*Proc. Natl. Acad. Sci. USA*
Vol. 91, pp. 12544–12548, December 1994
Biochemistry

# Subtiligase: A tool for semisynthesis of proteins

THOMAS K. CHANG*[†], DAVID Y. JACKSON*, JOHN P. BURNIER[‡], AND JAMES A. WELLS*[§]

Departments of *Protein Engineering and [‡]BioOrganic Chemistry, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, CA 94080

*Communicated by Harry B. Gray, August 1, 1994*

**ABSTRACT** **A variant of subtilisin BPN′, which we call subtiligase, has been used to ligate esterified peptides site-specifically onto the N termini of proteins or peptides in aqueous solution and in high yield. We have produced biotinylated or heavy-atom derivatives of methionyl-extended human growth hormone (Met-hGH) by ligating it onto synthetic peptides containing biotin or mercury. Polyethylene glycol (PEG)-modified atrial natriuretic peptide (ANP) was produced by ligating ANP onto peptides containing sites for PEG modification. We have established the N-terminal sequence requirements for efficient ligation onto proteins, using either synthetic substrates or pools of filamentous phage containing Met-hGH with random N-terminal sequences (substrate phage). To facilitate ligations involving proteins with highly structured or buried N termini, a more stable subtiligase was designed that effectively ligates peptides onto Met-hGH even in 4 M guanidine hydrochloride. The use of subtiligase should expand the possibilities for protein semisynthesis and rational protein design.**

Coupling of synthetic peptides onto a protein or protein fragment is a potentially powerful means of introducing natural or nonnatural constituents to probe and design protein function (for reviews, see refs. 1–3). For example, enzymatic and chemical peptide ligation methods have been used to produce a variety of semisynthetic proteins (4–14). However, in virtually all cases, coupling relied on noncovalent or covalent preassociation or some other special secondary structural feature of the fragments to be ligated. Thus, a more general method for the semisynthesis of any protein would be useful.

Recently, a variant of the serine protease subtilisin BPN′, in which the catalytic Ser-221 was converted to cysteine and Pro-225 was converted to alanine, was shown to ligate dipeptides onto tetrapeptide esters with high $k_{cat}$ values (20 $s^{-1}$) and with little hydrolysis of the ester substrate or proteolysis of the amide product (15). Here, we evaluate the sequence requirements for efficient peptide ligation using this double mutant of subtilisin BPN′, termed "subtiligase," and design more stable variants of it. These studies suggest that subtiligase may be generally useful for site-specific ligation of peptides containing affinity handles, isotopic labels, heavy atoms, or other nonnatural constituents onto the N terminus of proteins or protein fragments.

## MATERIALS AND METHODS

**Materials.** Enzymes for DNA manipulations were from New England Biolabs or BRL. Streptavidin–horseradish peroxidase (SAHRP) was from GIBCO/BRL, column resins were from Pharmacia, and atrial natriuretic peptide (ANP) was from Bachem. Oligonucleotides were synthesized by the Oligonucleotide Synthesis Group at Genentech.

**Peptide Synthesis.** All peptides were synthesized by standard methods (16). Peptides esterified with glycol-conjugated phenylalanylamide (glc-F-amide) were synthesized as described (15). For ligations onto immobilized supports, peptides were synthesized on 96-well (≈0.17 nmol per well) CovaLink ELISA plates (Nunc) by using *N*-(9-fluorenylmethoxycarbonyl) (Fmoc)-protected amino acids and dicyclohexylcarbodiimide (DCC) activation in dimethyl sulfoxide (DMSO). The plates were washed with 5% piperidine in methanol to neutralize the amino linkers and incubated with 100 μl of DMSO containing 1 mM Fmoc-Ala and 0.5 mM DCC for 10 min at 25°C. The plates were washed with DMSO, and the Fmoc protecting groups were removed with 5% piperidine in methanol. These steps were repeated to generate all 400 natural tripeptides having the form Xaa-Xaa-Ala. Side chains were deprotected by adding 150 μl of $CF_3COOH$/anisole, 95:5 (vol/vol) per well for 10 min, and the plates were washed three times with 100 mM Tricine (pH 8).

To each well was added 100 μl of 100 mM Tricine (pH 8) containing 10 μM biotinylated Lys-Gly-Ala-Ala-Pro-Phe-glc-Phe-amide (bio-KGAAPF-glc-F-amide), followed by addition of 10 μl of 10 μM subtiligase in Tricine buffer. The ligation proceeded for 15 min at 25°C, and the plate was washed with Tricine buffer. SAHRP (Sigma) was diluted 1:5000 in PBS (10 mM sodium phosphate, pH 7.4/150 mM NaCl) containing 0.01% Tween and added to the plates (100 μl per well), which were incubated for 30 min. Plates were washed five times with PBS/Tween. *o*-Phenylenediamine (60 mg) in 100 ml of 0.1 M citrate/phosphate buffer (pH 5) containing 50 μl of 30% $H_2O_2$ was added (100 μl per well). The plates were developed for 20 min at 25°C, and the reaction was quenched by the addition of 0.5 M $H_2SO_4$ (50 μl per well). The absorbance at 405 nm was read with an ELISA plate reader.

