

蛋白質ポケットとアミノ酸間の 距離分布分析

三好 真紀子[†] 伊藤 貴之[‡] 由良 敬[‡]

[†]お茶の水女子大学理学部 〒112-8610 東京都文京区大塚 2-1-1

E-mail: [‡]{iqams, itot}@itolab.is.ocha.ac.jp, yura.kei@ocha.ac.jp

あらまし 人体の主成分である蛋白質は、アミノ酸が多数連結して複雑に折りたたまれることで、固有の立体構造を形成する。薬は蛋白質表面上の窪み（ポケットと呼ぶ）に入り込みやすいことが知られている。そこで我々は、薬との親和性と、アミノ酸のポケット内で位置関係がどのように関係しているか分析している。その一環として本報告では、ポケットとアミノ酸の距離分布分析のための可視化ツールを提案する。本手法では、蛋白質表面形状から抽出された各ポケットに対して最深部を求め、アミノ酸を構成する各原子との距離を算出し、その分布を可視化する。本報告では可視化結果から得られた傾向を示し、その生物学的解釈の可能性について推察する。

キーワード 可視化, 蛋白質, アミノ酸, 距離分布分析, ポケット抽出

Distance Distribution Analysis in Pockets among Amino Acids

Makiko MIYOSHI[†] Takayuki ITOH[‡] and Kei YURA[‡]

[†] Faculty of Science, Ochanomizu University 2-1-1 Otsuka, Bunkyo-ku, Tokyo, 112-8610 Japan

E-mail: [‡]{iqams, itot}@itolab.is.ocha.ac.jp, yura.kei@ocha.ac.jp

Abstract Protein is – the major component of the organism. It has a unique typical three-dimensional structure determined by the sequence of, amino acids. A hole (pocket) on the surface of a protein is known to be the best target for a drug to react. We started analyzing how "druggability" of proteins related to the locations amino acids in a pocket. For a starter of this study, this paper presents a visualization tool for distance distribution analysis between two types of amino acids in a pocket. When protein surface is provided by triangular meshes, this tool first identifies a pocket from the protein surface, specifies the deepest position of the pockets, and calculates distances between atoms of an amino acid and the deepest positions of the pocket. This tool then visualizes the distribution of the distances by scatterplots. This paper proposes the a biological interpretation of the visualization results.

Keyword Visualization, Protein, Amino acid, Distance distribution analysis, Pocket identification

1. はじめに

蛋白質は生体を構成する主成分であり、20 種類のアミノ酸が多数連結し複雑に折りたたまれることで球状に近い特有な立体構造をもつ。アミノ酸が連結する順番は、暗号化されてゲノムに記述されている。蛋白質は数千万種以上存在するともいわれ、機能も多種多様である。蛋白質の表面には凸凹の形状があり、その中でも深くて大きな凹みはポケットと定義される。ポケットに薬成分が一定時間入り込むことで、薬が蛋白質に作用することが多いとされている。ここで薬との親和性の高さを *druggability* といい、親和性の高いポケットを *druggable* なポケット、そうではないポケットを *undruggable* なポケットという。*druggable* なポケットの発見は、創薬過程において開発費用の低減につな

がることから、これまでも数多くの研究が報告されてきた。

我々も先行研究として、蛋白質表面形状を粗い三角形の集合で近似することにより、表面形状の凹部を高速抽出する手法を提案した[1]。しかし、表面の凹部の中でも *druggable* なポケットとなる部位は限られていることから、やみくもに表面形状の凹部を抽出するのではなく、*druggable* なポケットだけを抽出するための分析手法が必要である。そこで我々は、凹部の形状特徴量の比較、あるいは化学的情報である疎水性、電位などを含める多変量を解析するための可視化ツールを発表した[2]。

一方で、薬を含む低分子とアミノ酸との結合しやすさには統計的傾向があることも知られている[6]。この

ことから、ポケットの形状特徴量や化学的情報だけでなく、ポケットとアミノ酸との距離も druggability を診断するための有効な情報となり得ることが示唆される。そこで本研究では先行研究を踏まえて、ポケット内のアミノ酸間距離分布を視覚的に分析することを試みた。これにより、アミノ酸と druggability の関係性を明らかにすることを目標にする。

