Combinatorics of aliphatic amino acids

Konrad Grützmann, Sebastian Böcker, Stefan Schuster

Friedrich Schiller University Jena
Jena Centre for Bioinformatics
Ernst-Abbe-Platz 2, D-07743 Jena, Germany
email: (konrad.g, sebastian.boecker, stefan.schu)@uni-jena.de

Abstract

On many scales of living systems, only relatively few building blocks are used. For example, out of more than 100 chemical elements, only six are mainly used. Only four nucleobases appear in DNA, and the number of encoded amino acids is relatively small in view of the enormous number of possible chemical structures of side chains.

This study combines biology and mathematics, showing that a relatively simple question from molecular biology can lead to complicated mathematics. The question is how to calculate the number of theoretically possible aliphatic amino acids, x_n, as a function of the number, n, of carbon atoms in the side chain. We use a graph-theoretical description of the molecular structure (Fig. 1).

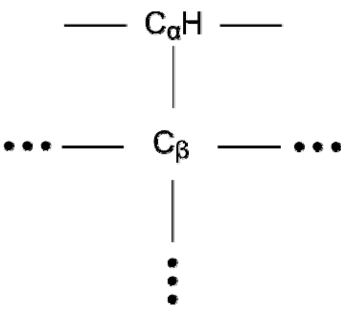

Figure 1. Schematic picture of the structure of aliphatic amino acids. Three dots stand for a tree subgraph. The minimum case is that only an H atom is attached.

The presented calculation is based on earlier results from theoretical chemistry concerning alkyl compounds [2,3]. The following recurrence formula with initialization x_0 = 1 can be used [4]:

$$x_n = \frac{1}{6}\left(\sum_{i+j+k=n-1} x_i x_j x_k + 3 \sum_{i+2j=n-1} x_i x_j + 2 \sum_{3i=n-1} x_i \right)$$

The sequence x_n increases very rapidly: 1, 1, 1, 2, 4, 8, 17, 39, 89, 211, 507, ... Otter [5] calculated the asymptotic behaviour of x_n to be

$$x_n \sim 0.5178760 \cdot 2.81546^n \cdot n^{-3/2} .$$

We show that, from relatively small *n* values on, this approximation is very reliable [1].

We examine which of the theoretically possible structures really occur in living organisms. Proteinogenic aliphatic amino acids include glycine ($n = 0$), alanine ($n = 1$), valine ($n = 2$), leucine, isoleucine and methionine ($n = 4$) [1]. Several other aliphatic amino acids are used for other physiological purposes, for example, 2-amino butanoic acid ($n = 2$) for chemical communication in *Globodera rostochiensis* [6]. L-norvaline ($n = 3$) is an anti-inflammatory arginase inhibitor in humans [7]. L-norleucine ($n = 4$) is a metabolite in orchid plants [8]. Moreover, we consider aliphatic amino acids for a less strict definition allowing sulfur, nitrogen and oxygen atoms in the side chain [1].

The results are in agreement with a general phenomenon found in biology: Usually, only a small number of molecules are chosen as building blocks to assemble an inconceivable number of different macromolecules as proteins. Thus, natural biological complexity arises from the multifarious combination of building blocks.

References

1. Grützmann K, Böcker S, Schuster S (2011) Combinatorics of aliphatic amino acids. *Naturwissenschaften* 98: 79-86
2. Henze HR, Blair CM (1931) The number of structurally isomeric alcohols of the methanol series. *J. Am. Chem. Soc.* 53: 3042-3046
3. Balaban AT, Kennedy JW, Quintas LV (1988) The number of alkanes having n carbons and a longest chain of length d. *J. Chem. Educ.* 65: 304-313
4. Pólya G (1937) Kombinatorische Anzahlbestimmungen für Gruppen, Graphen und chemische Verbindungen. *Acta Mathem.* 68: 145-254
5. Otter R (1948) The Number of Trees. *Ann. Mathem.* 49: 583-599
6. Riga E, Perry RN, Barrett J, Johnston MRL (1997) Electrophysiological responses of male potato cyst nematodes, *Globodera rostochiensis* and *G. pallida*, to some chemicals. *J. Chem. Ecol.* 23: 417-428
7. Ming XF, Rajapakse AG, Carvas JM, et al. (2009) Inhibition of S6K1 accounts partially for the anti-inflammatory effects of the arginase inhibitor L-norvaline. *BMC Cardiovasc. Disord.* 9: 12-18
8. Kikuchi T, Kadota S, Hanagaki S, et al. (1981) Studies on the constituents of orchidaceous plants. I. Constituents of *Nervilia purpurea* Schlechter and *Nervilia aragoana* Gaud. *Chem. Pharm. Bull.* 29: 2073-2078