**Substrate Phage Selections.** Phagemid phGHam-g3 (17) was modified to add a methionine codon at the mature N terminus of hGH. The first three residues of Met-hGH were randomized with an oligonucleotide containing NNS codons, where N represents all four bases and S represents G or C, by using site-directed mutagenesis (18). Phagemid particles were prepared from *Escherichia coli* strain XL1-Blue (Stratagene) (19) by using helper phage M13KO7 (20). About $10^{11}$ phage were bound to avidin-conjugated agarose (Pierce) in buffer A (phosphate-buffered saline containing 0.5% bovine serum albumin) to filter out the sequences that bind directly to avidin or the matrix. The remaining phage were incubated with 0.4 μM subtiligase and 1 mM iminobiotin-KGAAPF-glc-F-amide for 1 hr at room temperature. The iminobiotin-conjugated peptide was used because the biotin–avidin interaction is too strong to release under gentle elution conditions (21). Phage were precipitated twice by the addition of

Abbreviations: hGH, human growth hormone; Δ2 or Δ8 hGH, hGH with deletions of the first two or eight residues; ANP, atrial natriuretic peptide; PEG, polyethylene glycol; glc-F-amide, glycol-conjugated phenylalanyl amide; NHS, *N*-hydroxysuccinimide; SAHRP, streptavidin-horseradish peroxidase; bio-KGAAPF- and suc-KKKGAAPF-, biotinylated Lys-Gly-Ala-Ala-Pro-Phe- and succinylated Lys-Lys-Lys-Gly-Ala-Ala-Pro-Phe-.

[†]Present address: Division of Biotechnology, Syntex Discovery Research, 3401 Hillview Avenue, Palo Alto, CA 94303.
[§]To whom reprint requests should be addressed.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Table 1. Nucleophile specificity for subtiligase

| | P1′ | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| P2′ | M | F | K | L | R | A | S | H | W | Y | G | C | I | V | T | Q | N | P | D | E | Avg |
| T | 1.9 | 1.6 | 1.2 | 1.4 | 1.1 | 0.9 | 1.1 | 1.1 | 0.9 | 0.7 | 0.9 | 0.8 | 0.6 | 0.6 | **0.3** | 0.5 | **0.2** | **0.3** | **0.3** | 0.5 | 0.84 |
| K | 1.5 | 1.5 | 1.1 | 1.3 | 1.5 | 1.3 | 1.2 | 0.8 | 1.2 | 1.1 | 1.2 | 0.7 | **0.2** | 0.4 | 0.4 | **0.2** | **0.2** | **0.2** | **0.1** | **0** | 0.80 |
| V | 1.3 | 1.2 | 1.3 | 1.2 | 1.2 | 0.9 | 0.9 | 1.0 | 0.8 | 0.6 | 1.1 | 1.1 | 0.5 | 0.6 | 0.4 | 0.5 | **0.3** | **0.3** | **0.2** | 0.5 | 0.80 |
| H | 1.2 | 1.1 | 1.0 | 1.0 | 1.2 | 1.3 | 1.3 | 0.9 | 0.8 | 0.9 | 0.8 | 0.4 | **0.3** | 0.6 | **0.3** | **0.3** | **0.2** | **0.3** | **0.3** | **0.3** | 0.73 |
| Y | 1.1 | 0.9 | 1.3 | 1.1 | 0.7 | 1.1 | 1.2 | 0.5 | 1.2 | 1.0 | 0.7 | 1.1 | 0.5 | 0.4 | **0.2** | **0.3** | **0.3** | **0.1** | **0.3** | **0.2** | 0.70 |
| R | 1.5 | 1.1 | 1.3 | 1.3 | 1.0 | 1.1 | 1.0 | 1.0 | 0.8 | 0.7 | 0.7 | 0.5 | 0.5 | **0.2** | 0.4 | **0.2** | **0.2** | **0.2** | **0.1** | **0** | 0.70 |
| S | 1.3 | 1.0 | 1.1 | 1.0 | 1.1 | 1.2 | 1.0 | 1.0 | 1.1 | 0.6 | 0.8 | 0.9 | **0.1** | **0.3** | **0** | **0.3** | 0.4 | **0.1** | **0.2** | **0.2** | 0.68 |
| A | 1.0 | 0.9 | 1.0 | 1.2 | 1.0 | 1.2 | 0.7 | 1.0 | 1.0 | 1.1 | 0.7 | 0.5 | **0.2** | 0.5 | 0.5 | **0.3** | **0.2** | **0.2** | **0.2** | **0.1** | 0.67 |
| M | 0.9 | 0.9 | 1.0 | 1.5 | 0.7 | 1.1 | 0.7 | 0.5 | 1.5 | 1.1 | 0.5 | 0.9 | 0.7 | **0.3** | **0.3** | **0.1** | **0.3** | **0.2** | **0** | **0.1** | 0.66 |
| I | 1.6 | 1.5 | 1.3 | 0.6 | 0.7 | 0.8 | 1.3 | 0.6 | 0.8 | 0.8 | 0.9 | 0.4 | 0.4 | 0.4 | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | **0.2** | 0.66 |
| D | 1.2 | 1.1 | 0.9 | 1.1 | 1.0 | 0.6 | 1.2 | 0.9 | 0.7 | 0.8 | 0.9 | 0.4 | **0.3** | **0.2** | **0.3** | **0.3** | **0.3** | **0.3** | **0.3** | **0.3** | 0.65 |
| F | 0.8 | 1.6 | 1.2 | 1.1 | 1.0 | 0.9 | 1.4 | 0.5 | **0.2** | 0.9 | 0.5 | 0.9 | 0.5 | 0.4 | **0.3** | **0.2** | **0.1** | **0** | **0.1** | **0.1** | 0.63 |
| N | 0.9 | 1.0 | 1.1 | 1.4 | 0.9 | 1.0 | 0.8 | 0.8 | 0.7 | 1.1 | 0.8 | 0.4 | **0.2** | **0.2** | 0.4 | **0.2** | **0.3** | **0.1** | **0.1** | **0.2** | 0.62 |
| L | 1.3 | 0.9 | 1.0 | 1.0 | 1.0 | 1.1 | 0.7 | 0.9 | 0.7 | 0.7 | 0.7 | 0.5 | **0.1** | **0.1** | 0.4 | **0.1** | **0.3** | **0** | **0.1** | **0.1** | 0.58 |
| W | 1.2 | 0.5 | 0.9 | 0.7 | 0.7 | 0.5 | 0.7 | 0.6 | 0.6 | 1.2 | 0.5 | 1.0 | **0.3** | **0.3** | **0.2** | **0.2** | **0.1** | **0.2** | 0.4 | **0.3** | 0.56 |
| E | 0.8 | 1.0 | 0.9 | 1.0 | 1.0 | 0.5 | 0.7 | 0.8 | 0.7 | 0.6 | 1.0 | **0.3** | **0.2** | **0.1** | 0.6 | **0.3** | **0** | **0.3** | **0.2** | **0.2** | 0.55 |
| Q | 0.8 | 0.9 | 0.9 | 0.9 | 0.8 | 0.5 | 0.7 | 0.8 | 0.7 | 0.6 | 1.0 | 0.4 | **0.3** | **0** | 0.4 | 0.4 | **0** | 0.5 | **0.1** | **0.2** | 0.54 |
| G | 1.1 | 0.7 | 0.8 | 0.5 | 0.8 | 1.5 | 0.8 | 0.8 | **0.2** | **0.2** | 0.6 | 0.5 | **0.2** | **0.2** | 0.5 | 0.4 | **0.3** | 0.4 | **0.3** | **0.1** | 0.54 |
| P | 0.8 | 0.8 | 0.6 | 0.6 | 0.6 | 0.6 | 0.9 | 0.6 | **0.1** | **0.2** | 0.5 | **0.3** | **0.1** | **0.3** | **0.1** | **0.2** | **0** | **0.1** | **0.2** | **0.1** | 0.38 |
| C | **0.2** | **0.3** | **0.3** | **0.2** | 0.6 | 0.6 | **0.2** | 0.6 | **0.2** | **0.2** | **0.2** | **0.1** | **0.3** | **0.1** | **0.1** | **0.1** | **0.1** | **0** | 0.4 | **0.1** | 0.24 |
| Avg | 1.1 | 1.0 | 1.0 | 1.0 | 0.9 | 0.9 | 0.9 | 0.8 | 0.8 | 0.7 | 0.7 | 0.6 | 0.3 | 0.3 | 0.3 | 0.3 | 0.2 | 0.2 | 0.2 | 0.2 | |