2. 関連研究

2.1 蛋白質ポケット・結合部位の druggability 分析

Perot らは、druggable なポケットおよび結合部位を抽出し、その類似度比較評価結果や公開データベースをサーベイした論文[3]を発表している。この論文で紹介されている手法やシステムは、幾何学的な手法からエネルギー基準に基づく手法まで多種多様である。一例として、水素結合ドナーなどの物理化学の特性を使用した手法である SiteEngine[4]や、構造領域を見つけることにより、結晶構造がまだ決定されていない蛋白質の低分子結合部を推定するデータベース SMID-BLAST[5]など、化学的特徴量を用いた多様な手法が発表されている。

2.2 低分子結合部位のアミノ酸の傾向

Yamaguchi らは、すでに低分子が結合している部位に対して、低分子がどのアミノ酸種とよく相互作用しているのかの統計をとり分析している[6]。この研究では、Het-PDB[7]というデータベースを実装し、蛋白質に対してアミノ酸種に関する統計をとることで、知られていなかった低分子とアミノ酸の相互作用を発見している。図1は、ATP(左)またはGTP(右)と相互作用するアミノ酸を円グラフにより表している。しかし、この手法では蛋白質ポケットに分析対象を特定していない。そこで我々は、分析対象を蛋白質ポケットに特定し、アミノ酸との相互作用をポケットとアミノ酸間の距離分布を用いて分析することとした。

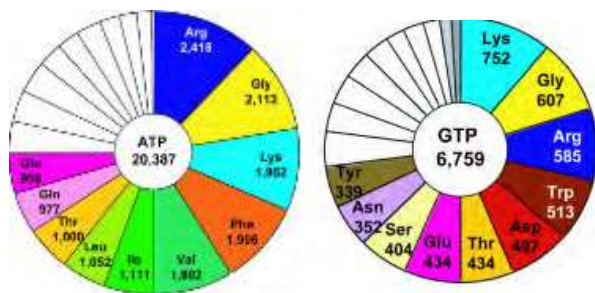


図 1.ATP(左),GTP(右)と相互作用するアミノ酸残基
(文献[6]から転載)

3. 処理手順

本章では、本手法の処理手順を述べる。3.1 節では本手法で扱う蛋白質表面形状を定義し、3.2 節では蛋白質表面形状からのポケット抽出手法を紹介する。3.3 節以降では、処理手順について述べる。

3.1 蛋白質表面形状データの取得

本手法では、蛋白質表面形状が三角メッシュで近似されることを前提としている。eF-site[8]は蛋白質の分子表面に関するデータベースであり、コノリー曲面抽出技術を用いて分子表面形状を算出している。本手法は、eF-siteに登録されている分子表面形状を利用する。eF-siteでは、PDB (Protein Data Bank)[9]に登録されている蛋白質を対象とし、頂点、2 頂点を両端とする稜線、3 稜線によって囲われた三角形領域の集合で構成される三角形メッシュを XML 形式で記述して提供している。また、このような幾何学的情報以外に、各頂点には疎水性、温度因子、電位などの物理的または化学的情報も同時に与えられている。

3.2 ポケット抽出

本手法では、まずポケットを蛋白質表面から抽出する。ポケット抽出手法は Nakamura[1]の手法を用いる。この手法は表面形状データから以下の手順を踏んでポケットを抽出している。図2は凝結因子 XA (PDB ID:1ezq) に適用した例である。

1. 三角形メッシュ単純化により、表面の細かい凸凹を除去して平滑された粗い形状を形成する。
2. 平滑化された表面形状から、凹部形状を抽出することで、ペプチドサイズの凹部を抽出する。
3. 抽出された凹部を、単純化前の三角形メッシュに投影する。
4. 投影された凹部形状の不要な部分を除去する。

