Non-Statistical Label Distribution in Biosynthetic $^{13}$C Enriched Amino Acids

E. Bengsch, J.-Ph. Grivet
Centre de Biophysique Moléculaire, C.N.R.S., 1A Avenue de la Recherche Scientifique, F-45045 Orleans Cedex, France

H.-R. Schulten
Institut für Physikalische Chemie der Universität Bonn, D-5300 Bonn, Wegelerstr. 12, Federal Republic of Germany

Z. Naturforsch. 36b, 1289–1296 (1981); received May 20, 1981

$^{13}$C Enriched Amino Acids, Biosynthesis, Label Distribution, NMR Spectra, Field Desorption Mass Spectrometry

A combined $^{13}$C nuclear magnetic resonance and field desorption mass spectrometric investigation of algae grown on $^{13}$CO$_2$ has shown that the isotopic enrichment of amino acids extracted therefrom is neither uniform nor statistical. The use of these two independent techniques allows a new, detailed and accurate insight into the label distribution resulting from biosynthesis. The observed deviations from the statistical abundances are systematic. A system for classifying each member of the complex ensemble of isotopic species has been devised, so that the isotopomers may be ordered according to their relative probability of occurrence.

Introduction

The most abundant elements found in biochemical compounds (H, C, N, O) all have a rare and heavy isotope ($^2$H, $^{13}$C, $^{15}$N, $^{17}$O). In each case, the magnetic properties of the rare nuclides are very different from those of the dominant species. These rare nuclei are increasingly used as labels in biochemical and medical research. Biosynthesis, starting with very simple enriched precursors, is the preferred method for the preparation of complex labelled molecules. However, the starting material is never isotopically pure, so that the resulting compounds such as amino acids, sugars, fatty acids, and even proteins and viruses must also be only partially enriched. It has been assumed in the past that isotopically labelled compounds of biosynthetic origin

i) are statistically enriched, and

ii) show a uniform enrichment factor [1–3].

We have tested these hypotheses for the special case of four amino acids namely serine, threonine, aspartic acid, and glutamic acid provided by Commissariat à l’Energie Atomique, Saclay, France. These compounds are extracted from algae which grow on $^{13}$CO$_2$ as their sole carbon source [1]. They have a nominal 13-C content of ca. 82%, 2 N isotopic

Table I. Pictorial display of the isotopomers of serine.

<table>
<thead>
<tr>
<th>$C'$</th>
<th>$C_{\alpha}$</th>
<th>$C_{\beta}$</th>
<th>horizontal projection</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>$X_3$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$X_2$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$X_1$</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>$X_0$</td>
</tr>
<tr>
<td>$Y_{CO}$</td>
<td>$Y_{\alpha}$</td>
<td>$Y_{\beta}$</td>
<td>vertical projection</td>
</tr>
</tbody>
</table>

species (or isotopomers), differing only in their $^{12}$C/$^{13}$C composition can be associated with an N-carbon atom molecule. Thus, in the series from serine (Table I) through threonine and aspartic acid to glutamic acid, the complexity of the isotopic mixture is seen to increase exponentially.

A quantitative analysis by mass number (or 13-C content) was carried out by field desorption mass
spectrometry (FD-MS)*. This soft ionisation technique is a sensitive, accurate and fast tool for isotopic analysis [5]. Usually, the intact molecular ion appears as base peak in the spectrum. The precision of direct isotope determination has been shown to be on the order of a few tenths of a percent. For biomolecules of high molecular weight and polarity, it can reach 0.01 % in favourable cases [6]. Nuclear magnetic resonance spectroscopy (NMR) [7], on the other hand, provided, first, approximate values of the local (or atomic) concentration of $^{13}$C, then, in a second stage, relative concentrations of each individual isotopomer. The poor sensitivity of NMR is offset by the large number of data that can be collected and by the numerous cross-checks that are thus made possible. A prerequisite for making a full use of the NMR spectra is, of course, a complete knowledge of all the spectral parameters involved, including the isotope effects. The details of this basic investigation have been and will be reported elsewhere [8, 9].

