Amino Acids (2010) 39:1381–1384
DOI 10.1007/s00726-010-0594-3

ORIGINAL ARTICLE

Synthesis of bioorthogonal and crosslinking amino acids for use in peptide synthesis

G. S. M. Sundaram · Ian R. Morgan · Eric M. Tippmann

Received: 12 March 2010 / Accepted: 6 April 2010 / Published online: 22 April 2010

Abstract The ability to incorporate non-canonical amino acids into proteins by genetic or chemical methods allows one to introduce novel chemical properties into a protein at a defined residue. Such a residue may then be modified using common organic transformations. In this way, the structure or function of the peptide may be altered without perturbing any of the other neighbouring amino acids in the peptide chain. Here, we describe the syntheses and potential applications of multiple *para*-substituted phenylalanine derivatives comprising an isothiocyanate, α-diazoketone, or nitrone functionality. In all, three novel amino acids were synthesized in good overall yields. These non-canonical amino acids permit the further development of in vitro and in vivo chemoselective and regioselective bioconjugate reactions not possible with other reagents.

Keywords Non-canonical amino acids · Expanded genetic code · Bioconjugate chemistry

Introduction

A covalent linkage generated in vivo or in vitro between a small molecule and a macromolecule can be used to characterize and map multiple structural features (Brunner 1996) or to probe for specific receptors (Bond et al. 2009; Hatanaka and Sadakane 2002). In this way, the development of bioconjugate reactions using affinity labels has permitted many elegant and structurally informative studies (Hino et al. 2005; Suchanek et al. 2005). With the advent of an expanded genetic code, many of these bioconjugate reagents can now be directly incorporated into proteins using the host's cell translational machinery (Xie and Schultz 2006).

Many photoaffinity probes are based on highly reactive functional groups such as carbenes (Platz et al. 1991) and nitrenes. Such reactive groups may be generated thermally or photochemically, through the liberation of $N_{2(g)}$, from diazo, diazirine, or aryl azide precursors (Buchmueller et al. 2003). As a result of the high reactivity of these reagents, crosslinking reactions are often prone to further rearrangement, which generates side products incapable of crosslinking functional groups common to biological systems (Ford et al. 1998). Furthermore, nucleophilic carbenes may insert indiscriminately into nascent XH bonds (X = C, N, O, S), as well as the ketenimine derived from aryl azide activation reacts with little selectivity with nucleophiles (Platz 2002). Other, more selective chemical coupling strategies exist, such as the global labelling of surface lysines on a protein, but these inherently suffer from a lack of chemo- and site-selectivity (Degani and Heller 1988). For these reasons, we have targeted the development of other functional groups, with varied reactivity, which may lead to enhanced use with chemoselective reagents useful for bioconjugate and other transformations when these amino acids are incorporated into proteins and peptides (Fig. 1).

Two of the functional groups that appear to be suitable for delivering chemoselective reactivity in bioconjugate transformations are the isothiocyanate and the ketene

Electronic supplementary material The online version of this article (doi:10.1007/s00726-010-0594-3) contains supplementary material, which is available to authorized users.

G. S. M. Sundaram · I. R. Morgan · E. M. Tippmann (✉)
Cardiff University School of Chemistry,
Main Building Park Place, Cardiff CF10 3AT, UK
e-mail: tippmann@cf.ac.uk

