Determination of Substrate Specificity for Peptide Deformylase through the Screening of a Combinatorial Peptide Library†

Yun-Jin Hu,‡ Yaoming Wei,§ Ying Zhou,‡ P. T. Ravi Rajagopalan,‡ and Dehua Pei*,‡,§

Department of Chemistry and Ohio State Biochemistry Program, The Ohio State University, 100 West 18th Avenue, Columbus, Ohio 43210

Received August 24, 1998; Revised Manuscript Received November 3, 1998

ABSTRACT: Peptide deformylase is an essential Fe^{2+} metalloenzyme that catalyzes the removal of the N-terminal formyl group from nascent polypeptides in eubacteria. In vivo, the deformylase is capable of deformylating most of the polypeptides in a bacterial cell, which contain diverse N-terminal sequences. In this work, we have developed a combinatorial method to systematically examine the sequence specificity of peptide deformylase. A peptide library that contains all possible N-terminally formylated tetrapeptides was constructed on TentaGel resin, with a unique peptide sequence on each resin bead. Limited treatment with the *Escherichia coli* deformylase resulted in the deformylation of those peptides that are the most potent substrates of the enzyme. By using an enzyme-linked assay, the beads containing the deformylated peptides were identified and isolated. Peptide sequence analysis using matrix-assisted laser desorption ionization mass spectrometry revealed a consensus sequence, formyl-Met-X-Z-Tyr (X = any amino acid except for aspartate and glutamate; Z = lysine, arginine, tyrosine, or phenylalanine), for the *E. coli* enzyme. The deformylase is also capable of efficient deformylation of formyl-Phe-Tyr-(Phe/Tyr) peptides. These results demonstrate that, despite being a broad-specificity enzyme, the peptide deformylase deformylates different peptides at drastically different rates. In addition, the selectivity of peptide deformylase for the *N*-formyl over the *N*-acetyl group has been studied with *N*-α-fluoroacetyl peptides, and the results suggest that both electronic and steric factors are responsible for the observed specificity. The deformylase was also shown to exhibit esterase activity. These results will facilitate the design of specific deformylase inhibitors as potential antibacterial agents. This combinatorial method should be generally applicable to the study of the substrate specificity of other acylases and peptidases.

A conserved, distinctive feature of prokaryotic (also chloroplast and mitochondrial) translation systems is that the methionyl-initiator tRNA, $tRNA_i^{Met}$, is Nα-formylated (*1*). Consequently, all nascent bacterial polypeptides bear an N-terminal formyl group (*2, 3*). Following the translational initiation, the *N*-formyl group is removed cotranslationally from most but not all of the polypeptides by peptide deformylase (*4*−*7*). This deformylase activity is essential for bacterial survival, as deletion of its gene (*def*) is lethal (*8*), suggesting that certain (at least one) essential bacterial proteins must be deformylated to function properly. At the meantime, there have also been observations that some proteins remain formylated in their functional states. Examples include several *Escherichia coli* ribosomal proteins (*9*), several *E. coli* chemotactic peptides (*10*), and *Salmonella typhimurium* aspartate chemoreceptor (*11*). There is also another group of proteins which normally undergo complete deformylation but become partially formylated when overproduced in *E. coli* cells (*12*). Increased production of peptide deformylase has been employed to achieve complete deformylation of these proteins (*12*). In one case, replacement of the penultimate residue (the residue following the initial Met) of bovine somatotropin from Phe to an Ala resulted in a shift from partial formyl retention to complete deformylation (*13*). An intriguing question is whether the state of formylation is a result of the substrate specificity of peptide deformylase. Also important is the question of how a single deformylase is capable of deformylating thousands of polypeptides which bear a diverse set of sequences at their N-termini. We felt that determining the substrate specificity of the deformylase would be a necessary first step to answer these questions. In addition, due to its essential character and unique existence in bacterial organisms, peptide deformylase is currently being pursued as a target for designing novel antibacterial drugs (*14*). Information on deformylase specificity would greatly facilitate the rational design of deformylase inhibitors.

To gain insight into deformylase specificity, we and others have previously carried out the kinetic characterization of peptide deformylase using N-formylated peptide substrates (*4, 15, 16*). While these studies have demonstrated some level of substrate specificity, they are practically limited by the number of peptides one can synthesize and test in a reasonable time frame as well as the availability of financial resources. In this work, we have developed a combinatorial approach for rapidly identifying the optimal substrates of

† This work was supported in part by grants from American Cancer Society (IRG-16-35) and the National Institutes of Health (AI40575).

