Site-Specific Isotopic Labeling (SSIL) – Access to High-Resolution Structural and Dynamic Information in Low Complexity Proteins

Annika N Urbanek, Carlos A Elena-Real, Matija Popovic, Anna Morató, Aurélie Fournet, Frédéric Allemand, Stéphane Delbecq, Nathalie Sibille, Pau Bernadó

▶ **To cite this version:**

Annika N Urbanek, Carlos A Elena-Real, Matija Popovic, Anna Morató, Aurélie Fournet, et al.. Site-Specific Isotopic Labeling (SSIL) – Access to High-Resolution Structural and Dynamic Information in Low Complexity Proteins. ChemBioChem, 2019, 10.1002/cbic.201900583 . hal-02359756

HAL Id: hal-02359756
https://hal.science/hal-02359756

Submitted on 12 Nov 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Site-Specific Isotopic Labeling (SSIL) – Access to High-Resolution Structural and Dynamic Information in Low Complexity Proteins

Annika Urbanek[1], Carlos A. Elena-Real[1], Matija Popovic[1], Anna Morató[1], Aurélie Fournet[1], Frédéric Allemand[1], Stephane Delbecq[2], Nathalie Sibille[1], Pau Bernadó[1,*]

[1] Centre de Biochimie Structurale (CBS), INSERM, CNRS, Université de Montpellier. 29, rue de Navacelles, 34090 Montpellier. France.

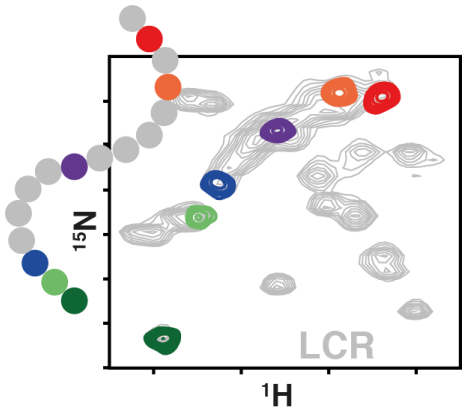
[2] Laboratoire de Biologie Cellulaire et Moléculaire (LBCM-EA4558 Vaccination Antiparasitaire), UFR Pharmacie, Université de Montpellier, Montpellier, France.

Corresponding Author: Pau Bernadó (pau.bernado@cbs.cnrs.fr)

Keywords: Low Complexity Regions (LCRs), Homo-Repeat (HR), Intrinsically Disordered Proteins (IDPs), Site-Specific Isotopic Labeling (SSIL), Nonsense Suppression, Nuclear Magnetic Resonance (NMR), Structural Biology, Protein Dynamics.

Table of Contents

Low complexity regions (LCRs) are strikingly simple sequences where only a limited number of amino acids is repeated. While LCRs are quite common, their high-resolution characterization is inherently difficult. Here we present Site-Specific Isotopic Labeling (SSIL) as a powerful tool to study these intriguing sequences and shed light on their structure/function relationships.



Abstract

Remarkable technical progress in the area of structural biology has paved the way to study previously inaccessible targets. For example, large protein complexes can now be easily investigated by cryo-electron microscopy, and modern high-field NMR magnets have challenged the limits of high-resolution characterization of proteins in solution. However, the structural and dynamic characteristics of certain proteins with important functions still cannot be probed by conventional methods. These proteins in question contain low complexity regions (LCRs), compositionally biased sequences where only a limited number of amino acids is repeated multiple times, which hamper their characterization. This *Concept* article describes a Site-Specific Isotopic Labeling (SSIL) strategy, which combines nonsense suppression and cell-free protein synthesis to overcome these limitations. An overview on how poly-glutamine tracts were made amenable to high-resolution structural studies is used to illustrate the usefulness of SSIL. Furthermore, we discuss the potential of this methodology to give further insights into the roles of LCRs in human pathologies and liquid-liquid phase separation, as well as the challenges that must be addressed in the future for the popularization of SSIL.

Introduction

Nuclear Magnetic Resonance (NMR) is the best-adapted technique to derive high-resolution structural and dynamic information of proteins in solution. A prerequisite for the NMR characterization of proteins in solution is the production of isotopically labeled (^{15}N or ^{15}N/^{13}C) samples, which is normally achieved by recombinant overexpression in bacteria[1]. Although less popular, eukaryotic cells and *in vitro* expression systems are nowadays-available tools for protein production[2–4]. Obtaining an (almost) complete NMR assignment of a protein relies on our capacity to discriminate between similar frequencies (chemical shifts), which depend on the environment that each individual nucleus experiences. Specifically designed NMR pulse sequences and the availability of high magnetic fields enable the straightforward assignment of proteins up to 40 kDa[5]. This also applies to Intrinsically Disordered Proteins (IDPs) that display reduced ^{1}H dispersion due to the lack of permanent secondary or tertiary structure (Fig. 1b)[6–8]. For large IDPs, the use of high dimensionality or carbon-detected experiments enhances spectral resolution, notably simplifying the assignment process[9–11].

