J IRAN CHEM SOC
DOI 10.1007/s13738-017-1158-1

ORIGINAL PAPER

Time–frequency approach in the cluster assignment of amino acids based on their NMR profiles

Abbas Shockravi[1] · Kaveh Kavousi[2] · Jafar Rezania[1] · Reza Jafari[1] · Mohammad Hossein Norouzi Beirami[2] · Shohreh Ariaeenejad[2,4] · Zahra Moosavi-Movahedi[2] · Parvaneh Maghami[2] · Amir M. Mortazavian[3] · Ali Akbar Moosavi-Movahedi[2]

Received: 29 October 2016 / Accepted: 12 June 2017

Abstract The "inversion recovery experiment" is used as a tool in nuclear magnetic resonance spectroscopy for characterization and identification of basic important amino acids. This methodology is based on determination of "longitudinal relaxation time (T_1)"of Carbon-13 of these molecular structures which relates to assignment of each carbon atom of amino acids. Not only the chemical shifts and frequencies of carbon atoms are different, but also the relaxation times of them in scale of seconds or less are different, without much overlapping. Due to larger shift effects in ^{13}C NMR spectra and larger paramagnetic origin for carbon-13 and other differences, it has made us to rely on ^{13}C nucleus as main clue in this work rather than ^{1}H nucleus. This procedure has helped us to identify the amino acids in terms of both "frequency" and "time" of relaxation for each carbon atom simultaneously. Applying "average linkage" as an agglomerative clustering method to the feature vectors extracted from NMR spectra of amino acids, a hierarchical clustering is provided. The obtained clusters reveal notable relationships between amino acids in a same cluster. After a time gap, the proposed clusters of amino acids which have similarities and differences with traditional grouping of amino acids provide a new perspective on amino acids characterization and related studies such as defining descriptors for proteins and peptides based on their sequence information.

Keywords Amino acids · Nuclear magnetic resonance · Relaxation time · Spectroscopy · Hierarchical clustering

Introduction

Amino acids are the main important components of proteins and are considered as repeating units in all living organisms. Approximately 20 amino acids are common constituents of proteins. There are numbers of researchers who have been involved in the isolation and characterization of amino acids since 1820 [1]. After World War II various analytical methods were improved for the isolation and characterization of these compounds. Microbiological assay using a lactic acid bacterium, chromatographic separation, ion-exchange resin chromatography and specially gas chromatography in the determination of racemic mixtures of D & L forms [2] are some examples of these methods. Acid or alkaline hydrolysis [3] and fermentation processes were other important methods in determination and production of amino acids, respectively. Physical and chemical properties of each amino acid are mostly affected by amino and carboxyl groups with opposite charges at the chiral (asymmetric) α-carbon atom. All amino acids except glycine have such carbon atom which generates enantiomeric isomerization and racemization phenomena in them. The X-ray crystallographic analyses of 23 amino acids

Electronic supplementary material The online version of this article (doi:10.1007/s13738-017-1158-1) contains supplementary material, which is available to authorized users.

✉ Abbas Shockravi
abbas_shockravi@yahoo.co.uk

[1] Faculty of Chemistry, Kharazmi University, Tehran, Iran

[2] Institute of Biochemistry and Biophysics (IBB), University of Tehran, Tehran, Iran

[3] Department of Food Science and Technology, Faculty of Nutrition Sciences, Food Science and Technology, Shahid Beheshti University of Medical Sciences and Health Services, Tehran, Iran

[4] Agricultural Biotechnology Research Institute of Iran (ABRII), Karaj, Iran

Published online: 22 June 2017

proved that crystal shapes of each of them vary widely. Raman spectra of amino acids were used to provide the first direct optical evidences for the dipolar structure. The IR studies were shown the overtone region on the NH-stretching frequencies in the neutral amino acids [4, 5]. Generally, the IR analysis does not lead us to precise assignment for amino acids.

A number of known correlation methods in NMR spectroscopy are important tools which are currently used in the structural analyses. Most of the others depend in one way or another on "conformation" and the way it affects "chemical shifts" and especially "coupling constants" [6]. A variation of the NMR method is to deduce configuration from the sense of shift produced in corresponding nuclei (^{1}H, ^{13}C, etc.) of enantiomers under the influence of chiral solvating agents (CSA) or upon conversion to one of a pair of diastereomers by combination with a chiral, enantiomerically pure reagent [7]. Groups may also be enantiotopic by external comparison that is groups in two different molecules are enantiotopic if they are related by reflection symmetry. Clearly, this can be so only if the two molecules themselves are enantiomeric such as the methyl groups in D- and L-alanine which are enantiotopics.