* Corresponding author: Department of Chemistry, The Ohio State University, 100 W. 18th Ave., Columbus, OH 43210. Telephone: (614) 688-4068. Fax: (614) 292-1532. E-mail: pei.3@osu.edu.

‡ Department of Chemistry.

§ Ohio State Biochemistry Program.

10.1021/bi9820412 CCC: $18.00
Published on Web 12/16/1998

peptide deformylase from a resin-bound library of N-formylated peptides. A substrate consensus sequence has been obtained for the *E. coli* deformylase. In addition, the origin of the enzyme's exquisite specificity for *N*-formyl vs other *N*-acyl groups has been studied using a set of α-fluoroacetylated peptides.

EXPERIMENTAL PROCEDURES

Materials. TentaGel S NH_2 resin and Wang resin were purchased from Advanced ChemTech (Louisville, KY). *N*-9-Fluorenylmethoxycarbonyl (Fmoc)-amino acids, 1-hydroxybenzotriazole (HOBt),[1] and 2-(1*H*-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) were from SynPep (Dublin, CA). *N*-Hydroxysuccinimidobiotin, acetylglycine, *N*-acetyl-D,L-alanine, X-gal, and streptoavidin−β-galactosidase conjugate were obtained from Sigma Chemical Co. (St. Louis, MO). Peptide deformylase was purified as described (*16*−*18*). All other chemicals were purchased from Aldrich Chemical (Milwaukee, WI).

Peptide Library Synthesis. TentaGel S NH_2 resin (80−100 μm, 0.26 mmol/g loading, 2.86×10^6 beads/g) was used as the solid support for the peptide libraries. Synthesis was typically carried out on a 1.0-g scale using standard Fmoc/HBTU/HOBt chemistry (*19*) on a homemade peptide synthesis apparatus. For the peptide libraries, a common four amino acid linker, BBRM (B = β-alanine), was first synthesized on the resin (*20*). The randomized positions were generated using a split-pool synthesis method to construct the one-bead−one-sequence peptide libraries (*21*, *22*). Specifically, after deblocking the *N*-Fmoc group with 20% (v/v) piperidine in DMF twice (5 min + 15 min) and washing 5 times with DMF, the resin was evenly divided into the desired number of aliquots and placed into separate reaction vessels. A different amino acid was coupled to the resin in each of the reaction vessels using a 5-fold excess of amino acid (2 h). The coupling reaction was repeated once to ensure complete coupling at each randomized position. After coupling is complete, the resin from all vessels was combined, thoroughly mixed, washed with 5 mL of DMF, deprotected twice with 20% piperidine (5 min and then 15 min), and redistributed into the reaction vessels. This process was repeated to produce the tetrapeptide libraries. To produce a small amount of chain termination for later sequencing analysis, 20% (mol/mol) acetylglycine was added to the amino acid solutions used for coupling at the randomized positions with the exception of norleucine, for which 10% *N*-acetyl-D,L-alanine was used instead. After coupling at all four randomized positions and deblocking with piperidine, a formyl group was added to the N-termini of the peptides using 96% formic acid, dicyclohexylcarbodiimide, and (dimethylamino)pyridine in CH_2Cl_2 (2×2 h). Deprotection of side chains was effected with a cocktail containing 4.75 mL of trifluoroacetic acid, 0.1 mL of anisole, 0.1 mL of ethanedithiol, and 0.25 mL of thioanisole for 1 h at room temperature. The resin was washed with CH_2Cl_2 (10×4 mL) and stored at 4 °C.

Soluble Peptides. Soluble peptides were synthesized on Wang resin using standard Fmoc chemistry on a 0.1-mmol scale with double coupling at each position. Deprotection and cleavage were effected using the same cocktail as described previously. The peptide solution was drained into a clean glass vial, and the solvent was evaporated with a flow of N_2. The residue was triturated with anhydrous diethyl ether (3×15 mL) and stored at 4 °C. HPLC analysis (C_{18} column and monitored at 214 nm) shows that most of the peptides each gave a single peak (estimated to be >80% purity) and were used in kinetic analysis without further purification. A few peptides had lower purity as judged by analytical HPLC; these peptides were purified by reversed-phase HPLC before use. The identities of all peptides were confirmed by matrix-assisted laser desorption ionization mass spectrometry (MALDI MS) analysis.