Low Complexity Regions (LCRs) in proteins represent a challenge for the above-mentioned NMR strategies. LCRs are compositionally biased protein sequences where one or more amino acids are repeated multiple times[12,13]. Different LCR patterns are known, ranging from homo-repeats (HRs) (Fig. 1a, right panel), in which one amino acid is consecutively appearing multiple times[14,15], to tandem repeats, large compositionally unbiased fragments repeated several times[16,17]. In LCRs, the chemical environment experienced by some nuclei along the sequence is very similar or equivalent and, as a consequence, their NMR frequencies overlap, hampering the sequential assignment. This phenomenon can be alleviated when repeats are not perfect, as the chemical shift perturbation expands to a few residues on both sides. When LCRs adopt a fully or partially folded structure, the chemical environment experienced by the nuclei is less efficiently averaged and some chemical dispersion is observed[18]. This is the case for the ^{1}H-^{15}N Heteronuclear Single Quantum Correlation (^{15}N-HSQC) experiment in a huntingtin construct with 16 consecutive glutamines (see Figure 1c,e) and the androgen receptor, where an elongated density of peaks appears in the ^{15}N-HSQC due to the formation of a transient α-helix[18–20]. Even under these circumstances, when the number of consecutive residues increases further, the dispersion induced by structure

formation is not enough to yield isolated peaks (see Fig. 1d,f for a huntingtin construct with 46 consecutive glutamines). As a consequence of this degeneracy of frequencies, it is impossible to obtain high-resolution information of LCRs and the structural bases of their biological function cannot be unveiled.

Site-Specific Isotopic Labeling (SSIL), also named Site-Directed Isotopic Labeling (SDIL)[21], provides a solution to the signal overlap problem in LCRs by introducing a single isotopically labeled residue into the sequence, dramatically reducing the complexity of the NMR spectra. This strategy is straightforward when employing traditional solid phase peptide synthesis and commercial Boc- or Fmoc-protected isotopically labeled amino acids[22,23]. In an elegant example of SSIL, several collagen-derived peptides containing ten consecutive copies of the proline-hydroxyproline-glycine tri-peptide $(POG)_{10}$ were synthesized by moving the position of a single ^{15}N-Gly within the chain, extremely simplifying resulting NMR spectra[23]. Using this strategy, the dynamics and the effects of mutations[24,25] and phosphorylation[26] on the structure and stability of the collagen triple helix were investigated. However, the systematic application of SSIL using peptide synthesis strategies presents several limitations, such as the cost of the protected isotopically labeled amino acids, the length of the peptides amenable to peptide synthesis, and the presence of contaminants when this length increases. In this *Concept* article we will describe the coupling of SSIL to the biochemical production of proteins as an efficient strategy to overcome the above-mentioned limitations. Then, some examples of potential applications of SSIL to address biomedically relevant questions and the remaining challenges of the approach will be discussed.

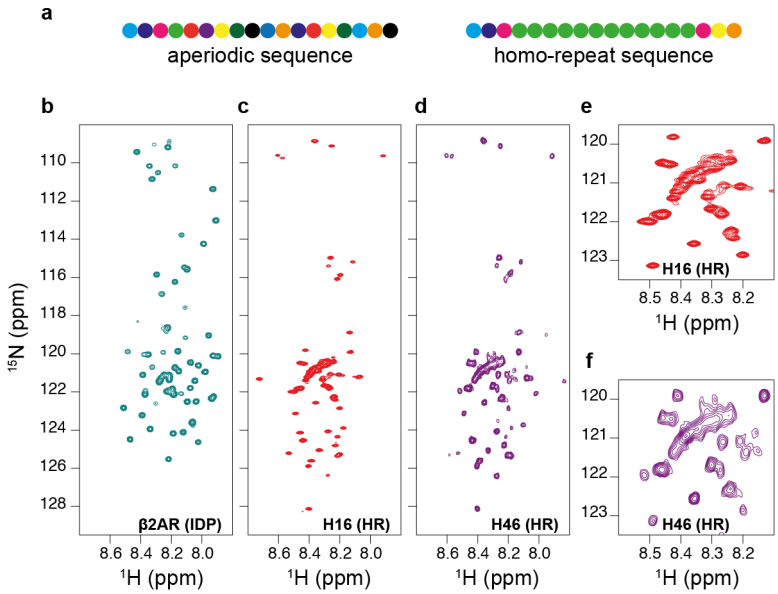


Figure 1. Compositional bias and frequency dispersion in NMR. (a) *Cartoon representation of an aperiodic compositionally unbiased sequence (left) and a homo-repeat (right). Amino acid types are represented with different colors.* ***(b)*** *^{15}N-HSQC spectrum of the 79-amino acid long C-terminal tail of the β2 GPCR, an IDP with an unbiased composition, displaying a good peak dispersion that enables straightforward NMR frequency assignment. ^{15}N-HSQC spectra of huntingtin exon1 containing 16* ***(c)*** *and 46* ***(d)*** *consecutive glutamines in the Poly-Q homo-repeat. A zoom of the glutamine region for both huntingtin versions is displayed in panels* ***(e)*** *and* ***(f)*** *with the same color code as in panels* ***(c)*** *and* ***(d)****. Notice that the capacity to discriminate between the different glutamine peaks is compromised when the length of the homo-repeat increases.*

Nonsense suppression for unnatural and natural amino acids

The general method for incorporating non-canonical (or unnatural) amino acids (ncAAs) into proteins *in vivo* is based on the nonsense suppression strategy[27,28]. This methodology uses a nonsense suppressor $tRNA_{CUA}$ that recognizes the amber stop codon (UAG) in the protein-encoding mRNA and elongates the nascent peptide chain with the ncAA previously loaded to the $tRNA_{CUA}$. To charge the $tRNA_{CUA}$ with the desired ncAA, several mutant aminoacyl-tRNA synthetases (mut-aaRSs) have been engineered to selectively recognize the ncAA and the corresponding $tRNA_{CUA}$[29]. To prevent non-specific charging of the tRNAs, these $tRNA_{CUA}$/mut-aaRS pairs have to be orthogonal to all endogenous tRNA/aaRS pairs of the host cells. In practice, *E. coli* cells are co-transformed with a plasmid coding for the $tRNA_{CUA}$/mut-aaRS pair, and a second plasmid, coding for the protein of interest, containing an amber stop codon at the desired position. Cells are then grown, and upon induction in presence of the unnatural amino acid, the ncAA is incorporated into the protein.

