

結晶スクリーニング結果を活用した機械学習による 結晶化条件の予測モデルアレイ

Developing a prediction array of crystallization condition models
from screening trial results by machine learning method

小西史一^{1 4} 竹本千重² 赤坂領吾² 行木信一^{2 3} 秋山 泰¹
Fumikazu Konishi Chie Takemoto Ryogo Akasaka Nobukazu Nameki Yutaka Akiyama
横山茂之^{2 4} 豊田哲郎⁵ 村山和隆^{2 6}
Shigeyuki Yokoyama Tetsuro Toyoda Kazutaka Murayama

東京工業大学 大学院情報理工学研究所¹ 理化学研究所 横浜研究所 生命分子システム基盤研究領域²
Tokyo Institute of Technology RIKEN Yokohama Institute, SSBC
群馬大学 大学院工学研究科³ 理化学研究所 横浜研究所 生命情報基盤研究部門⁵
Gunma Univ., Graduate School of Engineering RIKEN Yokohama Institute, BASE
東京大学 大学院理学系研究科⁴ 東北大学 大学院医工学研究科⁶
The Univ. of Tokyo, Graduate School of Science Tohoku Univ., School of Biomedical Engineering

1 はじめに

タンパク質の結晶構造解析は、生体高分子の働きを知る有用な手段であり、学術的にも応用的にも構造決定の重要性が高まっている。このようなタンパク質に関する基本構造の決定を目指したタンパク 3000 プロジェクト [1] では、高等生物の機能ドメインや小型タンパク質の構造解析には NMR を用い、より大きいタンパク質の構造解析には X 線による結晶回折を主に用いて解析を行ってきた。近年この X 線結晶構造解析分野でのソフト・ハードの発展は、顕著な効率化を実現しつつある一方で、結晶化のプロセスは、未だ試行錯誤的な経験的アプローチに終始しており、タンパク質一つの結晶化に数百から千を超える条件を実際に試すことも稀ではない。つまり、現在のタンパク質の結晶構造解析においては、その結晶化がボトルネックとなっており、結晶化プロセスの合理化は、多数の構造を知る上で急務の課題であると広く認識されている。

これまで、系統的に行われた精製実験や結晶化実験における成否情報（アミノ酸配列、発現、可溶化、結晶化溶液の条件検討、そして結晶化の有無等）は、質の高い実験データベースとして蓄積されており、この情報に対して Support Vector Machine (SVM)[2] などの数理情報科学手法を適用することで、未知のタンパクに対してスクリーニング条件を絞り込む支援を行うシステムの開発を実施することを目的としている。

2 学習データセット

本研究で使用した学習データセットは、タンパク 3000 プロジェクトにおいて、結晶化のスクリーニング実験として実施された結果の一部を利用した。一般的にタンパク質・核酸などの高分子を対象とした結晶化には、塩、pH、沈殿剤等のパラメータが関与することが知られており、その条件を探すためには初期スクリーニング実施することが有効とされている。そのため、結晶を必要とするタンパク質に対して、複数の系統だった市販のスクリーニングキットを初期段階のスクリーニングに使用している。本

研究での学習データは、Hampton Research 社の Crystal Screen[3] と Index[4]、そして Emerald BioSystems 社の Wizard での結果である。そのキットを構成する溶液の組成の違いがあるが、それぞれ初期スクリーニングとしての実績がある。そして、これらの 3 つのスクリーニングキットによる実験は、96well のプレート上で実施され、各 well では 1 つの独立した結晶化条件として試薬が調製されている。したがって、同一のタンパク質をプレート上にある各 well に規定量注入することで、96 種類の独立した結晶化条件を試すことができる。さらに、異なる 3 種類のキットを適宜組み合わせることで、1 つのタンパク質で最大、288 条件の結晶化条件についての結果を得ることができる。

