Acta Univ. Sapientiae, Informatica **12,** 1 (2020) 22–38

DOI: 10.2478/ausi-2020-0002

Formal concept analysis for amino acids classification and visualization

Adrian-Sorin TELCIAN
Babes-Bolyai University
Cluj-Napoca, Romania
email: adriant@cs.ubbcluj.ro

Daniela-Maria CRISTEA
Babes-Bolyai University
Cluj-Napoca, Romania
email: danielacristea@cs.ubbcluj.ro

Ioan SIMA
Babes-Bolyai University
Cluj-Napoca, Romania
email: sima.ioan@cs.ubbcluj.ro

Abstract. Formal concept analysis (FCA) is a method based on lattice theory, widely used for data visualization, data analysis and knowledge discovery. Amino acids (AAs) are chemical molecules that constitute the proteins. In this paper is presented a new and easy way of visualizing of the structure and properties of AAs. In addition, we performed a new Hydrophobic-Polar classification of AAs using FCA. For this, the 20 proteinogenic AAs were clustered, classified by hydrophobicity and visualized in Hasse-diagrams. Exploring and processing the dataset was done with Elba and ToscanaJ, some FCA tools and Conceptual Information System (CIS).

Formal concept analysis (FCA) is a method based on lattice theory and is used for data analysis, knowledge representation, information retrieval and knowledge discovery [12]. FCA is a semantic technology that targets a formalization of concepts for human understanding [3].

Computing Classification System 1998: F.4.1
Mathematics Subject Classification 2010: 03G10
Key words and phrases: formal concepts, lattices, proteins, amino acids, scales, organic compounds

FCA offers efficient algorithms for data analysis and data detection of hidden dependencies. It also makes it easy for the user to visualize the information.

Over the past decades, biological information has risen exponentially. The analysis and interpretation of these data remain a challenge for researchers [12].

The amino acids (AA) are monomers that form the proteins. From the hundreds of amino acids found in living organisms, only 20 AAs take part in the protein buildings. These are called proteinogenic AAs.

Biological data analysis is usually quantitative and based on mathematical statistics. A qualitative method based on Formal Concept Analysis (FCA) is used in this document. As far as we know, FCA has not been applied for the AAs study.

Using Elba and ToscanaJ, some known FCA tools, the 20 amino acid characteristics and hydrophobicity are evaluated and analyzed. FCA uses crosstables (a representation of the "formal contexts"), where rows represent amino acids, columns represent attributes of amino acids, and cells contain information about attribute values (i.e., number of atoms of molecules and other properties). The AAs are clustered into meaningful sets. The clusters, which form a hierarchy, are visually displayed in the Hasse diagrams [3].

FCA has been used for the following:

1. the automated classification of enzymes [2]; the authors used supervised and unsupervised classification, obtaining a correct classification for more than 50% percent from analyzed sequences.
2. knowledge identification, knowledge acquisition, knowledge development, knowledge distribution and sharing, knowledge usage, and knowledge sustainability concepts in the economic field. [17];
3. class hierarchy design in object-oriented programming (OOP) [6];
4. analysis, conception, implementation and validation of class (or object) hierarchies and component retrieval in the field of software engineering (SE) [7];
5. business intelligence (BI) as framework technologies that meaningfully reduce space of OLAP cube on a hierarchy of attributes [10];
6. membership constraints, a problem of consistency to determine if a formal concept exists whose object and attribute set include certain elements and exclude others; [15];
7. modeling and querying with a conceptual graph of data from RDBMS and XML databases. [11].

1 Formal concept analysis

Formal Concept Analysis (FCA) is a field of Applied Mathematics based on formalizing concepts and conceptual hierarchies from a lattice-theoretical perspective. FCA offers algorithms for data analysis and detection of hidden dependencies from sets of data and does this in a highly efficient manner. Representing data in FCA is done in the form of formal contexts, which is the easiest way of specifying what attributes are valid for which objects. Doing so, it makes data visualization an easily understandable way.

Some definitions are needed to understand the FCA.

Formal context - A *formal context* is a triplet (X, Y, I) where X and Y are non-empty sets, and I is a binary relation between X and Y, i.e., $I \subseteq X \times Y$.

In a formal context, items $x \in X$ are called objects and items $y \in Y$ are called attributes. $(x, y) \in I$ shows that object x has a y attribute.

Formal context is represented as a cross-table with n rows and m columns. The corresponding formal context consists of a set $X = \{x_1, x_2, ..., x_n\}$, a set $Y = \{y_1, y_2, ..., y_m\}$, and a binary relation I defined by: $(x_i, y_j) \in I$ if the table entry corresponding to row i and column j contains "$\times$" (see Fig. 1).