In order to apply this strategy to introduce isotopically labeled natural (or canonical) amino acids, the process of $tRNA_{CUA}$ loading has to be done in isolation *in vitro* (see below), because aaRSs cannot distinguish between an isotopically labeled and a non-labeled natural amino acid. The loaded $tRNA_{CUA}$ is then added to a cell-free (CF) system where the target protein is produced. CF is an *in vitro* protein synthesis method composed of a lysate containing the transcription-translation machinery, normally obtained from *Escherichia coli* cultures, which is supplemented with amino acids, nucleotides, salts and an energy regeneration system (Fig. 2)[20,21,28,30,31]. It is important to mention that, once used, the $tRNA_{CUA}$ cannot be reloaded inside the CF reaction, making this process a one-off reaction.

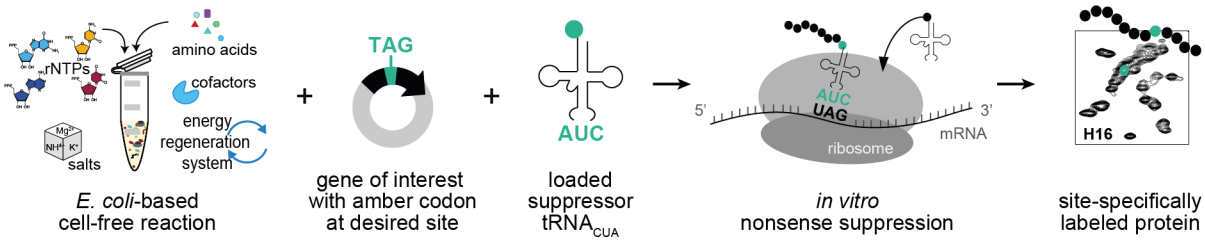


Figure 2. Scheme of the cell-free synthesis for SSIL. *A combination of three elements produces the desired sample where a single isotopically amino acid is incorporated into the protein, dramatically reducing the complexity of the NMR spectra. The* in vitro *cell-free reaction, containing the* E. coli *transcriptional and translational machineries, is supplied with amino acids, nucleotides, an energy regeneration system and other chemicals for efficient protein production. A second element required for SSIL is a plasmid coding for the protein of interest with an amber stop codon (TAG) in the position that is to be structurally investigated. Finally, a tRNA with the appropriate anticodon (CUA) and the isotopically labeled canonical amino acid (represented by a green dot) is also added to the CF reaction. This figure was adapted from Urbanek et al.[20].*

Strategies to load $tRNA_{CUA}$

Three different strategies have been described to load the $tRNA_{CUA}$ *in vitro* (Fig. 3):

The semisynthetic approach: This strategy uses an *in vitro* translated, truncated $tRNA_{CUA}$, lacking the last two nucleotides of the 3'-end. Then, a chemically synthesized aminoacylated dinucleotide is ligated to the truncated $tRNA_{CUA}$ with a T4-RNA ligase (Fig. 3a)[28,32]. This method is very versatile and can be used to attach any type of amino acid to any type of $tRNA_{CUA}$. It has been widely used in the past, prior to the emergence of the *in vivo* system. In a pioneering example, the semisynthetic approach was used to

load a $tRNA_{CUA}$ with a ^{13}C-Alanine, yielding a highly simplified ^{13}C-filtered ^{1}H spectrum of T4 lysozyme in native and denaturing conditions[30]. More recently, the Green Fluorescent Protein (GFP) was produced with a single ^{13}C,^{18}O-tyrosine[31]. The sample enabled the recording of time-resolved infrared absorption spectra in a site-specific manner and yielded novel information on the photodynamics of GFP. Despite these successful examples, when aiming to use isotopically labeled canonical amino acids, the synthesis of large quantities of the aminoacylated dinucleotides presents an important bottleneck.

Flexizyme: This method uses *in vitro* transcribed full-length $tRNA_{CUA}$, an excess of flexizyme (a specifically designed ribozyme), and an activated version of the desired amino acid, for example the dinitrobenzylester derivative (Fig. 3b)[33]. Similarly to the semisynthetic approach, the flexizyme strategy is very versatile in terms of chosen amino acids and tRNAs, although the acylation times and efficiencies are highly amino acid dependent, ranging from 2-72 hours and 17-91%, respectively[33]. The chemical activation of the amino acid is probably the main limitation for the application of this methodology in SSIL.

