From Chemical Mutagenesis to Post-Expression Mutagenesis: A 50 Year Odyssey

Tom H. Wright, M. Robert J. Vallée, and Benjamin G. Davis*

*S*ite-directed (gene) mutagenesis has been the most useful method available for the conversion of one amino acid residue of a given protein into another. Until relatively recently, this strategy was limited to the twenty standard amino acids. The ongoing maturation of stop codon suppression and related technologies for unnatural amino acid incorporation has greatly expanded access to nonstandard amino acids by expanding the scope of the translational apparatus. However, the necessity for translation of genetic changes restricts the diversity of residues that may be incorporated. Herein we highlight an alternative approach, termed post-expression mutagenesis, which operates at the level of the very functional biomolecules themselves. Using the lens of retrosynthesis, we highlight prospects for new strategies in protein modification, alteration, and construction which will enable protein science to move beyond the constraints of the "translational filter" and lead to a true synthetic biology.

1. Introduction

The detailed, mechanistic interrogation of protein function with residue-level precision requires methods for the conversion of one amino-acid side chain into another (one residue into another) and the assembly of proteins which contain such precise changes. To date, site-directed gene mutagenesis has provided the primary technology for achieving this goal.[1] Gene-level mutagenesis is founded on the central dogma of Francis Crick, which describes the directional flow of information in biological macromolecules from DNA to protein.[2] By using this form of site-directed mutagenesis, information is modified at the nucleotide level to ultimately effect a change in protein sequence, and hence at the functional biomolecule level. Through the suppression of certain codons[3] with tRNAs loaded with unnatural amino acids, unnatural amino acids can be introduced at a specific position of a protein, for example, by stop codon positioning[4]

if a specific, paired tRNA-tRNA synthetase system can be engineered.[5] This approach is an exciting and strikingly powerful method for certain structural motifs which are well-tolerated by the corresponding biosynthetic systems (e.g., synthetases, ribosome). However, one shortcoming of such (stop)codon suppression methodology is the lack of chemical diversity in the residues beyond those that can be incorporated with current technology, largely reflecting the evolved stringency of the amino-acid tRNA synthetase proteins. Most amino acids that have been incorporated thus far resemble lysine or tyrosine (by stop-codon) or methionine (by Met-codon) derivatives, thus reflecting the natural propensities of the most common synthetases that have been exploited.[3] Notably, there have been no synthetases reported for the direct incorporation of certain residues that are central to much current biological exploration. Such residues are the methylated derivatives of lysine, glutamine, and arginine, which are widely implicated in epigenetics or for any glycosylated amino acids. Indeed, whilst there have been some elegant indirect strategies adopted,[6] the difficulty of engineering a synthetase that can either discriminate a single methyl group and solely charge the modified residue, or accommodate even a single sugar on a side chain (let alone the larger complex structures found in glycobiology) is daunting. Despite these shortcomings, the success of these technologies is illustrated by their daily use in countless labs around the world.

In some respects, approaches such as these, which are transmitted (transcribed and translated = expressed) from the nucleotide level, are a fundamentally circuitous means of

[*] T. H. Wright, Dr. M. R. J. Vallée, Prof. B. G. Davis
    Department of Chemistry, University of Oxford, Chemistry Research
    Laboratory, Mansfield Road, OX1 3TA (UK)
    E-mail: ben.davis@chem.ox.ac.uk
    Homepage: http://users.ox.ac.uk/~dplb0149

achieving what is really sought after—a change in a specific side chain at the protein level. In principle, a method for post-translational or post-expression mutagenesis would be able to accomplish this transformation directly, at the level of the protein itself. Such a method would not be constrained by the evolutionary "filter" of the translational apparatus and would enable site-selective installation of any desired side chain, natural or designed, straight into the biomolecules that actually directly determine function in biology (the proteins and even the sugars and lipids).

In nature too, post-translational modification (PTM) is the key means by which protein side-chain diversity is extended beyond the proteinogenic amino acids. Chemically simple modifications such as methylation, and more complex structures, such as the attachment of sugars to glycoproteins, have profound consequences for biological processes in all kingdoms of life. The biochemical study of PTMs has been hampered by the lack of general and robust methods for their selective installation to a protein. The post-translationally modified side chains pose a distinct challenge for gene-level mutagenesis. An alternative approach to an expanded muta-genesis could draw synthetic inspiration from this natural strategy of PTM (Scheme 1, disconnection I) and modify proteins directly, without genetic mutation, in a test tube and ultimately inside cells and organisms.

A number of biochemists, in the mid-to-late 1960s, who were working quietly before the biotechnological revolution of molecular biology and amongst the more vocal backdrop of peptide synthesis, envisaged a synthetic approach to muta-genesis as a valuable tool for probing protein function. This work, which led to the first far-sighted conversion of one amino acid into another through the chemical alteration of a residue side chain (from Ser-to-Cys),[7] demonstrated a key principle. In this way, a powerful conceptual legacy for a more general method that could be applied directly to biomolecules was established. Indeed, this approach enabled the first point (site-directed protein) mutations of an enzyme, by any method, by using just such a synthetic, chemical approach. Polgar and Bender used the term "simulated mutations" and described a hypothetical method: "……by chemical or enzymatic means it is possible to substitute one amino acid of the protein molecule by another, simulating the effect of mutation."[7b]

2. The Early Years: β,γ-Carbon–Heteroatom Bond Formation

We can trace the first roots of the chemical mutagenesis concept to a 1965 paper from Wilchek et al., in which serine residues were quantitatively converted into cysteines in polypeptides (Scheme 2a).[7a] Serine tosylation and subsequent S_N2 displacement enabled the synthesis of both cysteine (reaction with thioacetate followed by hydrolysis) and unnatural analogues (reaction with other thiol nucleophiles), thus demonstrating the advantages of the chemical approach for rapid diversification of side-chain structures. The powerful strategic leap embodied in this work can be analyzed through simple retrosynthetic analysis of a generic polypeptide

Thomas Wright obtained his B.Sc(Hons) in chemical and biological sciences from the University of Auckland in 2012 working with Margaret Brimble. For his honors work, he developed the use of thiol-ene chemistry for the synthesis of lipidated peptides as vaccine candidates. In 2013 he began his D.Phil. with Prof. Benjamin G. Davis at the University of Oxford, where he is currently developing new reactions for the chemical mutagenesis of proteins.