I	y_1	y_2	y_3	y_4
x_1	×	×	×	×
x_2	×		×	×
x_3		×	×	×
x_4		×	×	×
x_5	×			

Figure 1: The cross-table corresponding to the formal context

Components of X are known as objects and refer to table rows, components of Y are known as attributes and refer to table columns, and for $x \in X$ and $y \in Y$, $(x, y) \in I$ suggests that object x has an attribute y, while $(x, y) \notin I$ implies that x does not have attribute y. For example, Fig. 1 displays a cross-table (aka logical attribute table) corresponding to triplet (X, Y, I), given by $X = \{x_1, x_2, x_3, x_4, x_5\}$, $Y = \{y_1, y_2, y_3, y_4\}$ and $I = \{(x_1, y_1), (x_1, y_2), ..., (x_2, y_1), \ldots, (x_5, y_1)\}$. Notice that $(x_1, y_1) \in I$, whereas $(x_2, y_2) \notin I$, etc.

In a formal context, always we have a pair of operators, called concept-forming operators.

Concept-forming operators - For a formal context (X, Y, I), operators $\uparrow: 2^X \to 2^Y$ and $\downarrow: 2^Y \to 2^X$ are defined for every $A \subseteq X$ and $B \subseteq Y$ by

$A^{\uparrow} = \{y \in Y \mid \text{for each } x \in A : (x, y) \in I\}$,

$B^{\downarrow} = \{x \in X \mid \text{for each } y \in B : (x, y) \in I\}$.

In FCA, the notion of a formal concept is essential, they are specific clusters in cross-tables defined through attribute sharing.

Formal concept - A formal concept in the formal context (X, Y, I) is a pair (A, B) of $A \subseteq X$ and $B \subseteq Y$ such that $A^{\uparrow} = B$ and $B^{\downarrow} = A$.

For a formal concept (A, B) in the formal context (X, Y, I), A and B are called the extent and intent of (A, B), respectively. Formal concepts can be described as: (A, B) is a formal concept iff A includes only objects sharing all attributes from B and B contains only attributes shared by all objects from A.

The formal concept as term can be seen as a mathematization of a well-known idea that evokes Port-Royal logic. In that logic, a concept is determined by a collection of objects (extent) that fall under the concept and a collection of attributes (intent) covered by the concepts. In the cross-table from Fig. 1 we can see some formal concepts. Thus, $(\{x_1\}, \{y_1, y_2, y_3, y_4\})$ is a formal concept with extent $\{x_1\}$ and intent $\{y_1, y_2, y_3, y_4\}$.

Concepts are usually ordered based on the subconcept-superconcept relation which is based on the inclusion relation defined on objects and attributes.

The subconcept-superconcept relationship is formally defined as follows:

Subconcept-superconcept ordering - For formal concepts (A_1, B_1) and (A_2, B_2) of formal context (X, Y, I), put $(A_1, B_1) \leq (A_2, B_2)$ iff $A_1 \subseteq A_2$ and $B_2 \subseteq B_1$.

A concept lattice, another fundamental notion in FCA, is the collection of all formal concepts of a specified formal context.

Concept lattice - Denote by $\beta(X, Y, I)$ the collection of all formal concepts of formal context (X, Y, I), i.e.

$\beta(X, Y, I) = \left\{(A, B) \in 2^X \times 2^Y \mid A^{\uparrow} = B, B^{\downarrow} = A\right\}$.

$\beta(X, Y, I)$ equipped with the $\leq$ subconcept-superconcept ordering is called a (X, Y, I) concept lattice.

$\beta(X, Y, I)$ is all (potentially interesting) clusters "hidden" in (X, Y, I) information.

According to Main theorem of concept lattices [16], $(\beta(X, Y, I), \leq)$ is a complete lattice.

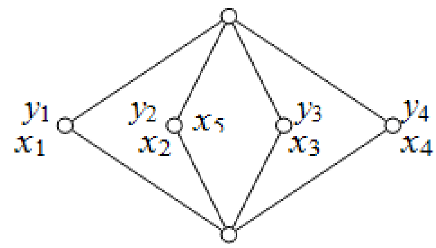
Denote the extent of concepts:

$\mathrm{Ext}(X, Y, I) = \left\{A \in 2^X \mid (A, B) \in \beta(X, Y, I) \text{ for some } B\right\}$

and intents of concepts:

	y_1	y_2	y_3	y_4
x_1	×			
x_2		×		
x_3			×	
x_4				×
x_5		×		

Formal Context



Concept Lattice

Figure 2: The concept lattice corresponding to the cross-table (formal context)

$Int(X, Y, I) = \{B \in 2^y \mid (A, B) \in \beta(X, Y, I)$ for some $A\}$.

