノマー連鎖および立体規則性などを反映してより複雑に分裂するため、多くの情報を簡便に得ることは難しくなる。

多変量解析は、複雑かつ大量の情報をいくつかの有用な情報へと変換する強力な手法である。合成高分子では、近赤外、赤外およびラマンスペクトルへの適用が数多く知られている¹。また、いわゆるメタボロミクスと呼ばれる生物代謝産物の網羅解析では、NMR スペクトルの多変量解析が役立っている²。

最近、著者らは、多くの分野で活用されている多変量解析を合成高分子の ^{13}C NMR スペクトルへ応用し、複雑なピークの帰属を行うことなく、構造因子に関する定量的な情報の取得に成功した。本発表では、メタクリル酸メチル(MMA)、メタクリル酸 *t*-ブチル(TBMA)、およびメタクリル酸 2-ヒドロキシエチル(HEMA)の二元・三元共重合体、ならびに 2 官能性のジメタクリル酸エチレングリコール(EGDMA)と TBMA からなる高分岐共重合体の ^{13}C NMR スペクトルへ適用した事例について紹介する。



Scheme 1. Chemical structures of monomers.

MMA-TBMA 二元共重合体^{3,4}

ラジカル重合で合成した PMMA および PTBMA, PMMA/PTBMA 混合物 9 種、組成や収率が異なる MMA-TBMA 共重合体 16 種の合計 27 種のサンプルを用意し、 ^{13}C NMR 測定を行った。得られたスペクトルを主成分分析(PCA)した結果を Figure 1 に示す。単独重合体とこれらの混合物の PC1 スコアは TBMA 組成の増加とともに単調に減少し、共重合体の PC1 スコアも同様の傾向を示した。PC1 スコアと TBMA 組成は相関係数 $R^2 = 0.998$ で直線関係となることから、PC1 スコアは組成を反映することがわかった。単独重合体 2 種とこれらの混合物 9 種の ^{13}C NMR スペクトルと組成を検量用データとした部分最小二乗回帰(PLSR)を行い、共重合体 16 種の TBMA 組成の決定を試みたところ、別途 ^1H NMR 測定から求め

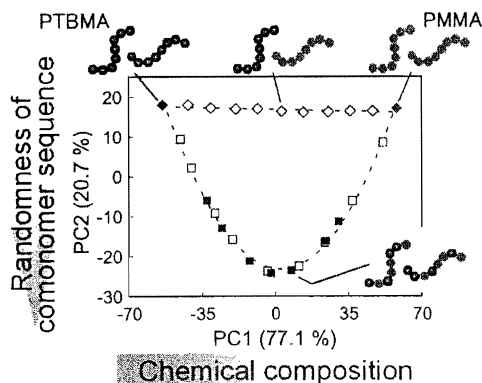


Figure 1. Principal component score plots for ^{13}C NMR signals of the carbonyl, backbone quaternary and α -methyl carbons of PMMA ($\blacklozenge$), PTBMA ($\blacklozenge$), their blends ($\circ$) and poly(MMA-*co*-TBMA)s obtained at early ($\square$) and end ($\blacksquare$) stages of copolymerization.

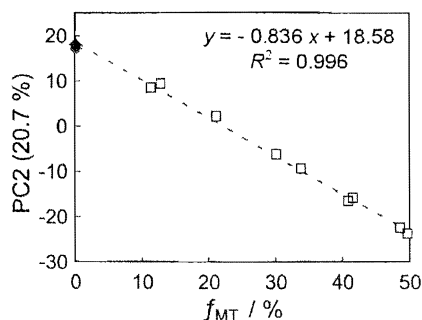


Figure 2. Relationship between the second principal component scores and the fraction of hetero-diad sequence, f_{MT} , calculated with monomer reactivity ratios (r_{MMA} and r_{TBMA}): PMMA ($\blacklozenge$), PTBMA ($\blacklozenge$) and poly(MMA-*co*-TBMA)s obtained at early stages of copolymerization ($\square$).