It is known that the anisochronous nuclei are much larger in ^{13}C than ^{1}H resonance; presumably because of the generally larger shifts effects in ^{13}C NMR spectra due to their larger paramagnetic origin [6], such facts have made us to consider ^{13}C nucleus as main clue in our work rather than ^{1}H nucleus. Valine is an amino acid which has diastereotopic methyl groups resonating at 1.39 and 1.44 ppm in ^{1}H NMR spectrum. Under an enzymatic transformation of the molecule it became of importance to determine which group was which [8] and corresponding assignments with ^{13}C NMR of labeled methyl groups have also been carried out [9].

Theoretically, the thermal equilibrium of the spin system is disturbed by irradiating the resonance frequency. In this process not only the population ratios alter, but also macroscopic magnetizations occur and cause transverse magnetic field components (M_x and M_y) to appear. The spin system relaxes, when perturbation ceases and once again reaches to its original thermal equilibrium. This process takes place by two mechanisms; one is characterized by longitudinal *or spin–lattice relaxation time* (T_1), and the other one is characterized by *transversal or spin–spin relaxation time* (T_2) [10]. The relaxation of nuclear systems is very slow (in contrast to rotational, vibrational and electronic excited states) especially in cases where the nuclear spin (I) is 1/2. Complete relaxation process may take seconds, minutes or even hours. We have used the ^{13}C relaxation times rather than ^{1}H nuclei as main clue for characterization of amino acids, because the relaxation process of protons (T_1) is of the order of the second and they do not vary greatly for protons in different bonding situations, while the corresponding phenomena in ^{13}C nuclei are different quietly. As a matter of a fact, T_1 values could be considered as an important clue for amino acids, because not only T_1s vary from milliseconds in large molecules to several hundred seconds in small ones, but also much larger differences in ^{13}C-chemical shifts and T_1 from one carbon atom to another are advantages for such analyses [10]. Furthermore, proton or ^{1}H NMR spectroscopy is the exploitation of ^{1}H nuclei to study the properties and interactions of matter [11–13]. Otherwise, ^{1}H NMR spectra are complicated and it can become ineffective due to signal overlap or noise.

Over the last few years there has been considerable interest in the new methods for the rapid identification [14, 15] and enantio discrimination [16] of α-amino acids owing to their highly biological and chemical relevance. Chromatographic methods are considered the most useful for enantio discrimination of analytes [17]. Although, the development of novel protocols has received considerable attention based on multiple diverse techniques such as carbon black-chiral polymer composite [18] circulardichroism [19, 20] fluorescence [21] UV and colorimetric methods [15–23], for more precise analysis, nuclear magnetic resonance (NMR) spectroscopy has been chosen by many researchers. The most of naturally found amino acids have been exhaustively studied by NMR spectra [24, 25].

However, these studies have primarily been concerned with investigations into the specific physicochemical properties and structural features of the amino acids. NMR spectroscopy was a worthwhile analytical tool to identification and characterization of the amino acids and consequently protein and peptides [26].

In this work we will confine our attention exclusively to NMR spectroscopy for such purpose and defining a similarity metric for comparative studies. In current study it is the first time that amino acids are clustered based on features extracted from NMR spectrum. These features are obtained from NMR signals of amino acids which are expected to represent their inherent properties. Clustering of amino acids means that they should be partitioned into "good" subsets, or clusters. In most of methods for the amino acids clustering, each amino acid is represented by a vector of numerical measurements for various biophysicochemical features such as hydrophobicity, volume, surface area, polarity, charge, hydrogen bonding and potentially many other features.

For clustering of 20 amino acids in m clusters, there are $\frac{1}{m!}\sum_{i=0}^{m}(-1)^{m-i}\binom{m}{i}i^{20}$ different ways. This value is named the Stirling number of the second kind. From this huge number of clustering ways, only very limited numbers satisfy the *goodness* condition [27]. It is important to note that the goodness of a partitioning strongly is related to how to measure the distance between the