In Fig. 1 is showed the concept lattice for a formal context represented by the cross-table. On the lattice the next formal concepts are extracted: $(\{x_1\}, \{y_1\})$; $(\{x_2, x_5\}, \{y_2\})$; $(\{x_3\}, \{y_3\})$ and $(\{x_4\}, \{y_4\})$.

Conceptual Scaling

Especially at the data collected from practical applications, the attributes no longer have binary values, of the type *yes / no*. These situations are formalized in the so-called *many-valued context*, in which a particular object has a certain attribute with a certain value.

Many-valued context - A *many-valued context* (G, M, W, I) consists of sets G, M and W and a ternary relation I between G, M and W (i.e., $I \subseteq G \times M \times W$) for which it holds that $(g, m, w) \in I$ and $(g, m, v) \in I$ always implies $w = v$.

The elements of G are called objects, those of M (many-valued) attributes and those of W attribute values. $(g, m, w) \in I$ is read as ”the attribute m has the value w for the object g”.

The many-valued attributes can be regarded as partial maps from G in W. Therefore, it seems reasonable to write $m(g) = w$ instead of $(g, m, w) \in I$.

The *many-valued context* is transformed into a *one-valued* one, trough process called *conceptual scaling* that is not at all uniquely determined.

In the process of scaling, first of all each attribute of a *many-valued context* is interpreted by means of a context. This context is called *conceptual scale*.

Conceptual scale - A *conceptual scale* is a *scale* for the attribute m of a

	⩽1	⩽2	⩽3	⩽4	⩾1	⩾2	⩾3	⩾4
1	×	×	×	×	×			
2		×	×	×	×	×		
3			×	×	×	×	×	
4				×	×	×	×	×

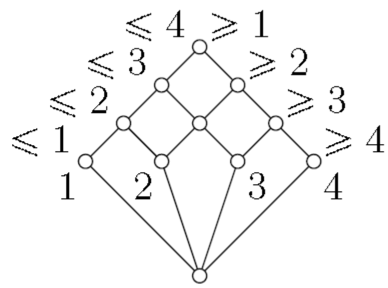


Figure 3: Interordinal scale

many-valued context. Thus, is a *one-valued* context $S_m = (G_m, M_m, I_m)$ with $m(G) \subseteq G_m$. The objects of a scale are called *scale values*, the attributes are called *scale attributes*.

The most commonly used scales also called *elementary scales* are nominal scales, ordinal scales, interordinal scales, biordinal scales and dichotomic scales. An example of interordinal scale is point out in Fig. 3.

ToscanaJ Suite

There is many software for FCA and most of them support the creation of contexts from scratch and the subsequent processing and presentation of the corresponding concept lattices. More than that, Elba and ToscanaJ are a set of mature FCA tools that allow us to query and navigate through data in databases. They are meant to be a *Conceptual Information System (CIS)*.

When implementing a CIS using methods of FCA, the data is modeled mathematically by a *many-valued context* and is transformed via *conceptual scaling* [5]. This means that is defined a formal context called *conceptual scale* for each of the many-valued attributes which has the values of the attribute as objects. Here, a CIS is an FCA-based system used to analyze data from one table.

Creating the *conceptual scale* is realized with a CIS editor (Elba) and usually is a highly iterative task. In the run-time stage, a CIS browser (ToscanaJ) allows us to explore and analyze the real data from the database with the CIS schema.

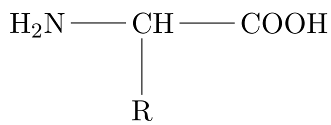
2 Amino acids – structure and properties

The monomers from which the proteins are created are called amino acids (AAs). The proteins are a large family of organic macromolecular compounds.

Only 20 AAs of the hundreds of AAs (which are found in the living organisms) take part in the protein assembly.

Chemically, AAs are organic molecules containing two types of antagonistic functional groups: carboxyl ($-COOH$) having an acid character; amine ($-NH_2$) having a base character. They also contains a side chain (R group or AA residue) specific to each amino acid [4].

The molecular formula for AA is:



All AAs can be classified according to different properties: hydrophobicity, R type, etc.

Hydrophobicity is the molecule's physical property to be repelled by water molecules. In fact, there is no repulsive force, but only the absence of attraction. In contrast, the hydrophilic (or polar) molecules are attracted to water molecules. Several hydrophobicity scales have been developed over time. Note that AA's hydrophobicity is crucial for understanding the protein folding. [1, 9, 14].