Experimental

Mass analyses were performed on a double focusing mass spectrometer (Varian MAT 731) with a combined electron impact/field desorption ion source. The emitter potential was +8 kV, while the counterelectrode was held at —3 kV. The ion signals were electrically recorded. The signal to noise ratio was improved by averaging in a multichannel analyser and the sensitivity optimized by selecting a narrow mass range. The presence of small amounts of alkali metal impurities leads to the attachment of a cation to the organic molecule and produces ions, which consist of a complex of the amino acid (M) with a metal ion. Such species are ideally suited for direct isotope determinations of compounds labelled with stable- [10, 11] and radio-isotopes [12, 13].

NMR spectra were collected at 22.63 MHz on a Fourier transform spectrometer (Bruker WH 90). Amino acids were dissolved in D$_2$O and a small amount of EDTA was added to each sample. Overlapping lines could often be resolved by changing the pH: this was accomplished using either NaOD or DCI. Spin-lattice relaxation of air-saturated solutions times were measured using the methods of Rüterjans et al. [14, 15]. T$_1$ was found to be less than 2 seconds for all proton-bearing carbon atoms, and, for carboxylic C atoms, ranged from 4 to 25 seconds, depending on the isotopic composition of the molecule and on its protonation state. Accordingly, the pulse repetition time was varied from 10 to 40 seconds depending on the nucleus to be examined. The flip angle was set to less than 45° and the digital resolution was better than 0.1 Hz. In order to test whether the nuclear Overhauser effect depended on isotopic composition, we have recorded spectra free of Overhauser enhancements [16], using repetition times of 60 to 100 seconds. These verifications were done for all carbon atoms of acetic acid, serine and aspartic acid, for all the carboxylic carbons and for some proton bearing carbon atoms of the remaining amino acids. Contrary to a previous report [17], spectra recorded under the conditions described above with and without Overhauser enhancement show closely similar relative abundances of the various isotopic species. We believe that in cases where the pulse repetition time can be considered as too short, weak partial saturation and Overhauser effects exert opposite effects on intensity ratios. A somewhat similar conclusion concerning the absence of selective Overhauser enhancements has been reached in connection with quantitative assays of tritium labelling by NMR [18]. Furthermore, since experimental spectra are always very similar to computed ones, it may be assumed that cross-correlation effects [19] are not important under our operating conditions. Area determinations are believed to be accurate to 0.1%.

Results and Discussion

A three step analytical procedure

A general idea of the path followed during the analysis can be obtained with reference to the Table, which has been drawn up for the case of serine. The carbon atoms, here and in the sequel, are numbered consecutively, starting with the carboxylic group [1], through the a carbon [2]. Let N be the total number of carbons in the molecule and p the number of $^{13}$C nuclei in a given isotopomer (0 < p < N). Peaks in the mass spectrum, such as that presented in Fig. 1, refer to a given mass number, i.e., a given p value or isotopic composition, irrespective of the nuclei's locations in the molecule. The variable p therefore labels the main rows of the Table, as well as the mass spectrum peaks. We denote by x$_p$ the relative intensity of peaks p. Since our goal is the determination of the relative concentration of each isotopomer, the mass spectrum can be considered as a rough horizontal projection of the data symbolized in the Table, or better, as a sort of coarse-graining of the information contained in the sample. Even though the x$_p$'s only partly describe the complete $^{13}$C distribution, they are so

