Genetic encoding of the post-translational modification 2-hydroxyisobutryl-lysine

William A. Knight* and T. Ashton Cropp*  
Department of Chemistry, Virginia Commonwealth University, 1001 W. Main Street, Richmond, Virginia 23284, USA.

Abstract

We report the synthesis and genetic encoding of a recently discovered posttranslational modification, 2-hydroxyisobutryl-lysine, to the genetic code of *E. coli*. The production of homogeneous proteins containing this amino acid will facilitate the study of modification in full-length proteins.

The number and complexity of identified lysine post-translational modifications (PTMs) continues to grow. Historically the most commonly observed protein modifications are located at lysine side chains, such as methylation or acetylation. In these examples the charge state of a lysine side chain is altered, which in turn can give rise to changes in interactions with DNA and other biomolecules. In the case of epigenetic control of gene expression is vital, and malfunction of these systems can be a hallmark of disease.

In addition to acetylation and methylation, it has been reported that lysine residues can be modified by malonylation, propionylation and butyrylation, succinylation, and crotonylation. These modifications are derived from intracellular acyl-CoA metabolites and provide wide spectrum of epigenetic control of gene expression. The extent to which these modifications are actively added and removed by enzymes is a current interest in deciphering the “histone code”. Recently proteomics profiling revealed a new lysine modification that was identified as 2-hydroxyisobutryl lysine (K_{hib})(1, Scheme 1). This modification appears to be conserved throughout evolution, appearing in human, mouse, *Drosophila*, and yeast cells. Intriguingly, unlike other histone marks which occur mainly at the N-terminus of histone proteins, K_{hib} is seen distributed throughout the globular protein structure, and includes many lysine residues that have not been previously reported to be modified. Moreover, in comparison to other lysine PTMs, K_{hib} is quite large and has the unique potential to serve as a hydrogen bond donor, raising the possibility of significant structural changes and alternative intermolecular interactions mediated by this modification. All of these point to an important role in chromatin function.
Critical to the study of PTMs and is the ability to produce pure protein samples containing homogeneous compositions of a modified residue. Lysine modifications in particular may offer the ability to generate designer chromatin to study the effects of PTMs either individually or in concert. Modified amino acids can be added to proteins using combinations of peptide synthesis and expressed protein ligation. This approach has the ability to precisely add PTMs, but it is technically challenging and cannot be scaled easily. As an alternative, some amino acids containing PTMs have also been added to the genetic codes of both prokaryotes and eukaryotes, allowing biosynthetic production of target proteins containing site-specific modifications. Indeed most lysine modifications have been incorporated using the pyrrolysine translational machinery. These in vivo production methods are very adaptable to biochemical laboratories. Moreover, biosynthetic production of proteins containing PTMs opens the door to more sophisticated experiments such as phage display, incorporation of isotopic labels, and a wide variety of in vivo experiments. Towards these goals we describe the synthesis and addition of 2-hydroxyisobutyryl-lysine to the genetic code of E. coli.

The synthesis of Khib was performed from lysine by first preparing the copper complex, which allows for selective acylation of the ε-nitrogen (Scheme 1). This complex was then treated with 2-hydroxyisobutryl-O-succinimide ester and the copper removed by chelating agent to generate the final product. This synthesis is simple and can provide gram-scale quantities of amino acid for protein expression studies.

We decided to test whether Khib is a substrate for several variants of the pyrrolysyl-tRNA synthetase (PylRS) derived from either Methanosarcina barkeri (Mb) or Methanosarcina mazei (Mm) (see ESI). These included the wild-type enzymes and variants that have been shown to have relaxed substrate specificity towards other, larger unnatural amino acids. The screen utilized an expression plasmid for superfolder green fluorescent protein (sfGFP) containing an amber stop codon, TAG, in place of the codon for Y151. The plasmid also contains the gene encoding the Mm-pyrrolysyl tRNA (pylT). Incorporation of unnatural amino acid leads to production of full-length protein and a corresponding increase in cellular fluorescence. Using a plate-based screening assay we first examined fluorescence in the presence and absence of Khib for any observable differences. As a positive control, we also used Nε-(tertbutyloxycarbonyl)-L-lysine (BocLys, Scheme 1), which is a known substrate for PylRS. Among the five variants we screened the most observable fluorescence difference was obtained using wild-type Mm PylRS. While the observable fluorescence was weak in comparison to BocLys, it did show clear differences when compared to controls (see ESI, Figure S2). Variants with larger active sites did not appear to accept Khib as substrate.

We chose to perform medium-scale expression of sfGFP in the presence and absence of 5mM Khib and purified the resulting His-tagged proteins proteins using Ni²⁺ affinity chromatography. As shown in Figure 1, we observed robust protein expression (~10mg/L) only the presence of Khib, indicating that this amino acid can serve as a substrate, without further evolution of PylRS. No protein is seen in the absence of Khib, verifying that endogenous amino acids are not substrates for Mm PylRS. In order to verify the position and identity of the mutation, the gel slice of the produced protein was excised and subjected to
Upon examining the tryptic fragments by LC/MS/MS, the spectrum of the expected fragment was trapped in a +2 charge state (Figure 2). Fragment masses of this ion are consistent with site-specific incorporation of K_hib at the correct position, in place of Y151. No masses that correspond to the same fragment containing other natural amino acids at position 151 were seen. In addition to tryptic peptide analysis, the protein samples were analyzed by ESI-MS on intact protein which also confirms incorporation of the amino acid (see ESI, Figure S3). Interestingly, we did not observe masses corresponding to a lysine residue at position 151 (or a resulting tryptic fragment), which would be indicative of active deacylation of K_hib in *E. coli*. Removal of other lysine PTMs has been previously observed and ascribed to bacterial sirtuins\(^{13,12}\), and can be prevented by the use of a nicotinamide enzyme inhibitor. It is possible that K_hib residues are not a substrate for these enzymes at all or when in the context of this mutation position in sfGFP.

**Conclusions**

In conclusion, we have demonstrated that the pyrrolysyl-tRNA synthetase (PyRS) has a suitably relaxed substrate specificity to accommodate 2-hydroxyisobutyrly-lysine (K_hib). This, coupled with a quick and high-yielding synthesis of this important amino acid makes this technology accessible to anyone working in the field of protein chemistry. We anticipate the ability to produce homogeneous samples of proteins containing K_hib will facilitate the study of the histone code and chromatin function.

**Supplementary Material**

Refer to Web version on PubMed Central for supplementary material.

**Acknowledgments**

This work was funded in part by NIH/NIGMS (GM084396).

**Notes and references**


Scheme 1.
Synthesis of Khib.
Figure 1.
Production of sfGFP containing 1 at position 151. No protein is produced in the absence of added unnatural amino acid.
Figure 2.
MS/MS spectrum of tryptic fragment of sfGFP bearing $K_{\text{hib}}$ at position 151.