Table 1 presents the structural and functional aspects of AAs. All 20 AAs are composed of carbon (C), hydrogen (H), nitrogen (N) and oxygen (O) atoms. Moreover, Cysteine and Methionine contain Sulphur (S) in addition to the other AAs. The first column shows the 3-letter name for AAs (biochemistry nomenclature). From two to six column there are carbon, hydrogen, nitrogen, oxygen and sulphur number of atoms, respectively. The last column returns the hydropathy index of AA residues. [9]. Hydropathy index is a measure of relative hydrophobicity. A higher hydropathic index value means more hydrophobic amino acid.

In literature, there are four known hydrophobicity scales taken from [20] (shown in Fig. 4). The most hydrophobic AA residues are at the top of the figure. References for the scales are: (1) Kyte and Doolittle [9]; (2) Rose, et al [13]; (3) Wolfenden, et al.[18]; and (4) Janin [8].

3 Experiments and results

FCA is suitable for binary attributes. In the real data type issue, the situation is slightly different. In our case, each attribute is assigned integer (numbers of atoms) or real values (hydropathic index), not binary values. A method called

AA	**C**	**H**	**N**	**O**	**S**	**Hydropathy index**
Gly	2	5	1	2	0	-0.4
Ala	3	7	1	2	0	1.8
Val	5	11	1	2	0	4.2
Leu	6	13	1	2	0	3.8
Ile	6	13	1	2	0	4.5
Phe	9	11	1	2	0	2.8
Pro	5	9	1	2	0	-1.6
His	6	9	3	2	0	-3.2
Trp	11	12	2	2	0	-0.9
Ser	3	7	1	3	0	-0.8
Thr	4	9	1	3	0	-0.7
Tyr	9	11	1	3	0	-1.3
Cys	3	7	1	2	1	2.5
Met	5	11	1	2	1	1.9
Asp	4	7	1	4	0	-3.5
Asn	4	8	2	3	0	-3.5
Glu	5	9	1	4	0	-3.5
Gln	5	10	2	3	0	-3.5
Lys	6	14	2	2	0	-3.9
Arg	6	14	4	2	0	-4.5

Table 1: List of proteinogenic amino acids

conceptual scaling is used for the FCA application to these data. Conceptual scaling converts the many-valued context into a standard formal context.

An advantage of FCA is that there is no standard attributes interpretation. The field expert chooses a suitable scale for attributes interpretation [3]. In our approach, for another ways of visualization of properties and classification of AAs, we used data relating to the molecular structure of the 20 AA, i.e. atoms number of the molecule. For data representation we used Elba and ToscanaJ tools.

Visualization 1. For transformation from many-valued context to one-valued context interordinal scale is used. In Figs. 5–8 can be visualized a clustering of AAs by the number of atoms from molecules.

In Fig. 5 we can see that have been retrieved 7 formal concepts. Thus, 5% from those 20 AAs (meaning 1 AA) have 11 carbon atoms, 10% (meaning 2 AAs) contains 9 carbon atoms (≥ 9 and < 10), and so forth.

Kyte and Doolittle (1)	Rose, et al (2)	Wolfenden , et al (3)	Janin (1979) (4)
Ile	Cys	Gly,Leu,Ile	Cys
Val		Val,ala	Ile
	Phe,Ile		Val
Leu	Val	Phe	Leu,Phe
	Leu,Met,Trp	Cys	Met
Phe		Met	Ala,Gly,Trp
Cys			
Met,Ala	His	Thr,Ser	
	Tyr	Trp,Tyr	His,Ser
Gly	Ala		Thr
Thr,Ser	Gly		Pro
Trp,Tyr	Thr		Tyr
Pro			Asn
		Asp,Lys,Gln	Asp
His	Ser	Glu,His	Gln,Glu
Asn,Gln	Pro,Arg	Asp	
Asp,Glu	Asn		
Lys	Gln,Asp,Glu		
			Arg
Arg	Lys	Arg	Lys

Figure 4: A comparison of four distinct scales for the hydrophobicity

In Fig. 7 notice 3 formal concepts: 65% monocarboxylic AAs (2 oxygen atoms), 10% AAs with 4 oxygen atoms, and 25% AAs have 3 oxygen atoms.

Similar information can also be extracted from Figs. 6 and 8. This is a faster way to visualize complex information than in the tables.