Use of purified aminoacyl-synthetases: Here the *in vitro* transcribed full-length $tRNA_{CUA}$ is mixed with catalytic amounts of the cognate aaRS, the isotopically labeled amino acid and a buffer containing ATP, resulting in nearly full turnover in less than one hour (Fig. 3c). The most important advantage of this methodology with respect to the previously mentioned ones is the direct use of any commercially available isotopically labeled natural amino acid. As previously mentioned, the $tRNA_{CUA}$/aaRS pair has to be orthogonal to the CF system used, and pairs from other organisms may be adapted for this reason. The relevance of this point is exemplified by the pioneering study by Yabuki *et al.* [21], who managed to isotopically label a single tyrosine in a protein by SSIL using the endogenous *E. coli* TyrRS. However, as a consequence of the lack of orthogonality, the isotopic labeling yield was only 50%, and the reaction time had to be optimized to minimize the undesired loading of the $tRNA_{CUA}$ with unlabeled tyrosine.

Overall, all these methods allow the specific aminoacylation of the chosen $tRNA_{CUA}$. Whereas the semisynthetic and the flexizyme approach may be preferable when applied to novel ncAAs without their corresponding engineered mut-aaRSs, enzymatic

aminoacylation is more efficient and cost effective when incorporating isotopically labeled natural amino acids.

An important chemical feature of loaded $tRNA_{CUA}$ is the high susceptibility to hydrolysis of the phosphoester bond between the $tRNA_{CUA}$ and the amino acid[34]. As a consequence of this lability, the yield of suppressed sample with respect to the total input of loaded $tRNA_{CUA}$ is relatively low[20,31]. This inherent limitation has to be addressed in the future for the efficient application of SSIL.

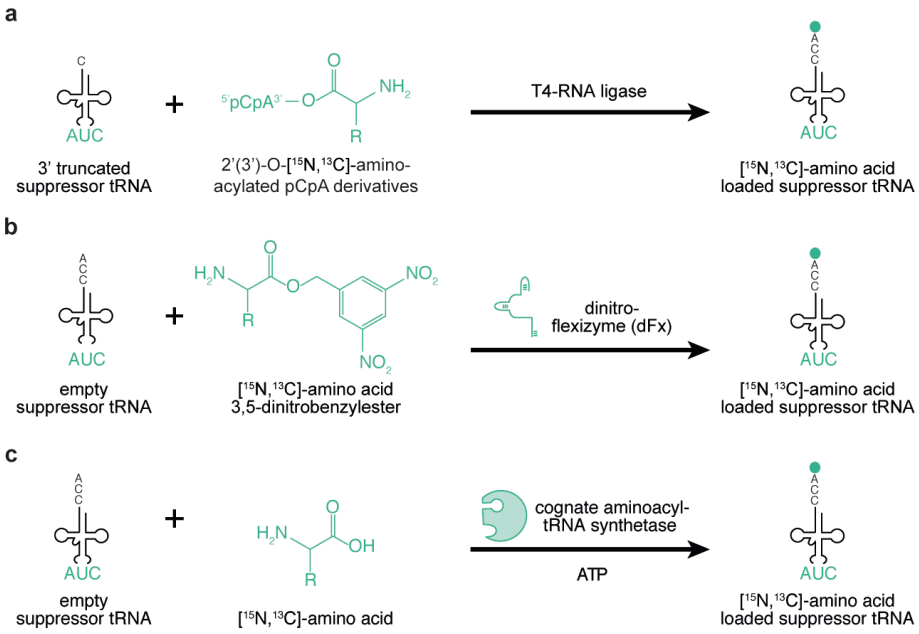


Figure 3. tRNA loading strategies. (a) *Semisynthetic approach. A truncated $tRNA_{CUA}$, lacking the last two nucleotides of the 3'-end is incubated with a chemically synthesized aminoacylated dinucleotide (cytidylyl (3'→5') adenosine 5'-phosphates, pCpA) and T4 RNA ligase, resulting in a full-length aminoacylated $tRNA_{CUA}$.* ***(b)*** *Flexizyme-mediated loading of $tRNA_{CUA}$. Dinitro-flexizyme (dFx) utilizes 3,5-dinitrobenzyl ester-activated amino acids and catalyzes the loading of $tRNA_{CUA}$.* ***(c)*** *The aminoacyl-synthetase charges the suppressor $tRNA_{CUA}$ with its cognate amino acid. R represents in principle any kind of side chain. Note that the final product of the three strategies is the same molecule.*

Application of SSIL to study the glutamine homo-repeat in huntingtin

We have recently demonstrated the use of SSIL to structurally characterize low complexity regions using the huntingtin protein as model protein[20]. The N-terminal region of huntingtin, also known as exon1 (httex1), is the causative agent of Huntington's Disease (HD), a deadly neurodegenerative pathology[35]. Interestingly, HD develops when the number of consecutive glutamines in httex1 exceeds 35. The

structural bases of this pathological threshold remain poorly understood due to the inherent problems that homo-repeats pose to traditional NMR approaches. Using a GlnRS/$tRNA_{CUA}$ pair derived from yeast[36], five glutamines of an httex1 construct with 16 consecutive glutamines (H16) were isotopically labeled with yields that allowed the NMR investigation (≈10 µM from 5 mL of CF reaction)[20]. The resulting samples provided isolated peaks in the highly congested Poly-Q region of the ^{15}N-HSQC, as well as in the side chain region (Fig. 4a,b). Using ^{13}C-HSQC experiments, the Cα and Cβ chemical shifts of the individual glutamines were also precisely measured, suggesting that the Poly-Q tract is enriched in α-helical conformations, in line with previous studies[19,37] (Fig. 4c,d,f). Interestingly, the ^{15}N-HSQC of Q48 displays a second low intensity peak that was attributed to the *cis* conformation of the neighboring proline 49 (Fig. 4e). The strategy was also successfully applied to a pathological httex1 construct with 46 consecutive glutamines, demonstrating the generality of the approach. Although only few positions were explored in this study, it demonstrates that the systematic structural investigation of homo-repeats is feasible.