表 1 学習データセットに含まれる生物種とスクリーニングキットの対応

Species	N	CS	Wizard	Index	sum
<i>Pyrococcus horikoshii</i> OT3	5	5	1	1	7
<i>Escherichia coli</i> W3110	6	5	4	4	13
<i>Thermus Thermophilus</i> HB8	6	4	2	5	11
<i>Aeropyrum pernix</i> K1	5	5	5	2	12
<i>Aquifex aeolicus</i> VF5	2	2	2	0	4
<i>Eukaryote</i> and the others	64	114	23	63	200
Total	88	135	37	75	247

表 1 に示したように、CS135 回の結果、Wizard37 回の結果、Index75 回の結果、そして合計 247 回のスクリーニング結果を基に結晶化の予測モデル作成を行う。なお、スクリーニング対象となったタンパク質の個数は、全 88 種類のタンパク質であった。このスクリーニング実験では、プレート上にある全ての条件について、5 段階（透明、沈殿、凝集、相分離、結晶）のスコアが人手によって付加される。通常では結晶を生じた条件を、さらに詳細に溶液条件として展開することで、良質な回折像を得ることができる品質にまで、結晶を成長させるなどの下流工程が実行される。また、本実験におけるタンパク質は全てアミノ酸配列が特定されている物を対象として

いる。したがって、本研究で対象としている学習データセットとは、3つのスクリーニングキットによる実験結果と、対象となったタンパク質の配列の2つの情報により教師付学習の枠組みとして活用する。

3 Support Vector Machine

サポートベクターマシン (Support Vector Machine: SVM) は、線形識別機で与えられた n 次元の特徴ベクトル ($x_i \in R^n, i = 1, \dots, l$) に、正例もしくは負例のラベル ($y_i \in R^1, y_i \in \{1, -1\}$) を付けて、特徴量空間上に配置する。そして、その2クラスを判定する分離平面を学習することで識別する数理手法である。定式化すると SVM は以下のような問題を解くことに帰着する。

$$\begin{aligned} \min_{w, b, \xi} \quad & \frac{1}{2} w^T w + C \sum_{i=1}^l \xi_i \\ & y_i(w^T \phi(x_i) + b) \geq 1 - \xi_i, \\ & \xi_i \geq 0, i = 1, \dots, l. \end{aligned} \quad (1)$$

また、問題が複雑な場合にカーネル関数により、非線形変換を施すことで、線形識別できるようにするなどの手法をとることができる。本研究ではガウシアンカーネルとして知られている RBF カーネルを採用している。

RBF(Radial Basis Function) Kernel:

$$K(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2)$$

4 モデル生成用データセット生成法

スクリーニングキットは、Hampton Research 社の Crystal Screen と Index, Emerald BioSystems 社の Wizard などでも販売されている物を使用した。添加後に、それぞれの well における結晶化の進行状況を観測できる。表 2 は、Crystal Screen の D02 位置における溶液条件を示している。実際には、96well の限られたスペースに配置する必要があるために、明らかに不要な条件などを削除して調製するなどの処置を行うことがあるが、どの位置にどのような組成条件があるのかはデータベースを使い管理されている。

表 2 Crystal Screen キットの D02 位置における結晶化溶液の組成値

Well	D02
kit	Crystal Screen 1
Salt	none
buffer	0.1M tri-Sodium citrate pH 5.6
precipitant	20% iso-Propanol. 20% PEG4000

つぎに、これらのスクリーニング実験はプレート単位で実行され、他の実験と区別するために ID が付加される。どのタンパク質が実験に使われたなどの情報が付加され、実験条件毎にマスターデータベースに登録される。したがって、本研究においては、専用のデータベーススキームを定義し、マスターデータベースからデー