Dr. M. Robert J. Vallée studied chemistry at the Universität Hamburg where he did his Diploma thesis in the group of Prof. Paul Margaretha. He obtained his Ph.D. in 2013 at the Freie Universität Berlin under the guidance of Prof. Christian P. R. Hacken-berger. In 2014 he joined the group of Prof. Benjamin G. Davis as a postdoctoral re-searcher at the University of Oxford, where he worked on different projects, one of which focused on the chemical mutagenesis of proteins.

Prof. Benjamin G. Davis received his B.A. (1993) and D.Phil. (1996) at Oxford where he studied carbohydrates with George Fleet. After influential periods at the Universities of Toronto and Durham, he moved to Oxford in 2001 and was promoted to Professor in 2005. His research centers on the chemical understanding and exploitation of biomolecular function with an emphasis on carbohydrates and proteins. He is Editor-in-Chief of Current Opinion in Chemical Biology and a Senior Editor for ACS Central Science (2014). In 2015 he was elected to the Royal Society.

structure. It was the case then (and remains the case now, some would argue) that the dominant disconnection of protein synthesis has been that of the peptidic bond (Scheme 1, disconnection F and see Scheme 3). This discon-nection reflects not only nature's primary synthetic method (amino-acid ester + peptide amine = amide) but also chem-ical methods through other activated carboxy derivatives of amino acids.[8] Wilchek et al. (Scheme 1, disconnection E) ushered in the notion that other disconnections would and should prove powerful, and would offer additional synthetic flexibility in the creation of proteins.[7a]

Building on this prescient work, in 1966, the groups of Koshland and Bender independently reported the chemical transformation of one specific amino-acid side chain into another on an intact protein, that is, the first point mutation, by any method (Scheme 2b).[7b,c] In both reports, the active-site serine of subtilisin was chemically converted into a cysteine residue. By comparison, the first mutation of a protein achieved by DNA-level site-directed mutagenesis was not to be achieved until 1978.[9] Sulfonylation of the active-site serine with phenylmethanesulfonyl fluoride, to create a useful leaving group, enabled displacement by thioacetate and subsequent acetyl hydrolysis, likely through
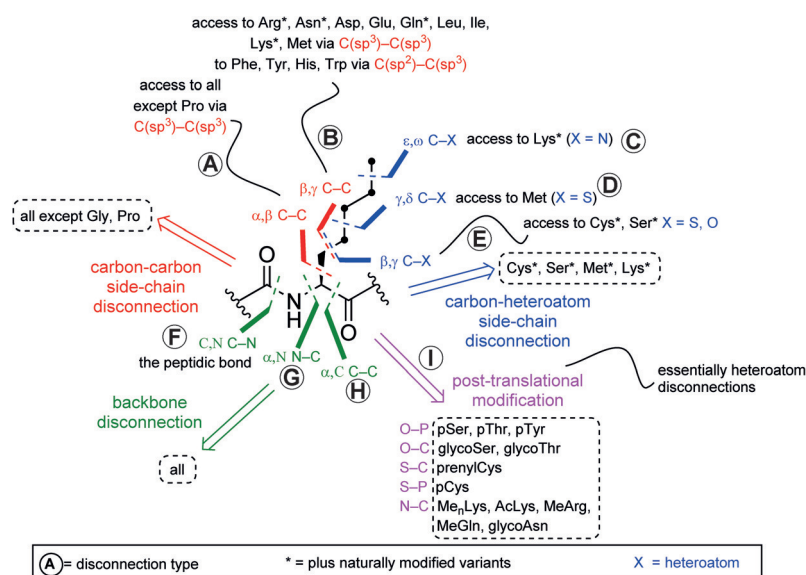
Scheme 1. Retrosynthetic analysis of protein construction.
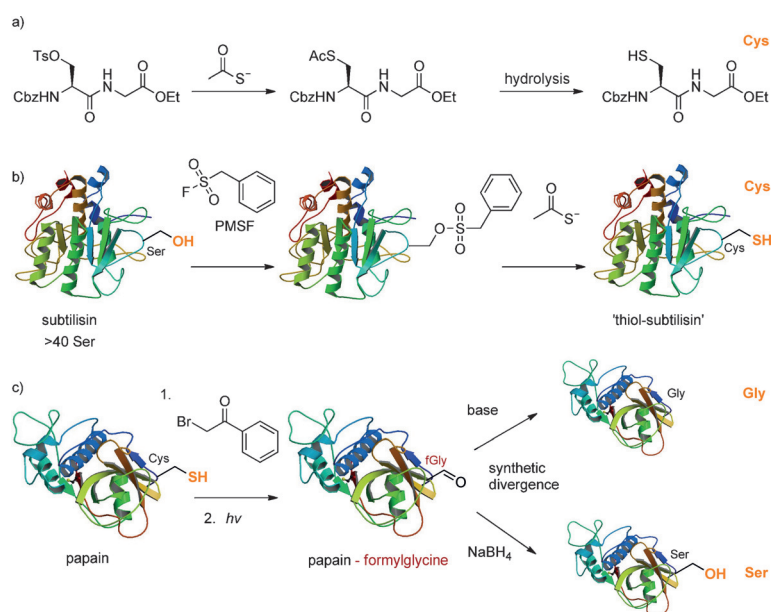
Scheme 2. Pioneering examples of β,γ-carbon–heteroatom bond formation.

the inherent activity of the enzyme. This reaction allowed the creation of a thiol-subtilisin, bearing a cysteine residue as the active-site nucleophile instead of Ser. The presence of the desired cysteine was confirmed by amino-acid analysis and colorimetric titration.