On-Bead Screening of Peptide Libraries. Library screening was carried out at room temperature in a 1-mL plastic chromatography column fitted with a filtering disk at the bottom (Bio-Rad Laboratories). Seventy milligrams of resin (~200 000 beads) was placed into the column and washed 5 times each with 1 mL of CH_2Cl_2, DMF, and 1× deformylase assay buffer (50 mM sodium phosphate, pH 7.0, 10 mM NaCl, and 1 mM EGTA). Deformylation was initiated by the addition of 5 μL of peptide deformylase (10 μg) to the resin in 1.0 mL of 1× deformylase buffer. After gentle shaking for 10 min, the solution was drained with vacuum suction, and the resin was washed 5 times each with 1 mL of doubly distilled H_2O and DMF. The resin was then incubated in 1 mL of 1.0 mg/mL *N*-hydroxysuccinimidobiotin in DMF for 1 h. The excess biotin reagent was removed by washing the resin with DMF (6×1 mL) and TBS buffer (25 mM Tris·HCl, pH 8.0, 150 mM NaCl, and 0.1% Tween 20) (2×1 mL). Any nonspecific protein binding sites were blocked by incubating the beads in 1 mL of TBS buffer plus bovine serum albumin (1 mg/mL) for 1 h. The buffer was then removed and replaced with 1 mL of TBS buffer containing 2 μg/mL streptavidin−β-galactosidase conjugate. After 1 h, the solution was removed, and the beads were washed 4 times with TBS buffer. The resulting beads were stained by the addition of 1 mL of TBS buffer containing 1.0 mg/mL X-gal and subsequent incubation (with shaking) for 2−4 h. Staining was terminated by removing the X-gal solution with vacuum and washing the resin 4 times with TBS buffer. Positive beads were readily identified by their intense green color and were manually removed with a micropipet under a low-power microscope. A control screening was carried out under the same conditions except for the exclusion of deformylase enzyme in the deformylation reaction. This screening resulted in no colored beads.

Peptide Sequencing by MALDI Mass Spectrometry. Individual beads were placed in microcentrifuge tubes and each treated with 20 μL of 100 mg/mL CNBr in 70% formic acid for 16−24 h in the dark. The released peptide in each tube was lyophilized to remove the excess CNBr and dissolved in 5 μL of 0.1% TFA in H_2O. For mass analysis, a 1-μL aliquot of the TFA solution was mixed with 1 μL of a saturated solution of α-cyano-4-hydroxycinnamic acid, and

[1] Abbreviations: Ac-Ala, *N*-acetyl-D,L-alanine; HBTU, 2-(1*H*-benzotriazol-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate; HOBt, 1-hydroxybenzotriazole; X-gal, 5-bromo-4-chloro-3-indolyl β-D-galactopyranoside; B, β-alanine; MALDI MS, matrix-assisted laser desorption ionization mass spectrometry; EGTA, ethylene glycol bis(β-aminoethyl ether)-*N*,*N*,*N*′*N*′-tetraacetic acid; TFA, trifluoroacetic acid; Ac-ML-*p*NA, *N*-acetylmethionylleucyl-*p*-nitroanilide; FAc-ML-*p*NA, *N*-(α-fluoroacetyl)methionylleucyl-*p*-nitroanilide; F_2Ac-ML-*p*NA, *N*-(α-difluoroacetyl)methionylleucyl-*p*-nitroanilide; F_3Ac-ML-*p*NA, trifluoroacetylmethionylleucyl-*p*-nitroanilide; Nle, norleucine.

1 μL of the resulting mixture was applied to the spectrometer plate. MALDI mass analysis was performed on a Kratos Kompact MALDI-III mass spectrometer in the positive ion mode.

Synthesis of N-(α-Fluoroacetyl)-Met-Leu-p-nitroanilide (FAc-ML-pNA). The HCl salt of Met-Leu-*p*NA (*23*) (63 mg, 0.15 mmol) and sodium monofluoroacetate (15 mg, 0.15 mmol) were suspended in 10 mL of CH_2Cl_2, and then dicyclohexylcarbodiimide (30 mg, 0.15 mmol) was added. The mixture was stirred for 3 h at room temperature. Dicyclohexylurea was then removed by filtration, and the filtrate was washed with 5% sodium carbonate (3 × 10 mL), 5% HCl (3 × 10 mL), water (10 mL), and saturated NaCl solution (10 mL). Evaporation of the solvent afforded 65 mg of a white solid. ^{1}H NMR (200 MHz, DMSO): δ 10.64 (s, 1H), 7.87−8.45 (m, 6H), 4.83 (d, J = 47.0 Hz, 2H), 4.35−4.55 (m, 2H), 2.45 (t, J = 7.9 Hz, 2H), 2.04 (s, 3H), 1.42−2.00 (m, 5H), 0.86−0.94 (m, 6H).