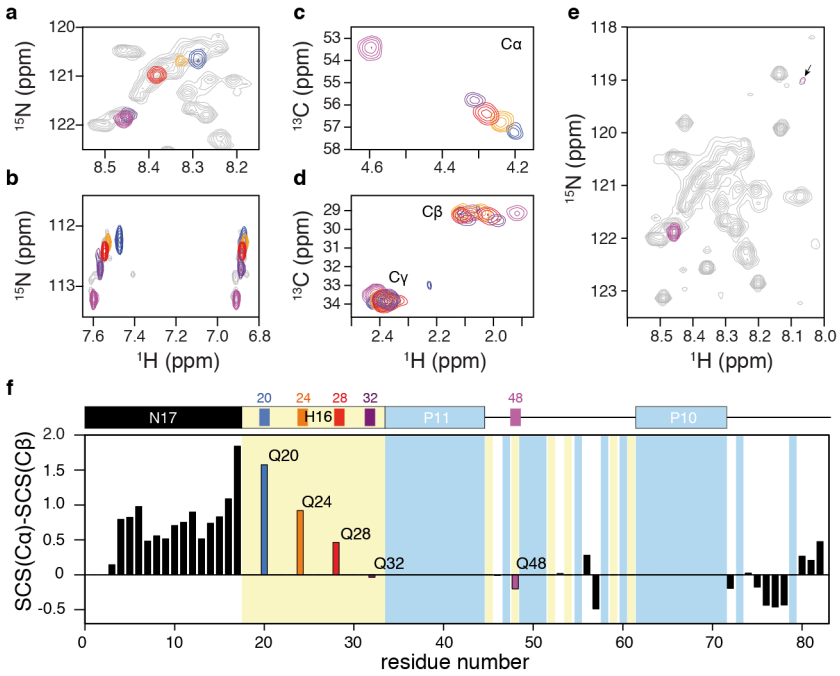


Figure 4. Application of SSIL to study the glutamine HR in huntingtin protein using NMR. (a) *Overlay of the ^{15}N-HSQC spectrum of fully labeled H16 (grey) with those of five SSIL samples: Q20 (blue), Q24 (orange), Q28 (red), Q32 (purple) and Q48 (magenta), focusing on the Poly-Q region. The same color code is used throughout the figure.* ***(b)*** *Zoomed view of the ^{15}N-HSQC spectra showing the glutamine side chains. ^{13}C-HSQC NMR*

spectra showing the Cα ***(c)*** *and the Cβ and Cγ* ***(d)*** *regions.* ***(e)*** *Overlay of the ^{15}N-HSQC spectra of fully labeled H16 and Q48. The arrow indicates the population of Q48 preceding the* cis *conformation of proline 49.* ***(f)*** *Secondary chemical shift analysis using the extracted Cα and Cβ chemical shifts and a random-coil library. A schematic representation of httex1 is shown to indicate the different domains, and the positions of glutamine and proline residues are highlighted in yellow and blue, respectively. This figure was adapted from Urbanek et al.[20].*

Applications of SSIL

SSIL represents a unique tool to access high-resolution information of biological systems that cannot be characterized by traditional means. Here, we briefly describe two cases that exemplify the kind of questions that could be addressed with SSIL.

Functional and pathological roles of homo-repeats (HRs): HRs, which are abundant in eukaryotes[14,15,38], represent the most obvious target for the application of SSIL. The structural features of these homopolymeric regions are largely unknown and possible structure/function relationships remain to be deciphered. This is especially relevant for a family of diseases linked to the pathological expansion of Poly-Q[39,40] and Poly-A[41] tracts, which lead to the formation of irreversible aggregates[42]. Structural perturbations exerted by these abnormal expansions are so far unknown, limiting our understanding of the underlying pathological mechanism and their potential remediation[43].

Low Complexity Regions and Liquid-Liquid Phase Separation: A large body of examples has identified LCRs inserted in some IDPs as the key elements for the formation of membrane-less compartments in cells through a liquid-liquid phase separation (LLPS) mechanism, a phenomenon that has been shown to be fundamental in a myriad of biological processes[44,45][46]. Although not exclusively, droplet-forming sequences include RG/RGG, FG, VPGV, Y-containing sequences and alternating charged blocks, which are able to establish multiple intermolecular low-affinity interactions. Several studies have focused on connecting protein sequences to their capacity to phase-separate[47]. Unfortunately, the role of protein structure in triggering or modulating this property is less clear. These structural investigations, too, are impaired by the congestion of the NMR spectra of LLPS-inducing LCRs that worsens within droplets[48–50].

Future challenges of SSIL

In order to address the structural questions described in the previous sections, several challenges will have to be addressed. In this section we will describe those that we think are the most relevant.