タをインポートすることで、モデル生成用のマスターデータベースを構築した。これにより、データベースの更新に応じて、高速に学習用データセットを生成することができる。次に、このように設計されたデータベースに関して、スコア情報とタンパク質のアミノ酸配列情報の2種類のデータを得る。取得した結晶化のスコアデータに関しては、2つのクラスをとして識別するラベリングを行う。そして、アミノ酸の配列データに関しては、そのアミノ酸組成から数値化する処理や、物理的・化学的特性値との加重平均等で構成される。このように、スクリーニングの実験結果から複数の特徴量に変換することで、特徴空間としてマップするデータ変換処理を行う。この変換されたデータは、スクリーニングキットの well 記号とスコアマッピング毎に独立して生成されて、学習データとして保存される。通常1つのキットは、96well あるとすると、それぞれの well 毎に、5つの結晶化状態を2つに分割する4つの分割点があることから、96x4=384 の独立した条件のデータセットを作成する。ここでは、3つのスクリーニングキットを対象としているため、384x3=1152 の独立したモデルが最終的に作成される。

アミノ酸配列情報から特徴空間へのマッピング処理法

SVM を行うためには、アミノ酸配列を活用して多次元から構成される、特徴空間の座標として表現することになる。比較的容易に利用することができる特徴空間として、全 20 種類のアミノ酸組成頻度があげられる。しかしながら、タンパク質を、それぞれのアミノ酸の頻度で構成される 20 次元空間の1点として表現する方法では、対象となるタンパク質のペプチド長が短い場合には、十分な識別面を構成することが困難となる可能性がある。そこで、同じアミノ酸組成から計算することができる特徴量を表す値として、アミノ酸のグループ (acidic, acyclic, aliphatic 等) を表す特徴量を 15 種類導入した。また、分子量や、各ペプチドが持つ疎水性に関する特性値などを基準とした、加重平均を使う手法を用いて、16 種類の物理化学特性値を特徴空間に導入した。

アミノ酸組成 (20 種類) A, C, E, D, G, F, I, H, K, M, L, N, Q, P, S, R, T, W, V, Y

アミノ酸グループ組成値 (15 種類) acidic, acyclic, aliphatic, basic, buried, charged, cyclic, hydrophobic, large, medium, negative, neutral, polar, positive, small

物理化学特性値 (16 種類) mass, pI, pK1, pK2, pKr, hydrophathy, residues, AO, AI, IO, RO, RI, RIO, logP, van der, Occurrence

結晶化スコアの2値化処理法

結晶化のスクリーニングの結果は、各条件毎に専門的な技術を持つ人の目によって確認し、スコアが付加される。そのスコアは、1 から 5 までの 5 段階 (透明 (1)、沈殿

(2)、凝集 (3)、相分離 (4)、結晶 (5) が与えられる。(1) の透明は、タンパク質を添加したにもかかわらず、一定期間スクリーニングしたにも関わらず変化が無い状態。(2) の沈殿は、溶液中になんらかの沈殿がおきているもので、それ以上の凝集などの変化が起こっていない状態。(3) は、沈殿が凝集されていることが認められるが、相分離までは判定できない状態 (4) は、相分離が起こっている状態。(5) は、何らかの結晶が認められた状態。以上のような状態がスコアによって付加される。この与えられた結晶化スコアを変換して、4つの新しい状態にマッピングし、2クラス判定の SVM に適用する教師データとして利用できるようになる。

M1 透明とそれ以外 (沈殿、凝集、相分離、結晶) の2クラスにマッピングする。これにより、溶液に何らかの変化が起きることを学習する。

M2 “透明、沈殿”と“凝集、相分離、結晶”の2クラスにマッピングする。これにより、溶液に相分離の兆候がおきる事を学習する。

M3 “透明、沈殿、凝集”と“相分離、結晶”の2クラスにマッピングする。つまり、相分離以上のことが起こることを学習する。

M4 “透明、沈殿、凝集、相分離”と結晶の2クラスにマッピングする。つまり、結晶化することを学習する。

表 3 結晶化スコアのマッピング機能によるサンプル数の推移 (CS/D02)

Map Method	Positive Sample	Negative Sample	N
M1	52	82	134
M2	18	116	134
M3	15	119	134
M4	7	127	134