The extent of ligation with bio-KGAAPF-glc-Phe-amide for 400 immobilized tripeptides having the sequence $H_2N$-P1′-P2′-Ala-plate as determined by a colorimetric assay with SAHRP as described in text. Each value indicates the change in $A_{405}$ above a nonligated control, which had the sequence Ala-Phe-Ala. In the absence of ligase, this sequence gave an $A_{405}$ value of 0.2. Residues are ordered according to the average extents of reaction for peptides having a fixed P1′ (shown at the bottom) or a fixed P2′ residue (shown on the right). Ligation reactions that had $A_{405}$ values ≥ 1.0 or ≤ 0.3 are shaded to highlight the best and worst dipeptide combinations, respectively.

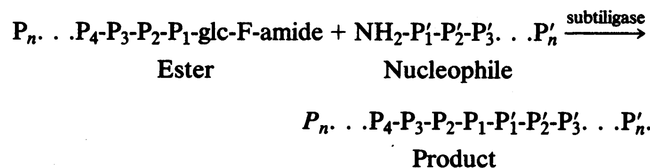
one-fifth volume of 2.5 M NaCl and 20% 8-kDa polyethylene glycol (PEG). The precipitated phage were resuspended in buffer A, bound to avidin-agarose for 15 min, centrifuged in a Microfuge, and washed four times with buffer A. The bound phage were eluted with 0.2 M glycine buffer (pH 2), and the phage titers were measured as colony-forming units with *E. coli* XL1-Blue.

**Mutagenesis and Expression of Variant Subtilisins.** Subtiligase and its variants were expressed and purified as described (15) except that no helper subtilisin was used. The plasmid for stabiligase was produced by using four different primers that incorporated five mutations (three different primers to introduce the single changes Met-50 → Phe, Asn-76 → Asp, or Asn-109 → Ser and another primer introducing the double mutation Lys-213 → Arg/Asn-218 → Ser). Once the sequences of the mutants were confirmed, component mutations were combined by ligation of DNA fragments containing them, and DNA propagated in *E. coli* JM101 was used to transform *Bacillus subtilis* strain BG2036 as before (15).

