Sugars and proteins: New strategies in synthetic biology*

Benjamin G. Davis

Department of Chemistry, University of Oxford, Chemistry Research Laboratory, Mansfield Road, Oxford OX1 3TA, UK

Abstract: The development of novel methodology for bond-forming processes that are compatible with biomolecules allows the assembly, alteration, or modification of proteins. Such synthetic proteins allow precise insight and investigation of function in a manner that has the potential for almost unlimited diversity.

Keywords: sugars; proteins; synthetic biology; synthetic proteins; post-translational modification; tag-modify strategy.

INTRODUCTION: THE NEED FOR SYNTHETIC PROTEINS

Protein structural diversity

The publication of the first draft of the human genome and the revelation of the relatively few genes associated with our species as compared with apparently less complex organisms raised an important paradox: a lack of correlation between number of genes in an organism and the consequent biological complexity, a conundrum referred to as the *c*-factor paradox. There are many potential explanations for the *c*-factor paradox, including, for example, alternative splicing of RNA, but one of the leading contenders is post-translational modification (PTM). PTM is the alteration of a protein after its biosynthesis, often after having been folded, typically through a change in the functional groups found in residue side chains. However, PTM is unlike transcription and translation, which are under tight genetic control and hence are effectively template-driven and therefore allow the prediction of protein product with good fidelity from the corresponding DNA "recipe". Instead, PTM is not such a templated process, it is often unpredictable, can give rise to complex mixtures of PTM proteins, and the different components in those mixtures often have different properties. Such mixtures make a full understanding of the PTM products that we obtain from biology and their resulting structure–activity relationships (SARs) difficult; as one potential solution to this difficulty, we have proposed that chemical assembly might provide an answer [1].

Chemistry allows, for example, either the assembly of pure PTM proteins or mimics of those PTMs in proteins. Thus, for example, the charged side chain of phosphoserine might be mimicked by simple mutagenesis to glutamate; this is a strategy that is successful in the p53 family of proteins [2]. Work by Waldmann et al. has shown that cysteine prenylation may be mimicked by simple alkylation in the Ras/Raf family of proteins [3], and we have shown that in certain circumstances the attachment of the monosaccharide *N*-acetylglucosamine (so-called GlcNAc-ylation) to serine can be mimicked by a disulfide analog [4].

By far, the most diverse of the hopes at PTMs is protein glycosylation; it has been estimated that some 70 % of cell surface proteins in humans are glycosylated, and in many cases we do not have a clear idea of the function of these glycosylations. Many of the examples in this review will, therefore, choose glycoprotein targets as an important dominant target. The synthesis of modified proteins, such as glycoproteins, may be broken down into disconnections and corresponding assembly according to three general modes, A, B, and C (Scheme 1) [5]. Mode A involves the linear assembly of modified amino acids or peptides in a growing peptide chain; this could correspond, for example, to the use of a modified amino acid building block in solid-phase peptide synthesis or the use of modified peptide fragments in thioester-mediated native chemical ligation. Mode C involves the alteration of an existing modification, often through site-selective methods that make use of the selectivity of biocatalysis, although such methods by definition must rely on an existing site of modification and a pre-existing modification; in the field of glycoprotein synthesis, this is often referred to as "glycoprotein remodeling". This review will feature select examples of both modes A and C, but the bulk of the strategic examples will focus on mode B, the site-selective convergent installation of a modification. This is a strategy that has more often been pursued within our research, largely since we view this as offering the broadest range of synthetic outcomes. This utilizes, typically, a prefolded protein platform, often derived from expression in a straightforward expression system, such as *E. coli* and modifications, even complex modifications, often accessed through target synthesis.
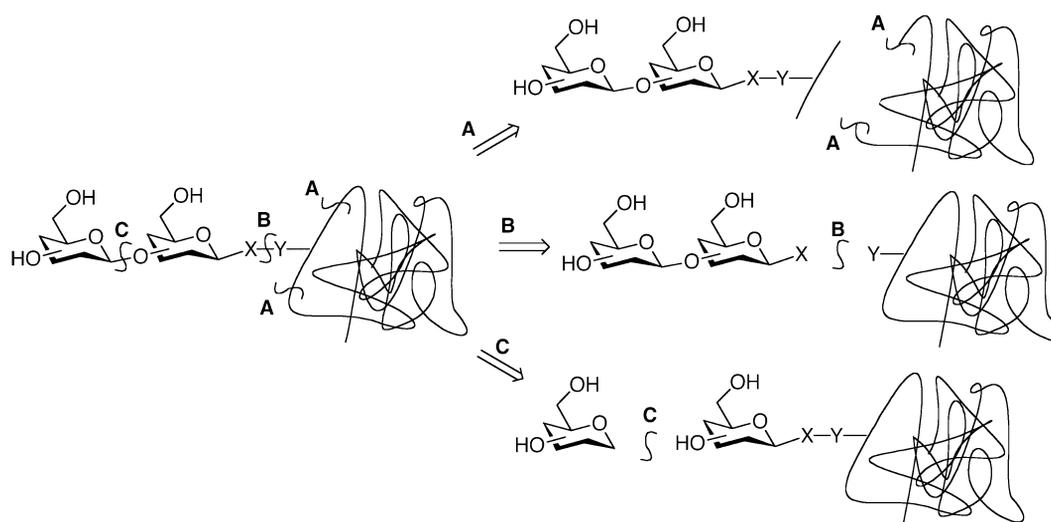
Scheme 1 The generalized retrosynthetic analysis of synthetic protein assembly. Mode A: linear assembly, Mode B: convergent assembly, Mode C: modification remodeling.

"Tag-modify" strategy

To a chemist, this type of Mode B site-selective, chemical protein modification boils down to an almost trivial two-step solution, something that we have referred to as the "tag-modify" strategy. Thus, to achieve site-selective chemical protein modification, all one needs to do is to introduce some form of chemical "tag" **Y** (Scheme 2) with control of position in the protein primary sequence and then simply perform a chemoselective reaction with the tag **Y** to form the bond **Y–X** with some form of reagent containing the modification of choice. This type of chemoselective reaction is often referred to as "ligation" reaction in chemical biology, probably originating from a distortion of the notion of general bond-forming processes performed by the class of biocatalysts referred to as ligases. In this way, through "tag"
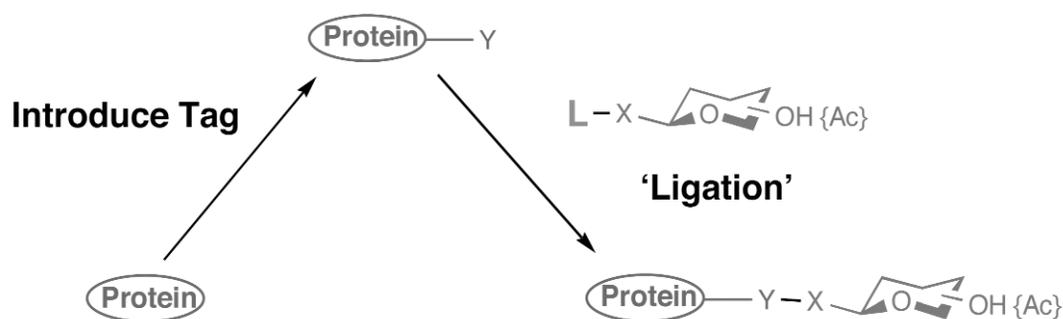
Scheme 2 The "tag-modify" strategy. A two-step process of (i) the site-controlled introduction of a chemical tag for further reaction and (ii) chemoselective formation of bond Y–X with that tag allows a versatile strategy for convergent, site-selective protein construction.