* For a basic introduction into the principle and technique of FD-MS see reference [4].
accurate that we shall consider these values as our primary data. We will use NMR to apportion concentrations of isotopomers within a p-manifold, without modifying the \( x_p \)'s. The position of a particular carbon atom in the molecule may be designated by subscript \( i \) \((1 < i < N)\). Fragments such as \( ...^{13}C_i-^{12}C_{i+1}... \) and \( ...^{12}C_i-^{13}C_{i+1}... \) (where the dots stand for any combination of atoms) generally give non-overlapping lines in the \( C_i \) spectral region because of the large vicinal C-C coupling constant. The relative probability of occurrence of these two moieties can then be simply determined by planimeter integration of the spectra. It is easy, from such NMR measurements, to derive an estimate of the relative \( ^{13}C \) abundance at position \( i \), irrespective of the particular isotopomer. The result is only approximate, since the method only senses one half of the species present in the sample, those for which \( C_i = ^{13}C \). The columns of the Table can be labelled by the parameter \( y_i \). This approach can be considered as a rough "vertical projection" of the information necessary to completely characterize the sample. At this point, we have in hand enough information to derive two independent estimates of the average enrichment factor (assuming an uniform statistical distribution): one from FD-MS and one from NMR measurements. Usually, the two values differ by less than 1%.

Fig. 1. Field desorption mass spectrum of \(^{13}C\)-enriched glutamic acid, \([\text{Glu} + \text{Na}]^+\) mass range, 30 scans, emitter heating current 25–27 mA.

A detailed, quantitative allocation of the \(^{13}C\) content amongst the various isotopomers can now be made in the third step. A sample made from an N-carbon molecule contains \( 2^{n-1} \) isotopomers which are, in principle, detectable by NMR. From the fact that a p-spin system gives rise to \( p2^{n-1} \) allowed transitions, it can be shown that the complete spectrum will comprise some \( N \times 3^{n-1} \) lines. Such a spectrum can easily be simulated by computer. We proceed in three stages. (i) Fit each isotopomer spectrum, using an iterative algorithm, such as LAOCOON-4 A [20] or NMRIT IV [21]. (ii) Compute the weighted sum of all these partial spectra, assuming a statistical distribution and the \( a \) value estimated previously. (iii) Adjust individual isotopomer concentrations until a best fit to experiment is obtained. These adjustments are effected only inside a p-manifold (in other words, inside a main row of the table), so that the accurate \( x_p \) values be not altered. The case of the C-5 position (\( \delta\)-carbonyl) of glutamic acid is shown in Fig. 2. From this spectral region, the \(^{13}C\) concentration at C-4 (\( \gamma\)-CH\(_2\)) can be estimated, simply from the ratio of the central peak area to that of whole spectrum. Then we show the central part of the previous trace, with the first (statistical abundances) and last (individually adjusted abundances) stages of the refinement.
Non-uniform and non-statistical enrichment

Fig. 3 shows, for given N and p, the experimentally measured relative concentrations, \( x_p^{\text{exp}} \), and also two sets of computed values, \( x_p' \) and \( x_p'' \). These have been derived, under the statistical distribution hypothesis, from two \( a \) values, obtained as follows:

For the mass spectrosocptist, the most natural definition of \( a \) is that of the mean atomic enrichment, defined formally as:

\[
\alpha' = \frac{1}{N} \sum_{p=0}^{N} p x_p^{\text{exp}}.
\]

This number gives the percentage of \(^{13}\text{C} \) nuclei among all carbon atoms. The \( x_p' \) are then defined by the binomial distribution:

\[
x_p' = \binom{N}{p} \alpha'^p (1-\alpha')^{N-p}.
\]

Another plausible definition of \( a \) is the following. We define \( \alpha'' \) as that value of \( a \) which, together with the assumption of uniform statistical enrichment,
minimizes the sum of the squared deviations between experimental and computed values. Formally:

\[ a_p'' = \left( \frac{N}{p} \right) a'^p (1-a'')N^{-p} \]
\[ \sigma^2 = \sum_{p=0}^{N} (x_p''-x_{p,\text{exp}})^2 \text{ minimum.} \]