One potentially attractive factor missing in these early reports was the element of synthetic divergence from a suitable precursor. The work of Clark and Lowe[10] on chemical mutations of papain provided the first example of mutational divergence in amino-acid side-chain conversion through the creation of both Gly and Ser from a common Cys precursor (Scheme 2c). The nucleophilic cysteine of papain was chemically mutated via the intermediate formylglycine. Notably, formylglycine has itself since been recognized as a desirable residue, both as a natural PTM (derived enzymatically from Cys or Ser) mediating sulfate ester hydrolysis at the active site of type I sulfatases, and as a uniquely reactive aldehyde tag for further modification.[11] Synthetically, chemical mutation of papain by Clark and Lowe commenced with selective alkylation of the reactive active-site Cys using phenacyl bromide. Repeated photolysis was necessary to convert the phenacyl-inhibited protein into the desired thioaldehyde product. The thioaldehyde slowly hydrolyzed, thus forming formylglycine with the release of hydrogen sulfide. From formylglycine, reduction with NaBH$_4$ provided Ser while incubation at pH 9.0 for prolonged periods led to Gly by a retroaldol reaction.[10]

Whilst undeniably prescient work, it should be noted that the chemistry performed by the groups of Koshland, Bender, and Lowe exploited the enhanced nucleophilicity of the

active-site nucleophiles in subtilisin (Ser) and papain (Cys). This approach provides a salient illustration of two attributes which can together create synthetic utility: both chemo- and regioselectivity. Although, therefore, a powerful strategy for achieving regioselectivity, its necessarily limited scope (to a single active-site residue in certain enzymes) does not provide a general solution with free-ranging control of site selectivity.[12]

3. Retrosynthetic Analysis of Protein Modification

To overcome the limitations in generality of early strategies for chemical approaches to mutagenesis (lack of generality in the side chains which can be accessed and also sometimes harsh reaction conditions), the application of retrosynthetic logic, as a convenient tool for allowing one to escape occasional subjective narrowness of design,[13] helps to identify further disconnections and thence synthons. For example, through the analysis of the early work on chemical mutagenesis described in the section above, we see that these E-type disconnections (Scheme 1) deliver, in both cases, electrophilic protein synthons at Cβ matched with suitable Xγ nucleophiles. The candidate precursors that are the synthetic equivalents of such synthons are typically called tags. We have used the terminology of tag-and-modify[14] to highlight how we feel a general disconnective strategy can be developed more broadly. Such an analysis often helps to highlight useful strategic similarities in apparently different protein chemistries. Thus, for example, whilst other Cβ electrophile tag-and-modify reactions have now been explored more broadly in proteins (e.g., Cβ-Seγ: new nucleophile[15] or Dha: as a Cβ electrophile for Sγ,[16] and Seγ[17]) the potential for a Cβ nucleophile to react with a heteroatom electrophile (an inverted polarity disconnection) has not yet been described. In this Minireview, we seek to explore the potential generality of future disconnections (and putative yet unrealized reactions) further. Perhaps strangely, the tools of retrosynthetic analysis have rarely been applied to protein chemistry in a systematic fashion, at least in print. We believe that undertaking such an analysis clarifies the challenges and opportunities inherent in chemical protein modification. By examining a generic polypeptide backbone two broadly defined disconnections are immediately apparent (Scheme 1): amide backbone (disconnections F–H) and side-chain disconnections (disconnections A–E and I).

3.1. Side-Chain Heteroatom–Carbon Disconnections: Further Developments

In nature, both of these amide backbone and side-chain disconnections are exploited and accomplished synthetically, essentially solely by carbon–heteroatom bond formation (Scheme 1, disconnections F and I). Hence, both nature and the early work in chemical mutagenesis exploit(ed) useful natural polarities and an oft-touted[18] cue to reactivity. This polarity can also provide benefits in chemoselectivity by allowing matched selectivities which are highly valuable in

the sea[19] of competing functional groups that can be found in nature (or even in a single biomolecule). Such an intent to use carbon–heteroatom bond-forming chemistry drives the extensive use of the natural nucleophilicity of cysteine thiol groups[20] in many bioconjugations (and indeed in natural PTM). The development of site-directed mutagenesis thus broadened the generality (site selectivity, regioselectivity) of protein chemical modification by allowing natural residues with particular reactivity, such as cysteine, to be introduced flexibly at predetermined sites. Today, sequential use of genetic mutagenesis to install/position a reactive cysteine tag residue, and then thiol-nucleophile chemistry (in some form) to modify the cysteine tag is perhaps, next to nonselective lysine conjugation, the most common of all synthetic protein modification procedures. In nature, cysteine is itself subject to a number of PTMs, some of which are amenable to installation by direct reaction of the Sγ nucleophile with a Cδ electrophile (e.g., prenylation[21]), and many which are not (e.g., phosphorylation[22]).