and then "modify" one can control the position and the nature/identity of the modification, respectively (Scheme 2).

There are various methods for tag introduction into proteins. One can use the functional groups found in the side chains of the 20 most readily occurring natural amino acids and control their position simply through mutagenesis. One can also think about the use of unnatural amino acids to expand the diversity of functional groups that can be used as tags. For example, the use of amber codon suppression methodology by Schultz et al. has allowed tag-modify glycosylation [6]. Unnatural amino acids may also be introduced into proteins using auxotrophs. Auxotrophs are species that are deficient in the biosynthesis of certain precursors, and one can often effectively fool them, when used as expression hosts, into incorporating analogs of those precursors simply by feeding them those analogs in their culture medium. Indeed, this is a strategy that has long been used by the crystallographic community for the replacement of Met by SeMet in proteins [7], and some leading work in this area has been performed by Tirell and coworkers [8].

It should be noted that in some rare cases the tag-modify strategy can be bypassed entirely and the ribosome can be used as synthetic catalyst for the incorporation of modified amino acids directly. However, such methods can be variable in their efficiency, and so while small changes from natural amino acids, typically based on the Phe or Tyr side chain motif, are very well incorporated, the direct incorporation of even just a monosaccharide-containing amino acid in this way meets with variable success [9].

The key to the tag-modify strategy relies on something that can be referred to as "bio-orthogonality"; this is essentially a need for the bond-forming modification to be selective for the tag **Y**. In this context, it is therefore useful to recognize that much of the competing functionality found in proteins is nucleophilic and so methods and new protein reactions that might rely on electrophiles, cycloadditions, and even radical chemistry may offer new opportunities for enhanced selectivity.

Finally, under the topic of our tag-modify approach, it is important to note that this strategy has its conceptual origins in the work of Bender [10] and Koshland [11] whose independent and near-simultaneous realization that the active site Ser nucleophile of the protease subtilisin could be chemically and *selectively* changed to cysteine led to not only the birth of "thiosubtilisin" but seleno- [12] and even telluro- [13] variants in a similar manner. This notion was also beautifully expanded by the work, in particular, of Jones and coworkers [14].

BUILDING SYNTHETIC PROTEINS

Site-selective convergent protein glycosylation using a thiol tag: An introductory case history

Nearly 10 years ago, we were able to realize the application of the tag-modify strategy for site-selective protein glycosylation [15]. Essentially, this was simply achieved through the combination of mutagenesis with chemical modification; we used the thiol in the side chain of Cys as our chemical tag and over the past decade we have developed a range of reagents that allow the formation of a mixed disulfide link between protein and modification: [4,16–19] methanethiosulfonates (MTSs) [15] reagents (originally in collaboration with Bryan Jones), phenylthiosulfonates [19], and, more recently, reagents based on the intermediacy of a mixed sulfur-selenium (SeS) linkage [4]. This type of tag-modify strategy through Cys-mutation and modification for site-selective protein glycosylation has been applied elegantly to Cys mutant proteins by a number of groups [20].

Of the varying methods, the Glyco-SeS [4] method is the most efficient. This typically involves the preactivation of a Cys mutant protein through the formation of a mixed Se–S bond followed by reaction with the appropriate thiol-appended modification of choice. This type of controlled conproportionation chemistry is based on the beautiful work in small model systems of Kice and Lee [21]. In this way, we can take a range of Cys-containing proteins, pre-activate with a solution formed from dissolution of PhSeBr, and then simply pick the thiol of choice to form the corresponding mixed-disulfide linked modified protein. This has allowed the incorporation not only of, for example, simple and complex glycans as single copies but also multiple copies of the same modification into proteins. Most striking is the efficiency. Typical protein modification reactions can require sometimes >1000 equiv of a reagent and then still not achieve anything close to complete conversion; in some cases, the Glyco-SeS method requires as little as 1 equiv. Mechanistic studies indicate that the success of this reaction may be a consequence of the electrophilic nature of S in the S–Se bond and the kinetic control that emerges from pre-equilibrating selenenyl exchange.

An additional utility of the method is that it can be combined with enzymatic extension. Thus, a "mode B" tag-modify strategy can be combined with "mode C" remodeling. This therefore allows one greater flexibility during synthetic protein assembly; one can either synthesize a complete modification *en bloc* and attach it to the protein intact or install what might be considered to be a "primer" unit and then extend that unit to the full structure. We have been able to use such a dual mode of prime-then-extend in the synthesis even of model glycoproteins in which we have site-selectively installed an 18-residue sugar (octadecasaccharide), the fully sialylated tetra-attennerary complex core-fucosylated *N*-glycan motif [22]. This synthesis has been applied to models of the biopharmaceutical erythropoietin [23].

Many of these methods use thiols as reagents, and while some good standard methods for preparing suitably thiol-appended variants of biomolecular modifications exist, these are often only preparatively effective on a multi-milligram scale, which is much larger than that which is typically isolated from many biological sources. As a means of incorporating biological samples more readily, we have recently developed direct thionation methods which allow the conversion of reducing sugars to glycosyl thiols without the need for any form of protecting group intervention [24]. Under these conditions, anomeric hydroxyls are selectively thionated in the presence of all other free hydroxyls. In this way, sugars may be isolated from natural sources, often on very small scales, even from other glycoproteins, directly thionated and then incorporated into alternative synthetic glycoprotein systems. In combination with Glyco-SeS, this, in effect, provides a one-pot glycoprotein synthesis strategy. The same strategy can also be applied to allylic alcohols to create, for example, prenyl thiols that allow the first examples of site-selective chemical protein prenylation [25].