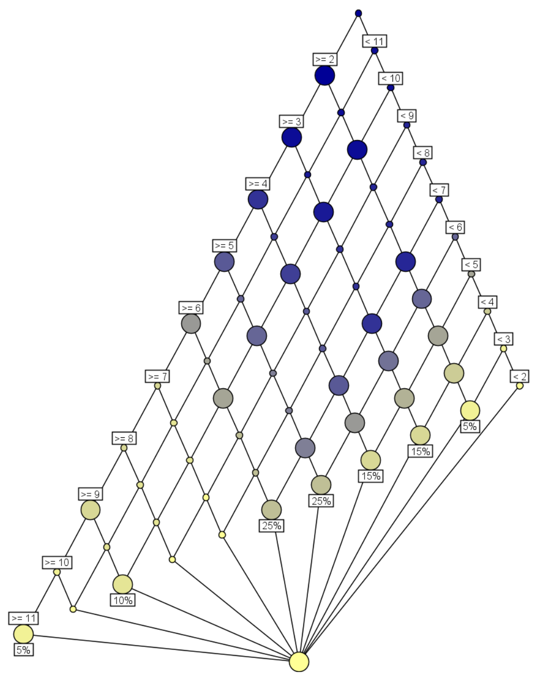


Figure 5: Diagram for number of Carbon atoms

Visualization 2. Subsequent, we used two attributes: the number of carbon atoms and the number of oxygen atoms from AA molecules and created a new scale based on these. The scaling was done as follows. AAs that have more than 3 atoms of the oxygen we called *AA Dicarboxylic*. The according to the number of carbon atoms distinguish: 1). *AAs Low* - AAs with less than 4 carbon atoms; 2). *AAs High* - AAs with more than 5 carbon atoms; and 3). *AAs Medium* - the others;

In Fig. 9 the concept lattice shows number of AAs belong to the abovementioned levels. For instance, we can read from the lattice that there are two AA Dicarboxylic, and these belong *AA Medium* group. This is one of the main distinguishing characteristics of using concept lattices to visualize information.

Visualization 3. A strong method of FCA is to "mix" many lattices together to provide a combined perspective of several lattices, called a nested line diagram. Fig. 10 displays a mix of diagrams from Fig. 11 and Fig. 9.

Hydrophobic-Polar Classification In addition, for classification, we have used data relating to the hydropathy index (*hi*). AAs classification in hydrophobic (H) and polar (P) is important for some protein folding models.

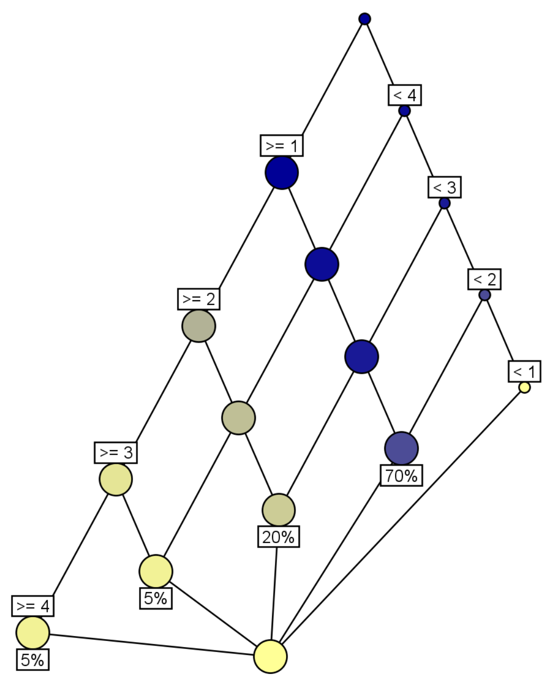


Figure 6: Diagram for number of Nitrogen atoms

Unfortunately, there is no single classification. For example, we have the ones four scale from Fig. 4.

Clustering algorithms most frequently used are: hierarchical, k-means, self-organizing maps, fuzzy k-means [12]. An alternative approach to grouping AAs can be FCA. The lattice concepts of AAs are assumed to show new or old biological relationships.

Relating to hydropathy index (*hi*) we have defined four hydrophobicity levels of AAs: i) if $\mathtt{hi} < -3$: *Polar with electrically charged propensity*; ii) if $\mathtt{hi} < -1.5$: *Polar*; iii) if $\mathtt{hi} < 0$: *Uncertain*; iv) if $\mathtt{hi} \leq 4.5$: *Hydrophobic*. Fig. 11 shows the concept lattice of this scale.

Our classification are presented in Table 2, column 4, compared to the classification taken from [4], page 26 (column 2) and Rosalind [19] (column 3).

Our classification:

Hydrophobic AAs: Ala, Val, Leu, Ile, Phe, Cys, Met.

Polar AAs: Pro, Asp, Glu, Arg, Lys, His, Asn, Gln.

The other five AAs (Gly, Trp, Ser, Thr and Tyr) it remains to be classified according to other criteria.