Expanding the panel of amino acids amenable to SSIL: The high throughput application of SSIL for structural purposes requires the use of scalable and highly efficient enzymatic $tRNA_{CUA}$ loading which, for the moment, has only been developed for glutamine[20]. Already existing orthogonal $tRNA_{CUA}$/aaRS pairs (for example for aspartate[51], glutamate[52], leucine[53], proline[54] and tyrosine[55]) may be adapted to this purpose. In the future, orthogonal pairs for the remaining amino acids will have to be found and validated if we want to expand the range of applications of SSIL. The use of promiscuous aaRSs with the capacity to load similar natural amino acids[56], which are present in some organisms, can be a simpler way to expand the panel of amino acids for SSIL.

Increasing the number of labeled sites: Present approaches only allow the incorporation of a single labeled residue within a protein chain. First, there is a limited amount of loaded $tRNA_{CUA}$ present in the reaction mixture due to the spontaneous deacylation of the $tRNA_{CUA}$ (see above). Note that for ncAAs, the engineered aaRS can be added to the reaction mixture, but this is not possible when the aim is to incorporate natural amino acids. Second, there is a systematic decrease of the translation yield when incorporating an amino acid from an orthogonal $tRNA_{CUA}$. This is mainly due to stalling, which increases the probability of ribosome disassembly, and translation abortion by the competitive action of release factor 1 (RF1). Several groups have proposed approaches to inactivate or remove RF1 from cell lysates (see ref [57] and references therein). For example, Loscha *et al*. generated an RF1-chitin binding domain chimera that can be removed during lysate preparation, which enabled the incorporation of 4-trifluoro-methyl phenylalanine in four positions of a protein[57]. Alternatively, *E. coli* strains with RF1 deletions and other complementing mutations have been engineered to optimize the incorporation of multiple ncAAs without compromising the final yield[58–60]. However, these strategies were validated with ncAAs and their corresponding mut-aaRS in the CF mixture. It remains to be demonstrated whether this strategy will be suitable

for the incorporation of isotopically labeled amino acids into multiple protein sites in a one-off reaction.

Incorporation of novel ncAAs: The extensive panel of ncAAs and their corresponding engineered mut-aaRSs that have been developed for *in vivo* incorporation[27] in the last decades can be directly added to CF systems as recently demonstrated for the incorporation of L-phosphoserine[61,62]. If the loading reaction is performed separately, as discussed previously, semisynthetic and flexizyme strategies offer nearly unlimited possibilities regarding the nature of the ncAA used. This is due to the fact that the latter strategies are only limited by the stability of the ncAAs under the conditions of the chemical reactions necessary to modify the amino acid and load the tRNAs. The use of cognate aaRSs to load ncAAs is more limited but offers very interesting possibilities. AaRS have evolved to be extremely specific for a single natural amino acid, thus the tolerance of structural modifications is restricted and only small chemical changes may be allowed. We anticipate that modifications such as halogenation (especially with fluorine), hydroxylation, or methylation could be incorporated into proteins using SSIL without engineering the aaRSs. In particular, the complete control over the experimental conditions of the loading reaction, including time, temperature, pH and concentration, may help to load such ncAAs onto the $tRNA_{CUA}$. Since the SSIL CF reaction can be tuned and scaled-up easily, a loading of the $tRNA_{CUA}$ of 100% is not necessary to obtain reasonable amounts of labeled protein (unpublished data). Some of these modifications are extremely interesting from biological and biophysical perspectives. For example, some amino acids are post-translationally methylated (arginine and lysine) or hydroxylated (proline and lysine), playing important roles in signaling and regulation. Moreover, fluorine is emerging as a very interesting probe to study protein structure and biomolecular interactions, thanks to the very interesting features of ^{19}F-NMR[63,64].

Concluding Remarks

Recent developments in structural biology allow the detailed study of biomolecules and biological processes that were not possible before. However, there are still families of biomolecules, such as LCRs, that cannot be structurally and dynamically characterized with the present technology. SSIL, which allows the placement of isotopes in specific positions within proteins, will enable the atomistic description of LCRs by NMR. The

general availability of high-magnetic fields and commercial cold-probes, which enhance NMR sensitivity, makes SSIL a very timely development that will complement knowledge derived from traditional NMR methodologies. We believe that use of SSIL will allow researchers to establish the structure/function relationships for biological systems that could not be addressed before. This way, we expect to reach a deeper understanding of biological processes with important medical and biotechnological relevance. However, the generalization of SSIL must be necessarily accompanied by the implementation of the appropriate technology in research laboratories. Surpassing the challenges described in this *Concept* article will expand the range of applications of SSIL and facilitate its generalization.

Acknowledgements

This work was supported by the European Research Council under the European Union's H2020 Framework Programme (2014-2020) / ERC Grant agreement n° [648030], and Labex EpiGenMed, an « Investissements d'avenir » program (ANR-10-LABX-12-01) awarded to PB. The CBS is a member of France-BioImaging (FBI) and the French Infrastructure for Integrated Structural Biology (FRISBI), 2 national infrastructures supported by the French National Research Agency (ANR-10-INBS-04-01 and ANR-10-INBS-05, respectively).

References

[1] S. Ohki, M. Kainosho, *Prog. Nucl. Magn. Reson. Spectrosc.* **2008**, *53*, 208–226.

[2] K. Ozawa, P. S. C. Wu, N. E. Dixon, G. Otting, *FEBS J.* **2006**, *273*, 4154–4159.