Applications of synthetic glycoproteins in functional study and chemical medicine

In this way, the use of thiol tag-modify site-selective protein glycosylation has allowed the investigation of many aspects of synthetic glycoprotein function. These have included the first examples of the systematic investigation of glycan–structure–protein–activity relationships [17], the alteration of biocatalyst substrate selectivity using prosthetic glycans [18], and dynamic glycopeptide libraries to delineate optimal protein ligand partners [26]. Such systems have also found potential biotechnological application in areas with potential therapeutic value (a field that might be termed "chemical medicine", the solution of medicinal problems with chemical methods). Thus, glycodendrimers, site-selectively mounted onto degradative enzymes, created the first examples of dendrimer-modified proteins, *glyco-dendriproteins*, that are able to target and selectively degrade pathogen adhesions, thereby inhibiting pathogen–host interactions [27]. A bipartite drug delivery system, called LEAPT (lectin-directed, enzyme-activated prodrug therapy), used glycosylated enzymes localized in tissues according to the nature of the sugars that they were decorated with to trigger tissue-selective drug release [28]. The same principles of sugar-targeted delivery have also been used for gene delivery using chemically glycosylated viruses [29].

Expanding the diversity of chemical protein modification

Using these methods, we had until recently developed methods for single-site modification and multi-site modification with the same modification type. However, to begin to approach the remarkable diversity that nature can display in protein structures accessed through PTM, we have recently developed methods that allow access to multiple different modifications and different sites, again with full control of site and modification identity [30]. To achieve this, all we effectively did was to expand the tag-modify approach. We increased the number of chemical tags that we introduced into our protein scaffolds and then developed corresponding modifications that were mutually compatible with (orthogonal to) each other.

In the first examples of "dual tag-modify", we used the thiol-tag-disulfide modify as one "ligation" reaction and as the second "ligation", a reaction that is now near-ubiquitous in organic chemistry, the Meldal [31]-Sharpless [32] Cu(I)-catalyzed variant of the Huisgen–Dimroth cycloaddition that takes place between an azide and an alkyne to yield 1,2,3-triazoles with good 1,4-regioselectivity. This <u>c</u>op-per-<u>c</u>atalyzed <u>H</u>uisgen <u>c</u>ycloaddition (CCHC) is a reaction that is often referred to as a CLICK reaction. Given the popularity of these types of reactions, it was somewhat surprising that when we started this work its application to site-selective protein modification had been limited to only two examples: both demonstrated by Schultz and coworkers [33] for the incorporation of a fluorophore in conversions that were judged by gel electrophoresis at ≤75 %. In the field of glycoconjugate synthesis, a single leading example had shown the conjugation of protected single amino acids to protected glycans [34].

We used this second ligation, the Glyco-CCHC, as part of the 5-step process that is shown in Scheme 3. We take a wild-type (WT) protein scaffold and mutate all existing Cys to near-isosteric Ser and all existing Met to near-isosteric Ile. This creates a protein sequence containing no Cys or Met. At first sight, this might seem quite a drastic approach, but there is a relatively low incidence of Cys and Met in proteins [35], and the work of people like Hilvert has shown that amino acid function may be achieved with other equivalents taken from within the typical proteinogenic 20, thereby allowing good access to proteins with near-equivalent function that may contain less than half of 20 possible amino acid types. This step 1 provides us with a "blank canvas", if you will, for the installation of our chemical tags. In step 2, we use site-directed gene mutagenesis and use, for example, the triplet codon for Cys to control the position of a thiol tag and, for example, the triplet codon for Met to control the position of, in this example, an azide tag. In step 3, we then express this gene sequence in an auxotroph of *E. coli* that is deficient in the biosynthesis of Met and we simply feed the auxotroph an analog of Met, azidohomoalanine (Aha) that Tirrell has nicely shown is incorporated instead of Met [36]. Thus,
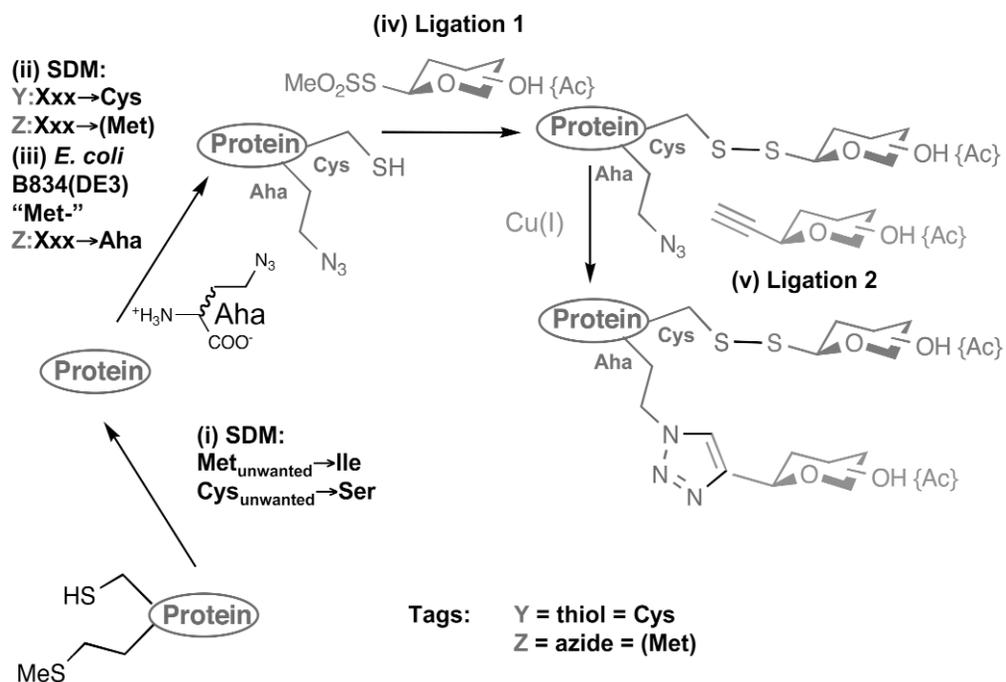
Scheme 3 Strategy for expanding the diversity of chemical protein modification. Application of a "dual" tag-modify strategy in a five-step process allows differential multisite-convergent site-selective protein modification. Step (i): WT protein sequence is mutated to convert unwanted to near-isosteric analogs Cys → Ser and Met → Ile. Step (ii): The Cys, Met-free protein is the starting point for the introduction of chemical tags. In this example, the positions of a thiol and azide tag are controlled by the triplet codons for Cys and Met, respectively. Step (iii): The mutated gene sequence is expressed in an auxotroph that is deficient in the biosynthesis of Met; this allows incorporation here of the azide tag using Met-analog Aha [36]. Steps (iv) and (v): Chemoselective and orthogonal reaction of the introduced thiol and azide tags, respectively.

we have controlled the position here of a thiol tag, and of an azide tag; in steps 4 and 5, we do a thiol-disulfide "ligation" and an azide-triazole "ligation". Using this dual tag-modify approach, we have access therefore to a synthetic protein in which we have controlled both the position and identity of two different modifications.