**Peptide Ligations.** Ligation of succinylated KKKGAAPF-glc-F-amide onto ANP was performed by using 160 μM human ANP (Bachem) in the presence of 0.7 μM subtiligase and 2 mM peptide in 100 mM Tricine buffer (pH 8) for 1 hr at room temperature. The product was separated from unreacted ANP and hydrolyzed peptide on a Vydac $C_{18}$ HPLC column with 0–50% acetonitrile gradient in aqueous 0.1% $CF_3COOH$. The ligated ANP was resuspended in 2 mM borate buffer (pH 8.5) containing $PEG_{5000}$-conjugated *N*-hydroxysuccinimide (NHS) at a 5-fold molar excess over ANP. After 1 hr at room temperature, the reaction was quenched by the addition of excess 1 M ethanolamine (pH 8). The "PEGylated" adduct was separated from non-PEGylated ANP on a Superose 12 column. The product's identity and mass were determined by matrix-assisted laser desorption MS. ANP activity was measured in whole-cell assay (22).

## RESULTS AND DISCUSSION

**Substrate Specificity of Subtiligase.** To effectively apply subtiligase for peptide ligation requires a detailed understanding of its specificity determinants. Subtilisin binds substrates in an extended conformation from four residues N-terminal of the cleaved bond (P1 to P4) to three residues C-terminal to the cleaved bond (P1′ to P3′) (23).¶ Therefore, we would expect subtiligase to interact with the last four residues of the peptide ester substrate and the first three residues of the amine nucleophile:




The requirements for binding the glc-F-amide ester have been extensively studied (15), but those for binding the amine nucleophile have not been. To establish the nucleophile specificity, 400 dipeptides were synthesized on a polystyrene plate to have the form $NH_2$-P1′-P2′-Ala-, where P1′ and P2′ are any of the 20 naturally occurring L-amino acids. The donor peptide ester, bio-KGAAPF-glc-F-amide, which was known to be a good substrate, was added along with subtiligase. After a fixed time, the extent of ligation was measured from the binding of a SAHRP conjugate to the bio-KGAAPF-peptide product (Table 1).

Some P1′ residues (methionine, phenylalanine, lysine, leucine, arginine, alanine, serine, and histidine) gave consistently higher $A_{405}$ values (≥1.0) when paired with virtually any P2′ residue, whereas other P1′ residues (glutamic acid, aspartic acid, proline, asparagine, glutamine, threonine, valine, and isoleucine) gave consistently lower values ($A_{405}$ ≤ 0.3). Thus, it appears that the P1′ residue exerts a greater effect on ligation efficiency than does the P2′ residue. Although most P2′ residues are tolerated, some are less preferred (cysteine, proline, glycine, glutamine, glutamic acid,

¶By analogy to protease substrates (24), ligation substrates are designated $NH_2$-Pn . . . P2-P1 + P1′-P2′ . . . Pn′-COOH, where the ligated bond is between the P1 and P1′ residues.

and tryptophan), and three in particular are consistently poor, cysteine, proline, and glycine. A good P1′ residue can overcome problems with a poor P2′ residue, further suggesting that the requirements at P1′ were more important for ligation. For example, a P2′ glycine is generally poor, but some ligation occurred for Met-Gly and Ala-Gly nucleophiles.

Next, we investigated the ability of subtiligase to attach a biotin affinity peptide onto a folded protein. For this we chose Met-hGH, which has a favorable N-terminal sequence, Met-Phe-. Indeed, the biotinylated peptide was ligated effectively onto hGH at ≈95% yield (Table 2). Ligation resulted in an expected increase in molecular weight by SDS/PAGE (Fig. 1 *Upper*) or MS ($M_r$ expected, 23,059.3; found, 23,060.0). Moreover, hGH variants containing a poor N-terminal sequence do not ligate (see below). These results show that ligation was specific for the N-terminal $\alpha$-amine and did not occur at any of the nine $\epsilon$-amino groups of lysine in hGH. The biotin could be readily detected by staining with SAHRP (Fig. 1 *Lower*), and the conjugate had the same $K_d$ for the hGH receptor (1 ± 0.8 nM as determined by the BIAcore analysis; ref. 25) and identical bioactivity in a cell-based assay (26). Thus, the labeling procedure is efficient and specific, and it preserved the biochemical and biological activities we have measured for the hormone.

**Substrate Phage Selection for Identifying Optimal Ligation Sequences.** We evaluated further the optimal N-terminal sequences for ligation onto a folded protein using a modification of substrate phage display (27). A library of phagemid particles was produced in which the first three residues of hGH were randomly mutated, and variant proteins were displayed in a monovalent fashion. We incubated the phagemid particles with iminobiotin-KGAAPF-glc-F-amide and subtiligase, and purified those particles containing the iminobiotin-peptide by binding them to avidin-agarose. A streptavidin-binding motif, His-Pro-Gln, has been isolated from phage display of random peptides (28). To avoid selection on the basis of this motif, the phagemid particles were prebound to avidin-agarose beads to remove those particles that may bind the matrix independent of biotin. Control experiments showed that Met-hGH-carrying phage, which has a good ligation sequence, Met-Phe-Pro, was enriched 20- to 30-fold more than hGH-carrying phage, which has a poor ligation sequence, Phe-Pro-Thr. Thus, the other phage coat proteins were not good substrates for subtiligase. Moreover,