[3] A. Meola, C. Deville, S. A. Jeffers, P. Guardado-Calvo, I. Vasiliauskaite, C. Sizun, C. Girard-Blanc, C. Malosse, C. van Heijenoort, J. Chamot-Rooke, T. Krey, E. Guittet, S. Pêtres, F. A. Rey, F. Bontems, *J. Struct. Biol.* **2014**, *188*, 71–78.

[4] B. Hoffmann, F. Löhr, A. Laguerre, F. Bernhard, V. Dötsch, *Prog. Nucl. Magn. Reson. Spectrosc.* **2018**, *105*, 1–22.

[5] M. Sattler, J. Schleucher, C. Griesinger, *Prog. Nucl. Magn. Reson. Spectrosc.* **1999**, *34*, 93–158.

[6] H. J. Dyson, P. E. Wright, *Chem. Rev.* **2004**, *104*, 3607–3622.

[7] R. L. Narayanan, U. H. N. Dürr, S. Bibow, J. Biernat, E. Mandelkow, M. Zweckstetter, *J. Am. Chem. Soc.* **2010**, *132*, 11906–11907.

[8] M. R. Jensen, M. Zweckstetter, J. Huang, M. Blackledge, *Chem. Rev.* **2014**, *114*, 6632–6660.

[9] I. C. Felli, R. Pierattelli, *J. Magn. Reson.* **2014**, *241*, 115–125.

[10] J. Nováček, L. Žídek, V. Sklenář, *J. Magn. Reson.* **2014**, *241*, 41–52.

[11] K. Grudziąż, A. Zawadzka-Kazimierczuk, W. Koźmiński, *Methods* **2018**, *148*, 81–87.

[12] J. C. Wootton, *Comput. Chem.* **1994**, *18*, 269–285.

[13] J. C. Wootton, S. Federhen, *Methods Enzymol.* **1996**, *266*, 554–571.

[14] J. Jorda, A. V Kajava, *Adv. Protein Chem. Struct. Biol.* **2010**, *79*, 59–88.

[15] M. Y. Lobanov, O. V Galzitskaya, *Mol. Biosyst.* **2012**, *8*, 327–337.

[16] M. A. Andrade, C. Perez-Iratxeta, C. P. Ponting, *J. Struct. Biol.* **2001**, *134*, 117–131.

[17] M. H. Schaefer, E. E. Wanker, M. A. Andrade-Navarro, *Nucleic Acids Res.* **2012**, *40*, 4273–4287.

[18] B. Eftekharzadeh, A. Piai, G. Chiesa, D. Mungianu, J. García, R. Pierattelli, I. C. Felli, X. Salvatella, *Biophys. J.* **2016**, *110*, 2361–2366.

[19] M. Baias, P. E. S. Smith, K. Shen, L. A. Joachimiak, S. Żerko, W. Koźmiński, J. Frydman, L. Frydman, *J. Am. Chem. Soc.* **2017**, *139*, 1168–1176.

[20] A. Urbanek, A. Morató, F. Allemand, E. Delaforge, A. Fournet, M. Popovic, S. Delbecq, N. Sibille, P. Bernadó, *Angew. Chem. Int. Ed. Engl.* **2018**, *57*, 3598–3601.

[21] T. Yabuki, T. Kigawa, N. Dohmae, K. Takio, T. Terada, Y. Ito, E. D. Laue, J. A. Cooper, M. Kainosho, S. Yokoyama, *J. Biomol. NMR* **1998**, *11*, 295–306.

[22] R. Warrass, J.-M. Wieruszeski, C. Boutillon, G. Lippens, *J. Am. Chem. Soc.* **2000**, *122*, 1789–1795.

[23] A. M. Acevedo-Jake, A. A. Jalan, J. D. Hartgerink, *Biomacromolecules* **2015**, *16*, 145–155.

[24] A. M. Acevedo-Jake, K. A. Clements, J. D. Hartgerink, *Biomacromolecules* **2016**, *17*, 914–921.

[25] K. A. Clements, A. M. Acevedo-Jake, D. R. Walker, J. D. Hartgerink, *Biomacromolecules* **2017**, *18*, 617–624.

[26] A. M. Acevedo-Jake, D. H. Ngo, J. D. Hartgerink, *Biomacromolecules* **2017**, *18*, 1157–1161.

[27] C. C. Liu, P. G. Schultz, *Annu. Rev. Biochem.* **2010**, *79*, 413–444.

[28] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182–188.

[29] L. Wang, P. G. Schultz, *Angew. Chem. Int. Ed. Engl.* **2005**, *44*, 34–66.

[30] J. A. Ellman, B. F. Volkman, D. Mendel, P. G. Schultz, D. E. Wemmer, *J. Am. Chem. Soc.* **1992**, *114*, 7959–7961.

[31] S. Peuker, H. Andersson, E. Gustavsson, K. S. Maiti, R. Kania, A. Karim, S. Niebling, A. Pedersen, M. Erdelyi, S. Westenhoff, *J. Am. Chem. Soc.* **2016**, *138*, 2312–2318.

[32] T. G. Heckler, L. H. Chang, Y. Zama, T. Naka, M. S. Chorghade, S. M. Hecht, *Biochemistry* **1984**, *23*, 1468–1473.

[33] Y. Goto, T. Katoh, H. Suga, *Nat. Protoc.* **2011**, *6*, 779–790.