To accomplish this process, we needed to enable the CCHC reaction as one that could be applied to the complete conversion of tags on protein surfaces. In one example, we took the family 1 glycosyl-hydrolase from *Sulfolobus solfataricus* (SSβG) which has 10 WT Met and 1 WT Cys. We created a 10-point mutant (Cys → Ser plus Met × 9 → Ile × 9) gene sequence and then expressed the gene sequence in the auxotroph *E. coli* B834(DE3) whilst feeding it Aha racemate. This generated a mono-azido protein in which we had controlled the position of the single azide tag. This then became the substrate for synthetic methodology optimization to create a viable glyco-CCHC reaction. These optimization studies led to the development of the first examples of complete (>95 % conversion), benign site-selective triazole ligation reactions on proteins [37]. Although much has been made of the "toxicity" of Cu(I), in our hands when properly complexed (using, e.g., tetradentate tris-triazole ligands [38]) Cu(I) is no more toxic than lanthanide salts properly complexed (indeed, this is the way that Cu(I) is found and used in nature).

Building a synthetic protein mimic of P-selectin glycoprotein ligand 1 (PSGL-1)

Having developed methods for multisite differential protein modification, we have explored a number of things including SARs of modification effect [30]. However, we wanted to use these systems to create synthetic protein probes for investigating an intriguing aspect of innate immunity in mammals. When most mammals receive some form of poison or wound, a process is initiated in the endothelial cells that line the blood vessels near that wound that is part of the so-called inflammatory cascade. Chemokines/cytokines lead to presentation of a family of proteins, the CD62 proteins, and, in particular, a key player in that family CD62P, otherwise known as P-selectin. P-selectin looks out for a cognate protein ligand found on the surface of leukocytes, PSGL-1. Binding of CD62P to PSGL-1 leads to catch bond mediated tether-rolling of leukocytes as they slow and bind to the vessel internal wall in the rapid flow of blood. Eventually, once halted, leukocytes extrovase and pass through the vessel wall to the site of the poison/wound.

Some beautiful work [39,40] has delineated some of the putative structural requirements for this binding using peptidic fragments of the *N*-terminal region of PSGL-1. Of these papers, the most striking was an X-ray structure [40] of domains of CD62P co-crystallized with a peptide section of PSGL-1 that showed the presence of two PTMs in PSGL-1: sulfation of Tyr (Tys48) binding at one site in CD62P and the presence of an *O*-glycan containing the tetrasaccahride sialyl-Lewis-x (sLe^x) attached to Thr57 binding at another, the distance between these two sites being approximately 12 Å.

On the basis of these structures, we posed a broad and possibly naïve question. Could we take a generic protein scaffold, such as a triosephosphate isomerase (TIM) barrel (representative of a significant fraction of all protein structures), and simply attach to that scaffold using chemistry two mimics of those PTMs (Tys and sLe^x) at the right distance? If we did so, could we create an effective functional mimic of PSGL-1? Having evaluated possible sites in TIM barrel structures, we picked two sites that were ~13 Å apart. We used the dual tag-modify strategy to attach an aryl sulfonate as a non-hydrolyzable mimetic of Tys using disulfide "ligation" and then mono-, di-, tri- fragments of sLe^x as well as the full tetrasaccharide sLe^x itself using the triazole "ligation" (Scheme 4).

The resulting constructs displayed in vitro activity, for example, in binding to human CD62P in enzyme-linked immunoabsorbent assays (ELISAs). This binding was synergistically dependent on the presence of both a Tys and sLe^x mimic in the scaffold; advantageously since these constructs had been assembled convergently using chemistry, then we were able to build up SARs exploring the underlying contributions of each PTM and fragments of those PTMs.

Excitingly, not only did these proteins show in vitro activity, they also showed in vivo activity. Inflammation is a hallmark of pathology in a number of diseases and, in particular, diseases of the brain, such as multiple sclerosis or cerebral malaria. In models of these diseases, we were able to show that not only do our synthetic PSGL-1 mimics bind to CD62P displayed at sites of disease-associated inflammation, but also by choosing an appropriate reporter function for our TIM-barrel scaffold we could create a system for visualizing location in vivo (such biomarker-ligand-dependent strategies may be termed "molecular imaging" or "functional imaging"). Thus, choosing enzymatic activity (β-galactosidase) that corresponds to *lacZ* activity we were able to produce reporter protein PSGL-LacZ. We suggest that such protein probes [30], which use molecular design to mediate and explore the recruitment of protein–protein interfaces, are perhaps more representative of the types of processes that occur in vivo. They may thus be considered to be useful alternatives for the now standard use of monoclonal antibody protein probes [41].
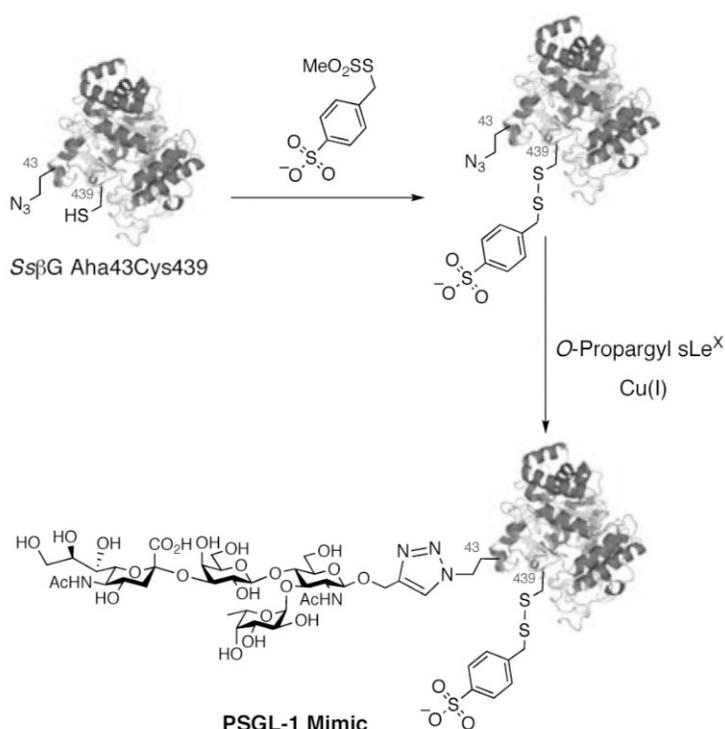
Scheme 4 Dual tag-modify strategy allows the creation of effective functional protein mimics, here the human protein PSGL-1 bearing sulfoTyr-mimic and glycan modifications.