I. R. Morgan
e-mail: morganir@cf.ac.uk

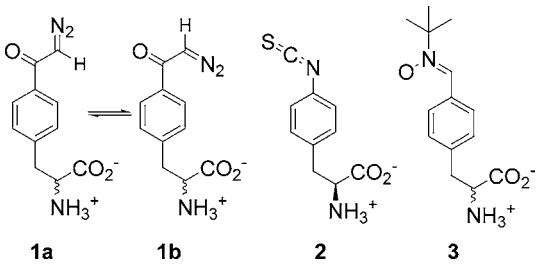


Fig. 1 Structures of targeted amino acids

(Tidwell 2006). Taken together, their respective reactivity profiles span several orders of magnitude with regards to electrophilic character (Hermonson 1996). They have the added feature of permitting spatial temporal control over the crosslinking reaction through the modulation of pH (isothiocyanate), or through activation by ultraviolet light (diazoketone to ketene). The addition of these compounds to the battery of other crosslinking amino acids should permit novel bioconjugate reactions not currently possible. A third functional group, the nitrone, has found widespread use in both chemistry and biology. In chemistry, a nitrone is a useful reagent for several organic transformations due to its ability to participate in 1,3-dipolar cycloaddition reactions and this should translate to bioorthogonal activity in a biological system (Gothelf and Jørgensen 1998). In biology, nitrone chemistry is closely related to that of the nitroxyl group, which is now commonly used as a spin label (Hubbell et al. 2000; Zhao et al. 2000). Nitrones are commonly used in biological research as radical scavengers. Studying the product of a nitrone-radical reaction has led to many in vivo and in vitro electron paramagnetic resonance studies (Floyd et al. 2008).

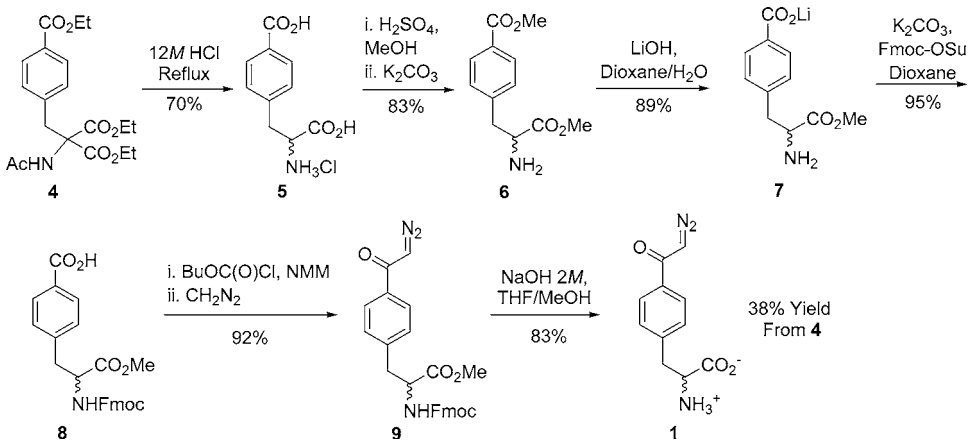
The main impetus in pursuing these three functional groups is that they have already found such widespread use in biological applications, and we can draw upon these many citations for inspiration in new applications. For example, the use of phenylisothiocyanate in the Edman degradation to sequence of *N*-terminus of peptides and proteins (Stryer 1995) and the use fluorescein isothiocyanate as a highly useful bioconjugate reagent are clear examples of the power of a bioconjugate transformations to impact the study of biology (Heymer et al. 1970; Vermes et al. 1995).

Results and discussion

Our approach to the synthesis of the α-diazoketone **1** is shown in Scheme 1. The synthesis of **1** was accomplished with a 38% yield over six steps from the literature compound **4**, with individual yields ranging from 70 to 95%. A key step was the late introduction of the diazo moiety, **9**, which avoids problems with the sensitivity of the diazoketone to acidic conditions (foot note, aryl diazoketones decompose instantly with neat TFA). Interestingly, the title compound **1** was observed to be present as a mixture of conformers (supplementary information). Amino acid **1** was ultimately synthesized via a synthetic route that resulted in racemic **1**. Using a racemic mixture for in vivo applications stipulates that only 50% of the material, the (S)-enantiomer for these designer amino acids, is available for use by the host organism. A general modification to the synthesis in Scheme 1 should allow an enzymatic resolution step to preferentially obtain the (S)-enantiomer.

Although the in vivo incorporation of a nonstandard amino acid is only possible with an (S)-enantiomer, the method is equally successful with racemic mixtures (Lee et al. 2009). If an amino acid is deemed to be nontoxic, then the added cost associated with preparing the (S)-enantiomer is often overwhelmed by the efficacy of obtaining the racemic form. Our preliminary results suggest that the diazoketone **1** is, in fact, non-toxic to *Escherichia coli* (data not shown). This stands in contrast to the diazoketone derived from glutamic acid, which was observed to be highly toxic to *E. coli* (Liu and Schultz 1999).