Synthesis of N-(α-Difluoroacetyl)-Met-Leu-p-nitroanilide (F_2Ac-ML-pNA). This compound was prepared in a similar manner as described above. ^{1}H NMR (200 MHz, DMSO): δ 10.68 (s, 1H), 8.97 (d, J = 7.3 Hz, 1H), 8.42 (d, J = 7.3 Hz, 1H), 8.22 (d, J = 9.3 Hz, 2H), 7.85 (d, J = 9.3 Hz, 2H), 6.24 (t, J = 53.8 Hz, 1H), 4.36−4.50 (m, 2H), 2.44 (t, J = 7.9 Hz, 2H), 2.03 (s, 3H), 1.80−2.00 (m, 2H), 1.42−1.80 (m, 3H), 0.86−0.94 (m, 6H). FABMS: m/z 461 (M + H^+).

Synthesis of 2-Formyloxycaproylleucyl-p-nitroanilide. L-Leucyl-*p*-nitroanilide (522 mg, 2 mmol), (*S*)-2-hydroxycaproic acid (264 mg, 2 mmol), and triethylamine (182 mg) were dissolved in 20 mL of CH_3CN, and HBTU (800 mg, 2 mmol) was then added. The mixture was stirred at room temperature for 30 min, and 50 mL of ethyl acetate was added. The organic solution was washed with 50 mL each of 5% HCl, 5% sodium bicarbonate, and water, and dried over $MgSO_4$. Evaporation of the solvent gave a white solid, which was used directly in the following reaction. To a solution of the white solid (200 mg, 0.53 mmol) and formic acid (49 mg, 1.1 mmol) in 5 mL of CH_2Cl_2 was added dicyclohexylcarbodiimide (330 mg, 1.6 mmol). The mixture was stirred for 4 h at room temperature and filtered to remove the precipitate formed. After evaporation of the solvent, the residue was purified by silica gel chromatography to afford 213 mg of a white solid (quantitative yield). ^{1}H NMR (250 MHz, $CDCl_3$): δ 9.28 (s, 1H), 8.17 (s, 1H), 8.09 (d, J = 9.1 Hz, 2H), 7.61 (d, J = 9.1 Hz, 2H), 6.70 (d, J = 7.9 Hz, 1H), 5.31 (t, J = 6.0 Hz, 1H), 4.73 (m, 1H), 1.60−1.92 (m, 5 H), 1.25−1.35 (m, 4H), 0.83−1.07 (m, 9H). FABMS: m/z 394 (M + H^+).

Deformylase Assays. Three assays were employed in this work. Method A, which monitors the release of formate by a formate dehydrogenase as previously described (*16*), was used for most of the N-formylated substrates. In method B, the deformylase reaction was coupled to an aminopeptidase, which further processes the deacylated peptides to release a chromogenic product (*23*). This method was used for all of the N-acetylated peptides. For those peptides that are either poor substrates of the deformylase or have limited solubility, method C was used. Briefly, the deformylase reaction was terminated by the addition of 10 mM H_2O_2 or an equal volume of 100 mM TFA, and the reaction mixture was analyzed by reversed-phase HPLC (monitored at 214 or 315 nm). Integration of the product peak and the remaining substrate peak gave the percentage conversion, from which the initial rates were calculated.