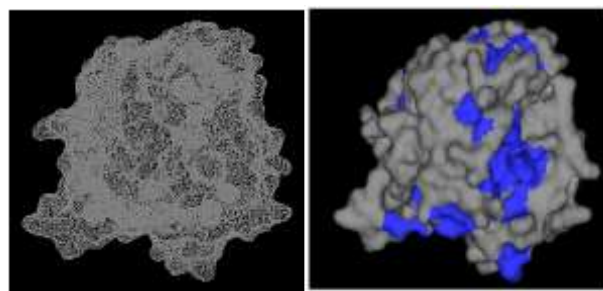


図 2.蛋白質表面形状(左)とポケット(右)の例

この手法ではメッシュ単純化により、非常に小さい計算量で的確にポケットを抽出する。

3.3 ポケットとアミノ酸間の距離算出

続いて、蛋白質を構成する各アミノ酸とポケットの

距離を算出する．現時点での我々の実装では，最初の試みとしてポケットとアミノ酸間の距離を「ポケットの最深部から，アミノ酸を構成する原子のうち最も近い原子まで」と定義している．この定義をイラストにしたものを図 3 に示す．ただし，この定義はあくまでも一例であり，他の定義にしたがった距離算出も今後の課題として取り組みたい．

現時点での我々の実装ではまず，抽出した各ポケットに対して最深部を算出する．最初に，外周を近似する平面の方程式 $pv+k=0$ を定義する．ここで p は平面上の点であり v は平面の法線ベクトルである．次に，式(1)を適用してポケットを構成している各頂点 p について平面との距離 d を算出する．そしてこの距離が最も大きい頂点を最深部とする．

$$d = \frac{pv+k}{|v|} \quad (1)$$

続いて，最深部である頂点とアミノ酸を構成している各原子とのユークリッド距離を算出する．ここで我々の実装では，Protein DataBank [9]にて取得した PDB ファイルを開き，蛋白質を構成する各原子とその座標値，および各原子がどのアミノ酸に属するかなどの情報をここから参照する．

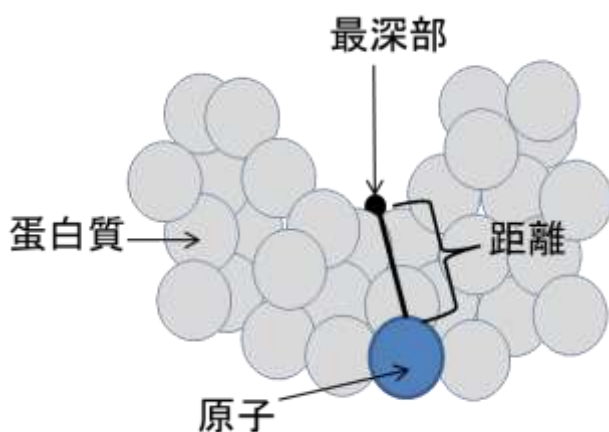


図 3. 距離算出手法

3.4 視覚的分析 1

我々はまず特定のアミノ酸の近傍に druggable なポケットは多いのかどうか調べるため，3.3 節で算出したポケットと最深部間の距離が 5 Å 未満のものの個数を druggability ごとに数え，折れ線グラフで表示し分析している．横軸にアミノ酸 20 種類，縦軸にポケットの個数を割り当てている．また，各折れ線の色はポケットの druggability を表しており，赤い折れ線が druggable なポケット，青い折れ線が undruggable なポケットである．

3.5 視覚的分析 2

次に前節を踏まえ，特定のアミノ酸の組み合わせの近傍にどんなポケットが多いのかを調べた．我々は 3.3 節までの処理を，金子ら[2]が提案した可視化ツール上で実装し，ポケットとアミノ酸間の距離に関する傾向を視覚的に分析している．現時点で我々が利用している可視化ツールでは，druggable なポケットの基準を発見するために，各ポケットが持つ特徴量と druggability の関係性を散布図として可視化している．そこで我々は現在，個々のポケットと 2 種類のアミノ酸との距離を 2 変数として散布図表示を試みている．

図 4 は，可視化ツールの表示例である．表示画面右側は，散布図の軸選択領域となっている．選択された二項目についての散布図を画面の左側に表示される．ここで散布図の各プロットはポケットを表しており，赤いプロットが druggable なポケット，青いプロットが undruggable なポケットである．散布図の中の各プロットをクリックすると画面中央にポケットの形状が表示されるようになっている．散布図とポケットの形状を同時に表示・観察できるようになっている．