Mimicry of a target side-chain structure (such as a PTM) has been oft-employed in protein science when the desired residue is inaccessible by current chemical methods. This approach too, has often exploited heteroatom–carbon bond formation to create thia-Xxx analogues such as thia-Lys,[23] thia-Arg,[24] or even thia-homo-Glu,[25] thereby allowing sometimes unique mechanistic insight (into for example, pK_a or geometry) beyond that offered by the limited palette of classical mutagenesis. In some cases, the trace Cys left from native chemical ligation (NCL; see below) has been advantageously exploited,[25b] and even the degree of the mimicry has been investigated.[26] Notably, however, careful control of reaction conditions may prove necessary to prevent unwanted over-reaction on non-Cys residues (e.g., His. Met[27]).

Such side-chain mimicry has not been restricted to the use of Cys as a nucleophile. The same principle of exploiting inherent polarity also underpinned the creation of what might be viewed as a near-chemical mutagenesis, that is, the conversion of Lys into homoArg using isoureas through reaction of an Nω nucleophile with a Cω electrophile,[28] an unnatural disconnection which shares close similarity to natural disconnections (e.g., Lys acylation).

Notably, despite retrosynthetic analysis of C−X (and all) bonds allowing, in theory, three different synthons from this type of disconnection [two heterolytic (positive-negative and vice versa) and one homolytic], to date most approaches have focused on the use of charged synthons with natural polarity.[29] An important and usefully provocative exception is the use of thiol-ene (strictly thiyl-ene) chemistry, which takes advantage of the single-electron character of a thiyl-radical for alkylation with somophiles (somophiles or SOMO-philes are used here as a catch-all for chemical species with affinity for ′open-shell′ free-radical species that possess singularly occupied MOs (SOMOs) such as terminal alkenes. In this system, inherent biases of polarity (electrophilic thiyl) allow regioselectivity (terminal thiylation), but unlike heterolytic disconnections do not dictate the mode of disconnection per se.[30] This feature has thus allowed disconnection not only to a typically unreactive double-bond Cδ somophile,[30] but also access to a rare example of a C-type disconnection

(Scheme 1). The installation of an *N*-acetyl thialysine mimetic of acetyllysine (using an Sγ radical), which is inaccessible by direct alkylation (using an Sγ nucleophile),[31] also exploits the altered selectivity of radicals in protein chemistry. A fascinating recent report from the group of Payne provides a tantalizing hint at the potential use of C radicals in carbon–heteroatom bond formation in protein modification.[32] Generation of an alanyl radical on a peptide substrate by deselenization of synthetically introduced selenocysteine and subsequent trapping with an excess of oxone provides access to Ser through a Cβ radical and an Oγ somophile. This reaction allows an alternative homolytic access to an E-type disconnection (Scheme 1).

It should be noted that in recent years a number of reactions, often formal cycloadditions that are carbon–heteroatom bond-forming, have been developed for protein modification, and were largely inspired by the click-chemistry concept.[18] A number of different goals have informed the design of such reactions, and often those of "labeling" have dominated many aspects of synthetic protein research in chemical biology. The use of mutually selective, compatible reagent pairs with rapid kinetics for heteroatom bond formation has evolved as a potent solution for this application. However, the limited range of natural structures that can be accessed by such an approach still necessitates the investigation of different disconnections to allow access to structures that might more usefully mimic those needed to recapitulate function.

3.2. Amide Bond Disconnections: State-of-the-Art and Prospects

The ribosome, for many the most remarkable and useful biological machine in nature, relies upon the amide bond disconnection F (Scheme 3). At the ribosomal center the condensation reaction of an amino-acid ester with an amine is enhanced and controlled. The logic of designed processes for synthetic manipulation at the translational level (stop codon suppression, auxotrophic incorporation, etc.) thus relies on the amide bond disconnection approach applied through any associated biosynthetic stringency. Thus the fidelity of the system responsible for error-free protein synthesis, necessary for cellular functioning, also places grave constraints on such protein engineering because of the evolved stringency of both the ribosome and associated amino-acid/tRNA synthetases. Whilst cell-free systems and those with removed or modulated error mechanisms may benefit from greater plasticity, this tension is ever present in direct co-translational methods for the incorporation of artificial amino acids. Nonetheless, some very useful PTM examples (nitro-Tyr,[33] sulfo-Tyr,[34] various acylated Lys,[35] phosphoSer[36]) have been successfully accomplished by exploiting associated plasticities and/or selectivities. At their best these successful systems for incorporation can, of course, benefit from the self-same translational stringency as a mode of selectivity and quality control, but looser systems which create plasticity may also generate unwanted infidelity.[37]

Moreover, stop-codon suppression can of course allow the smuggling of a ʹstumpʹ for further ʹgraftingʹ as a tag for chemical modification. Opportunities exist in this regard for synthetic divergence by encoding residues which act as precursors to multiple modifications. For example, *p*-borono-phenylalanine can be chemically transformed, post-expression, to either phenylalanine or tyrosine.[38]