FUTURE DIRECTIONS IN SYNTHETIC PROTEIN DESIGN FOR SYNTHETIC BIOLOGY

Expanding synthetic protein diversity yet further

It is not an enormous conceptual leap to realize that the addition of further tags in multiple tag-modify strategies will allow the chemical mapping of whole swathes of "protein functional diversity space". To this end, we have been developing further examples of important protein chemistries. This is an exciting and challenging area for synthetic chemistry. The criteria for such design are stringent: aqueous solvent reaction systems compatible with biological milieu, benign reagents that do not degrade biomolecules, temperatures compatible with life and/or function, and high selectivities and efficiencies that allow complete conversion at (and only at) the chosen tag site(s). This work is driving the expansion of new areas of synthetic chemistry exploring fundamental new strategies for bond-forming processes in water and areas that might be termed in vivo chemistry.

Dehydroalanine (Dha): A powerful tag

We have recently developed a novel protocol that has allowed us to desulfurize disulfide-linked protein conjugates to their corresponding thioether-linked counterparts [42]. Although disulfide-linked glycoproteins can show good stability in vivo, these new conjugates are advantageously fully resistant to reductive cleavage. During the course of the investigations that have allowed us to develop these reactions, we identified two mechanistic extremes for the desulfurization of glycosyl(cysteinyl)disulfides: one akin to thio-Mitsunobu reaction involves attack of the phosphine desulfurizing agent upon the γ-sulfur atom of the glycosyl disulfide followed by reaction of the resulting glycosyl thiophosphonium as an electrophile with a Cys thiol nucleophile to form a thioether [43], the other involves attack on the β-sul-

fur and then elimination of the resulting Cys thiophosphonium to form a dehydroalaninyl derivative (Dha) followed by conjugate addition of glycosyl thiolate. In the case of many glycosyl-protein disulfides, we believe that it is the Dha mechanism, and its intermediacy has been observed in a number of model systems. Such a mechanism is also consistent with previously observed, but unexplained, epimerization of cysteinyl-containing peptides by tris(amino)phosphines [44].

These observations indicated two key aspects of Dha's chemistry: firstly, that in order to recapture the low concentrations of thiol released in such desulfurizations, Dha must be a remarkable conjugate electrophile, and, secondly, that Dha can be accessed from Cys through sequential oxidation (e.g., disulfide formation) followed by elimination. It should be noted that the electrophilic properties of Dha in chemically synthesized peptides [45], biosynthetically derived peptides [46], and proteins [47] have been noted before. Thus, a new tag-modify strategy could be envisaged in which Dha could be used as an electrophilic tag, the position of which is controlled by the triplet codon for Cys, and that is introduced into proteins following some form of tandem oxidation-elimination process.

A fully compatible method for creating Dha in proteins

In our search for a novel chemical method of converting Cys to Dha, we discovered the utility of the reagent *O*-(mesitylenesulfonyl)hydroxylamine (MSH) [CAUTION: MSH has been reported as being explosive]. MSH has previously been used as a possible source of electrophilic nitrogen in amination reactions and has also been used in the conversion of aryl thioethers to the corresponding olefins [48]. Yet it is its reaction with Cys derivatives that is most remarkable. Early model reactions [49] led to the formation of near equal amounts of Dha and cystinyl disulfide, a key result that illustrated not only that we could access Dha from Cys but that MSH's inertness to disulfides might allow its application in proteins bearing free Cys residues without affecting internal disulfides. Model reactions for MSH were readily optimized to conditions suitable for conversion of Cys to Dha in proteins (phosphate buffer 50 mM, pH 8.0, 20 min) [49]. The reaction requires 2 equiv of MSH, and we therefore speculate that it proceeds via an intermediate sulfenamide that undergoes Cope-type elimination (or perhaps simple β-elimination).

The utility of MSH has many interesting features. As noted above, we have not been the first to consider the utility of Dha in proteins. Previous strategies [46,47] for introduction of Dha into proteins have made use of derivatives of the unnatural amino acid SeCys such as Ph–Se–Cys incorporated, for example, using amber codon suppression methods [47]. Subsequent treatment with hydrogen peroxide allows oxidative elimination via an intermediate selenoxide. This treatment, however, not only yields Dha but also causes concomitant and unavoidable oxidation of any Met residues in the protein to Met-sulfoxides and sulfones. This therefore raises the question of the compatibility of Met residues with MSH. It turns out that Met residues do react with MSH to form sulfilimines; however, these sulfilimines revert to the precursor thiother, i.e., they revert to Met under basic conditions with appropriate buffer or in the presence of suitable thiols (such as those even used for subsequent conjugate additions to Dha) [48,50]. This reversion means that MSH treatment effects not only the conversion of Cys to Dha but also provides an "NH_2^+" source that temporarily protects the thioether moiety of Met and returns unaltered Met at the end of the reaction. The upshot is that MSH allows (i) the first benign, chemical conversion of a *natural* amino acid to Dha in proteins and (ii) the first method that is compatible with all other natural amino acids including Met.

Dha in the tag-modify strategy allows wide-ranging and switchable mimicry of PTMs

Once Dha can be accessed in a protein, then many mimics of PTMs can be readily introduced (Scheme 5). Protein glycosylation may be achieved using corresponding glycosyl thiols in conjugate addition to Dha [47,49]; use of a two-step [(i) MSH, (ii) glycosyl thiol] protocol therefore allows Cys to be fully converted to a glycosylated variant in 90 min total reaction time. Using appropriate thiols,
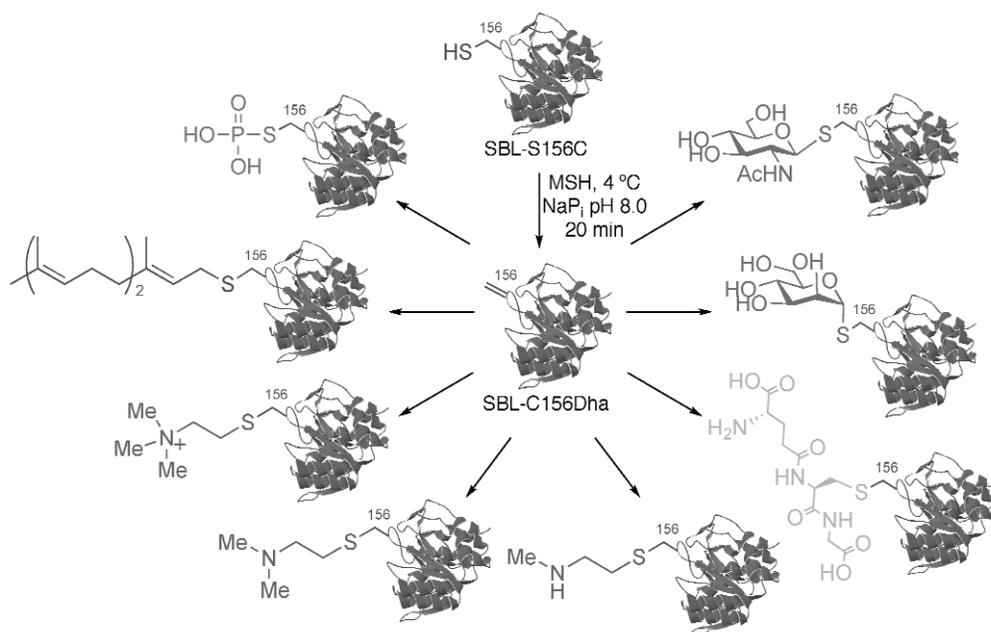
Scheme 5 Dha is a powerful tag for the subsequent introduction of many PTM mimics. The unnatural residue may be directly accessed in proteins from a Cys residue using the reagent MSH [49].