amino acid representative vectors and also to the nature of these vectors. One of the earliest efforts to cluster amino acids was that of Sneath. He used 134 different and possibly independent features and introduced a dissimilarity index for clustering amino acids. This dissimilarity measure was used as input to the average linkage method for clustering. Another different approach for clustering process was originated from the evolutionary differences between proteins. Proteins can be classified in a variety of functional categories. A number of substitutions of amino acids in protein sequences are used to compute distances or similarities between them [28]. The more substitution of one amino acid with the other, the more similarity score for those two amino acids. Based on this idea, and without using any formal clustering approach, Dayhoff and colleagues partitioned amino acids such that the members within each group are more probable to substitute with each other than with members of the other partitions. In many approaches, dissimilarities between amino acids are calculated based on polarity, polarizability, side-chain molecular weight, hydrophobicity, hydrophilicity and many other physicochemical properties. In [29], polarity indices and side-chain molecular weight were used for the first time to define distance between amino acids. Grantham considered polarity, volume and the ratio of atomic weights of non-carbon elements in rings or end groups to carbon inside chain as three chemical properties, and a weighted Euclidean distance was used as a dissimilarity measure for clustering [30]. Combining this distance and the correlation coefficients obtained from the procedure of applying linear regression on the log of the relative substitution frequencies of amino acids (introduced by McLachlan [31]) a single correlation value was calculated for each amino acid. Regardless of details, the magnitude of these coefficients was interpreted to show the significance of mentioned chemical attributes in substitution rate of amino acids in evolution. Much of studies have concentrated on clustering of physicochemical and biochemical properties of amino acids [32–36]. Among them, Saha and colleagues [37] performed one of the most recent comprehensive studies in this area. They categorized available theoretic and experimental knowledge of a variety of biophysicochemical features of amino acids, as collected in the AAindex database for known 544 amino acid indices. Using their proposed method, three high-quality indices were provided to the research community which can be used for characterizing amino acids and subsequently protein and peptide sequences.

Different extracted features and indices for amino acids have been used in many applications including protein structure prediction [34], protein subcellular localization [38], domain prediction [39], drug-target interaction prediction [40], peptide function prediction [41] and many others. The mentioned references are only some samples among a variety related of research works.

To the best of our knowledge among all the techniques used in the characterization and identification of amino acids, there are no evidences using the "relaxation times" of spin nuclei of amino acids based on ^{13}C NMR spectroscopy. This is the main reason that we have studied the relaxation times of specifically ^{13}C nuclei in series of important amino acids in this work using NMR spectroscopy. This novel vision to characterizing amino acids and defining new indices for them hopefully is supposed to use in the above-mentioned applications.

Materials and methods

Data set

Based on the Felix Bloch equations; the rate at which the equilibrium condition ($M_z = M_0$) reasserts itself, the *spin–lattice relaxation time* T_1 can be determined by equation; $dM_z/dt = -M_z - M_0/T_1$, in which in the chemical kinetically point of view $1/T_1$ is the rate constant of the relaxation process. To determine the T_1 values, we have used "inversion recovery experiment" (IRE) for all amino acids at room temperature and D_2O as solvent. In this method we recorded a series of ^{13}C NMR spectra of amino acids by using pulse sequence $180°_{x'}$–τ–$90°_{X'}$—FID. After the pulse $180°_{x'}$, M_o lies along the (−Z) against an external magnetic field ($M_z = -M_0$). During the time "τ" the system relaxes with the rate constant $k = T_1^{-1}$. Between these two stages, there is an intermediate stage in which the amount of M_z is zero ($M_z = 0$) and this is the stage at which no signal can be detected and its time "τ" is called "τ_{zero}" ($\tau = \tau_{zero}$). This stage is important for the quantitative measurement of *longitudinal relaxation* (T_1). After this stage, M_z *starts to become positive*, and finally return to the equilibrium in which after complete evolution the magnetization vector stays along the external magnetic field, in another words; $M_z = M_0$. At the time "τ_{zero}" the signal intensity is just zero called as "zero-crossing point", which is related to T_1 by equation; $\tau_{zero} = T_1 \ln 2$. The "τ_{zero}" indicates how greatly the T_1 value may vary from one carbon to another in each amino acid. Based on the signal processing carried out for each amino acid it is demonstrated that all these molecules have a different diagram produced by using the "τ_{zero}" values versus the frequencies took place. The diagrams obtained for these molecules have convinced us that such data may open a new area for identification and characterization of amino acids. One typical spectroscopic analysis is appeared in results and discussion section.

Due to precise data obtained from NMR spectroscopy, we were encouraged to cluster amino acids based on new NMR-based features. In clustering which is known as unsupervised learning, training data, of known class labels, are not available. In clustering problems, the goal is to unravel the similarities (and also dissimilarities) between objects (in this study amino acids). The similar objects are grouped together in a cluster.