Table 2. Yields for ligating activated ester peptides onto proteins with variable N-terminal sequences

| Protein | N-terminal sequence | % ligation | Peptide |
|---|---|---|---|
| Met-hGH | MFPTIP | 95 | 1 |
| hGH | FPTIP | <2 | 1 |
| P2A hGH | FATIP | 90 | 1 |
| Δ2 hGH | TIP | <2 | 1 |
| Δ8 hGH | LFDN | 5 | 1 |
| | LFDN | 15* | 1 |
| | LFDN | 45† | 1 |
| Met-hGH | MFPTIP | 80 | 2 |
| ANP | SLRRSS | 70 | 3 |

Peptides used were bio-KGAAPF-glc-F-amide (1); suc-AAHYZ-glc-F-amide (2), where Z was a mercurated cysteine; and suc-KKKGAAPF-glc-F-amide (3). The ligation with Δ8 hGH (hGH with eight residues deleted) was achieved by using subtiligase (*) or stabiligase (†) in the presence of 0.1% SDS and at 55°C. All ligations were for 1 hr with 1 mM peptide ester, 0.1 mM protein, and 0.7 μM enzyme in 100 mM Tricine (pH 8.0). The extent of ligation was determined by SDS/PAGE and laser densitometry as described in Fig. 1. The detection limit of the assay was 2% ligation product. Δ2 hGH, hGH with two residues deleted.

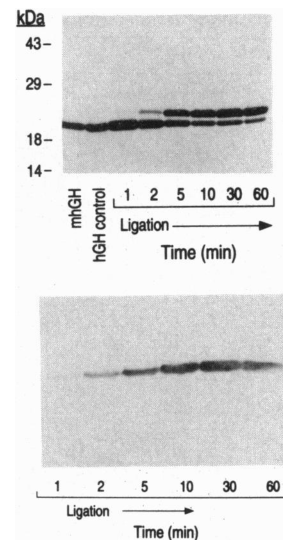



FIG. 1. Time-course of ligation of bio-KGAAPF-glc-F-amide onto Met-hGH (mhGH). (*Upper*) Coomassie-stained gel. (*Lower*) Protein bands transferred onto a nitrocellulose membrane and then stained with SAHRP, with 4-chloro-1-naphthol as the substrate. Ligation of peptides onto Met-hGH was performed with 1 mM peptide ester, 100 μM Met-hGH, and 0.7 μM subtiligase in 100 mM Tricine (pH 8). Aliquots from the reaction mixture were removed at fixed time points, and the enzyme was inactivated by the addition of equal volume of 100 mM *N*-ethyl maleimide. It was essential to avoid Tris because its primary amine group competes with that of the protein in the nucleophilic displacement of the enzyme from the acyl–enzyme intermediate.

the enrichment required treatment with subtiligase (data not shown).

We sequenced 20 clones prior to selection and found the first three codons of hGH to be essentially random, indicating no obvious bias of codon choices in the starting library. After three or six rounds of selection, a total of 74 clones was sequenced (Table 3). Although a much larger number of sequences would be necessary to acquire a statistical sampling of ligation products, some clear trends were evident (Table 3). At the N terminus (the P1′ residue), some residues were grossly underrepresented in the selected pool such as threonine, proline, asparagine, isoleucine, glutamic acid, and aspartic acid; others were overrepresented such as methionine, tyrosine, alanine, arginine, and leucine. At the second position (the P2′ residue), aspartic acid and proline were grossly underrepresented, while phenylalanine, histidine, tryptophan, leucine, serine, asparagine, and tyrosine showed higher abundance than expected. At the third position (the P3′ residue), virtually all amino acids were found except aspartic acid and cysteine. Selection appeared to be dominated by the residue at P1′ followed by the P2′ and P3′.

Ligations are performed on a pool of substrates with the substrate phage, and thus substrates are selected competitively based on relative catalytic efficiencies (27). This identifies good N-terminal sequences for ligation plus ones that are secreted, processed, and stable to endogenous proteases. In particular, the selection identified five different Met-Leu-Xaa sequences as well as three different Met-Phe-(Pro or Asn) or Arg-Ser-Xaa sequences (Table 3). These sequences may be good choices to incorporate into a secreted protein or protein fragment if one only cared about having the ligation proceed efficiently.

The sequences recovered by the substrate phage selections can be biased by the requirement that the sequences be secreted, processed, and stable. Nonetheless, the ligation preferences seen for the P1′ residue from the substrate phage selections were similar to those found from the reactions with

Table 3. Sequences obtained from substrate phage after three and six rounds of selection with subtiligase and iminobiotin-KGAAPF-glc-F-amide