5 予測モデルアレイ

教師付の学習モデルである SVM による学習とは、正負のラベリングがされた特徴空間上の点を、分離することができる超平面を見つけることになる。したがって、与えられた学習データから作成される適切な平面を表現できるモデルが、未知のアミノ酸配列に対して適用されるモデルとなる。1つの実験条件を表現するモデルには、スコアを4つの異なる解釈にマッピングしたことから、4つの2クラス判定モデルとして生成される。そして、その判定は4つのモデルからの結果を合成することで総合的に解釈される。したがって、これらの学習モデルを仮想的なスクリーニングキットのように見立て、構築された全てのモデルによる判定から、結晶化の条件を絞り込むことができるようになる。このモデル間は、基本的に独立しているが、結晶化条件キットは、条件間に溶液濃度などの勾配が付けられていることから、同一の溶媒によって支持された結果がでた場合には、その溶媒条件が結晶化にとって有利に働くことが予想される。したがっ

て、このようなモデルを複数集合的に利用するシステムを構築した。

システム実装

本研究のシステムは、モデル生成のためのデータファイルを生成する部分と、それらのデータファイルを一連の処理び実行によりモデルアレイを構築する部分、そして、構築されたモデルアレイを使い、未知のタンパク質の対して、その評価を行い結果を提示する3つの部分から構成される。データファイル生成部分では、高速でコンパクトな特徴を持つ、データベースエンジンである sqlite3[5] と Python によって実装されており、高速化と大規模な分散処理に耐えられるソフトウェア構造を持つ。また、学習用のデータファイルは、ディレクトリ階層で管理されており、モデルアレイを蓄積するルートディレクトリ以下は、結晶化キット階層、マッピング階層、Well 階層の3層で管理されている。このようなファイルシステム上のディレクトリ階層として実装することで、共有メモリ型の計算機からクラスター型の計算機の広い範囲で実装することができる。ディレクトリ毎の異なるモデルの学習プログラムの粗粒度並列性を持つことから、多数の CPU コアを利用したシステムの利用が可能となる。次に、モデル構築実行を行う部分に、libSVM[6] のライブラリを使用している。libSVM では、svm-train コマンドに 10-fold cross-validation などの機能が実装されていることや、カーネルの追加方法が容易であること、Python が標準で用意されているなどの利便性から判断して採用している。さらに、libSVM は C 言語で実装されており多くのプラットフォームでも高速に動作する。最後に、構築されたモデルアレイを使って、未知の配列の結晶化を予測するシステムは、Django と呼ばれる Python のフレームワークにより構築されており、高度なプロトタイプ生成システムにより生産性の向上とパフォーマンス向上を実現することができる。

5.1 単一モデル生成に必要なプロセス

3層のディレクトリによって管理されたデータに関し、以下にあげる手順によってモデルが構築される。まず、特徴量の次元ごとでスケーリング (正規化) を行う。現状では、特徴量の次元ごとに重みを変えるなどの処理を行っていない。つぎに RBF カーネルがもつ2つのパラメータに関するチューニングを実施する。図1に示のように、RBF カーネルのパラメータとして、実行する試行は、それぞれ20通りの値について試し、400通りの組み合わせを実行する。そして、その際に得られた最適なパラメータを採用する。モデル生成時には、10-fold cross-validation を行い、複数の試行を繰り返し最適なモデル生成を行う。以上までの処理により、モデルファイルが作成される。最後にこの選ばれたモデルの評価を ROC 曲線と AUC を使い評価する。