Data analysis

In hierarchical clustering, the distances between pairs of sets of feature vectors are computed. In a more general scheme, the proximity between subsets of the set of features is measured and used to cluster objects. Hierarchical clustering strategies fall into two general types, divisive and agglomerative clustering methods. In agglomerative algorithms, starting from each single object as a cluster, pairs of clusters are merged to build a new level of hierarchy in a bottom-up process. Divisive methods start with all objects in one cluster and then in a recursive top down process, each cluster is broken up into two new clusters to form a new level of hierarchy. In both divisive and agglomerative methods, a tree diagram named dendrogram is formed which is an effective means of representing the produced clusters in each stage of the algorithm. Three of the main agglomerative approaches are single linkage, complete linkage and average linkage. In these linkage hierarchical methods, a set of clusters are formed based on a threshold value. At each step, the two clusters with the shortest distance are merged. Differentiation between these methods is in relation with the definition of the shortest distance. In single linkage hierarchical clustering, the dissimilarity between two clusters is defined as the least interconnecting distance (dissimilarity) between a member of one and a member of the other [42]. In the complete linkage the distance between clusters is defined as the distance between the two elements (one in each cluster) that are farthest away from each other. In the average linkage or so-called UPGMA (Unweighted Pair Group Method with Arithmetic Mean) the distance between clusters G and H is defined as average of all distances between all possible pairs of objects one in cluster G and one in H as defined by the following equation:

$$d_{\text{average}}(G,H) = \frac{1}{n_G.n_H} \sum_{i \in G, j \in H} d_{ij} \quad (1)$$

In the above equation, n_G and n_H are the cardinality of clusters G and H, respectively. Also d_{ij} is the distance between sample i from cluster G and sample j from cluster H. There are seven variable features which are used in our ^{13}C NMR spectrum analysis process for 19 naturally amino acids in terms of clustering methodology. The NMR-based extracted features are listed in Table 1.

Table 1 Features extracted from ^{13}C NMR spectrum for clustering of amino acids and their dimensionality

Feature group name	Dimension
Relative spectral power (RSP) [43]	10
Slow-wave index (SWI)-(alpha, theta and delta slow-wave index) [44]	3
Harmonic parameters (HP)	15
Hjorth (activity, mobility, complexity) [45]	3
Entropy [46]	1
Midpoint of the peak of α-carbon and the next peak (this study)	1
Midpoint of the peak of α-carbon and the previous peak (this study)	1
Integral (the area under the signal curve) (this study)	1
Standard deviation, skewness, kurtosis [47]	1
Longitudinal relaxation time (T_1) (this study)	1

Experimental section

T_1-measuring of all amino acids is carried out in D_2O as solvent, and their almost saturated solutions are prepared for such process. The Bruker-AVANCE 300 NMR and pulse program *t1irpg* (T_1 measurement using inversion recovery with power gated decoupling) were used. Because of different solubility of these acids in D_2O, number of scanning was also different for better resolution purposes, especially for those which had lower solubility. Tyrosine was not only insoluble in D_2O, but also was not soluble in other polar solvents such as DMSO, methanol, acetone-d6 and $CDCl_3$. Due to extremely low solubility of L-Tyrosine the related data were excluded from our analysis and related features are not considered neither in Table 1 nor in hierarchical clustering. It is also known that relaxation time of carbons in each amino acid is different from others and sometimes very close to each other. To overcome such problems and approaching better resolution and detection, we carried on numbers of recovery delay time (RDT) program in the T_1-measurements for each acid. The RDT data for the L-Histidine are summarized in Table 2. The RDT data for other amino acids are illustrated in supplementary Table S1.

A 180° pulse is applied at the beginning of pulse sequence with a recycle delay which inverts the magnetization. The recovery delay follows to allow varying of T_1 relaxation. The final 90° pulse then converts any Z-magnetization into observable transverse magnetization which is detected during the acquisition period immediately following the final pulse.