| Round 3 | | Round 6 | |
|---|---|---|---|
| DNA | P1′-P2′-P3′ | DNA | P1′-P2′-P3′ |
| GCG TGC AGC | A-C-S | GCG TGC CTG | A-C-L |
| GCG TTC GAC | A-F-D | GCC AGG GCG | A-R-A |
| GCG TTG AGC | A-L-S | GCC AGG CAG | A-R-Q |
| GCG GTC CCG | A-V-P | GCG ACC GTG | A-T-V |
| TGC ACG CCC | C-T-P | TTC TAC CGC | F-Y-R |
| TTC TTC TTG | F-F-L | GGG ATC GAG | G-I-E |
| TTC GTC AAG | F-V-K | GGG ATG TAC | G-M-Y |
| GGG GCC TCG | G-A-S | GGG GTG CTG | G-V-L |
| GGG GCC TCG | G-A-S | AAG GAA GCG | K-E-A |
| GGG AAA AAG | G-K-K | CTG TTC GCC | L-F-A |
| GGG TGG AGC | G-W-S | TTG TTC GTG | L-F-V |
| CAC GCC CAC | H-A-H | TTG TTG CCG | L-L-P |
| AAG TTG TAC | K-L-Y | TTG TCA ACG | L-S-T |
| CTC CTC ATC | L-L-I | CTG TCG GTC | L-S-V |
| TTG TCA ACG | L-S-T | CTG TGG CTC | L-W-L |
| CTG TGG CCC | L-W-P | TTG TAC CCG | L-Y-P |
| ATG TTC CCA | M-F-P | ATG TTC AAC | M-F-N |
| ATG TTC CCC | M-F-P | ATG GGG AAC | M-G-N |
| ATG CAC TGG | M-H-W | ATG CAC ATC | M-H-I |
| ATG TTG CGG | M-L-R | ATG TTG GCC | M-L-A |
| ATG TTG AGG | M-L-R | ATG TTG TAG | M-L-Am |
| ATG TTG GTC | M-L-V | ATG AAC AAC | M-N-N |
| ATG CAA AGG | M-Q-R | ATG AAC CCG | M-N-P |
| AGG GCC CTC | R-A-L | ATG TGG TAC | M-W-Y |
| CGC CCA TTG | R-L-L | CAG TTC TAC | Q-F-Y |
| AGG AGG TCG | R-R-S | AGG GGC ATG | R-G-M |
| AGG TCG GGG | R-S-G | CGC TTG AGG | R-L-R |
| CGC TCA CAC | R-S-H | CGC CAG GAG | R-Q-E |
| AGG TGG CCC | R-W-P | AGG TCG AGC | R-S-S |
| CGG TGG CCC | R-W-P | CGG GTG TTG | R-V-L |
| TCT GAG TGG | S-E-W | TCG CGG TCC | S-R-S |
| AGC CAC CTC | S-H-L | TCC TAC CGG | S-Y-R |
| TCG AGC TAC | S-S-Y | GTG AGC TTC | V-S-F |
| GTC TCC TTG | V-S-L | TGG AGC TCC | W-S-S |
| TGG CTG CTG | W-L-L | TAC AAC GGG | Y-N-G |
| TAC TTC AGC | Y-F-S | TAC ACC AGC | Y-T-S |
| TAC CAC TAC | Y-H-Y | | |
| TAC CGG GCC | Y-R-A | | |

Both the nucleotide and the corresponding amino acid sequences are shown. Am denotes an amber stop codon, which is suppressed as glutamine in XL1-Blue strain (17).

synthetic substrates (Table 4). In general, the most frequently selected P1′ residues were the strongest reacting synthetic peptides, and the least selected P1′ residues were the weakest reacting. The correspondence between the two methods was worse for the P2′ position, which probably reflects the fact that the P2′ position is less important, and a much larger sampling of sequences from substrate phage selectants would be necessary to distinguish the mild preferences here.

Several variants of hGH were expressed, and ligations were performed to directly link these specificity studies with ligation yields on a folded protein. Wild-type hGH gave no detectable ligation product (Table 2). This is explained by the presence of the P2′ proline residue; indeed, when this was changed to alanine, high ligation efficiency was restored. Deletion of the first two residues in hGH to give a Thr-Ile-Pro sequence also resisted ligation presumably because of the presence of the P1′ threonine.

**Improving Ligation Yields Through Protein Engineering of Subtiligase.** We were surprised that the Δ8 variant of hGH was such a poor substrate (Table 2) because this protein has a favorable N-terminal sequence, Leu-Phe-Asp. Helix-1 in hGH begins at position 6 (29). Thus, it is possible that Leu-9 is in an α-helical conformation, which will not fit the requirement of an extended substrate structure from P1′ to P3′ (23). We tested the ability of subtiligase to ligate bio-KGAAPF-glc-F-amide onto Met-hGH with increasing concentrations of guanidine hydrochloride. However, ligation efficiency decreased by 50% even in 1 M guanidine hydrochloride (Fig. 2).

Table 4. Summary of best and worst substrates carrying specific P1′ and P2′ residues ordered according to their extent of reaction with synthetic substrates or frequency of selection from the substrate phage library

| Substrate | Substrate residue: P1′ | Substrate residue: P2′ |
|---|---|---|
| | Best | |
| Synthetic | MFKLRASHWYGC | TKVHYRSAMIDFN |
| Phage | MYRALGFKSWQ | FHWLSYNVACEQ |
| Common | MFKLRASWY | VHYSAFN |
| | Worst | |
| Synthetic | EDPNQTVI | CPGQEWL |
| Phage | EDPNTIVHC | PDGIKMTR |
| Common | EDPNTIV | PG |

From the reactivities with the synthetic substrates, we arbitrarily chose the best or worst substrates carrying specific P1′ and P2′ residues to be those having average extents of reaction of above or below 0.6, respectively. From the substrate phage selection, we considered the substrates carrying specific P1′ and P2′ residues to be those whose observed and expected frequencies were above or below 0.8, respectively.