we have also been able to demonstrate site-selective chemical protein peptidylation [49], prenylation [47,49] (creating the natural Cys-prenylated connectivity [51]), phosphorylation [49], and even the introduction of analogs of modified Lys (mono-, di-, and trimethyl) [49]. Such modified Lys derivatives may allow dissection of the role of PTMs in epigenetics [52]. The latter thiol addition method provides a usefully complementary electrophile-based approach to the nucleophilic Cys-alkylation strategies used by Shokat and coworkers [53]. Since our first publication [49] of this thiol addition method, it has also been successfully applied by Schultz and coworkers to introduce such modified Lys derivatives into H3 histone protein [54].

The thioether linkage in the resulting PTM-mimic proteins is, as discussed above, resistant to reduction. However, this resistance does not imply permanence, and, remarkably, use of an excess of MSH allows the β-thioether products of conjugate addition of thiols to be "switched" back to Dha [49]. Presumably this "switch" back proceeds through a sulfilimine that by virtue of its more acidic α-proton eliminates more readily than those formed from, for example, Met. Regardless, this means that not only can MSH be used to install Dha in the first place (starting from Cys) but it also enables reagent-controlled switchable iterative modification of protein structure from a single starting protein scaffold. This therefore represents a novel protein engineering strategy that we believe will find a number of useful applications in selection- or screen-driven design of function.

Protein olefin cross-metathesis (OCM) enabled by the discovery of a privileged unnatural amino acid motif, *S*-allylcysteine (Sac)

The MSH-Dha protocol has also allowed us to survey the utility and reactivity of a number of previously unknown amino acid motifs within protein architectures. This, in turn, as illustrated by one recent case has allowed us to solve some central problems in synthetic protein chemistry. Ever since some of the very first installations of C=C double-bond-containing amino acids into proteins [55], it has been suggested that such olefin handles in proteins might provide a suitable point at which to conduct olefin

metathesis. Yet, despite the widespread utility of OCM in small-molecule synthetic organic chemistry, typically in organic solvents, its application to protein modification has, until recently, not been achieved, and this goal has become an identified and important one in the field.

The lack of success has been due, in part, to somewhat of a gap in structural complexity between what one would like to use as substrates in aqueous OCM (e.g., proteins, carbohydrates) and what one could until recently use [largely reactive simple unsaturated alcohol (allylic)], a lack of substrate breadth and tolerance that persisted despite the elegant efforts of some leading groups in reaction condition and catalyst design [56]. We initiated a project in this area from the standpoint of investigating the types of olefin-containing amino acid side chains that we might be able to ultimately introduce into a protein. Initial surveys of reactivities under a range of conditions included some more immediately obvious targets, such as homoallylglycine (Hag) [55]; yet in our hands, these derivatives yielded little or no product. However, as our survey expanded we discovered the remarkable properties of one amino acid, *S*-allyl-Cys (Sac) and in one fell swoop were able to achieve, even in model amino acid systems, some of the most complex examples of aqueous OCM achieved to date [57]. The key to the success of Sac in OCM appears to be due to its allyl sulfide motif, which is a privileged one in our hands that enhances the rate of aqueous OCM. While the nature of the *N*- and *C*-terminal substitution around Sac can be varied widely, the distance between olefin and thioether linkage is critical; *S*-butenyl-Cys and *S*-pentenyl-Cys fail as substrates. The privileged nature of the allyl sulfide motif is all the more notable given the quite generally perceived "incompatibility of the ruthenium-based metathesis catalysts with substrates containing sulfur(II) donor sites" [58]. We speculate that its striking reactivity might be due to sulfur affinity for Ru that assists metathesis provided that any intervening stable chelate does not sequester catalyst. In effect, this might be a "sulfur-relayed" metathesis passing through intermediate sulfur-coordinated metallacyclobutanes; in this context, it is interesting to note that carbonyl-coordinated metallacyclobutanes have been identified from tungsten-containing alkylidenes [59].

Having established the striking reactivity of Sac in OCM in model amino acid and peptide systems, we used the MSH-Dha-thiol addition protocol to introduce it into a protein environment. Once installed, we then tested the same OCM conditions that had been successful in smaller models. Initially, we were disappointed to see a lack of OCM, but careful analysis revealed the formation of a protein mass signal equivalent in mass to an unproductive alkylidene adduct. Similar unproductive adducts have, in fact, been observed previously in small-molecule ring-closing metathesis (RCM) [60]; in these systems, disruption of hard Lewis base coordination through the use of a sacrificial hard Lewis acid led to productive turnover. Heartened by these observations, we added Mg(II) chloride to the reaction system and saw consequent formation in >90 % of OCM product with allyl alcohol as a partner [57]. Not only can such simple olefins be used as partners, but olefins of more relevance to protein modification can be employed, including allyl glucosides and mannosides (allowing protein glycosylation using OCM) and allyl-poly(ethylene glycol) (PEG) variants (allowing protein PEG-ylation through OCM). All the proteins remain active and functional following what is a remarkable series of three very different, sequential, site-selective modifications: (i) MSH-Dha, (ii) allyl thiol Sac formation, and (iii) OCM, thereby highlighting the benign and biocompatible nature of all of the methods used. Very recently, the strategic utility of Sac and other allyl sulfide-containing unnatural amino acids has been augmented by methods developed in our laboratory for their direct biosynthetic incorporation.