Table 2 Times (T_1)–frequencies correlations in L-Histidine

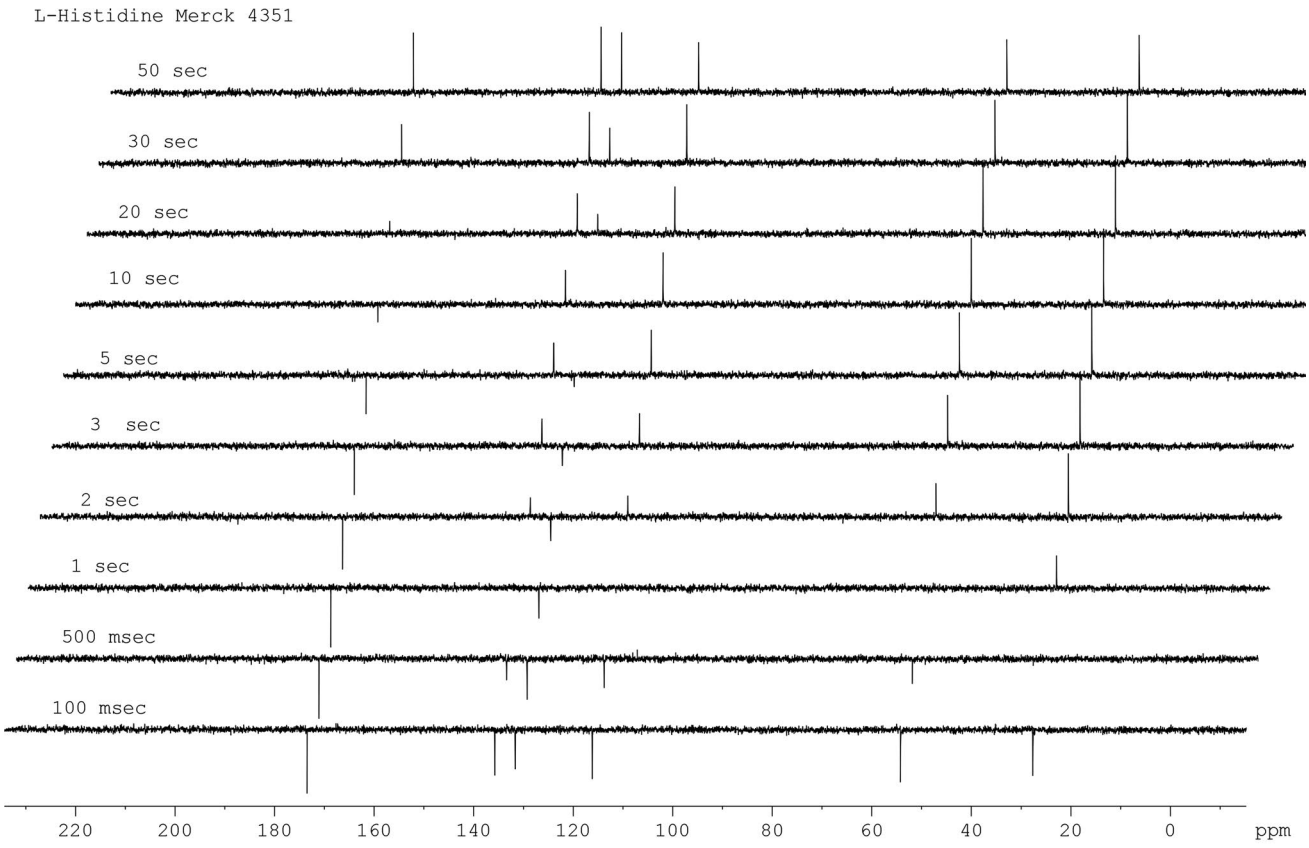
L-Histidine Carbon no.	Time (s)	Chem. shift (ppm)	Frequency (Hz)
1	0.779	27.788	2097.1604
2	1.371	54.350	4101.7945
3	1.682	116.218	8770.9725
4	18.5	131.692	9938.7952
5	1.730	135.805	10249.203
6	19	173.699	13109.064

RDT (s) 50–30–20–10–5–3–2–1–0.5–0.1

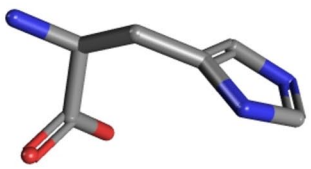
Results and discussion

A typical analysis using this methodology is shown in Fig. 1 for L-Histidine as one of the amino acids. As it can be seen, one axis is based on the "time" in which "zero-crossing" has occurred and the other axis is defined as "frequency" (in ppm or Hz) at which this phenomenon takes place. These experiments for all other 18 amino acids are appeared in supplementary Table S2.

As mentioned in the introduction section, different clustering methods have been applied to group amino acids



(a)



(b)

Fig. 1 **a** Inversion recovery experiment on the L-Histidine as a model study. **b** 3D structure of the molecule

Table 3 Amino acids distribution in three clusters proposed by Stanfel [48]

Cluster No.	Amino acids	Hhdrophobic Polar Positively Charged Negatively Charged
1	A,C,G,I,L,M,P,S,T,V	
2	D,E,N,Q	
3	F,W,Y,H,K,R	

based on their properties. One of the most accepted groupings of amino acids proposed by Stanfel [48] is shown in Table 3. He introduced an information theoretic objective function for optimizing the clustering process and minimized it by converting the problem to an integer programming structure. His clustering method shows that it is possible to group amino acids which are similar from the evolutionary aspects into one cluster solely based on their physicochemical properties.