AA	**Dinu**	**Rosalind**	**FCA**
H	Ala, Val, Leu, Ile, Pro Phe, Trp, Met	Ala, Val, Leu, Ile, Phe Trp, Met, Tyr	Ala, Val, Leu, Ile, Phe Cys, Met
P neutral	Ser, Thr, Tyr, Cys, Asn Gln, Gly	Ser, Thr, Asn, Gln	Pro
P charged	Asp, Glu, Arg, Lys, His	Asp, Glu, Arg, Lys, His	Asp, Glu, Arg, Lys, His, Asn, Gln
Uncertain	-	Gly, Cys, Pro	Gly, Trp, Ser, Thr, Tyr

Table 2: HP AAs classification

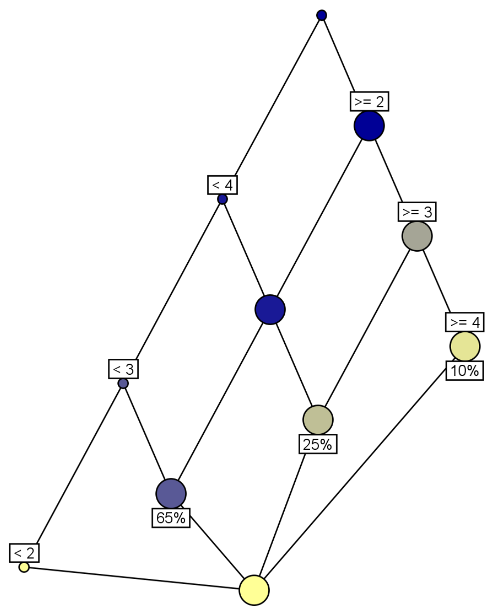


Figure 7: Diagram for number of Oxygen atoms

It can be notice that the classes of AAs identified by applyed FCA, taking into account that hydropathy index are similar to those in the classification taken from Dinu and Rosalind respectively.

Of those seven "Hydrophobic" AAs found, six are found in the same class and in the other two classifications. The seventh AA, Cysteine (Cys), is considered "Polar neutral" AA in Dinu, and in Rosalind it is considered a special case.

Through FCA, we find a single "Polar neutral" AA: Proline (Pro). In Rosalind, this AA is classified in special cases and is considered a hydrophobic AA by Dinu.

In the "Polar charged" class were found the five AAs from the Dinu and Rosalind classifications. Additionally, by our method, we found two new AAs in this class: Asn and Gln. Both AAs are classified as "Polar neutral" AAs in both Dinu and Rosalind.

In contrast, in our classification, there are 5 AAs difficult to classify, which we called "Uncertain AAs". The characteristic of the classification using FCA is that it is sensitive to defining hydrophobicity levels.