RESULTS

Design, Construction, and Screening of Resin-Bound Peptide Library. A tetrapeptide library was generated on TentaGel S resin, using the split-synthesis method (*21*, *22*). This resulted in the synthesis of approximately 100 pmol of a single peptide on each resin bead; a large collection of these beads thus produced a "one-bead−one-sequence" library. This method ensures an equal representation of all amino acids used at each of the randomized positions. In this work, each randomized position included 15 natural amino acids, Ala, Asn, Asp, Gln, Glu, Gly, His, Leu, Phe, Pro, Ser, Thr, Trp, Tyr, and Val, and norleucine (Nle), as a replacement of Met. Therefore, the theoretical diversity of the library is 16^4 or 65 536 different sequences. The decision to randomize four residues was based on our previous observation that only the four N-terminal residues are important for deformylase binding and catalysis; extension of the peptide beyond the fourth position does not increase the kinetic constant, k_{cat}/K_M (*16*). A peptide linker, BBRM, was added to the C-terminus of the random region for the following reasons: (1) the C-terminal methionine allows for the bound peptides to be cleaved from the resin by CNBr treatment prior to sequencing; (2) the arginine residue makes the peptides more soluble in aqueous medium and ensures a greater efficiency of ionization, improving the sensitivity for MS analysis; and (3) the β-alanine residues add some flexibility to the peptides, making them more accessible to the deformylase enzyme (*20*, *24*). Arginine and lysine were omitted from the randomized positions because their side chains could be labeled by *N*-hydroxysuccinimidobiotin and generate false positive sequences.[2] Isoleucine, cysteine, and methionine were omitted to simplify the sequence determination, prevent the formation of disulfides, and eliminate internal cleavage by CNBr, respectively.

A critical step in combinatorial library screening is how to identify the active compound on a single resin bead. Although the amount of peptide on a single bead (~100 pmol) was sufficient for conventional Edman sequencing, its high cost and time-consuming nature prevented the analysis of a large number of beads, which would be necessary in order to arrive at a consensus sequence for the deformylase. Therefore, in this work we adopted an ingenious method recently developed by Youngquist et al., which enables rapid determination of the sequences of biologically active compounds isolated from support-bound combinatorial libraries by mass spectrometry (*24*). In essence, this method encodes the peptide sequence on each bead by generating a set of chain-termination products as the peptide is being synthesized. This was achieved in our case by the addition

[2] Initially, arginine was included at the randomized positions, and as a result, most of the selected peptides contained at least one arginine in the randomized region. Kinetic analysis of those peptides individually, however, showed that they either had no activity (when R is at the N-terminus) or were poor substrates of the deformylase (R at other positions). Therefore, arginine was excluded from the libraries used in this work. Fortunately, the arginine at the C-terminus did not present a problem, perhaps because this position is not readily accessible to the screening enzyme, streptavidin−β-galactosidase conjugate.

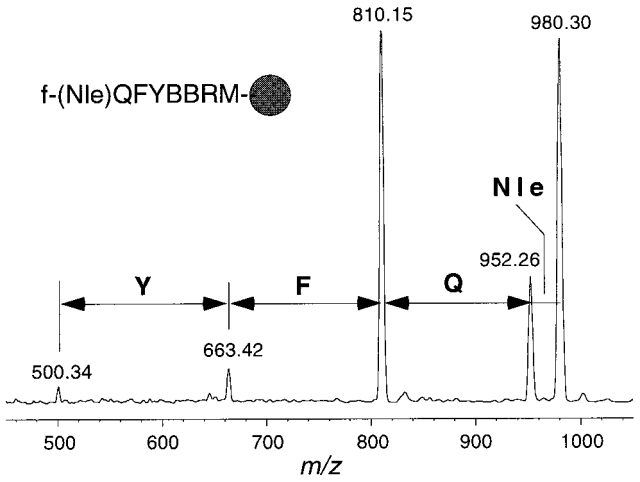

FIGURE 1: MALDI mass spectrum of a peptide and its four chain termination products, obtained with 20% of the material from a single bead (No. 11). M, methionine prior to CNBr cleavage and homoserine lactone after CNBr cleavage.

of a small amount of a capping agent, 10% *N*-acetyl-D,L-alanine (for norleucine) or 20% *N*-acetylglycine (for all other amino acids), to each reaction vessel along with the individual amino acid (90% or 80%) during the synthesis of the randomized region. For the tetrapeptide library, each bead contained a full-length peptide and four chain-termination products. The sequence of the full-length peptide was determined from the mass differences between these termination products (vide infra).

Typically, 70 mg of resin (~200 000 beads) was used in each screening. Statistically, each sequence should appear 3 times in the library, and the confidence level for the presence of that sequence was at 95%. Screening for positive beads was based on that the deformylase reaction would expose a free NH_2 group at the N-terminus of a peptide. The exposed amino group was selectively labeled with a biotin by incubating the library with *N*-hydroxysuccinimidobiotin (arginine and lysine were excluded from the library). The biotin molecule subsequently recruited a streptavidin−β-galactosidase conjugate to the positive beads. Hydrolysis of X-gal by the bound β-galactosidase resulted in intense green color at the surface of the positive beads. To select for the optimal substrates, the proper amount of deformylase enzyme and the reaction time are critical. In this work, we empirically determined that the addition of 10 μg of the Fe^{2+} deformylase (*17*) to 70 mg of resin suspended in 1.0 mL of the deformylase buffer resulted in ~0.1% (~200) colored beads in 10 min. A control experiment without deformylase treatment produced no colored beads.