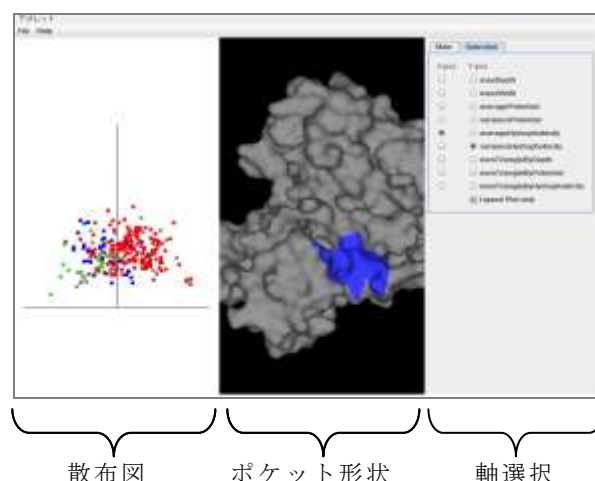


図 4. 可視化ツール

現時点での我々の実装では，軸選択領域において 20 種のアミノ酸から 2 種類を選択し，その 2 種類のアミノ酸とポケット間の距離を可視化できるようにした．しかしこの実装では，2 種類のアミノ酸とポケットとの距離分布を視覚的分析できるが，3 種類以上のアミノ酸との距離を同時に視覚的分析するのは困難である．そこで散布図以外の可視化手法の適用を今後の課題にしたい．

4. 実行結果

本章では Halgren が文献[10]にて druggable, difficult, undruggable のいずれかの結果を付与した 60 個の蛋白

質を対象とし、Nakamura らの手法[1]で抽出した 274 個のポケットについて距離分布を可視化した結果を示す。

4.1 分析 1

各々のアミノ酸の近傍に位置する druggable なポケットの個数を集計した。図 5 は、横軸に 20 種類のアミノ酸、縦軸にポケットの個数を割り当てた折れ線グラフである。undruggable なポケットや difficult なポケットはどのアミノ酸種に対しても 20 個未満であることが分かった。反対に、druggable なポケットの数はアミノ酸によって大きな差があることが分かった。システインとポケット最深部との間が 5 Å 未満である druggable なポケットの数は約 20 個だったが、ロイシンの場合は 100 個以上だった。