た TBMA 組成に対して、相対標準偏差 $RSD = 3.4\%$ 、 $R^2 = 0.999$ で一致することがわかった。

一方、共重合体の PC2 スコアは組成が等モル付近で最小となり、組成が偏るにつれ単独重合体に近づくことから、PC2 はモノマー連鎖のランダム性を反映すると考えられる。モノマー連鎖のランダム性を定量的に表す指標として異種モノマーの 2 連子含有量 f_{MT} (=ランナンバ⁵) を採用し、PC2 スコアと f_{MT} との関係を調べた。初期共重合体 9 種の仕込みモノマー組成と共重合体の組成から求めたモノマー反応性比($r_{\text{MMA}} = 0.81 \pm 0.06$, $r_{\text{TBMA}} = 1.26 \pm 0.03$)を用いて算出した f_{MT} と PC2 スコアとの関係を Figure 2 に示す。両者は良い相関($R^2 = 0.996$)を示すことがわかった。また、単独重合体 2 種と初期共重合体 9 種の ^{13}C NMR スペクトルと算出した f_{MT} を検量用データとした PLSR を行い、高重合率の共重合体 7 種について平均の f_{MT} を決定できることもわかった。

MMA-TBMA-HEMA 三元共重合体⁶

前述の MMA-TBMA 二元共重合体での検討と同様に、PMMA, PTBMA, PHEMA と、3 つの単独重合体の混合物 33 種、組成の異なる MMA-TBMA 二元共重合体 12 種と MMA-HEMA 二元共重合体 14 種に、MMA-TBMA-HEMA 三元共重合体 4 種を加えた合計 66 種の ^{13}C NMR スペクトルの PCA で得られた三次元スコアプロットと、2 つの主成分に射影した二次元スコアプロットを Figure 3 に併せて示す。単独重合体とこれらの混合物のスコアは、PC1-PC2 平面にはほぼ平行な平面上に、単独重合体 3 種を各頂点とした三角相図として表された(Figure 3a, 3d)。二元共重合体のスコアは、単独重合体とそれらの混合物を底面とし、垂線が PC3 軸と平行な三角すいの側面上に分布した(Figure 3b, 3c)。また、二元共重合体の PC3 スコアは、各共重合体の組成が等モル付近で極大となった。三元共重合体のスコアは、三角すいの内部に分布し、組成が等モルに近い共重合体の PC3 スコアが最も大きくなった。以上の結果から、MMA-TBMA-HEMA 三元系のスコアプロットは、PC1 と

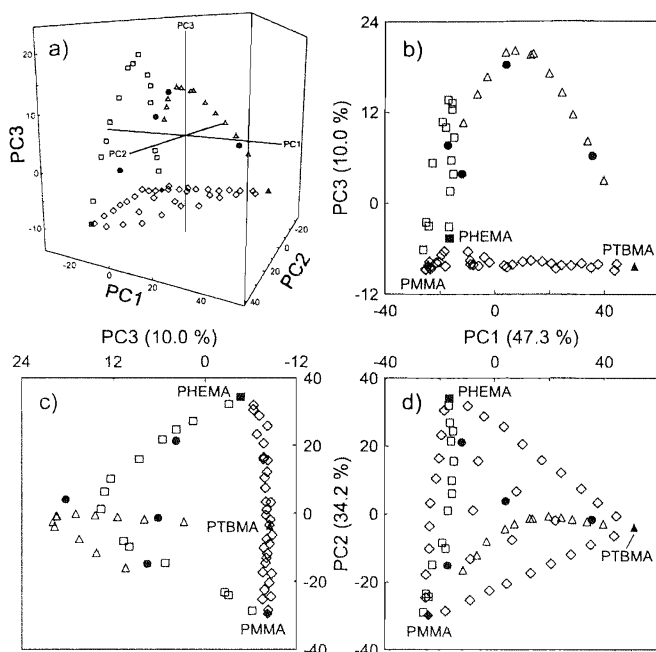


Figure 3. Principal component score plots for ^{13}C NMR signals of the carbonyl, backbone quaternary and α -methyl carbons of PMMA (◆), PTBMA (▲), PHEMA (■), homopolymer blends (◇), poly(MMA-co-TBMA)s (△), poly(MMA-co-HEMA)s (□), and poly(MMA-co-TBMA-co-HEMA)s (●).