The results of clustering analysis based on features extracted from NMR spectrum are illustrated in Fig. 2.

The vicinity of amino acids in hierarchical clustering illustrated in Fig. 2 can be analyzed from two aspects. First, in many cases, the NMR-based features put similar amino acids together but somehow scattered. For example five amino acids with hydrophobic properties out of seven are clustered into two separate groups. Phe and Leu are not detected very near as well as groups (A,I) and (V,P,G). Similarly, polar amino acids are categorized in three groups (W,T,H), (C,S,M) and (N,Q). Negatively charged amino acids (E,D) are very well clustered and positively charged amino acids (R,K) are almost so. This way of grouping amino acids shows that NMR-based features have a kind of association with physicochemical properties.

From another aspect, we can further interpret the clustering diagram to reveal some more facts. Being in the same group for alanine and isoleucine which are benefited by similarity in aliphatic structure, and vicinity of threonine and isoleucine could be due to having an additional chiral center. From structural and stereochemical aspects, this observation is in accordance with the clustering result.

Histidine and tryptophan are only amino acids with a nitrogen-containing five-member aromatic ring which are found in good proximity in this diagram. Occupying one subset by two acidic amino acids (aspartic acid and glutamic acid) and also similar electrically charged among nineteen amino acids is a good evidence that such

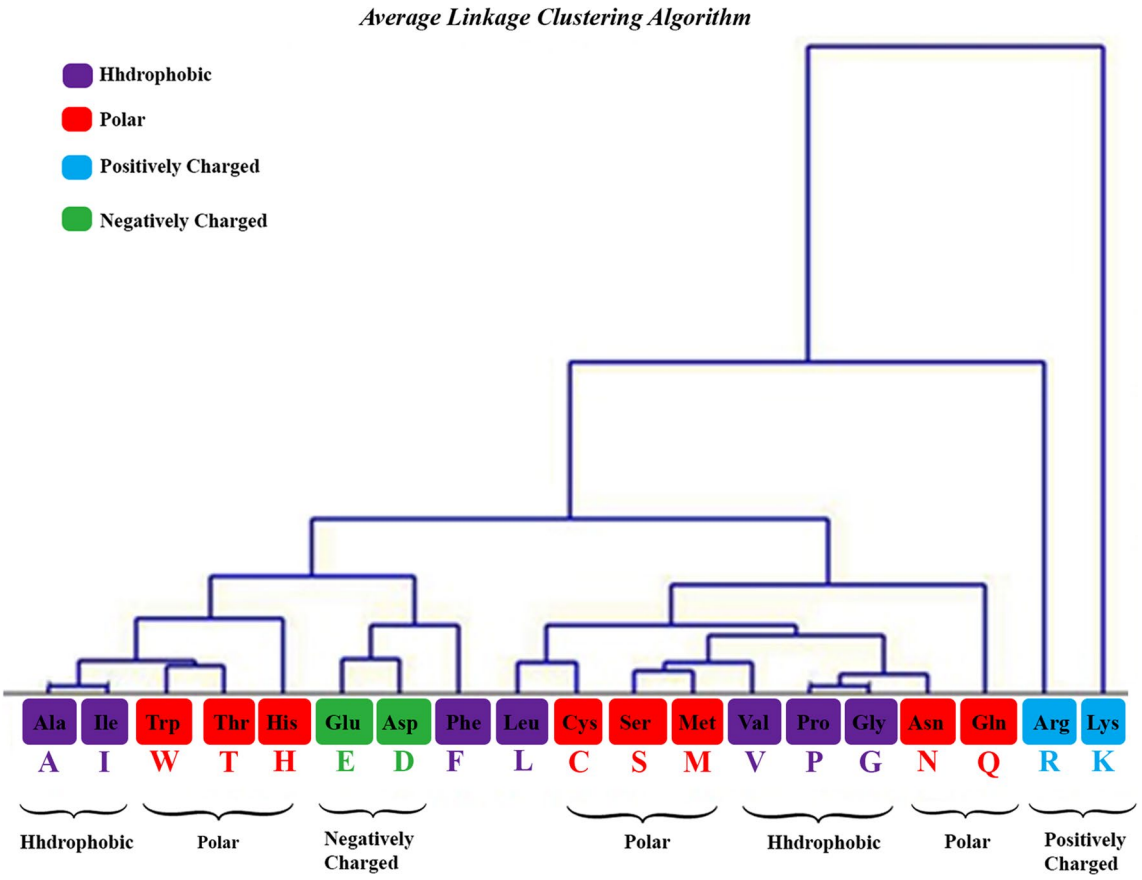


Fig. 2 Hierarchical clustering diagram based on 10 feature groups obtained from NMR spectroscopy

clustering could be improved in characterization analysis. In the next branch there are cysteine and leucine which are two amino acids with a "positive" hydrophobicity index (respectively, 2.5 and 3.8) [49].