CONCLUSIONS

The development of novel synthetic protein chemistry methodology has the potential to enable an era of synthetic protein and synthetic biomolecule development in the field of synthetic biology akin to the rapid and remarkable development of synthetic organic chemistry. Both fields started (nearly a century apart) with the precise isolation and characterization of their respective key molecules. Synthetic organic chemistry's development via natural product synthesis into analog design and drug discovery using unnatural molecules perhaps foretells the likely development of synthetic biology. Total gene,

protein, and large oligosaccharide synthesis is already within grasp; the next phase will be to redesign such structures at the molecular level for intended function.

ACKNOWLEDGMENTS

I would like to thank the work of the BGD group over the past almost 10 years; this review is a selective overview of one aspect of their work. All in the group have worked fantastically hard, driven by a heartening and inspiring desire to solve questions that interest them, me, and, we hope, others. I feel privileged to have had such a group. I also feel honored to have had the support of faithful and like-minded sponsors and collaborators who have understood our approach and goals.

REFERENCES

1. B. G. Davis. *Science* **303**, 480 (2004).
2. T. R. Hupp, D. P. Lane. *J. Biol. Chem.* **270**, 18165 (1995).
3. B. Bader, K. Kuhn, D. J. Owen, H. Waldmann, A. Wittinghofer, J. Kuhlmann. *Nature* **403**, 223 (2000).
4. D. P. Gamblin, P. Garnier, S. van Kasteren, N. J. Oldham, A. J. Fairbanks, B. G. Davis. *Angew. Chem., Int. Ed.* **43**, 827 (2004).
5. B. G. Davis. *Chem. Rev.* **102**, 579 (2002).
6. H. T. Liu, L. Wang, A. Brock, C.-H. Wong, P. G. Schultz. *J. Am. Chem. Soc.* **125**, 1702 (2003).
7. V. Ramakrishnan, J. T. Finch, V. Graziano, P. L. Lee, R. M. Sweet. *Nature* **362**, 219 (1993).
8. J. C. M. van Hest, D. A. Tirrell. *Chem. Commun.* 1897 (2001).
9. (a) R. R. Schmidt, J. C. Castro-Palomino, O. Retz. *Pure Appl. Chem.* **71**, 729 (1999); (b) Z. Zhang, J. Gildersleeve, Y.-Y. Yang, R. Xu, J. A. Loo, S. Uryu, C.-H. Wong, P. G. Schultz. *Science* **303**, 371 (2004); (c) R. Xu, S. R. Hanson, Z. Zhang, Y.-Y. Yang, P. G. Schultz, C.-H. Wong. *J. Am. Chem. Soc.* **126**, 15654 (2004); (d) C. H. Röhrig, O. A. Retz, L. Hareng, T. Hartung, R. R. Schmidt. *ChemBioChem* **6**, 1805 (2005).
10. L. Polgar, M. L. Bender. *J. Am. Chem. Soc.* **88**, 3153 (1966).
11. K. E. Neet, D. E. Koshland. *Proc. Natl. Acad. Sci. USA* **56**, 1606 (1966).
12. Z. P. Wu, D. Hilvert. *J. Am. Chem. Soc.* **112**, 5647 (1990).
13. S. Mao, Z. Dong, J. Liu, X. Li, X. Liu, G. Luo, J. Shen. *J. Am. Chem. Soc.* **127**, 11588 (2005).
14. P. Berglund, G. DeSantis, M. R. Stabile, X. Shang, M. Gold, R. R. Bott, T. P. Graycar, T. H. Lau, C. Mitchinson, J. B. Jones. *J. Am. Chem. Soc.* **119**, 5265 (1997).
15. (a) B. G. Davis, R. C. Lloyd, J. B. Jones. *J. Org. Chem.* **63**, 9614 (1998); (b) U.S. Patent appl. 60/091,687 (1998).
16. (a) B. G. Davis, J. B. Jones. *Synlett* 1495 (1999); (b) B. G. Davis, S. J. Ward, P. M. Rendle. *Chem. Commun.* 189 (2001); (c) B. G. Davis. *Chem. Commun.* 351 (2001).
17. (a) R. C. Lloyd, B. G. Davis, J. B. Jones. *Bioorg. Med. Chem.* **8**, 1527 (2000); (b) R. C. Lloyd, B. G. Davis, J. B. Jones. *Bioorg. Med. Chem.* **8**, 1537 (2000).
18. (a) K. Matsumoto, B. G. Davis, J. B. Jones. *Chem. Commun.* 903 (2001); (b) K. Matsumoto, B. G. Davis, J. B. Jones, *Chem.—Eur. J.* **8**, 4129 (2002).
19. D. P. Gamblin, P. Garnier, S. J. Ward, N. J. Oldham, A. J. Fairbanks, B. G. Davis. *Org. Biomol. Chem.* **1**, 3642 (2003).
20. (a) D. Macmillan, R. M. Bill, K. A. Sage, D. Fern, S. L. Flitsch. *Chem. Biol.* **8**, 133 (2001); (b) G. G. M. Watt, J. Lund, M. Levens, V. S. K. Kolli, R. Jefferis, G.-J. Boons. *Chem. Biol.* **10**, 807 (2003); (c) J. Müllegger, H. M. Chen, R. A. J. Warren, S. G. Withers. *Angew. Chem., Int. Ed.* **45**, 2585 (2006).
21. J. L. Kice, T. W. S. Lee. *J. Am. Chem. Soc.* **100**, 5094 (1978).
22. B. G. Davis, A. J. Fairbanks et al. Unpublished results.