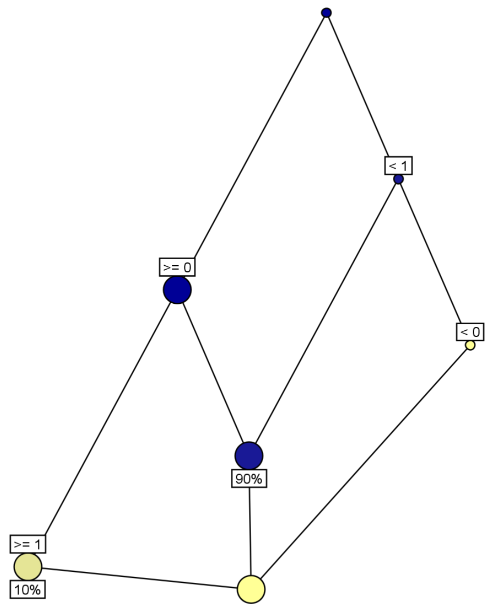


Figure 8: Diagram for number of Sulphur atoms

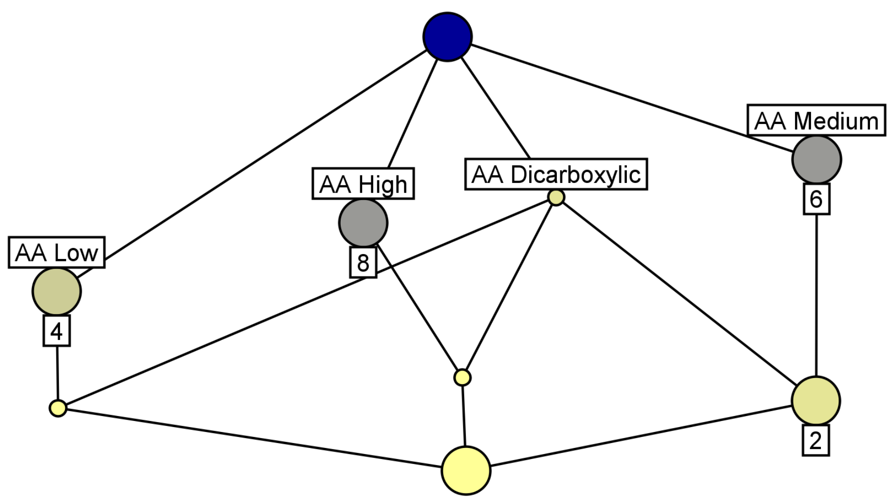


Figure 9: Diagram of the scale based on number of carbon and oxygen atoms from AAs

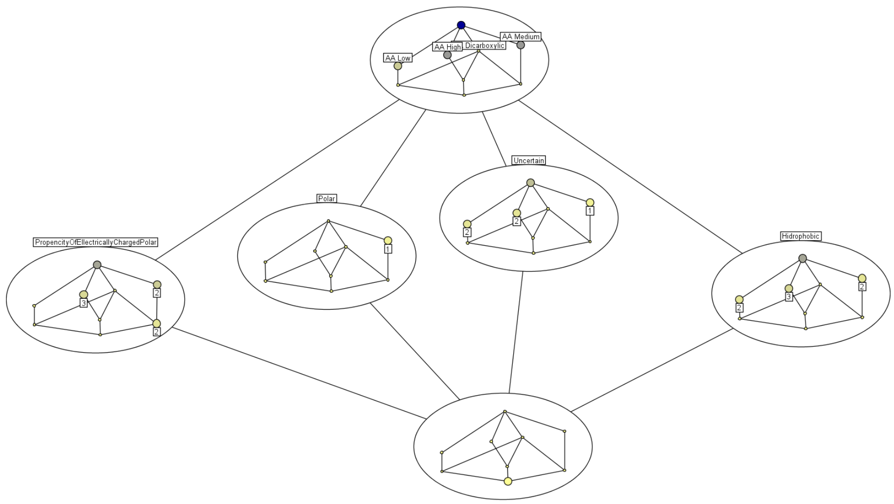


Figure 10: Nested diagram

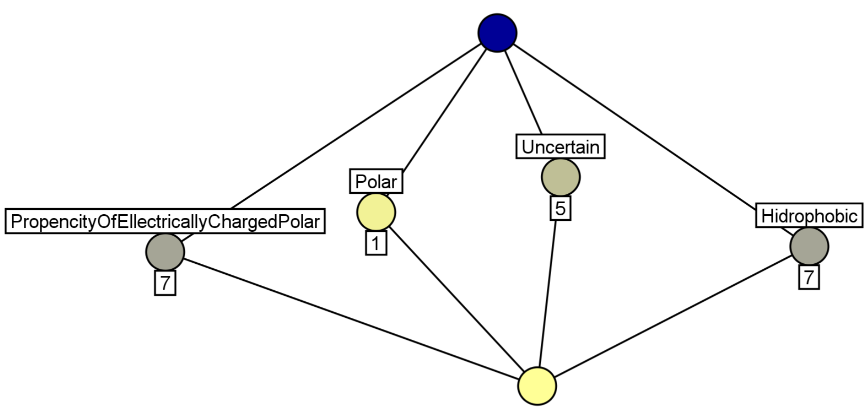


Figure 11: Hydrophobicity scale

Finally, it is difficult to say whether this method adds to the other types of classifications because the hydrophobicity depends on the physical and chemical conditions in which the measurement was taken.

4 Conclusion

In the presented paper, we relied on lattice theory commonly used to analyze and visualize data with formal concept analysis (FCA).

The purpose of this work was to realize a new Hydrophobic-Polar classification and to visualize structure information of AAs in perspective of the Hasse diagrams. Elba and ToscanaJ tools were used to process and explore the dataset.

We defined the hydrophobicity index based on the Kyte and Doolittle scale. In the future, AAs can be classified considering others three known scales of hydrophobicity: Rose, Wolfenden and Janin.

Acknowledgements

The authors would like to thank Prof. Univ. Dr. Bazil Pârv, our research project supervisor, for his professional guidance and valuable support.