PC2 が主として組成, PC3 が主としてモノマー連鎖のランダム性をそれぞれ反映していると考えられる。

単独重合体とこれらの混合物, 合計 36 種の ^{13}C NMR スペクトルと組成を検量用データ (Set A) とした PLSR により, 三元共重合体の組成を求めた結果を Table 1 に示す。Set A に 2 系列の二元共重合体を加えた合計 62 種を検量用データ (Set B) とした場合の結果も Table 1 に併せて示す。予め ^1H NMR 測定から求めた組成と PLSR で求めた組成とは, Set A および Set B の検量用データを用いた場合, RSD が 10.1 および 8.7 % でそれぞれ一致した。

EGDMA-TBMA 高分岐共重合体⁷

多量の重合開始剤 [2, 2'-アゾビスイソ酪酸ジメチル(MAIB)] を用いて得られた可溶性高分岐ポリマー PEGDMA, および組成の異なる EGDMA-TBMA 高分岐共重合体 11 種と, PTBMA の合計 13 種の ^{13}C NMR スペクトルの PCA で得られた結果を Figure 4 に示す。線状の MMA-TBMA 共重合体では, TBMA 組成の増加とともに PC1 スコアが単調に変化した³, 高分岐共重合体では TBMA 組成の増加とともに PC1 の変化

Table 1. Comparison of predicted and measured compositions of poly(MMA-co-TBMA-co-HEMA)s by the PLSR using training data sets A and B.

Code	Monomeric unit	Composition / mol%		
		Measured ^a	Set A ^b	Set B ^c
T-811	MMA	77.0	71.5	76.3
	TBMA	11.7	14.2	10.7
	HEMA	11.3	14.3	13.0
T-181	MMA	8.4	11.0	12.2
	TBMA	80.7	77.2	78.9
	HEMA	10.9	11.8	8.9
T-118	MMA	8.8	12.2	14.1
	TBMA	8.5	7.8	5.6
	HEMA	82.7	80.0	80.3
T-111	MMA	30.9	28.5	31.9
	TBMA	34.2	35.5	34.3
	HEMA	34.9	36.0	33.8

^a Determined by ^1H NMR. ^b Training data sets: three homopolymers and 33 homopolymer blends. ^c Training data sets: 26 copolymers in addition to Set A.

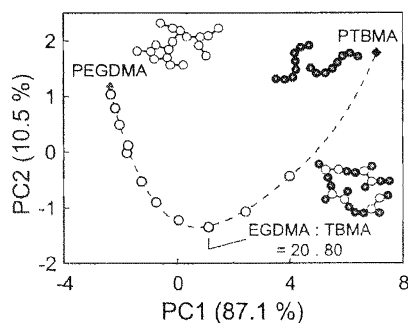


Figure 4. Principal component score plots for ^{13}C NMR signals of the carbonyl carbon of PEGDMA (○), PTBMA (◆) and poly(EGDMA-co-TBMA)s (●).

率が増加した。また、線状共重合体では共重合組成が等モル付近で PC2 スコアが最小になったが、高分岐共重合体では、EGDMA : TBMA = 20 : 80 (mol 比)の組成で最小となった。分岐構造とともに増加する MAIB 由来の末端基の影響と考えられる。

まとめ

アクリル系共重合体の ^{13}C NMR スペクトルを多変量解析することで、共重合組成とモノマー連鎖に関する定量的な情報を得ることができた。本手法は、二元系から三元系へ拡張できることも確認でき、多成分共重合体の新しいキャラクタリゼーション法として多方面での応用が期待される。今後は、立体規則性についても検討を進め、最終的には、多成分共重合体の組成、モノマー連鎖、立体規則性などを同時に定量できるようにしたい。

文献

1. 総説として, a) J. M. Chalmers, N. J. Everall, *TrAC*, **15**, 18 (1996). b) 下山 昌彦, 日本赤外線学会誌, **8**, 78 (1998).
2. 総説として, a) M. R. Viant, C. Ludwig, U. L. Gunthe, “Metabolomics, metabonomics and metabolite profiling”, ed. W. J. Griffiths, Cambridge: Royal Society of Chemistry; p. 44 [chapter 2] (2008). b) 吉田 欣史, 久原 とみ子, 菊池 淳, ぶんせき, **7**, 371 (2009).
3. H. Momose, K. Hattori, T. Hirano, K. Ute, *Polymer*, **50**, 3819 (2009).
4. 百瀬 陽, 前田 智也, 服部 康祐, 平野 朋広, 右手 浩一, 高分子討論会予稿集, **58**, 3211 (2009).
5. H. J. Harwood, W. M. Ritchey, *J. Polym. Sci. B Polym. Lett.*, **2**, 601 (1964).
6. 百瀬 陽, 前田 智也, 服部 康祐, 平野 朋広, 右手 浩一, 高分子分析討論会講演要旨集, **14**, 199 (2009).
7. 上池 亮太, 百瀬 陽, 平野 朋広, 右手 浩一, 高分子分析討論会講演要旨集, **14**, 149 (2009).