The biocatalytic manipulation of amides is, of course, not restricted to ribosomal catalysis: proteolytic activity and its reversal can in principle be harnessed to achieve similar ends. Such a series of enzymatic reactions has, in fact, been suggested as a generalized means for the conversion of a single amino-acid precursor into a number of different side
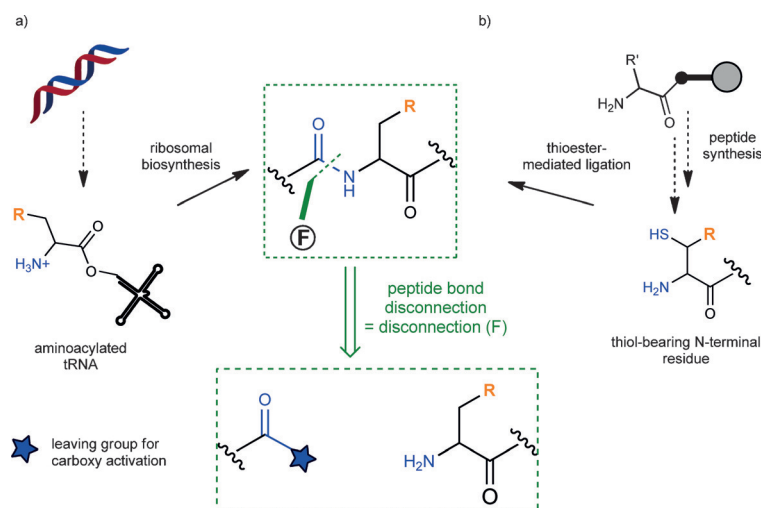
Scheme 3. Peptide bond disconnection for inserting an amino acid of choice. As the dominant mode of protein synthesis (chemically and biosynthetically), strategic methods for creating sources of fragments are widespread. Here are two contrasting methods for generating an amino acid of choice. Both exploit the nucleophilic C-terminal synthetic equivalent fragment which is revealed by disconnection F: a) A "charged" tRNA used during N-to-C ribosomal biosynthesis (and so must be tolerated by the ribosome). b) A C-terminal peptide fragment that bears a thiol auxiliary which may participate in thioester-mediated NCL with a suitably activated N-terminal carboxy donor fragment.

chains.[39] Thus, mutation from Arg to Lys on the protein soybean trypsin inhibitor through incubation with trypsin (leading to selective cleavage of the Arg64-Ile65 peptide bond) with subsequent addition of an excess of a free amino acid and use of carboxypeptidase B as a catalyst (or use of chemical activation) allowed the ligation of free Lys to the C-terminus of the N-terminal fragment. The use of chemical methods valuably broadened substrate scope to any suitably activated amino acid, albeit with the danger of loss of selectivity.

These and other protease-catalyzed ligations[40] as enzymatic mutations can be seen as a forerunner to contemporary chemical protein ligation methods, thus directly altering the protein backbone by a cut-and-paste approach. Creative application of such an approach, aided by the rapid maturation of ligation technologies, has significant potential for new approaches in chemical mutagenesis. Native-chemical ligation (NCL; Scheme 3b)[8c] is perhaps the seminal example of the utility of chemical approaches to protein synthesis (and thus the incorporation of modifications). Although our focus is on varied methods, some key features that lie behind the success of the reaction are worth considering as potential guidelines for the design of future reactions that rely on other disconnections. Amongst these, the requirement for two chemical recognition events, the initial capture by a free thiol followed by nucleophilic attack of the N-terminal amine, lies behind the notable selectivity of the reaction. Thus, dual motifs (if found in suitably productive reaction manifolds) may prove beneficial.

Moreover, the wonderful suitability of this dual motif means that, despite more than thirty years of research, NCL typically relies on this reactivity alone (notwithstanding variations of homologated or altered thiol/selenol etc). Therefore, in practice, the strategy is still largely limited to either a Xaa-Cys or Xaa-Ala backbone disconnection although numerous dechalcogenative variations from bespoke thiol/selenol amino acids provide other disconnections in principle.[41] Thus, the chief obstacle to extending the approach lies also in the reliance on this dual motif, that is, the requirement for an N-terminal thiol peptide moiety as well as the C-terminal thioester of a second peptide, which can be challenging to synthesize. While many new thiolated amino-acid building blocks have been reported for peptide synthesis, few are commercially available and their syntheses tend to be involved (although are ever-improving). The conversion of cysteine into alanine by desulfurization is currently the only extension to the original method that has been widely applied to protein substrates.[41] Furthermore, expressed protein ligation (EPL),[42] one of the most useful applications of NCL for the study of medium to large proteins, remains limited to Xaa-Cys and Xaa-Ala disconnections when performed with an expressed C-terminal fragment. New ligations at non-Cys sites (e.g., Met[43]), as well as new methods for converting cysteine residues into other natural residues post-ligation (e.g., akin to prior methods used to create mimics,[25b] using other disconnections shown in Scheme 1) would thus significantly extend the usefulness of ligation strategies. Also, examples wherein an N-terminal thiol/selenol auxiliary is installed post-synthetically, are particularly noteworthy[44]

since they too may prove accessible on recombinant proteins, thus extending access to the C-terminal partner fragment in the EPL methodology.

3.3. Prospects in Chemical Mutagenesis: Overlooked Disconnections

Notably, in seeking to highlight current methods, this Minireview has touched on but a handful of the possible disconnections identified in Scheme 1: essentially disconnections E, F, and I. The others remain largely absent or populated by rare examples. To date, side-chain modification has also largely been limited to carbon–heteroatom linkages, and driven by the perceived need for fast labeling kinetics in applications such as fluorescent labeling in cells. In these applications, where bolting-on of a label is paramount, the precise nature of the linkage in the side chain used for modification, and thus the identity of the resultant product, are not necessarily of interest. Such carbon–heteroatom chemistry is thus understandably used because it is rapid and operationally simple. Perhaps an even greater limitation is that currently, only a very few carbon–carbon forming reactions which are potentially protein compatible can be performed in aqueous media. However, the chemical construction of the carbon-rich native amino-acid side chains in situ as well as most of their post-translationally modified variants, in a target-driven fashion, will necessitate carbon–carbon bond disconnections (such as disconnections A or B (Scheme 1), amongst others).

The C_α–C_β disconnection (A) results in a glycine-type species as the protein-associated synthon. At present, no reactions have been reported for the functionalization of glycine in proteins. However, glycine modification is well-established in the field of peptide chemistry, with reported approaches utilizing glycine cations, anions/enolates, and radicals, including asymmetric variants of the reaction. Unfortunately, few of the reported reactions are sufficiently benign to be readily transferred to a protein context, with most requiring organic solvent, strong base, or elevated temperatures well beyond those either accessible in water or in proteins. Furthermore, as glycine is one of the most abundant amino acids in proteins, opportunities for site-specific modification are likely to be limited in part by the sheer statistics of design and/or selectivity (a problem that also often plagues Lys modification methods). In this context, functionalization at N-terminal glycine residues or the use of sequon-based approaches to position a metal catalyst or activating reagent may provide some requisite chemoselectivity.