The proximity of cysteine and methionine in clustering tree can be explained by the fact that both amino acids contain sulfur atom and also have positive hydrophobicity index. Also, Proline and glycine which both have negative hydrophobicity indices (i.e., hydrophilic) with nonpolar properties are grouped together.

Among these compounds, the most similarity found in structure with additional amidic group is for glutamine and asparagine which are classified already as polar amino acid with the same negative hydrophobic index(−3.5). Moreover, vicinity of two basic amino acids (arginine and lysine) and also their similarity in electrically charged structures with a negative hydrophobicity index are other good evidences that introduced such clustering process being regarded as a good characterization candidate methodology.

Literature survey for clustering of amino acids shows that there is no algorithm being able to cluster all amino acids quite meaningful. In this approach, with assistance of NMR data spectroscopy including T_1 (which is benefited by high precision) we have faced with acceptable and meaningful results which have convinced us to consider such methodology as good candidate in amino acids characterization.

Conclusion

Characterization and identification of amino acids based on "simultaneous time–frequency" relationship toward obtaining the relaxation time could be one of the good methodologies in assignment of these important acids as building blocks of life. By applying the "Inversion Recovery Experiment," with continuous ^{1}H broad banding (^{1}H BB decoupling) we are convinced that this method is a good candidate for such purpose. There are two main reasons for relying on the relaxation of carbon atom rather than hydrogen atom. First is the larger shifts effects of carbon atoms in amino acids (~10–200 ppm) in ^{13}C NMR spectra due to their larger paramagnetic origin, and second is the relaxation process of ^{13}C nuclei which are quietly different, while the corresponding phenomena for protons are of the order of the seconds or so and they do not vary greatly for them in different bonding situations. These two physicochemical characteristic properties (T_1 and chemical shift) are precise enough and open new views for identification and distinction of one amino acid from another without much overlapping. This methodology may be modified and developed for more complex molecules (in terms of inter-, intramolecular interactions and, etc.) such as peptides and proteins. Hierarchical clustering of amino acids based on NMR data spectroscopy was useful procedure in the classification of varieties of data obtained in such analysis.

The authors believe that the newly introduced clusters which have many similarities and overlaps with traditional grouping of amino acids provide a new perspective on characterizing these important molecules and consequently on characterizing proteins. Based on the NMR-based features for amino acids, many novel features and descriptors can be introduced for machine learning tasks on proteins and peptides. Structure and function prediction, determination of localization, determination of activities, etc., are examples of such analyses.

Acknowledgements This research was supported by a grant from Iran National Science Foundation (INSF) (funding reference number 93026701). The supports from Kharazmi University on NMR spectroscopy and related software facilities and University of Tehran are acknowledged.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical approval This article does not contain any studies with human participants or animals performed by any of the authors.

References

1. S. Kinoshita, Adv. Appl. Microbiol. **1**, 201 (1959)
2. M. Hasegawa, I. Matsubara, Anal. Biochem. **63**, 308 (1975)
3. R.J. Simpson, M.R. Neuberger, T. Liu, J. Biol. Chem. **251**, 1936 (1976)
4. A.W. Herlinger, T.V. Long, J. Am. Chem. Soc. **92**, 6481 (1970)
5. J. Pearson, M. Slifkin, Spectrochim. Acta Part A **28**, 2403 (1972)
6. W. Kemp, *NMR in chemistry: a multinuclear introduction* (Macmillan, New York, 1986)
7. P.L. Rinaldi, Prog. Nucl. Magn. Reson. Spectrosc. **15**, 291 (1982)
8. R. Hill, S. Yan, S. Arfin, J. Am. Chem. Soc. **95**, 7857 (1973)
9. J. Baldwin, J. Löliger, W. Rastetter, N. Neuss, L. Huckstep, N. De La Higuera, J. Am. Chem. Soc. **95**, 3796 (1973)
10. R.M. Silverstein, F.X. Webster, D.J. Kiemle, D.L. Bryce, *Spectrometric identification of organic compounds* (Wiley, New York, 2014)
11. S. Porto, J.M. Seco, J.F. Espinosa, E. Quinoá, R. Riguera, J. org. chem. **73**, 5714 (2008)
12. X. Lei, L. Liu, X. Chen, X. Yu, L. Ding, A. Zhang, Org Lett **12**, 2540 (2010)
13. G. Uccello-Barretta, L. Vanni, M.G. Berni, F. Balzano, Chirality **23**, 417 (2011)
14. A. Buryak, K. Severin, J. Am. Chem. Soc. **127**, 3700 (2005)
15. A.T. Wright, E.V. Anslyn, J.T. McDevitt, J. Am. Chem. Soc. **127**, 17405 (2005)
16. J.F. Folmer-Andersen, M. Kitamura, E.V. Anslyn, J. Am. Chem. Soc. **128**, 5652 (2006)