1. モデル生成用データファイルのスケーリング

2. RBF カーネルのパラメータ探索 (C, γ パラメータの探索 20x20)
3. モデルファイル生成 (10-fold cross-validation)
4. ROC 曲線の作成 (AUC の算出)

RBF カーネルのパラメータ探索

RBF カーネルの利用には、C と γ の 2 つのパラメータ調整を行う必要がある。本研究では、教師データを使い、これらの二つのパラメータを以下の範囲の 400 点での評価を行い探索し最適な C と γ を決定する。

$$C = 2^i, \quad i = -5, \dots, 15 \quad (3)$$

$$\gamma = 2^j, \quad j = -15, \dots, 5 \quad (4)$$

図 1 から図 4 までに示した 4 枚のグラフには、それぞれ、マッピング処理されたデータセットに対して、パラメータの探索を行った結果を示している。CS/D02 データの正例・負例の個数は表 2 にあるように、マッピングによって推移する。したがって、M1 に関するデータである図 1 は、全 134 点の内で、52 点の正例があり、且つそれらの関係性があまり認められない事が起因し、探索時におけるモデルの 10-fold cross-validation による精度をプロットは、全体的に低い値となっている。これらの 4 つのグラフからもわかるように、マッピングによる影響は大きく、それぞれのモデル毎にパラメータのチューニングが必要であることを示している。

6 予測モデルの評価

予測モデルに関する評価は、ROC 曲線 (Receiver Operating characteristic Curve) により行った。図 5 から図 8 のグラフは、Crystal Screen キットにおける D02well で作成したモデルに対する ROC 曲線である。図 5 の M1 による変換基準では、溶液に起こる変化全てをまとめることになるため、その現象と特徴空間との間に適切な分離平面を見いだすことができなかった事が予想される。その効果をより定量的にとらえるために、ROC 曲線の下部面積を表す AUC 値を導入することで比較を行った。M1 の場合には、AUC が 0.6337 となっており、M2 の 0.7045 と比べてもみても低いことがわかる。この例では、M2 で構成されたモデルの方が、より適切であることが伺える。よって、本研究で対処とした 3 つの結晶化スクリーニングキットにおける SVM モデルの全てについて、AUC による評価を行い、このモデルを使った予測を行う場合の信頼性を評価する指標とした。表 4 には、Crystal Screen を対象として作成した全 384 個の SVM モデルの結果をまとめたものを提示している。表は、その結晶化スコアのマップモデル毎に AUC を算出してある。また、AUC 値が 0.7 以上ものはボールドで表現しており、0.9 以上ものにはさらに下線を付加した。最後に M1 から M4 までの結果から、AUC が 0.7 以上モデルを Overlay した値を表 4 の最後に示した。これは、96well のモデル中で信頼性の高いモデルを使うことで、96well 中の 52well のモデルを採用できることを示している。そ

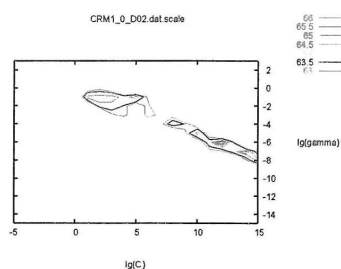


図 1 CS/D02/M1

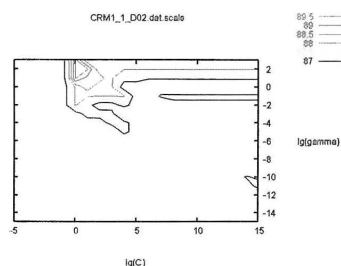


図 2 CS/D02/M2

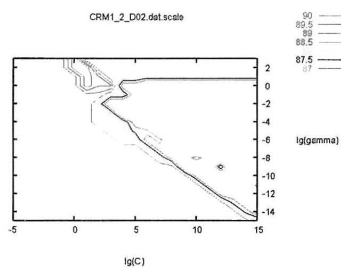


図 3 CS/D02/M3

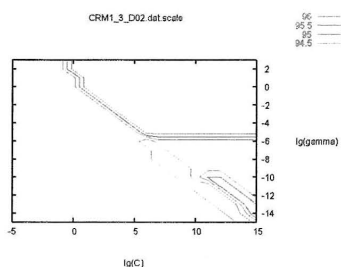


図 4 CS/D02/M4

して M1 から M4 までのマップは、それぞれ解釈の異なるモデルであるが、このようにアレイ上に配置することで、仮想結晶化条件スクリーニングが実現できる。

7 結論

結晶化スクリーニング結果から、SVM のモデルアレイを構築を行った結果次のような知見を得ることができた。

- 結晶化スコアをマップ処理することで、2 クラス判定の SVM モデルに合うシステム設計スキームを構築した。
- ROC 曲線の AUC を評価尺度として、モデルアレイの構築と評価を行うことを示した。
- SVM による結晶化条件の予測モデルアレイシステムの構築を示した。