The C_β–C_γ disconnection (B) results in consideration of an alanine-type species. Alanine itself could conceivably be modified by selective C_{sp3}–H activation. This activation would likely also require either the identification or design of a polypeptide sequence to facilitate metal coordination and specific C–H activation. Recent work on palladium-catalyzed transformations, such as Suzuki–Miyaura couplings,[45] has shown that transition-metal-catalyzed reactions can be used for protein modification, and more applications of these

powerful metal catalysts can be envisioned, although it is likely that C−C bond formation in the creation and manipulation of non-aromatic side chains will prove a tougher challenge than those shown so far for $C(sp^2)$−$C(sp^2)$ generation.

Particularly, research into the prospects for C−C bond formation using such an approach should be a high priority and make use of the full gamut of disconnective pathways from new and existing tags as synthetic equivalents. Whilst C_{sp3}−C_{sp3} formation would provide useful direct access, approaches such as metathesis could conceivably provide indirect access by C_{sp2}−C_{sp2} formation and functional-group interconversion (i.e. reduction) to exploit the same disconnection. Although efficient cross-metathesis on proteins has, to date, exploited chalcogen-mediated relay effects[17,46] (and hence access to longer side chains), multiply relayed methods could be envisaged, for example, for the C_β−C_γ disconnection (B).

One tantalizing opportunity (amongst many) awaits the discovery of synthetic methods for these C−C disconnections. Building the carbon scaffold that dominates amino-acid side chains will allow the proper synthesis of not only many natural amino acids but also their modified variants. Interestingly, functional chemical recapitulation of the decoration of proteins with PTMs by writer enzymes can be demonstrated with, for example, thia-analogues of either methylated Lys[22,23d] or GlcNAcylated[22,47] or phosphorylated Ser[48]. However, native (e.g., C−C) linkages would allow true emulation using tag-and-modify chemistry, in which a uniquely reactive functional group is introduced into the protein (the synthon tag), which is then modified by selective chemical functionalization to bring in, through the reagent, not only the amino-acid side chain but also any chosen attached modification. Ready methods for diversely creating many variants at one site from a single tag/synthon protein intermediate can then be envisaged. The synthetic challenges for any method that addresses disconnections A and B (Scheme 1) are significant and exciting in their own right. Thus, any method that allows these disconnections, either A or B, will also face multiple challenges: 1) sufficient reactivity in aqueous/biological milieu; 2) selective bond formation; 3) sufficiently mild reaction conditions under which both protein function and modification function are retained after reaction. Many existing reactions that one can envisage, fail. The solution to such a challenge will likely entail dual development of entirely new reactions with retrofitting of partially forgotten approaches (as highlighted by this Minireview).

4. Conclusion

The field of chemical mutagenesis stands poised for a resurgence. The explosion of interest in protein modification chemistry, coupled with the rapid maturation of ligation and carbon–heteroatom bond-forming approaches, indicates the field is ripe for a new challenge. Natural PTMs are one of the outstanding challenges in protein modification, with the current methods to access these residues in a target-driven fashion still lacking in key respects. Currently, it does not

seem possible to efficiently introduce very small (such as monomethylation of Lys) or overly large (such as GlcNAcylation of Ser) modifications by intervention in the translational machinery. This range leaves enormous opportunity for the discovery of new chemical synthetic methods in the mechanistic dissection of biology. We hope that the somewhat simplistic (but perhaps nonetheless informative) retrosynthetic analyses suggested herein highlight just how many opportunities remain unexplored, and will encourage the investigation of new methods.

The utility of the products of such synthetic methodology seem immediately apparent: not only the creation of a new generation of artificial protein drugs,[49] but it may not be too much to envisage the development of a vibrant, long-lived, associated chemical industry, much as current chemistry still bears the strategic hallmarks of the utilization of coal tar. In our own opinion, although discussion of synthetic biology is currently dominated by the exploitation of indirect methods from nucleotides, only when the vision of chemical mutagenesis is achieved and when we can flexibly reprogram the functional molecules of life (proteins, sugars, lipids), can we speak of a true synthetic biology to match the development of synthetic chemistry in the last century.

How to cite: *Angew. Chem. Int. Ed.* **2016**, *55*, 5896–5903
Angew. Chem. **2016**, *128*, 5994–6002

[1] G. Winter, A. R. Fersht, A. J. Wilkinson, M. Zoller, M. Smith, *Nature* **1982**, *299*, 756–758.

[2] F. Crick, *Nature* **1970**, *227*, 561–563.

[3] A. Dumas, L. Lercher, C. D. Spicer, B. G. Davis, *Chem. Sci.* **2015**, *6*, 50–69.

[4] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498–500.

[5] a) J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026–9027; b) S. K. Blight, R. C. Larue, A. Mahapatra, D. G. Longstaff, E. Chang, G. Zhao, P. T. Kang, K. B. Church-Church, M. K. Chan, J. A. Krzycki, *Nature* **2004**, *431*, 333–335.

[6] a) D. P. Nguyen, M. M. Garcia Alai, P. B. Kapadnis, H. Neumann, J. W. Chin, *J. Am. Chem. Soc.* **2009**, *131*, 14194–14195; b) H.-w. Ai, J. W. Lee, P. G. Schultz, *Chem. Commun.* **2010**, *46*, 5506–5508; c) D. Groff, P. R. Chen, F. B. Peters, P. G. Schultz, *ChemBioChem* **2010**, *11*, 1066–1068; d) D. P. Nguyen, M. M. G. Alai, S. Virdee, J. W. Chin, *Chem. Biol.* **2010**, *17*, 1072–1076.