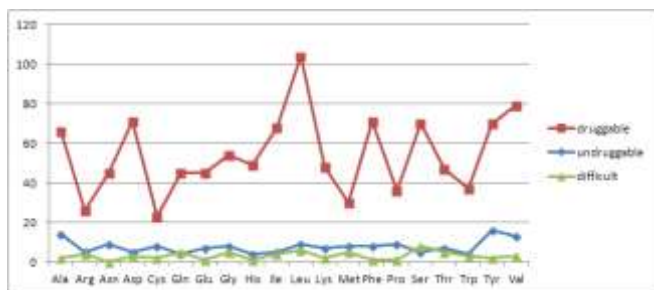


図 5.折れ線グラフの観察結果

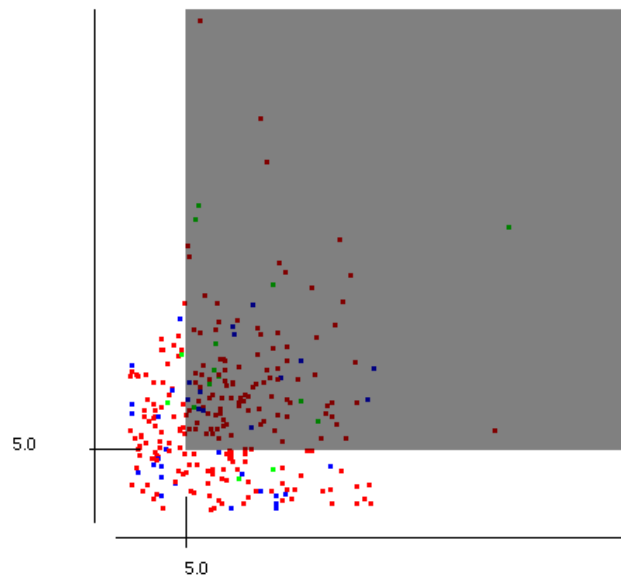
4.2 分析 2

前章で述べた可視化ツールを用いて、ポケットとアミノ酸間の距離と druggability の関係性を観察した結果を示す。本章でのコマンドの実行に際して、ポケットから距離の近いアミノ酸に着目するために、2 種類のアミノ酸の両方と距離が遠い部分に陰を作ることによって、アミノ酸との距離が近いものだけを注目できるようにした。本章が示す結果では、2 種類のアミノ酸の両方とも 5Å 以上離れている部位に陰をつけた。

図 6 は、X 軸にポケット最深部とアラニンとの距離、Y 軸にポケットとチロシンとの距離を割り当てた場合の結果である。ポケットとの距離が近いものには druggable なポケットと undruggable なポケット両方が混じった状態である。したがって、アラニンとチロシンが近いから druggable なポケットであるという予測は出来ないことがわかった。このような結果を以後「結果 A」と称する。

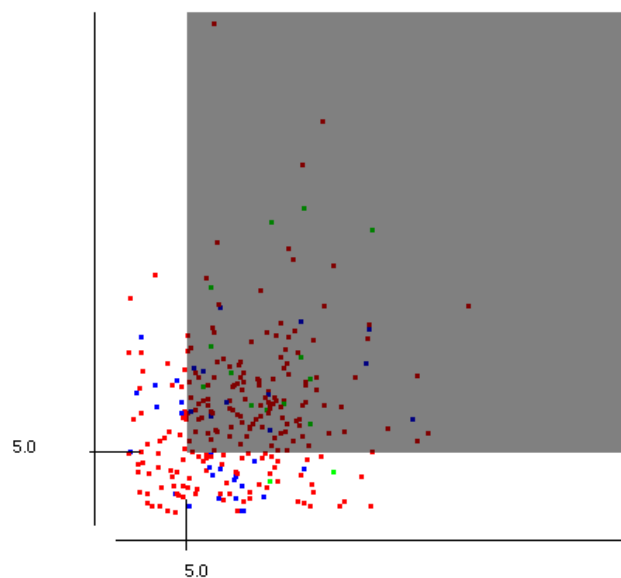
図 7 は X 軸にアスパラギンとの距離、Y 軸にポケットとチロシン間の距離を割り当てた結果である。この結果をみると、二つのアミノ酸に近いポケットは druggable なポケットが多いことがわかった。したがって、まだ druggable なポケットか undruggable なポケットかどうかわかっていないポケットについても、ヒス

チジンとチロシンとの距離が近いことが分かれば、druggable なポケットであるかもしれないという推測が立てられる。このような結果を以後「結果 B」と称する。



ポケットとアラニン間の距離(X 軸)
ポケットとチロシン間の距離(Y 軸)

図 6.散布図の観察結果 1



ポケットとアスパラギン間の距離(X 軸)
ポケットとチロシン間の距離(Y 軸)

図 7.散布図の観察結果 2

4.3 分析まとめ

分析 1 と分析 2 を総合した知見を得るために、折れ線グラフと散布図を並べて図 8 のように表示した。格

子状に色分けされた領域には、縦方向、横方向ともに図5と同じ順番で20種類のアミノ酸種が割り当てられており、各格子領域が2種類のアミノ酸種の組み合わせを示している。図8において各格子領域には、散布図の結果が結果Aと同じ傾向を示したものを青色、結果Bと同じ傾向を示したものをオレンジ色で色付けた。このグラフから、分析1での個数を観察しながら、分析2の結果を観察できる。

まず、二軸にどんなアミノ酸種を割り当てたときに結果Aが見られるかについて示す。図9では、結果Aと同じ傾向を示したアミノ酸ペア（青色）と、その中でも分析1でdruggableなポケットの個数が平均よりも多かったものを二軸のどちらかにでも割り当てているペアを水色で色付けた。分析1で個数の多かったアミノ酸種を二軸に割り当てているにも関わらず、結果Aが同じ傾向を示すことがわかった。