References

[1] D. Bandyopadhyay, E. L. Mehler. Quantitative expression of protein heterogeneity: Response of amino acid side chains to their local environment. *Proteins: Structure, Function, and Bioinformatics*, **72,** 2 (2008) 646–659. ⇒28

[2] F. Coste, G. Garet, A. Groisillier, J. Nicolas, T. Tonon, Automated enzyme classification by formal concept analysis, In C. V. Glodeanu, M. Kaytoue, and C. Sacarea, editors, *Formal Concept Analysis*, **8478**, pp. 235–250, Cham, 2014. Springer International Publishing. ⇒23

[3] F. Dau, B. Sertkaya, Formal concept analysis for qualitative data analysis over triple stores, In O. De Troyer, C. Bauzer Medeiros, R. Billen, P. Hallot, A. Simitsis, H. Van Mingroot, editors, *Advances in Conceptual Modeling. Recent Developments and New Directions. ER 2011. Lecture Notes in Computer Science*, **6999**, pp. 45–54, Berlin, Heidelberg, 2011. Springer Berlin Heidelberg. ⇒22, 23, 29

[4] V. Dinu, E. Trutia, E. Popa-Cristea, A. Popescu, *Medical Biochemistry – small treated (in Romanian)*. Ed. Medicală, București, 2006. ⇒28, 32

[5] B. Ganter, R. Wille, Conceptual scaling. In F. Roberts, editor, *Applications of Combinatorics and Graph Theory to the Biological and Social Sciences*, pp. 139–167, Berlin–Heidelberg–New York, 1989. Springer. ⇒27

[6] R. Godin, P. Valtchev, Formal concept analysis-based class hierarchy design in object-oriented software development, In B. Ganter, G. Stumme, and W. R., editors, *Formal Concept Analysis. Lecture Notes in Computer Science*, **3626**, pp. 304–323. Springer Berlin Heidelberg, 2005. ⇒23

[7] W. Hesse, T. Tilley, Formal concept analysis used for software analysis and modelling, **3626**, In *Formal Concept Analysis Used for Software Analysis and Modelling*, pp. 288–303. Springer-Verlag, Berlin, Heidelberg, 2005. ⇒23

[8] J. Janin, Surface and inside volumes in globular proteins, *Nature*, **277** (1979) 491–492. ⇒28

[9] J. Kyte, R. F. Doolittle, A simple method for displaying the hydropathic character of a protein, *Journal of Molecular Biology*, **157,** 1 (1982) 105–132. ⇒ 28

[10] J. Macko, Formal concept analysis as a framework for business intelligence technologies II, In *CUBIST Workshop*, 2013. ⇒23

[11] A. Molnar, V. Varga, C. Săcărea, D. Cîmpan, B. Mocian, Conceptual graph driven modeling and querying methods for RDBMS and XML databases, In *2017 13th IEEE International Conference on Intelligent Computer Communication and Processing (ICCP)*, pp. 55–62, 09 2017. ⇒23

[12] K. Raza. Formal concept analysis for knowledge discovery from biological data, *International Journal of Data Mining and Bioinformatics*, **18,** 4 (2017) 281–300. ⇒22, 23, 32

[13] G. D. Rose, A. Geselowitz, G. Lesser, R. Lee, M. Zehfus, Hydrophobicity of amino acid residues in globular proteins, *Science*, **229,** 4716 (1985) 834–838. ⇒ 28

[14] G. D. Rose. R. Wolfenden, Hydrogen bonding, hydrophobicity, packing, and protein folding. *Annual Review of Biophysics and Biomolecular Structure*, **22,** 1 (1993) 381–415. PMID: 8347995. ⇒28

[15] S. Rudolph, C. Săcărea, D. Troancă, Membership constraints in formal concept analysis, In *Proc. 24th Int. Conf. on Artificial Intelligence*, IJCAI'15, pp. 3186–3192. AAAI Press, 2015. ⇒23

[16] R. Wille, Restructuring lattice theory: An approach based on hierarchies of concepts, In I. Rival (ed.), *Ordered Sets*, pp. 445–470, Dordrecht, 1982. Springer Netherlands. ⇒25

[17] R. Wille, Conceptual knowledge processing in the field of economics, In W. R. Ganter B., Stumme G. (eds.), *Formal Concept Analysis. Lecture Notes in Computer Science*, **3626,** pp. 226–249. Springer Berlin Heidelberg, 2005. ⇒23

[18] R. Wolfenden, L. Andersson, P. Cullis, C. Southgate, Affinities of amino acid side chains for solvent wate, *Biochemistry*, **20** (1981) 849–855. ⇒28

[19] ∗ ∗ ∗ Rosalind project, `http://rosalind.info/glossary/amino-acid/`, 2019. ⇒32

[20] ∗ ∗ ∗ Wikipedia: Hydrophobicity scales, `https://en.wikipedia.org/wiki/hydrophobicity_scales`, 2019. ⇒28

Received: November 11, 2019 • Revised: January 6, 2020