Archives of Biochemistry and Biophysics xxx (2017) 1–7

Contents lists available at ScienceDirect

Archives of Biochemistry and Biophysics

journal homepage: www.elsevier.com/locate/yabbi

Advances in stable isotope assisted labeling strategies with information science

Takanori Kigawa [a, b, *]

[a] RIKEN Quantitative Biology Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan
[b] School of Computing, Tokyo Institute of Technology, 4259 Nagatsuta-cho, Midori-ku, Yokohama 226-8503, Japan

ARTICLE INFO

Article history:
Received 20 January 2017
Received in revised form
15 June 2017
Accepted 18 June 2017
Available online xxx

Keywords:
Amino acid-selective stable isotope labeling
Cell-free protein synthesis
Coding theory
Combinatorial selective labeling
Signal assignment

ABSTRACT

Stable-isotope (SI) labeling of proteins is an essential technique to investigate their structures, interactions or dynamics by nuclear magnetic resonance (NMR) spectroscopy. The assignment of the main-chain signals, which is the fundamental first step in these analyses, is usually achieved by a sequential assignment method based on triple resonance experiments. Independently of the triple resonance experiment-based sequential assignment, amino acid-selective SI labeling is beneficial for discriminating the amino acid type of each signal; therefore, it is especially useful for the signal assignment of difficult targets. Various combinatorial selective labeling schemes have been developed as more sophisticated labeling strategies. In these strategies, amino acids are represented by combinations of SI labeled samples, rather than simply assigning one amino acid to one SI labeled sample as in the case of conventional amino acid-selective labeling. These strategies have proven to be useful for NMR analyses of difficult proteins, such as those in large complex systems, in living cells, attached or integrated into membranes, or with poor solubility. In this review, recent advances in stable isotope assisted labeling strategies will be discussed.

1. Combinatorial selective labeling

Amino acid-selective stable isotope (SI) labeling (AASIL) facilitates the discrimination of the amino acid type of each signal, without the need for a triple resonance experiment-based sequential assignment. Therefore, it is especially useful for the signal assignment of difficult targets, such as large complex systems [1], low-solubility proteins [2], and proteins in living cells [3]. The dual selective labeling method, which utilizes both the amide nitrogen and carbonyl carbon labeling of selected amino acids, narrows down the assignment possibilities even further [4], and as a consequence leads to the assignment of amino acid pairs occurring only once in the sequence, without the need for a series of time-consuming triple resonance experiments.

For discriminating all amino acids, however, these simple AASIL schemes require a large number of samples, which are typically the same as the number of amino acids (19 for nitrogen or 20 for carbon). To reduce the number of labeled samples required in AASIL, various combinatorial selective labeling (CSL) methods were proposed [5–19] as more sophisticated labeling strategies, as summarized in Table 1.

In the CSLs, amino acids are represented as combinations of SI labeled samples, rather than simply assigning one amino acid to one SI labeled sample as described above. For instance, the CSL scheme developed by Parker et al. [5] can discriminate 16 amino acids with one uniformly ^{13}C- and ^{15}N-labeled reference and four selectively (100% or 0% for ^{13}C and 100% or 50% for ^{15}N, respectively) labeled samples. In order to ensure that the ^{13}C information is obtained from the HN(CO) spectrum, irrespective of the ^{15}N labeling ratio, a labeling ratio of 100% or 50% for ^{15}N, rather than 100% or 0%, was used. The simpler CSL scheme reported by Otting and colleagues used five samples [9] employing the single selective ^{15}N-labeling approach, in which one amino acid with high occurrence and at most three amino acids with low occurrence in each sample were labeled in order to diminish spectral overlaps. Dötsch and colleagues developed CSL [7,11] focused on membrane proteins, in which 6 or 7 frequently appearing amino acids were labeled with ^{15}N or 1-^{13}C. They further improved CSL to discriminate up to 20 amino acids with a number of samples labeled with ^{15}N and/or ^{13}C based on a dual selective approach [20], or to discriminate 12 amino

* RIKEN Quantitative Biology Center, 1-7-22 Suehiro-cho, Tsurumi, Yokohama 230-0045, Japan.
E-mail address: kigawa@riken.jp.

http://dx.doi.org/10.1016/j.abb.2017.06.014
0003-9861/

acids with only 3 samples [14], in which the samples were labeled with combinations of ^{15}N, 1-^{13}C, and ^{13}C/^{15}N. Choe and colleagues improved membrane protein-focused CSL [10,15] to discriminate up to 19 amino acids except for glutamate with 6 samples, in which the samples were simply labeled with ^{13}C and/or ^{15}N. In order to maximize the assignable residues, computational methods for designing the labeling patterns for CSL were employed [10,15,20]. Almost all of the CSL schemes simply rely on the presence or absence of cross peaks, mainly because the spectral analysis can be easily performed by visual inspection.

We realized that AASIL can be regarded as an "encoding-and-decoding" process, which is frequently utilized in digital communication (Fig. 1a). In the process, a sender converts a letter to a codeword, according to a predefined table associating each letter with a codeword. The converted codeword, generally consisting of binary digits, is transmitted through a communication channel, and then the receiver converts it back to the letter according to the table. Likewise, in the AASIL process, the information of the amino acid type of each residue is converted to the SI labeling ratio, which corresponds to a codeword if the SI labeling pattern is considered as a codeword table, and then it is retrieved from the observed NMR spectra according to the table.