MALDI Sequencing of Peptides. Peptide sequencing was carried out as previously described (*24*). Positive beads carrying the selected sequences were individually removed from the library, washed with water, and treated with cyanogen bromide. This cleaved the peptide off the solid support at the C-terminal methionine residue, generating a homoserine lactone. Since each bead carried a unique full-length peptide and four chain-termination products, MALDI analysis of the cleavage mixture generated a peptide ladder in the mass spectrum. Figure 1 shows a typical MALDI spectrum, obtained with 20% of the peptide mixture isolated from a single resin bead. The protonated full-length peptide gives a peak at *m*/*z* 980.30. The four protonated termination products produce four peaks at *m*/*z* 952.26, 810.15, 663.42, and 500.34. The mass difference between the full-length peptide and the highest-mass termination product (980.30 − 952.26 + 85 = 113.04 Da) indicates that the N-terminal residue is norleucine. The addition of 85 Da to the mass difference is due to the fact that the full-length peptide is N-formylated, whereas the termination product is capped with Ac-Ala (a difference of 85 Da). This mass difference is very close to the actual mass of norleucine or leucine (113.08 Da). Leucine was ruled out as the N-terminal residue, however, because during the coupling of leucine to the support, Ac-Gly was used as the capping agent; the mass difference between the two highest-mass peaks in the spectrum would be 113.08 − 71 = 42.08 Da (observed 980.30 − 952.26 = 28.04 Da). The penultimate residue is assigned as a glutamine because the mass difference between the first and second termination products (952.26 − 810.15 − 14 = 128.11 Da) closely matches the actual residue weight of glutamine (128.06 Da). The subtraction of 14 Da from the mass difference is due to the different capping agents used, Ac-Ala during the coupling of Nle vs Ac-Gly for Gln. The third and fourth residues are identified as phenylalanine (residue weight 147.07 Da) and tyrosine (residue weight 163.06 Da), respectively, by the simple mass differences between the adjacent peaks (146.73 and 163.08 Da, respectively). Therefore, the selected peptide has the sequence of formyl-(Nle)QFYBBRM* (B = β-alanine; M* = homoserine lactone). Note that the biotinylated full-length peptide, which should give a peak at *m*/*z* = 1178.6, is not observed. This is because only a very small fraction of the support-bound peptides was deformylated during the deformylase treatment as a result of both the limited amount of deformylase added and the inaccessibility of the deformylase to the peptides buried inside the solid support (*25*, *26*).

Selected Sequences. To test the reproducibility of the library screening method, two libraries (each containing 70 mg of resin) were screened separately but under similar conditions. Both libraries produced ~0.1% (~200 each) colored beads. Twenty and sixty positive beads were removed from the first and the second libraries, respectively. The beads were picked in the order of their color intensity as judged by the eye with the aid of a low-magnification microscope, with the brightest beads picked first. Out of the 80 isolated beads, 58 (72.5%) produced high-quality mass spectra, and their sequences were unambiguously determined (Table 1). The sequences on the other 22 beads could not be determined due to their poor spectra. Comparison of the selected sequences from the two libraries demonstrates a good level of reproducibility, with a similar set of preferred sequences obtained from both libraries. For example, peptides f-MHFY and f-MMHY were selected from both libraries. Many peptides from the two libraries share the same amino acids at the P_1', P_3', and P_4' positions (e.g., f-MNYY vs f-MQYY, f-MLYY, f-MAYY, f-MMYY, and f-MPYY; f-GAFY vs f-GFFY and f-GNFY). We believe that many identical sequences would likely be obtained from the two libraries if all of the positive beads (~200 from each library) were to be sequenced. Twenty beads were also randomly picked from the library prior to screening. Sequence analysis of these beads revealed an essentially random distribution of amino acids at the P_1'−P_4' positions.