Scheme 1 Synthesis of *p*-DKPhe **1**



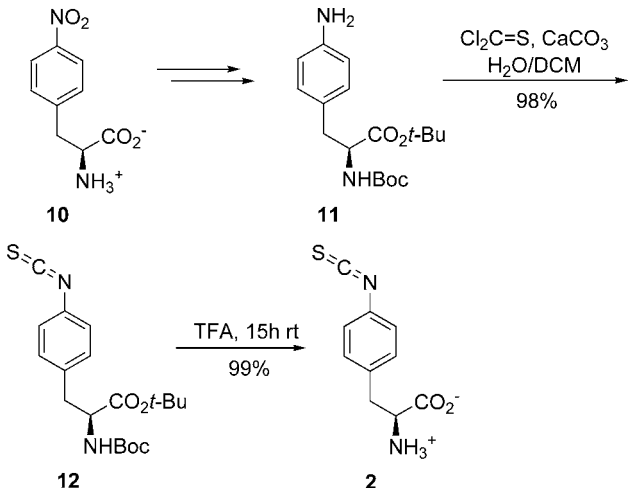
This difference in reactivity between an aryldiazoketone and an aliphatic diazoketone is also supported by other work, which incorporated the diazoketone moiety into protease inhibitors (Delpierre and Fruton 1966).

The precursor **12** of the isothiocyanate amino acid **2** was obtained from a documented procedure (Atkinson et al. 1999). This intermediate was never isolated and characterized by King and co-workers, but these data are reported here. Concomitant deprotection of the *tert*-Butyl ester and Boc-amine using trifluoroacetic acid furnished the desired stable product **2**, essentially in a quantitative yield (Scheme 2).

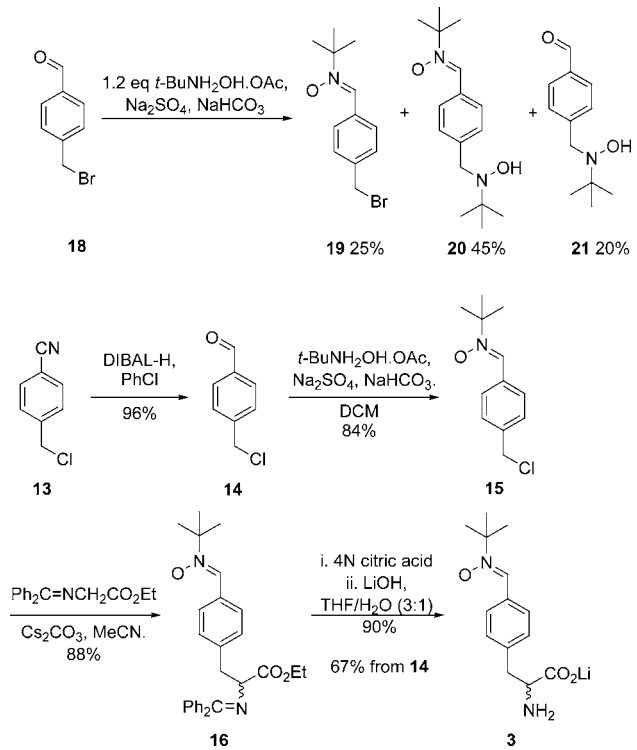
It should be mentioned that isothiocyanates have significant electrophilic character and would be expected to react readily with activated nucleophiles (Hermonson 1996). This presents new challenges for attempting to genetically encoded this amino acid. In *E. coli*, which is the more common host organism to showcase expanded genetic codes, is known to buffer its intracellular at ~pH 7.4 (Slonczewski et al. 1981). In this work, the use of trifluoroacetic acid deprotection avoided any complications with activating the isothiocyanate in situ. It is also intriguing that, in an appropriate host or application, **2** has the capacity to provide spatio-temporal control over crosslinking reactions through modulation of pH, and could therefore allow the isothiocyanate to be used in chemistry reminiscent of the Edman degradation (Stryer 1995) or for the synthesis of amino acid functionalized ligands amenable to biological imaging.