Many mutants of subtilisin have been produced that enhance the stability of the enzyme to heat, alkali, and organic solvents (30–35). Five of these mutations (Met-50 → Phe/Asn-76 → Asp/Asn-109 → Ser/Lys-213 → Arg/Asn-218 → Ser) were incorporated into subtiligase (which carries two mutations), creating a heptamutant of subtilisin BPN′, which we call "stabiligase." Indeed, stabiligase retained nearly 50% of its activity to ligate bio-KGAAPF-glc-F-amide onto Met-hGH in up to 4 M guanidine hydrochloride (Fig. 2). Furthermore, stabiligase was able to ligate bio-KGAAPF-glc-F-amide onto the Δ8 variant of hGH in a yield of 45% in 0.1% SDS, compared with a 15% yield for subtiligase under the same conditions (Table 2). Thus, the use of stabiligase makes it possible to ligate fragments onto proteins that would otherwise be very poor because of structural features of the substrate. Alternatively, in cases where the sequence of the ligation junction is not critical, one can engineer a good N-terminal ligation sequence (such as Met-Leu-) followed by

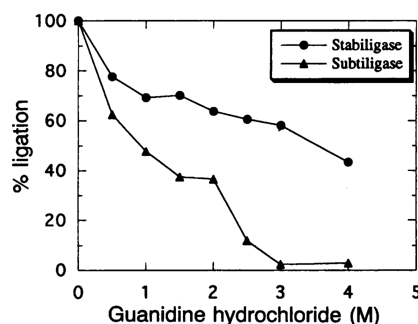



FIG. 2. Relative ligase activities of subtiligase and the more stable version of it, stabiligase, for ligating bio-KGAAPF-glc-F-amide onto Met-hGH in increasing concentrations of guanidine hydrochloride. The extent of ligation was measured by running the protein samples on a SDS/14% polyacrylamide electrophoresis gel, staining it with Coomassie blue, and analyzing the band intensity with an LKB 2202 laser densitometer. The ligation efficiency was normalized to the yield achieved in the absence of denaturant.

a di- or tri-glycine sequence to promote flexibility and accessibility to subtiligase.

**Additional Examples of Protein Semisynthesis and Incorporation of Nonnatural Amino Acids.** In addition to providing ways of introducing affinity handles such as biotin or isotopic labels, subtiligase may be useful for incorporation of heavy atom-laden peptides to create derivatives for x-ray crystallography. For example, a peptide was prepared with the sequence succinyl-Ala-Ala-His-Tyr-Zaa-glc-Phe-amide, where Zaa was a mercurated cysteine. This peptide was ligated onto Met-hGH in 80% yield (Table 2).

A common method of increasing the retention of proteins in serum as well as reducing their immunogenicity is to attach PEG polymers (for a recent review, see ref. 36). PEG modification is usually achieved by random chemical modification of the $\varepsilon$-amino group of lysine residues with a PEG-(NHS) ester. However, the specificity and stoichiometry of modification are difficult to control. ANP is a small hormone secreted from the heart that is important in regulating salt balance and blood pressure (37). ANP is rapidly cleared from serum, and one of the clearance routes is through kidney filtration. ANP does not contain any lysine residues, but we could ligate onto it a peptide having the sequence suc-KKKGAAPF- with a yield of 70% (Table 2) and then PEGylate it with $PEG_{5000}$-NHS as described. The bulky nature of PEG groups limited the accessibility of the $\varepsilon$-amine groups so that the products consisted of only mono- and di-pegylated ANP. The PEGylated ANP was 50% as active as unmodified ANP in a cell-based assay (22).

## CONCLUSIONS

These studies have been designed to explore the applications and limitations of subtiligase for protein modification and semisynthesis. In other examples (4–14), successful ligations were performed mostly on preassociated fragments. The advantage of the present approach is that it does not require the ligation fragments to have high affinity for each other prior to coupling. Proteins or protein fragments only need to possess a favorable sequence and accessible structure at the N terminus for efficient and site-specific modification with subtiligase. In general, sequence limitations can be overcome by using ligases with altered specificity (15) or by changing the sequence of the ligation junction. Structural limitations may be overcome by catalyzing the reaction with stabiligase in denaturants.

In addition to site-specific incorporation of affinity handles, immunoprobes, isotopic labels, heavy-atom derivatives, and PEG moieties, other nonnatural constituents may be incorporated into peptides by using subtiligase. In fact, we have recently shown that subtiligase can be practically applied to the synthesis of a 124-residue enzyme, ribonuclease A, to incorporate unnatural catalytic residues (38) and for cyclization of peptides (39). New technologies are often fueled by the discovery of new enzyme activities. These rationally tailored peptide ligases should expand the possibilities for protein engineering and molecular design.

We thank Jeff Tom, Martin Struble, and Kathryn Chan for peptide synthesis and purification, the protein chemistry department for amino acid sequencing and MS analysis, the oligonucleotide group for DNA synthesis, Germaine Fuh for providing PEG-NHS, Henry Lowman and David Matthews for advice on phage sorting. David Lowe for ANP cell-stimulation assay, Herman Chu for hGH whole-cell assay, and Marcus Ballinger, Paul Carter, Henry Lowman, and Kevin Judice for critical reading of the manuscript.