Table 1: Sequences of Selected Peptides from Combinatorial Libraries[a]

peptide sequence	bead no.	peptide sequence	bead no.
class I			
f-MNYY	21, 31, 52	f-FYFH	**3**
f-MQYY	**18**	f-FYYF	59
f-MLYY	76	f-YPMY	22
f-MAYY	**6**, **17**	f-YLGY	41
f-MAFY	35, 38		
f-MMYY	29	class III	
f-MPYY	46	f-GHYY	47
f-MQFY	**11**	f-GPYY	54
f-MSFY	51	f-GNFY	63
f-MHFY	**4**, 53	f-GFFY	58
f-MYFY	61	f-GAFY	**13**
f-MYHY	40	f-GMYF	32
f-MMHY	**8**, 37	f-GSFL	55
f-MSHY	56	f-GMYM	25
f-MYHF	42	f-GYMM	24
f-MLFM	**16**	f-GYFT	**14**
f-MLFQ	43	f-GWYM	36
f-MYYH	44		
f-MQYH	**2**	class IV	
f-MMHM	49	f-HVYY	30
f-MHYL	**1**	f-HVYM	77
f-MPLY	33	f-HYWY	**7**
f-MFAY	45	f-HQVF	48
f-MTQY	50		
f-MNQF	67	class V	
f-MLQF	69	f-LHYF	78
f-MHMM	28	f-DVLF	**15**
		f-EHFH	79
class II		f-AYFY	80
f-FYFL	23	f-APYY	**10**

[a] Sequences with boldfaced numbers were selected from the first library, whereas the rest were from the second library, M, norleucine.

The selected sequences were grouped into five classes, based on the identity of their N-terminal residues (Table 1). The majority of the selected peptides (33 out of 58) belong to class I, with norleucine as the N-terminal residue. This is consistent with that methionine is the N-terminal residue of all nascent polypeptides, the physiological substrates of peptide deformylase. Earlier studies based on a small set of peptides have also shown that methionine and norleucine are the preferred amino acids at the N-terminus among the amino acids tested (*4*, *16*). Somewhat surprising is that glycine (11 beads), histidine (5 beads), phenylalanine (3 beads), tyrosine (2 beads), and a few other amino acids (total 5 beads) were also selected at the N-terminal position (classes II−V). However, peptides from all five classes have similarities at P_3' and P_4' positions. Analysis of the frequency for each amino acid to occur at a given position resulted in a clear consensus sequence (Figure 2). At the P_1' site, methionine (norleucine) is the most preferred amino acid, followed by glycine, histidine, and phenylalanine or tyrosine. At the P_2' site, there appears to be little selectivity. All amino acids, with the exception of glycine and the negatively charged aspartate and glutamate, appeared at similar frequencies (with slightly higher frequency for tyrosine). The P_3' site is highly selective for an aromatic residue; 47 out of 58 beads had either tyrosine, phenylalanine, or histidine at this position. The P_4' position strongly prefers a tyrosine residue (34 beads). Even phenylalanine occurred at much lower frequency (8 beads). Histidine or hydrophobic residues such as norleucine and leucine are acceptable substitutions. Other

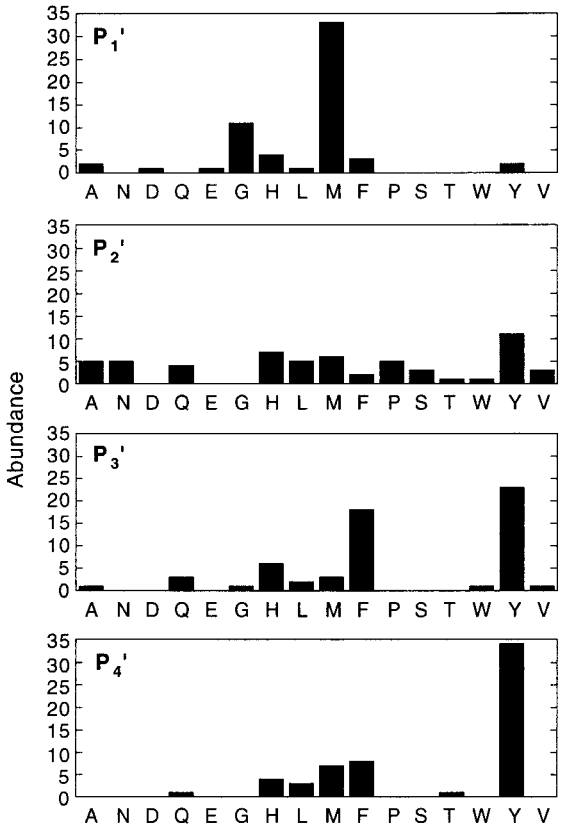

FIGURE 2: Substrate specificity of peptide deformylase. Displayed are the amino acids detected at each of the four N-terminal positions (P_1' is the N-terminal residue). Abundance on the *y* axis represents the number of occurrence of an amino acid at a certain position (maximum 58). M, norleucine.

amino acids are clearly disfavored. Therefore, the consensus sequence for an optimal deformylase substrate is f-MX(F/Y)Y (X = any amino acid except for Asp and Glu).