It can easily be shown that \( a' = a'' \) for a true statistical distribution. The results to be presented are essentially unchanged if the above sum of squares is weighted to take into account the number \( \binom{N}{p} \) of isotopomers found in a \( p \)-manifold. The various \( a \) values computed according to the above definitions are as follows:

- **serine:** \( a' = 0.879 \), \( a'' = 0.876 \)
- **threonine:** \( a' = 0.765 \), \( a'' = 0.796 \)
- **aspartic acid:** \( a' = 0.852 \), \( a'' = 0.872 \)
- **glutamic acid:** \( a' = 0.726 \), \( a'' = 0.780 \)

The abundances computed with these average enrichment factors are displayed in Fig. 3. The first fact to be noticed is that the \( a \)'s are different for each amino acid: with respect to \( ^{13}\text{C} \) content, some compounds are favoured by the biosynthetic pathways. A second fact is also readily apparent from Fig. 3: the theoretical distributions do not reproduce the experimental ones. In fact, we have been unable to find \( a \) values which will lead to a good fit of the experimental results, if we insist on statistical enrichment. Part of the discrepancy, as Fig. 3 shows, is due to the zero-spin (all \( ^{12}\text{C} \)) species, the concentration of which is way off any reasonable prediction. It might be suggested that this is an artefact of the preparation procedure. We have looked into the possibility of disregarding this species. If we set \( x_0,\text{exp} \) to zero, \( a' \) changes and moves closer to \( a'' \), but our conclusions are essentially unchanged. As could have been conjectured, we observe that the deviation from statistical enrichment (as measured by \( \sigma^2 \)) increases when the enrichment factor itself decreases. Further, when the deviations \( x_p,\text{exp} - x_p'' \) are plotted versus \( p \), (data not shown), two general trends seem to emerge: The concentrations of high-\( p \) molecules are too low and those of the low-\( p \) species are too high, when compared to the binomial law \( x_p'' \). Moreover, these deviations seem to alternate in magnitude: even \( p \) isotopomers tend to be favoured for threonine, aspartic and glutamic acid. The opposite seems to be true for serine: it is the odd numbered \( x_{p,\text{exp}} \) and \( x_{p,\text{exp}} - x_{p''} \) which are higher than the general trend would predict:

Consideration of the \( y_i \) values shows that the enrichment is different for different positions in the molecule. Even the approximate values derived from the graphical integration of the NMR spectra show this rather clearly. In Fig. 4, we have plotted the deviations of the \( y_i \) from their mean as a function of position (dots indicate a hydrophilic atom).

The opposite seems to be true for serine: it is the odd numbered \( x_{p,\text{exp}} \) and \( x_{p,\text{exp}} - x_{p''} \) which are higher than the general trend would predict:

![Fig. 4. Deviations of the \( y_i \) from their mean as a function of position (dots indicate a hydrophilic atom).](image-url)
These general conjectures are quantitatively confirmed by the detailed assay of individual isotopomers which can be arrived at through spectral simulation. Fig. 2 shows an example of the quality of fit than can be obtained. The complete results for glutamic acid are displayed in Fig. 5. It can be seen that quite large concentration fluctuations occur among the members of a p-manifold.

**Rules for relative abundances allow to predict isotopomer rank**

The rules presented above can be used the other way round: knowing the structure of an isotopomer, we may rank this species according to concentration among its (N, p) congeners in a p-manifold. Let us restate these rules in a slightly different form. An isotopomer will be favoured (that is more abundant than the binomial law predicts) if (i) hydrophilic positions are occupied by $^{13}\text{C}$ nuclei, and (ii) if hydrophobic and α positions contain $^{12}\text{C}$ nuclides. The method proposed amounts to polling the N atoms in the molecule for a positive or negative answer to the above criteria. Let us illustrate the procedure with the case $N = 5$, $p = 2$. The compound symbolized by $^{13}\text{C}_1-^{13}\text{C}_2-^{12}\text{C}_3-^{12}\text{C}_4-^{13}\text{C}_5$ returns a unanimous positive vote (both carboxylic atoms