[7] a) C. Zioudrou, M. Wilchek, A. Patchornik, *Biochemistry* **1965**, *4*, 1811–1822; b) L. Polgar, M. L. Bender, *J. Am. Chem. Soc.* **1966**, *88*, 3153–3154; c) K. E. Neet, D. E. Koshland, *Proc. Natl. Acad. Sci. USA* **1966**, *56*, 1606–1611.

[8] a) V. du Vigneaud, C. Ressler, J. M. Swan, C. W. Roberts, P. G. Katsoyannis, *J. Am. Chem. Soc.* **1954**, *76*, 3115–3121; b) R. B. Merrifield, *J. Am. Chem. Soc.* **1963**, *85*, 2149–2154; c) P. Dawson, T. Muir, I. Clark-Lewis, S. Kent, *Science* **1994**, *266*, 776–779.

5902 www.angewandte.org

© 2016 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Angew. Chem. Int. Ed. **2016**, *55*, 5896–5903

[9] C. A. Hutchison, S. Phillips, M. H. Edgell, S. Gillam, P. Jahnke, M. Smith, *J. Biol. Chem.* **1978**, *253*, 6551–6560.

[10] P. I. Clark, G. Lowe, *Eur. J. Biochem.* **1978**, *84*, 293–299.

[11] M. J. Appel, C. R. Bertozzi, *ACS Chem. Biol.* **2015**, *10*, 72–84.

[12] In principle, both chemistries could by applicable to a more general approach if sufficient reactivity were present for either non-active-site Ser or Cys. Sadly, Cys alkylation with α-halo carbonyls is often nonselective (His, Lys compete) and Ser sulfonylation is too sluggish for non-activated Ser.

[13] a) R. Robinson, *The Structural Relations of Natural Products*, **1955**; b) E. J. Corey, X.-M. Cheng, *The Logic of Chemical Synthesis*, Wiley, New York, **1995**.

[14] J. M. Chalker, G. J. L. Bernardes, B. G. Davis, *Acc. Chem. Res.* **2011**, *44*, 730–741.

[15] Z. P. Wu, D. Hilvert, *J. Am. Chem. Soc.* **1989**, *111*, 4513–4514.

[16] G. J. L. Bernardes, J. M. Chalker, J. C. Errey, B. G. Davis, *J. Am. Chem. Soc.* **2008**, *130*, 5052–5053.

[17] Y. A. Lin, O. Boutureira, L. Lercher, B. Bhushan, R. S. Paton, B. G. Davis, *J. Am. Chem. Soc.* **2013**, *135*, 12156–12159.

[18] H. C. Kolb, M. G. Finn, K. B. Sharpless, *Angew. Chem. Int. Ed.* **2001**, *40*, 2004–2021; *Angew. Chem.* **2001**, *113*, 2056–2075.

[19] E. M. Sletten, C. R. Bertozzi, *Angew. Chem. Int. Ed.* **2009**, *48*, 6974–6998; *Angew. Chem.* **2009**, *121*, 7108–7133.

[20] J. M. Chalker, G. J. L. Bernardes, Y. A. Lin, B. G. Davis, *Chem. Asian J.* **2009**, *4*, 630–640.

[21] M. J. Brown, P. D. Milano, D. C. Lever, W. W. Epstein, C. D. Poulter, *J. Am. Chem. Soc.* **1991**, *113*, 3176–3177.

[22] J. M. Chalker, L. Lercher, N. R. Rose, C. J. Schofield, B. G. Davis, *Angew. Chem. Int. Ed.* **2012**, *51*, 1835–1839; *Angew. Chem.* **2012**, *124*, 1871–1875.

[23] a) H. Lindley, *Nature* **1956**, *178*, 647–648; b) J. M. Messmore, D. N. Fuchs, R. T. Raines, *J. Am. Chem. Soc.* **1995**, *117*, 8057–8060; c) L. M. Gloss, J. F. Kirsch, *Biochemistry* **1995**, *34*, 3990–3998; d) M. D. Simon, F. Chu, L. R. Racki, C. C. de la Cruz, A. L. Burlingame, B. Panning, G. J. Narlikar, K. M. Shokat, *Cell* **2007**, *128*, 1003–1012.

[24] D. D. Le, A. T. Cortesi, S. A. Myers, A. L. Burlingame, D. G. Fujimori, *J. Am. Chem. Soc.* **2013**, *135*, 2879–2882.

[25] a) M. Lukac, R. J. Collier, *J. Biol. Chem.* **1988**, *263*, 6146–6149; b) G. G. Kochendoerfer, S.-Y. Chen, F. Mao, S. Cressman, S. Traviglia, H. Shao, C. L. Hunter, D. W. Low, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wilken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Horvat, L. Savatski, J. W. Adamson, A. Kung, S. B. H. Kent, J. A. Bradburne, *Science* **2003**, *299*, 884–887.

[26] a) G. Jia, W. Wang, H. Li, Z. Mao, G. Cai, J. Sun, H. Wu, M. Xu, P. Yang, W. Yuan, S. Chen, B. Zhu, *Cell Res.* **2009**, *19*, 1217–1220; b) D. Seeliger, S. Soeroes, R. Klingberg, D. Schwarzer, H. Grubmüller, W. Fischle, *ACS Chem. Biol.* **2012**, *7*, 150–154.

[27] C. E. Hopkins, P. B. O'Connor, K. N. Allen, C. E. Costello, D. R. Tolan, *Protein Sci.* **2002**, *11*, 1591–1599.