1. Offord, R. E. (1990) in *Protein Design and the Development of New Therapeutics and Vaccines*, eds. Hook, J. B. & Poste, G. (Plenum, New York), pp. 253–282.
2. Wallace, C. J. A. (1991) in *Peptides 1990*, eds. Giralt, E. & Andreu, D. (ESCOM Science, Leiden, Netherlands), pp. 260–263.
3. Chaiken, I. M. (1981) *CRC Crit. Rev. Biochem.* **11**, 255–301.
4. Sealock, R. W. & Laskowski, M., Jr. (1973) *Biochemistry* **12**, 3130–3146.
5. Leary, T. R. & Laskowski, M., Jr. (1973) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **32**, 465.
6. Wuttke, D. S., Gray, H. B., Fisher, S. L. & Imperiali, B. (1993) *J. Am. Chem. Soc.* **115**, 8455–8456.
7. Wallace, C. J. A. (1993) *FASEB J.* **7**, 505–515.
8. Homandberg, G. A. & Laskowski, M., Jr. (1979) *Biochemistry* **18**, 586–592.
9. Gráf, L. & Li, C. H. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 6135–6139.
10. Morihara, K. (1990) *J. Mol. Recognit.* **3**, 181–186.
11. Simmerman, H. K. B., Wang, C.-C., Horwitz, E. M., Berzofsky, J. A. & Gurd, F. R. N. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 7739–7743.
12. Roy, R. P., Khandke, K. M., Manjula, B. N. & Acharya, A. S. (1992) *Biochemistry* **31**, 7249–7255.
13. Bongers, J., Lambros, T., Liu, W., Ahmad, M., Campbell, R. M., Felix, A. M. & Heimer, E. P. (1992) *J. Med. Chem.* **35**, 3934–3941.
14. Gaetner, H. F., Offord, R. E., Cotton, R., Timms, D., Camble, R. & Rose, K. (1994) *J. Biol. Chem.* **269**, 7224–7230.
15. Abrahmsén, L., Tom, J., Burnier, J., Butcher, K. A., Kossiakoff, A. & Wells, J. A. (1991) *Biochemistry* **30**, 4151–4159.
16. Barany, G. & Merrifield, R. B. (1980) in *The Peptides*, eds. Gross, E. & Meinenhoffer, J. (Academic, New York), Vol. 2, pp. 1–284.
17. Lowman, H. B., Bass, S. H., Simpson, N. & Wells, J. A. (1991) *Biochemistry* **30**, 10832–10838.
18. Kunkel, T. A., Bebenek, K. & McClary, J. (1991) *Methods Enzymol.* **204**, 125–139.
19. Bullock, W. O., Fernandez, J. M. & Short, J. M. (1987) *BioTechniques* **5**, 376–379.
20. Vieira, J. & Messing, J. (1987) *Methods Enzymol.* **153**, 3–11.
21. Heney, G. & Orr, G. A. (1981) *Anal. Biochem.* **114**, 92–96.
22. Chang, M.-S., Lowe, D. G., Lewis, M., Hellmiss, R., Chen, E. & Goeddel, D. V. (1989) *Nature (London)* **341**, 68–72.
23. McPhalen, C. A. & James, M. N. G. (1988) *Biochemistry* **27**, 6582–6598.
24. Schechter, I. & Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27**, 157–162.
25. Cunningham, B. C. & Wells, J. A. (1993) *J. Mol. Biol.* **234**, 554–563.
26. Fuh, G., Cunningham, B. C., Fukunaga, R., Nagata, S., Goeddel, D. V. & Wells, J. A. (1992) *Science* **256**, 1677–1680.
27. Matthews, D. J. & Wells, J. A. (1993) *Science* **260**, 1113–1117.
28. Devlin, J. J., Panganiban, L. C. & Devlin, P. E. (1990) *Science* **249**, 404–406.
29. de Vos, A. M., Ultsch, M. & Kossiakoff, A. A. (1992) *Science* **255**, 306–312.
30. Cunningham, B. C. & Wells, J. A. (1987) *Protein Eng.* **1**, 319–325.
31. Pantoliano, M. W., Whitlow, M., Wood, J. F., Dodd, S. W., Hardman, K. D., Rollence, M. L. & Bryan, P. N. (1989) *Biochemistry* **28**, 7205–7213.
32. Wong, C.-H., Chen, S.-T., Hennen, W. J., Bibbs, J. A., Wang, Y.-F., Liu, J. L.-C., Pantoliano, M. W., Whitlow, M. & Bryan, P. N. (1990) *J. Am. Chem. Soc.* **112**, 945–953.
33. Narhi, L. O., Stabinsky, Y., Levitt, M., Miller, L., Sachdev, R., Finley, S., Park, S., Kolvenbach, C., Arakawa, T. & Zukowski, M. (1991) *Biotechnol. Appl. Biochem.* **13**, 12–24.
34. Bryan, P. N. (1992) in *Stability of Protein Pharmaceuticals, Part B: In Vivo Pathways of Degradation and Strategies for Protein Stabilization*, eds. Ahern, T. J. & Manning, M. C. (Plenum, New York), pp. 147–181.
35. Chen, K. & Arnold, F. H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 5618–5622.
36. Francis, G. E., Delgado, C. & Fisher, D. (1992) in *Stability of Protein Pharmaceuticals, Part B: In Vivo Pathways of Degradation and Strategies for Protein Stabilization*, eds. Ahern, T. J. & Manning, M. C. (Plenum, New York), pp. 235–263.
37. Chinkers, M. & Garbers, D. L. (1991) *Annu. Rev. Biochem.* **60**, 553–575.
38. Jackson, D. Y., Burnier, J., Quan, C., Stanley, M., Tom, J. & Wells, J. A. (1994) *Science* **266**, 243–247.
39. Jackson, D. Y., Burnier, J. P. & Wells, J. A. (1994) *J. Am. Chem. Soc., in press.*