Fig. 5. Relative isotopomer concentrations for glutamic acid. Values for singly and non labelled species, which could not be represented conveniently on this scale are as follows:

$[1-^{13}\text{C}]\text{Glu} = [1] = 0.2\%$
$[2] = 0.35\%$
$[3] = 0.18\%$
$[4] = 0.2\%$
$[5] = 0.38\%$

O-spin isotopomer 7.0 %
carry $^{13}\text{C}$, the others $^{12}\text{C}$). As Fig. 5 shows, and as should be expected if our generalizations are correct, this species is indeed the most abundant of the ten $N = 5$, $p = 2$ isotopomers. As a further example, consider the molecule whose carbon backbone is $^{13}\text{C}_1-^{13}\text{C}_2-^{13}\text{C}_3-^{13}\text{C}_4-^{12}\text{C}_5$, a member of the $N = 5$, $p = 3$ manifold. The poll result is now unanimously negative (the two carboxylic groups have a $^{13}\text{C}$, all others carry $^{12}\text{C}$). It can be checked from Fig. 5 that this species is one of the least probable of its group. Two isotopomers which carry the same number of positive and negative factors can be further ordered using the following additional rules. Aliphatic and carboxylic atoms have more influence than hydroxylic positions, which, in turn, exert a greater effect than a carbons. As Fig. 5 shows, these tentative rules indeed seem to hold for the isotopomers of glutamic acid. They also stand up well to experimental verification in the cases of threonine, serine, and aspartic acid. Obviously, a definitive validation can only stem from the successful prediction of concentrations for a different amino acid. A seemingly related result has been noted recently concerning whole plant organs [22, 23]: hydrophobic constituents are depleted and the hydrophilic constituents enriched in $^{13}\text{C}$ relative to the average abundance. Moreover, our results are in perfect agreement with earlier combustion analysis experiments of amino acids from plants grown in natural isotopic abundances [24]. This method, however, allows only a distinction of the carboxyl carbons and the remainder of the molecule: The first are always markedly heavier.

**Conclusion**

The label distribution in the biosynthetic, $^{13}\text{C}$-enriched, aliphatic amino acids serine, threonine, aspartic acid, and glutamic acid is not random, either with respect to the number of heavy nuclides in the molecule or with respect to the position of $^{13}\text{C}$ nuclei within the molecule. As a consequence, deviations from statistical occurrence are found for the concentrations of individual isotopomers. The size of these deviations seems to depend on the following parameters: (i) distance from global isotopic purity ($\sigma^2$ increases with $1-\alpha$); (ii) number of $^{13}\text{C}$ nuclei ($p$); (iii) parity of $p$; (iv) distribution of $^{13}\text{C}$ among hydrophilic and hydrophobic sites. We may remark that the $p = 0$ isotopomer of glutamic acid is singular with respect to the first three points: low $\alpha''$ (0.78), small and even $p$. Thus the high concentration of this species is perhaps not entirely due to a contamination.

It is reasonable to assume that the rules derived in this work apply to other aliphatic amino acids as well. If such were the case, then isotopomer abundances could be predicted from the sole knowledge of the global enrichment factor, a quantity that can be experimentally determined in minutes, using microgram amounts of material, by field desorption mass spectrometry [25, 26]. The type of effects described here should be qualitatively identical and perhaps quantitatively more important in the case of $^{14}\text{C}$ labelling, for which one could, in the same manner, predict sites of preferential labelling. Finally, these data in some way reflect the metabolism of carbon in photosynthetic algae. Although the decipherment of such a message is a formidable task, we believe that the method and results presented here constitute an important step in the right direction.

This work was supported in part by the Deutsche Forschungsgemeinschaft (Schu 416/1-5).

---