[34] J. R. Peacock, R. R. Walvoord, A. Y. Chang, M. C. Kozlowski, H. Gamper, Y.-M. Hou, *RNA* **2014**, *20*, 758–764.

[35] F. O. Walker, *Lancet (London, England)* **2007**, *369*, 218–228.

[36] D. R. Liu, P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96*, 4780–4785.

[37] E. A. Newcombe, K. M. Ruff, A. Sethi, A. R. Ormsby, Y. M. Ramdzan, A. Fox, A. W. Purcell, P. R. Gooley, R. V Pappu, D. M. Hatters, *J. Mol. Biol.* **2018**, *430*, 1442–1458.

[38] P. Mier, G. Alanis-Lobato, M. A. Andrade-Navarro, *Proteins* **2017**, *85*, 709–719.

[39] J. Shao, M. I. Diamond, *Hum. Mol. Genet.* **2007**, *16*, 115–123.

[40] A. J. Williams, H. L. Paulson, *Trends Neurosci.* **2008**, *31*, 521–528.

[41] J. Amiel, D. Trochet, M. Clément-Ziza, A. Munnich, S. Lyonnet, *Hum. Mol. Genet.* **2004**, *13 Spec No*, R235-243.

[42] A. L. Darling, V. N. Uversky, *Molecules* **2017**, *22*, 2027.

[43] X. Feng, S. Luo, B. Lu, *Trends Biochem. Sci.* **2018**, *43*, 424–435.

[44] L.-P. Bergeron-Sandoval, N. Safaee, S. W. Michnick, *Cell* **2016**, *165*, 1067–1079.

[45] V. N. Uversky, *Curr. Opin. Struct. Biol.* **2017**, *44*, 18–30.

[46] Y.-H. Lin, J. D. Forman-Kay, H. S. Chan, *Biochemistry* **2018**, *57*, 2499–2508.

[47] R. M. Vernon, J. D. Forman-Kay, *Curr. Opin. Struct. Biol.* **2019**, *58*, 88–96.

[48] K. A. Burke, A. M. Janke, C. L. Rhine, N. L. Fawzi, *Mol. Cell* **2015**, *60*, 231–241.

[49] J. P. Brady, P. J. Farber, A. Sekhar, Y.-H. Lin, R. Huang, A. Bah, T. J. Nott, H. S. Chan, A. J. Baldwin, J. D. Forman-Kay, L. E. Kay, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E8194–E8203.

[50] S. E. Reichheld, L. D. Muiznieks, F. W. Keeley, S. Sharpe, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E4408–E4415.

[51] M. Pastrnak, T. J. Magliery, P. G. Schultz, *Helv. Chim. Acta* **2000**, *83*, 2277–2286.

[52] S. W. Santoro, J. C. Anderson, V. Lakshman, P. G. Schultz, *Nucleic Acids Res.* **2003**, *31*, 6700–6709.

[53] J. C. Anderson, P. G. Schultz, *Biochemistry* **2003**, *42*, 9598–9608.

[54] A. Chatterjee, H. Xiao, P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109*, 14841–14846.

[55] S. Ohno, T. Yokogawa, I. Fujii, H. Asahara, H. Inokuchi, K. Nishikawa, *J. Biochem.* **1998**, *124*, 1065–1068.

[56] M. A. Swairjo, P. R. Schimmel, *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 988–993.

[57] K. V Loscha, A. J. Herlt, R. Qi, T. Huber, K. Ozawa, G. Otting, *Angew. Chem. Int. Ed. Engl.* **2012**, *51*, 2243–2246.

[58] M. J. Lajoie, A. J. Rovner, D. B. Goodman, H.-R. Aerni, A. D. Haimovich, G. Kuznetsov, J. A. Mercer, H. H. Wang, P. A. Carr, J. A. Mosberg, N. Rohland, P. G. Schultz, J. M. Jacobson, J. Rinehart, G. M. Church, F. J. Isaacs, *Science* **2013**, *342*, 357–360.

[59] S. H. Hong, I. Ntai, A. D. Haimovich, N. L. Kelleher, *ACS Synth. Biol.* **2014**, 1–6.

[60] R. W. Martin, B. J. Des Soye, Y. Kwon, J. Kay, R. G. Davis, P. M. Thomas, N. I. Majewska, C. X. Chen, R. D. Marcum, M. G. Weiss, A. E. Stoddart, M. Amiram, A. K. Ranji Charna, J. R. Patel, F. J. Isaacs, N. L. Kelleher, S. H. Hong, M. C. Jewett, *Nat. Commun.* **2018**, *9*, 1203.

[61] H.-S. Park, M. J. Hohn, T. Umehara, L.-T. Guo, E. M. Osborne, J. Benner, C. J. Noren, J. Rinehart, D. Söll, *Science* **2011**, *333*, 1151–1154.

[62] J. P. Oza, H.-R. Aerni, N. L. Pirman, K. W. Barber, C. M. ter Haar, S. Rogulina, M. B. Amrofell, F. J. Isaacs, J. Rinehart, M. C. Jewett, *Nat. Commun.* **2015**, *6*, 8168.

[63] H. Chen, S. Viel, F. Ziarelli, L. Peng, *Chem. Soc. Rev.* **2013**, *42*, 7971–7982.

[64] H. Arthanari, K. Takeuchi, A. Dubey, G. Wagner, *Curr. Opin. Struct. Biol.* **2019**, *58*, 294–304.