17. N.M. Maier, P. Franco, W. Lindner, J. Chromatogr. A **906**, 3 (2001)
18. E.J. Severin, R.D. Sanner, B.J. Doleman, N.S. Lewis, Anal. Chem. **70**, 1440 (1998)
19. S. Nieto, V.M. Lynch, E.V. Anslyn, H. Kim, J. Chin, J Am Chem Soc **130**, 9232 (2008)
20. S. Nieto, J.M. Dragna, E.V. Anslyn, Chem Eur J **16**, 227 (2010)
21. H.L. Liu, X.L. Hou, L. Pu, Angew. Chem. Int. Ed. **48**, 382 (2009)
22. S.H. Shabbir, L.A. Joyce, G.M. da Cruz, V.M. Lynch, S. Sorey, E.V. Anslyn, J. Am. Chem. Soc. **131**, 13125 (2009)
23. S.H. Shabbir, C.J. Regan, E.V. Anslyn, Proc. Natl. Acad. Sci. **106**, 10487 (2009)
24. O. Jardetzky, C.D. Jardetzky, J. Biol. Chem. **233**, 383 (1958)
25. F. Bovey, G. Tiers, J. Am. Chem. Soc. **81**, 2870 (1959)
26. B. Sen, W.C. Wu, Anal. Chim. Acta **46**, 37 (1969)
27. P. Sneath, J. Theor. Biol. **12**, 157 (1966)
28. M. Dayhoff, R. Schwartz, B. Orcutt, Natl Biomed Res Found, Washington, DC (1978)
29. C.J. Epstein, Nature **215**, 355 (1967)
30. R. Grantham, Science **185**, 862 (1974)
31. A. McLachlan, J. Mol. Biol. **64**, 417 (1972)
32. K. Nakai, A. Kidera, M. Kanehisa, Protein Eng. **2**, 93 (1988)
33. K. Tomii, M. Kanehisa, Protein Eng. **9**, 27 (1996)
34. K.-C. Chou, Curr. Proteomics **6**, 262 (2009)
35. D. Georgiou, T.E. Karakasidis, J. Nieto, A. Torres, J. Theor. Biol. **257**, 17 (2009)
36. D. Georgiou, T. Karakasidis, J.J. Nieto, A. Torres, J. Theor. Biol. **267**, 95 (2010)
37. I. Saha, U. Maulik, S. Bandyopadhyay, D. Plewczynski, Amino Acids **43**, 583 (2012)
38. K. Laurila, M. Vihinen, Amino Acids **40**, 975 (2011)
39. P. Chen, C. Liu, L. Burge, J. Li, M. Mohammad, W. Southerland, C. Gloster, B. Wang, Amino acids **39**, 713 (2010)
40. Y. Yamanishi, M. Araki, A. Gutteridge, W. Honda, M. Kanehisa, Bioinformatics **24**, i232 (2008)
41. A. Tyagi, P. Kapoor, R. Kumar, K. Chaudhary, A. Gautam, G. Raghava, Sci. Rep. **3**, 2984 (2013)
42. F. Murtagh, P. Contreras, Wiley Interdisciplinary Rev. Data Min. Knowl. Disc. **2**, 86 (2012)
43. F. Mormann, R.G. Andrzejak, C.E. Elger, K. Lehnertz, Brain **130**, 314 (2007)
44. W.M. Herrmann, *Electroencephalography in drug research* (G. Fischer, Frankfurt, 1982)
45. B. Hjorth, Electroencephalogr. Clin. Neurophysiol. **29**, 306 (1970)
46. N. Kannathal, M.L. Choo, U.R. Acharya, P. Sadasivan, Comput. Methods Programs Biomed. **80**, 187 (2005)
47. C. Yucelbas, S. Ozsen, S. Gunes, S. Yosunkaya, IADIS Int. J. Comput. Sci. Inf. Syst. **8**, 119–131 (2013)
48. L.E. Stanfel, J. Theor. Biol. **183**, 195 (1996)
49. J. Kyte, R.F. Doolittle, J. Mol. Biol. **157**, 105 (1982)