The initial attempt to synthesize the nitrone amino acid **3** began with 4-(bromomethyl)-benzaldehyde **18**. However, installation of the nitrone via aldehyde **18** resulted in the significant competing displacement of the bromide, and thus a mixture of products was obtained (Scheme 3). A second route was developed based on 4-(chloromethyl)-benzaldehyde synthesized by a DIBAL-H reduction of commercially available 4-(chloromethyl)-cyanobenzene. This method resulted in excellent yields of nitrone **15** and then coupling with a protected glycinate derivative (Scott et al. 2002) followed by an optimized deprotection protocol, afforded the nitrone amino acid **3** in a high yield, albeit as a racemic mixture. It also should be mentioned that the deprotection of the nitrone can often be accompanied by hydrolysis of the nitrone to the aldehyde, especially at elevated temperature, but this can be minimized by the mild deprotection protocol reported here (supporting information). The low lying oxidation potential of the nitrone is well suited to impart antioxidant properties to a peptide or protein and could thus complement other studies that worked to genetically encode oxidative damage (Neumann et al. 2008).

Here, we have described the synthesis of α-diazoketone *p*-(diazoacetyl)phenylalanine **1**, isothiocyanate *p*-(isothiocyanyl) phenylalanine **2**, and nitrone *p*-(*tert*-butylnitronyl) phenylalanine **3** which, to our knowledge, are all novel amino acids. The added utility of each amino acid is that they should be amenable to most methods that incorporate non-canonical amino acids site-selectively into proteins. One methodology uses orthogonal tyrosyl tRNA/aminoacyl tRNA synthetase pairs (tRNA/aaRS), and has been shown to be especially well suited for para-substituted phenylalanines (Xie and Schultz 2006). Furthermore, the novel activity of **1** and **3** are compatible with semi-synthetic



Scheme 2 Synthesis of *p*-ITCPhe **2**



Scheme 3 Synthesis of *p*-PBNphenylalanine **3**

preparations (Pellois and Muir 2006) of proteins, or in solid-phase synthesis. The current list of genetically encoded amino acids is largely dominated by commercially available amino acids. To continue to enhance the utility of these powerful methods to translationally modifying proteins, the development of amino acids with diverse structures and unique reactivity is clearly warranted. To this end, we have detailed the synthesis of three novel amino acids. We are currently investigating the feasibility of these molecules as candidates to further expand the genetic code.

Materials and methods

NMR spectra were recorded at 400 MHz (^{1}H NMR) and 100 MHz (^{13}C NMR) in $CDCl_3$ solution at 25°C, if not otherwise indicated, with the solvent signals as internal reference. Reactions which required an inert atmosphere were carried out under dry N_2 with flame-dried glassware. All commercial reagents were purchased from Sigma-Aldrich Company Ltd. (Poole, United Kingdom). High resolution mass spectra were recorded on a ESI-Orbitrap instrument.

Acknowledgments This research was supported by the Engineering and Physical Sciences Research Council in UK. We would like to thank Dr R. Jenkins for his expert advice and help with the analytical analysis.

References

Atkinson RN, Moore L, Tobin J, King SB (1999) Asymmetric synthesis of conformationally restricted L-arginine analogues as active site probes of nitric oxide synthase. J Org Chem 64(10):3467–3475

Bond MR, Zhang H, Vu PD, Kohler JJ (2009) Photocrosslinking of glycoconjugates using metabolically incorporated diazirine-containing sugars. Nat Protoc 4(7):1044–1063

Brunner J (1996) Use of photocrosslinkers in cell biology. Trends Cell Biol 6(4):154–157

Buchmueller KL, Hill BT, Platz MS, Weeks KM (2003) RNA-tethered phenyl azide photocrosslinking via a short-lived indiscriminant electrophile. J Am Chem Soc 125(36):10850–10861

Degani Y, Heller A (1988) Direct electrical communication between chemically modified enzymes and metal electrodes. 2. Methods for bonding electron-transfer relays to glucose oxidase and D-amino-acid oxidase. J Am Chem Soc 110:2615

Delpierre GR, Fruton JS (1966) Specific Inactivation of pepsin by a diazo ketone. Proc Natl Acad Sci USA 56(6):1817–1822