23. Glycoform Ltd. Unpublished results.
24. G. Bernardes, D. P. Gamblin, B. G. Davis. *Angew. Chem., Int. Ed.* **45**, 4007 (2006).
25. D. P. Gamblin, S. I. van Kasteren, G. J. L. Bernardes, J. M. Chalker, N. J. Oldham, A. J. Fairbanks, B. G. Davis. *Mol. Biosyst.* **4**, 558 (2008).
26. T. Hotchkiss, H. B. Kramer, K. J. Doores, D. P. Gamblin, N. J. Oldham, B. G. Davis. *Chem. Commun.* 4264 (2005).
27. (a) P. M. Rendle, A. P. Seger, J. Rodrigues, N. J. Oldham, R. R. Bott, J. B. Jones, M. M. Cowan, B. G. Davis. *J. Am. Chem. Soc.* **126**, 4750 (2004); (b) P. M. Rendle, A. P. Seger, J. Rodrigues, N. J. Oldham, R. R. Bott, J. B. Jones, M. M. Cowan, B. G. Davis. *Nature* **428**, 907 (2004); (c) C. M. Henry. *Chem. Eng. News* 12 April 2004; (d) C. Cookson. *Financial Times* 30 April 2004.
28. M. A. Robinson, S. T. Charlton, P. Garnier, X. Wang, S. S. Davis, A. C. Perkins, M. Frier, R. Duncan, T. J. Savage, D. A. Wyatt, S. A. Watson, B. G. Davis. *Proc. Natl. Acad. Sci. USA* **101**, 14527 (2004).
29. O. M. T. Pearce, K. D. Fisher, J. Humphries, L. W. Seymour, A. Smith, B. G. Davis. *Angew. Chem., Int. Ed.* **44**, 1057 (2005).
30. S. I. van Kasteren, H. B. Kramer, H. H. Jensen, S. J. Campbell, J. Kirkpatrick, N. J. Oldham, D. C. Anthony, B. G. Davis. *Nature* **446**, 1105 (2007).
31. C. W. Tornøe, C. Christensen, M. Meldal. *J. Org. Chem.* **67**, 3057 (2002).
32. V. V. Rostovtsev, L. G. Green, V. V. Fokin, K. B. Sharpless. *Angew. Chem., Int. Ed.* **41**, 2596 (2002).
33. (a) A. Deiters, T. A. Cropp, M. Mukherji, J. W. Chin, J. C. Anderson, P. G. Schultz. *J. Am. Chem. Soc.* **125**, 11782 (2003); (b) A. Deiters, P. G. Schultz. *Bioorg. Med. Chem. Lett.* **15**, 1521 (2005).
34. B. H. M. Kuijpers, S. K. Groothuys, A. R. Keereweer, P. J. L. M. Quaedflieg, R. H. Blaauw, F. L. van Delft, F. P. J. T. Rutjes. *Org. Lett.* **6**, 3123 (2004).
35. M. N. Fodje, S. Al-Karadaghi. *Protein Eng.* **15**, 353 (2002).
36. K. L. Kiick, E. Saxon, D. A. Tirrell, C. R. Bertozzi. *Proc. Natl. Acad. Sci. USA* **99**, 19 (2002).
37. S. I. van Kasteren, H. B. Kramer, D. P. Gamblin, B. G. Davis. *Nature Protocols* **2**, 3186 (2007).
38. T. R. Chan, R. Hilgraf, K. B. Sharpless, V. V. Fokin. *Org. Lett.* **6**, 2853 (2004).
39. (a) A. Leppänen, R. P. McEver, R. D. Cummings, P. Mehta, Y. B. Ouyang, K. L. Moore, T. Ju, W. M. Canfield, J. Helin, I. Van Die. *J. Biol. Chem.* **274**, 24838 (1999); (b) A. Leppänen, S. P. White, J. Helin, R. P. McEver, R. D. Cummings. *J. Biol. Chem.* **275**, 39569 (2000).
40. W. S. Somers, J. Tang, G. D. Shaw, R. T. Camphausen. *Cell* **103**, 467 (2000).
41. G. Köhler, C. Milstein. *Nature* **256**, 495 (1975).
42. G. J. L. Bernardes, E. J. Grayson, S. Thompson, J. M. Chalker, J. C. Errey, F. El Oualid, T. D. W. Claridge, B. G. Davis. Angew. *Chem., Int. Ed.* **47**, 2244 (2008).
43. D. N. Harpp, J. G. Gleason, J. P. Snyder. *J. Am. Chem. Soc.* **90**, 4181 (1968).
44. D. Crich, V. Krishnamurthy, T. K. Hutton. *J. Am. Chem. Soc.* **128**, 2544 (2006).
45. (a) K. Hashimoto, M. Sakai, T. Okuno, H. Shirahama. *Chem. Commun.* 1139 (1996); (b) M. Sakai, K. Hashimoto, H. Shirahama. *Heterocycles* **44**, 319 (1997); (c) N. M. Okeley, Y. Zhu, W. A. van der Donk. *Org. Lett.* **2**, 3603 (2000); (d) Y. Zhu, W. A. van der Donk. *Org. Lett.* **3**, 1189 (2001); (e) R. Quaderer, D. Hilvert. *Chem. Commun.* 2620 (2002).
46. F. P. Seebeck, J. W. Szostak. *J. Am. Chem. Soc.* **128**, 7150 (2006).
47. J. Wang, S. M. Schiller, P. G. Schultz. *Angew. Chem., Int. Ed.* **46**, 6849 (2007).
48. (a) W. Franek, P. K. Claus. *Monatsh. Chem.* **121**, 539 (1990); (b) J. Matsuo, T. Kozai, H. Ishibashi. Org. *Lett.* **8**, 6095 (2006).
49. G. J. L. Bernardes, J. M. Chalker, J. C. Errey, B. G. Davis. *J. Am. Chem. Soc.* **130**, 5052 (2008).
50. For sulfilimine reversion to sulfide, see: N. Furukawa, T. Omata, T. Yoshimura, T. Aida, S. Oae. *Tetrahedron Lett.* **13**, 1619 (1972).
51. F. L. Zhang, P. J. Casey. *Annu. Rev. Biochem.* **65**, 241 (1996).
52. P. Trojer, D. Reinberg. *Cell* **127**, 213 (2006).

53. 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* **128**, 1003 (2007).
54. J. Guo, J. Wang, J. S. Lee, P. G. Schultz. *Angew. Chem., Int. Ed.* **47**, 6399 (2008).
55. (a) J. C. M. van Hest, D. A. Tirrell. *FEBS Lett.* **428**, 68 (1998); (b) J. C. M. van Hest, K. L. Kiick, D. A. Tirrell. *J. Am. Chem. Soc.* **122**, 1282 (2000).
56. (a) S. H. Hong, R. H. Grubbs. *J. Am. Chem. Soc.* **128**, 3508 (2006); (b) J. P. Jordan, R. H. Grubbs. *Angew. Chem., Int. Ed.* **46**, 5152 (2007); (c) J. B. Binder, J. J. Blank, R. T. Raines. *Org. Lett.* **9**, 4885 (2007); (d) L. Gulajski, A. Michrowaska, J. Naroznik, Z. Kaczmarska, L. Rupnicki, K. Grela. *ChemSusChem* **1**, 103 (2008).
57. Y. A. Lin, J. M. Chalker, N. Floyd, G. J. L. Bernardes, B. G. Davis. *J. Am. Chem. Soc.* **130**, 9642 (2008).
58. A. Fürstner, G. Seklel, N. Kindler. *Tetrahedron* **55**, 8215 (1999).
59. J. Feldman, J. S. Murdezek, W. M. Davis, R. R. Schrock. *Organometallics* **8**, 2260 (1989).
60. A. Fürstner, K. Langemann. *J. Am. Chem. Soc.* **119**, 9130 (1997).