次に結果Bの場合でも、同様にグラフを作成した。分析1でdruggableな個数が多かったものを二軸のどちらかにでも割り当てているものをピンク色で色付けたものを図10で示す。そうしたところ、結果Bと同じ傾向を示すものは、軸のどちらかには分析1でdruggableなポケットの個数が多かったアミノ酸種を割り当てていることが分かった。

図9と図10をまとめたものを図11に示す。分析1でdruggableの個数が多かったアラニン（片方の軸に割り当てた場合、結果Aと結果Bどちらの傾向を示すものが結果としてでることが分かった。ゆえに分析1と分析2それぞれから、独立の知見を得ることが出来た。

5. まとめ

本報告では、蛋白質ポケットのdruggability分析の一環として、ポケット最深部とアミノ酸間の距離が近いものの個数を折れ線グラフで、またポケットとアミノ酸間の距離分布を、金子[2]の可視化ツールを使用して表示した結果を示した。その結果として、druggableなポケットには特定のアミノ酸との距離が近い傾向があることが示唆された。

今後の課題として、以下について検討したい。

- ポケットの最深部以外の部位とのアミノ酸の距離分布を視覚的分析する。
- 3種類以上のアミノ酸との距離分布を同時に分析するために、散布図以外の可視化手法を導入する。
- アミノ酸との距離分布と、その他の化学的特性（電位や疎水性など）を同時に視覚的分析する。
- 可視化ツールだけでなく判別分析などの統計的手法を導入する。
- Halgren[10]が示した60種類の蛋白質だけでなく、

もっと多数の蛋白質を対象として距離分布の可視化を試みる。

参考文献

- [1] Y. Nakamura, T. Itoh, Fast Detection and Visualization of Protein Surface Pockets for Drug Design, The 11th Asian Symposium on Visualization, 2011.
- [2] 金子彩香, 中村友香理, 伊藤貴之, “蛋白質ポケットとdruggability分析のための可視化ツール”, NICOGRAPH 2012, 1-8.
- [3] S. Perot, O. Sperandio, M. A. Miteva, A. -C. Camproux, B. O. Villoutreix, Druggable Pockets and Binding Site Centric Chemical Space: A Paradigm Shift in Drug Discovery, Drug Discovery Today, 15(15-16), 656-667, 2010.
- [4] “SiteEngine”, <http://bioinfo3d.cs.tau.ac.il/SiteEngine/>
- [5] A. Shulman-Peleg, et al. Recognition of functional sites in protein structures, J. Mol. Biol. 339, 607-633, 2004.
- [6] A. Yamaguchi, K. Iida, N. Matsui, S. Tomoda, K. Yura, M. Go, Het-PDB Navi. :A Database for Protein-Small Molecule Interactions, J. Biochem, 135, pp. 79-84, 2004.
- [7] “Het-PDBNavi.” <http://hetpdbnavi.nagahama-i-bio.ac.jp/index.php>
- [8] K. Kinoshita, eF-site, <http://ef-site.hgc.jp/eF-site/index.jsp>
- [9] “ProteinDataBank”, <http://www.rcsb.org/pdb/home/home.do/>
- [10] T. A. Halgren, Identifying and Characterizing Binding Sites and Assessing Druggability, Journal of Chemical Information and Modeling, 49(2), 377-389, 2009.

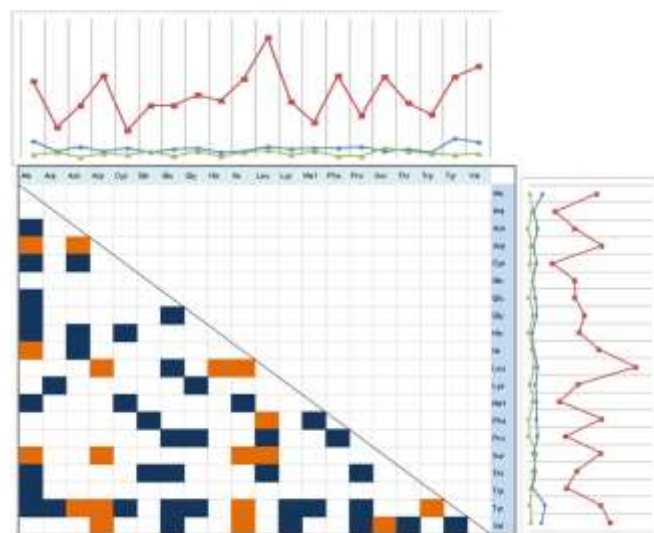


図8. 分析まとめ結果(1)