現状のシステムでは、特徴空間の構成要素としては、配列の組成情報と、化学物理特性から得られる値だけであるが、アミノ酸の 2 次構造の情報や、コンタクトナンバーや、ディスオーダーなどの付加的な情報を取り入れることで、規定を満たすモデルを増やす試みを検討している。

謝辞

本研究は、平成 17 年度理化学研究所理事長ファンド・戦略型研究の準備調査「タンパク質結晶化条件最適化プログラムの開発」のサポートにより実施した内容が含まれている。

参考文献

- [1] Yokoyama S, Hirota H, Kigawa T, Yabuki T, Shirouzu M, Terada T, Ito Y, Matsuo Y, Kuroda Y, Nishimura Y, Kyogoku Y, Miki K, Masui R, and Kuramitsu S. Structural genomics projects in Japan. *Nat Struct Biol.*, 7 Suppl, 2000.
- [2] V. Vapnik. *The Nature of Statistical Learning Theory*. Springer-Verlag, 1995.
- [3] E. F Garman and E. P Mitchell. Glycerol concentrations required for cryoprotection of 50 typical protein crystallization solutions. *Appl. Cryst.*, Volume 29 Issue 5:584–587, 1996.
- [4] A. D'Arcy, A. Mac Sweeney and M. Stihle, and A. Haber. The advantages of using a modified microbatch method for rapid screening of protein crystallization conditions. *Appl. Cryst.*, D59:396–399, 2003.
- [5] Hwaci. *sqlite3*, 2004. Software available at <http://www.sqlite.org>.
- [6] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at <http://www.csie.ntu.edu.tw/~cjlin/libsvm>.

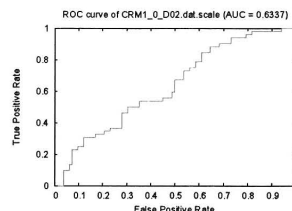


図 5 ROC Curve: CS/D02/M1

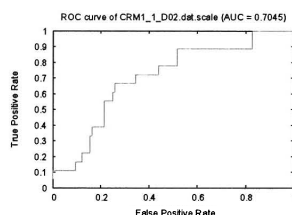


図 6 ROC Curve: CS/D02/M2

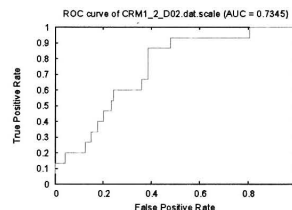


図 7 ROC Curve: CS/D02/M3

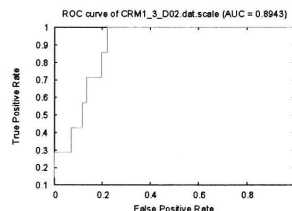


図 8 ROC Curve: CS/D02/M4

表 4 SVM 予測アレイモデルにおける ROC 曲線の AUC による評価結果 (Crystal Screen 1,2)