Floyd RA, Kopke RD, Choi CH, Foster SB, Doblas S, Towner RA (2008) Nitrones as therapeutics. Free Radic Biol Med 45(10):1361–1374

Ford F, Yuzawa T, Platz MS, Matzinger S, Fulscher M (1998) Rearrangement of dimethylcarbene to propene: study by laser flash photolysis and ab initio molecular orbital theory. J Am Chem Soc 120(18):4430–4438

Gothelf K, Jørgensen K (1998) Asymmetric 1, 3-dipolar cycloaddition reactions. Chem Rev 98(2):863–910

Hatanaka Y, Sadakane Y (2002) Photoaffinity labeling in drug discovery and developments: chemical gateway for entering proteomic frontier. Curr Top Med Chem 2(3):271–288

Hermonson GT (1996) Bioconjugate techniques. Elsevier, San Diego

Heymer B, Haferkamp O, Schmid WC (1970) Preparation and use of fluorescein isothiocyanate-labelled streptococcal cell-wall carbohydrates. Z Immunitatsforsch Allerg Klin Immunol 140(5):457–467

Hino N, Okazaki Y, Kobayashi T, Hayashi A, Sakamoto K, Yokoyama S (2005) Protein photo-cross-linking in mammalian cells by site-specific incorporation of a photoreactive amino acid. Nat Methods 2(3):201–206

Hubbell WL, Cafiso DS, Altenbach C (2000) Identifying conformational changes with site-directed spin labeling. Nat Struct Biol 7(9):735–739

Lee HS, Spraggon G, Schultz PG, Wang F (2009) Genetic incorporation of a metal-ion chelating amino acid into proteins as a biophysical probe. J Am Chem Soc 131(7):2481–2483

Liu D, Schultz P (1999) Progress toward the evolution of an organism with an expanded genetic code. Proc Natl Acad Sci USA 96(9):4780–4785

Neumann H, Hazen JL, Weinstein J, Mehl RA, Chin JW (2008) Genetically encoding protein oxidative damage. J Am Chem Soc 130(12):4028–4033

Pellois JP, Muir TW (2006) Semisynthetic proteins in mechanistic studies: using chemistry to go where nature can't. Curr Opin Chem Biol 10(5):487–491

Platz MS (2002) Comparison of phenylcarbene and phenylnitrene. Acc Chem Res 28(12):487–492

Platz M, Admasu AS, Kwiatkowski S, Crocker PJ, Imai N, Watt DS (1991) Photolysis of 3-aryl-3-(trifluoromethyl)diazirines: a caveat regarding their use in photoaffinity probes. Bioconjug Chem 2(5):337–341

Scott W, O'Donnell M, Delgado F, Alsina J (2002) A solid-phase synthetic route to unnatural amino acids with diverse side-chain substitutions. J Org Chem 67(9):2960–2969

Slonczewski JL, Rosen BP, Alger JR, Macnab RM (1981) pH homeostasis in *Escherichia coli*: measurement by 31P nuclear magnetic resonance of methylphosphonate and phosphate. Proc Natl Acad Sci USA 78(10):6271–6275

Stryer L (1995) Biochemistry, 4th edn. Freeman and Company, New York

Suchanek M, Radzikowska A, Thiele C (2005) Photo-leucine and photo-methionine allow identification of protein-protein interactions in living cells. Nat Methods 2(4):261–267

Tidwell T (2006) Ketene chemistry after 100 years: ready for a new century. Eur J Org Chem 2006(3):563–576

Vermes I, Haanen C, Steffens-Nakken H, Reutelingsperger C (1995) A novel assay for apoptosis. Flow cytometric detection of phosphatidylserine expression on early apoptotic cells using fluorescein labelled Annexin V. J Immunol Methods 184(1):39–51

Xie J, Schultz PG (2006) A chemical toolkit for proteins—an expanded genetic code. Nat Rev Mol Cell Biol 7:775–782

Zhao M, Kalai T, Hideg K, Altenbach C, Hubbell WL, Kaback HR (2000) Binding of spin-labeled galactosides to the lactose permease of *Escherichia coli*. Biochemistry 39(37):11381–11388