[28] W. L. Hughes, H. A. Saroff, A. L. Carney, *J. Am. Chem. Soc.* **1949**, *71*, 2476–2480.

[29] D. Seebach, *Angew. Chem. Int. Ed. Engl.* **1979**, *18*, 239–258; *Angew. Chem.* **1979**, *91*, 259–278.

[30] N. Floyd, B. Vijayakrishnan, J. R. Koeppe, B. G. Davis, *Angew. Chem. Int. Ed.* **2009**, *48*, 7798–7802; *Angew. Chem.* **2009**, *121*, 7938–7942.

[31] F. Li, A. Allahverdi, R. Yang, G. B. J. Lua, X. Zhang, Y. Cao, N. Korolev, L. Nordenskiöld, C.-F. Liu, *Angew. Chem. Int. Ed.* **2011**, *50*, 9611–9614; *Angew. Chem.* **2011**, *123*, 9785–9788.

[32] L. R. Malins, N. J. Mitchell, S. McGowan, R. J. Payne, *Angew. Chem. Int. Ed.* **2015**, *54*, 12716–12721; *Angew. Chem.* **2015**, *127*, 12907–12912.

[33] H. Neumann, J. L. Hazen, J. Weinstein, R. A. Mehl, J. W. Chin, *J. Am. Chem. Soc.* **2008**, *130*, 4028–4033.

[34] C. C. Liu, P. G. Schultz, *Nat. Biotechnol.* **2006**, *24*, 1436–1440.

[35] a) H. Neumann, S. Y. Peak-Chew, J. W. Chin, *Nat. Chem. Biol.* **2008**, *4*, 232–234; b) C. H. Kim, M. Kang, H. J. Kim, A. Chatterjee, P. G. Schultz, *Angew. Chem. Int. Ed.* **2012**, *51*, 7246–7249; *Angew. Chem.* **2012**, *124*, 7358–7361; c) H. Xiao, W. Xuan, S. Shao, T. Liu, P. G. Schultz, *ACS Chem. Biol.* **2015**, *10*, 1599–1603; d) B. J. Wilkins, L. E. Hahn, S. Heitmüller, H. Frauendorf, O. Valerius, G. H. Braus, H. Neumann, *ACS Chem. Biol.* **2015**, *10*, 939–944.

[36] a) C. M. Zhang, C. Liu, S. Slater, Y. M. Hou, *Nat. Struct. Mol. Biol.* **2008**, *15*, 507–514; b) S. Lee, S. Oh, A. Yang, J. Kim, D. Söll, D. Lee, H.-S. Park, *Angew. Chem. Int. Ed.* **2013**, *52*, 5771–5775; *Angew. Chem.* **2013**, *125*, 5883–5887; c) D. T. Rogerson, A. Sachdeva, K. Wang, T. Haq, A. Kazlauskaite, S. M. Hancock, N. Huguenin-Dezot, M. M. K. Muqit, A. M. Fry, R. Bayliss, J. W. Chin, *Nat. Chem. Biol.* **2015**, *11*, 496–503.

[37] H. R. Aerni, M. A. Shifman, S. Rogulina, P. O'Donoghue, J. Rinehart, *Nucleic Acids Res.* **2015**, *43*, e8.

[38] E. Brustad, M. L. Bushey, J. W. Lee, D. Groff, W. Liu, P. G. Schultz, *Angew. Chem. Int. Ed.* **2008**, *47*, 8220–8223; *Angew. Chem.* **2008**, *120*, 8344–8347.

[39] a) R. W. Sealock, M. Laskowski, *Biochemistry* **1969**, *8*, 3703–3710; b) H. R. Wenzel, H. Tschesche, *Angew. Chem. Int. Ed. Engl.* **1981**, *20*, 295–296; *Angew. Chem.* **1981**, *93*, 292–293.

[40] a) V. Schellenberger, H.-D. Jakubke, *Angew. Chem. Int. Ed. Engl.* **1991**, *30*, 1437–1449; *Angew. Chem.* **1991**, *103*, 1440–1452; b) A. Goswami, S. G. Van Lanen, *Mol. BioSyst.* **2015**, *11*, 338–353.

[41] P. E. Dawson, *Isr. J. Chem.* **2011**, *51*, 862–867.

[42] T. W. Muir, D. Sondhi, P. A. Cole, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 6705–6710.

[43] T. Tanaka, A. M. Wagner, J. B. Warner, Y. J. Wang, E. J. Petersson, *Angew. Chem. Int. Ed.* **2013**, *52*, 6210–6213; *Angew. Chem.* **2013**, *125*, 6330–6333.

[44] L. R. Malins, K. M. Cergol, R. J. Payne, *Chem. Sci.* **2014**, *5*, 260–266.

[45] J. M. Chalker, C. S. C. Wood, B. G. Davis, *J. Am. Chem. Soc.* **2009**, *131*, 16346–16347.

[46] Y. A. Lin, J. M. Chalker, N. Floyd, G. J. L. Bernardes, B. G. Davis, *J. Am. Chem. Soc.* **2008**, *130*, 9642–9643.

[47] L. Lercher, R. Raj, N. A. Patel, J. Price, S. Mohammed, C. V. Robinson, C. J. Schofield, B. G. Davis, *Nat. Commun.* **2015**, *6*, 7978.

[48] K. P. Chooi, S. R. G. Galan, R. Raj, J. McCullagh, S. Mohammed, L. H. Jones, B. G. Davis, *J. Am. Chem. Soc.* **2014**, *136*, 1698–1701.

[49] B. G. Davis, in *Biotherapeutics: Recent Developments using Chemical Biology* (Eds.: L. H. Jones, A. J. McKnight), Royal Society of Chemistry, Cambridge, **2013**, pp. 130–144.