Based on this consideration, we proposed a novel SI-labeling strategy based on coding theory, which we call "stable isotope encoding" (SiCode) [21]. In this strategy, the amino acid information encoded into the SI-labeling ratio pattern of the samples is decoded from the signal intensity ratio of the NMR spectra (Fig. 1b and c). We have designed a novel scheme to use ternary digits as codewords in the simplest implementation, and an example codeword table based on a dual selective approach is shown in Fig. 1b. In this scheme, the ternary digits, "0", "1", and "2", are represented by SI-labeling grades of 50%, 75%, and 100% (for ^{15}N) or 0%, 50%, and 100% (for ^{13}C), respectively. In addition, we only use codewords with at least one "2", so that the signal with the largest intensity can be used as a fully-labeled reference. The number of assignable codewords based on this scheme is 19, which is the exact number required for representing non-proline amino acids. By employing this scheme, we can discriminate 19 kinds of non-proline amino acids with only 3 labeled samples, and the additional uniformly labeled reference sample used in the above-mentioned CSL [5] can be omitted.

Almost all of the CSL methods employ the cell-free system for sample preparation in order to achieve accurate SI-labeling by avoiding SI scrambling and dilution, which are problematic in the *in vivo* protein expression systems. For SiCode, we employed an *Escherichia coli*-based cell-free protein synthesis system [22–25] supplemented with metabolic inhibitors [26] for the sample preparation, to achieve the highly accurate SI-labeling ratios we designed. The cell-free system we developed for the accurate SI-labeling is commercially available as the "in vitro Protein Expression (iPE) kit" (MilliporeSigma/Sigma-Aldrich, Ohio, USA) and "Musaibou-kun" (Taiyo Nippon Sanso, Tokyo, Japan).

SiCode can be performed using the protein expression system with a manageable level of SI scrambling (see Supplementary Information of [21] for a detailed discussion). If the SI scrambling profile for a specific expression system must be evaluated, then a customized labeling pattern addressing SI scrambling could be designed based on the information distance between amino acids, as described later. In the future, SiCode will be performed with *in vivo* expression systems; for example, by the combination of the single protein production system [27,28] and amino acid auxotrophic *E. coli* strains. Thus, even more complicated labeling patterns can be easily achieved, as long as the cell-free system without SI scrambling is used.

Table 1
Summary of combinatorial selective labeling methods.

Reference	Purpose	No. of assignable amino acids	No. of samples	Labeling	Spectra	Comment
[5]	General	16	5	u-^{13}C/^{15}N or 50%^{15}N	HSQC, HN(CO)	
[6]	General	19	4	u-^{13}C/^{15}N, u-^{13}C, or non-labeled	HSQC, HN(CA), HN(CO), CABA (CO)NH, HN(CACB)	Conventional triple resonance experiments with u-^{13}C/^{15}N-labeled sample are also used.
[7]	Specific protein	7	3	u-^{15}N, u-^{13}C or non-labeled	HSQC, HNCO	Conventional triple resonance experiments with u-^{13}C/^{15}N-labeled sample are also used.
[8]	Specific protein	8	5	u-^{13}C/^{15}N, u-^{15}N, or non-labeled	HSQC, HN(CO)	Conventional triple resonance experiments with u-^{13}C/^{15}N-labeled sample are also used.
[9]	General	19	5	u-^{15}N, or non-labeled	HSQC	No information about the preceding residue.
[10,15][a]	Membrane protein	17	5	u-^{15}N, u-^{15}C or non-labeled	HSQC, HN(CO)	
[11]	Membrane protein	6	3	u-^{15}N, u-^{15}C or non-labeled	HSQC, HN(CO)	
[12,13][b]	General	7	11	u-^{13}C/^{15}N, u-^{15}N, or non-labeled	HSQC, {^{12}CO$_i$–^{15}N$_{i+1}$}-filtered HSQC	
[14]	Membrane protein	12	3	u-^{13}C/^{15}N, u-^{15}N, or 1-^{13}C	HSQC, HNCO, HN(CA), (CO) HN(CA), HN(COCA), HN(COCA), DQ-HN(CA)	
[16]	Membrane protein	7	1	u-^{13}C/^{15}N, u-^{15}N, or 1-^{13}C	ts-HN(CO)/HN(CA) (a) and ts-HN(CO)/HN(COCA)	
[17]	Membrane protein	13	3	u-^{13}C/^{15}N, u-^{15}N, 1-^{13}C, or 2-^{13}C	HSQC, HN(CO) HN(COCA), HN(CA), CO-filt. HN(CA), DQ-HN(CA)	
[18][b]	General	20	1	u-^{13}C/^{15}N, u-^{15}N, or non-labeled	HSQC, HN(CO), HN-XU	
[19]	Ligand-binding site	19	3	u-^{13}C/^{15}N, u-^{15}N,^{13}C/^{15}N, 1-^{13}C, or 2-^{13}C	C'-filtered HSQC, HN(CO), HN(COCA), HN(CA), C'-filtered HN(CA), DQ-HN(CA)	
[21,29][c]	General	19	3	0%, 50%, or 100%^{13}C/50%, 75%, 100%^{15}N		

[a] Combinatorial dual-labeling (CDL).
[b] Selective unlabeling.
[c] Stable isotope encoding (SiCode).

Please cite this article in press as: T. Kigawa, Advances in stable isotope assisted labeling strategies with information science, Archives of Biochemistry and Biophysics (2017), http://dx.doi.org/10.1016/j.abb.2017.06.014