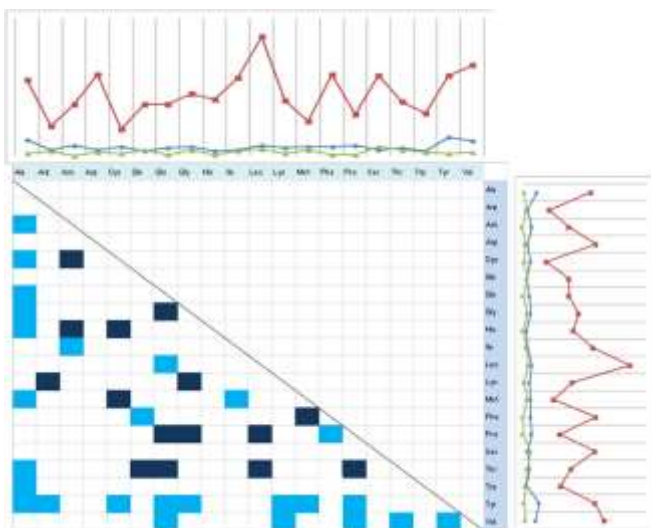


図 9. 分析まとめ結果(2)

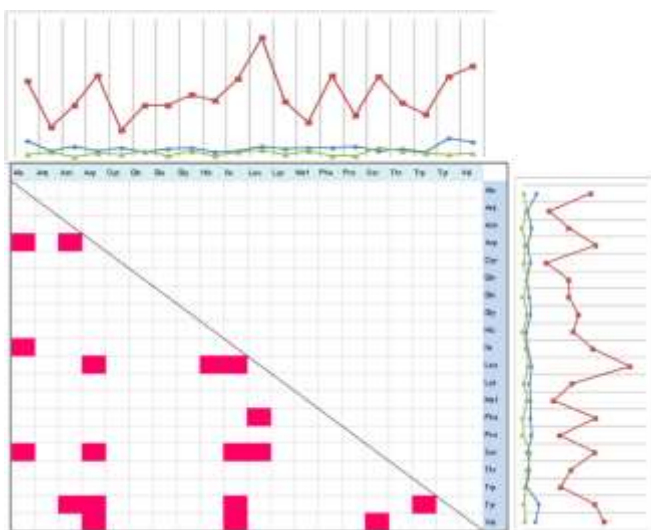


図 10. 分析まとめ結果(3)

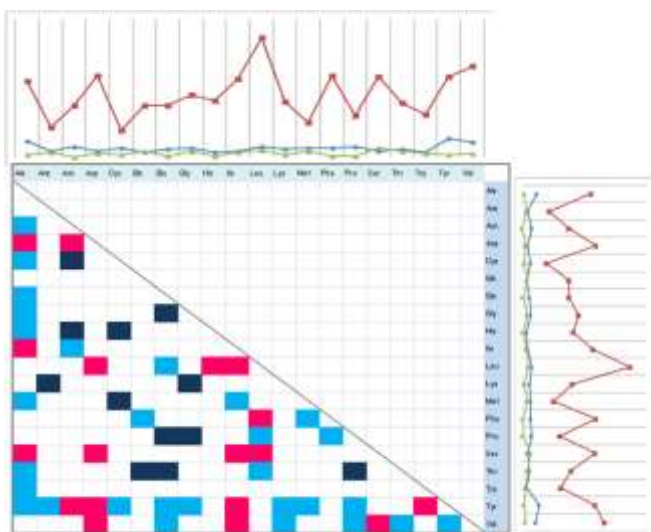


図 11. 分析まとめ結果(4)