M1	1	2	3	4	5	6	7	8	9	10	11	12
A	0.684	0.572	0.679	0.675	0.791	0.711	0.620	0.690	0.655	0.662	0.783	0.602
B	0.621	0.654	0.547	0.601	0.687	0.770	0.639	0.651	0.491	0.655	0.589	0.657
C	0.718	0.633	0.389	0.697	0.702	0.762	0.689	0.697	0.792	0.620	0.728	0.679
D	0.541	0.634	0.512	0.732	0.671	0.694	0.646	0.784	0.654	0.676	0.570	0.697
E	0.722	0.596	0.650	0.575	0.788	0.574	0.698	0.532	0.779	0.599	0.698	0.676
F	0.677	0.678	0.600	0.655	0.670	0.656	0.697	0.589	0.655	0.688	0.514	0.671
G	0.689	0.619	0.762	0.688	0.457	0.571	0.650	0.562	0.628	0.660	0.735	0.551
H	0.601	0.791	0.477	0.493	0.697	0.675	0.722	0.688	0.733	0.745	0.658	0.652
M2	1	2	3	4	5	6	7	8	9	10	11	12
A	0.460	0.456	0.688	0.663	0.478	0.634	0.415	0.526	0.527	0.456	0.181	0.321
B	0.674	0.115	0.606	0.669	0.719	0.716	0.584	0.669	0.521	0.479	0.677	0.538
C	0.706	0.613	0.286	0.495	0.601	0.577	0.546	0.420	0.669	0.677	0.588	0.561
D	0.310	0.705	0.672	0.586	0.599	0.424	0.560	0.682	0.319	0.698	0.438	0.636
E	0.672	0.486	-	0.648	0.513	0.516	0.284	0.916	0.637	0.545	0.844	0.762
F	0.419	0.692	0.466	0.764	0.476	0.597	0.365	0.748	0.467	0.294	0.625	0.487
G	0.423	0.613	0.553	0.614	0.460	0.610	0.556	0.639	0.642	0.358	0.829	0.486
H	0.598	0.730	0.406	0.398	0.361	0.539	0.548	0.444	0.673	0.491	0.405	0.525
M3	1	2	3	4	5	6	7	8	9	10	11	12
A	-	0.575	0.685	0.715	0.909	0.646	-	0.913	0.719	0.858	0.783	0.346
B	0.659	-	0.636	0.666	0.748	0.737	0.799	0.662	0.388	0.676	0.905	0.573
C	0.723	0.572	0.470	0.493	0.606	0.272	0.763	0.420	0.477	0.868	-	0.725
D	0.478	0.734	0.627	0.616	0.508	0.397	0.751	0.712	0.652	0.831	0.815	0.586
E	0.756	0.474	-	0.659	0.761	0.516	0.544	-	0.690	0.667	0.844	0.614
F	0.393	0.647	0.626	0.787	0.656	0.408	0.497	0.288	0.606	0.470	0.485	0.441
G	0.756	0.727	0.502	0.572	0.425	0.652	0.600	0.288	0.502	0.693	0.706	0.481
H	0.550	0.539	0.300	0.556	0.526	0.177	0.484	0.550	0.466	0.608	0.434	0.716
M4	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	0.784	-	0.411	-	-	0.873	-	-	-
B	-	-	0.829	0.947	0.537	0.374	-	0.955	-	0.370	-	-
C	-	0.125	-	0.797	0.555	-	-	-	-	1.000	-	-
D	0.223	0.894	0.660	0.560	0.962	-	-	-	-	-	-	-
E	-	-	-	-	-	0.773	0.220	-	-	-	-	-
F	-	-	-	-	-	-	-	-	-	0.508	-	0.117
G	0.538	0.561	1.000	0.550	0.822	0.705	-	-	-	-	0.902	0.710
H	0.639	-	-	-	0.399	0.057	0.306	-	-	-	-	0.865
O	1	2	3	4	5	6	7	8	9	10	11	12
A	-	-	-	M4	M3	M1	-	M3	M4	M3	M3	-
B	-	-	M4	M4	M3	M3	M3	M4	-	-	M3	-
C	M3	-	-	M4	M1	M1	M3	-	M1	M4	M1	M3
D	-	M4	-	M1	M4	-	M3	M3	-	M3	M3	-
E	M3	-	-	-	M3	M4	-	M2	M1	-	M3	M2
F	-	-	-	M3	-	-	-	M2	-	-	-	-
G	M3	M3	M4	-	M4	M4	-	-	-	-	M4	M4
H	-	M2	-	-	-	-	M1	-	M1	M1	-	M4