Kinetic Characterization of Selected Peptides. Representative sequences from each class (Table 2, entries 1−7) were resynthesized individually and assayed against the *E. coli* deformylase in the solution phase. As expected, all class I peptides characterized (f-MAFY, f-MAYY, and f-MNYY) are excellent substrates of the *E. coli* enzyme, with k_{cat}/K_M values of 1.08×10^6, 9.9×10^5, and 3.2×10^5 M^{-1} s^{-1}, respectively. These represent some of the best peptide substrates so far identified, and their activities are 2−3 orders of magnitude higher than those of some randomly chosen *N*-formylmethionyl peptides, which generally have k_{cat}/K_M values in the range of 10^3-10^4 M^{-1} s^{-1} (*16*). These results indicate that the class I peptides were selected because of their high activity, demonstrating the validity of the combinatorial method. Based on the sequence similarities, we believe that the other class I peptides in Table 1 are also good substrates of the *E. coli* deformylase. Surprisingly, a class II peptide (f-FYFH) with an N-terminal phenylalanine, which is structurally distinct from methionine, is also an excellent substrate ($k_{cat}/K_M = 4.2 \times 10^5$ M^{-1} s^{-1}), whereas peptides with other amino acids that are more structurally similar to methionine (e.g., leucine) are much poorer substrates and were not selected from the library. *N*-Formyltyrosyl peptides, also selected from the library, are significantly poorer substrates ($k_{cat}/K_M = 1.6 \times 10^3$ M^{-1} s^{-1} for f-YPMY). The class III and IV peptides that contain N-terminal glycine or histidine (e.g., f-GAFY and f-HVYY)

Table 2: Kinetic Properties of Selected Peptides[a]

entry	substrate	k_{cat} (s^{-1})	K_M (μM)	k_{cat}/K_M ($\times 10^4$ M^{-1} s^{-1})
1	f-MAFYBR	1030 ± 100	950 ± 150	108
2	f-MAYYBR	−	−	99
3	f-MNYYBR	−	−	32
4	f-FYFHBR	204 ± 12	490 ± 60	42
5	f-YPMYBR	1.7 ± 0.1	1070 ± 100	0.16
6	f-HVYYBR	0.26 ± 0.01	1390 ± 130	0.019
7	f-GAFYBR	0.51 ± 0.08	3300 ± 650	0.015
8	f-IAYYBR	11 ± 1	2420 ± 270	0.44
9	f-KAYYBR	0.014 ± 0.001	560 ± 150	0.0024
10	f-RRYABBRM	−	−	ND
11	f-MNYLBBRM	136 ± 18	920 ± 180	15
12	f-MDYLBBRM	−	>2000	2.6
13	f-MAKYBR	580 ± 56	320 ± 90	180
14	f-MARYBR	−	−	150
15	f-MARLBR	360 ± 9	404 ± 26	89
16	f-MAYKBR	680 ± 34	780 ± 70	86
17	f-MAYRBR	280 ± 15	310 ± 54	90
18	f-MAYLBBRM	−	−	16
19	f-MAFLBBRM	−	−	16
20	f-HVYLBBRM	0.058 ± 0.003	470 ± 36	0.012
21	f-GAFLBBRM	0.051 ± 0.003	1420 ± 150	0.0036
22	f-FYFSBBRM	−	−	10
23	f-MDDDBR	−	−	0.0090

[a] All kinetic constants were determined in 50 mM sodium phosphate, pH 7.0. M, methionine; B, β-alanine; −, limited solubility prevented determination of k_{cat} and K_M values; ND, no detectable activity.

are very poor substrates (k_{cat}/K_M < 200 M^{-1} s^{-1}). The class V peptides were not individually characterized due to their sequence diversity but are likely to be poor substrates based on the earlier studies by this laboratory and others (*4*, *6*, *15*, *16*). Thus, our results show that the deformylase has strict structural requirement at the P_1' site, with phenylalanine being the only natural amino acid that can replace methionine while retaining substantial deformylase activity (*4*). Note that the *N*-formylphenylalanyl peptides appear to have a different consensus sequence from the *N*-formylmethionyl peptides [f-FY(F/Y)X vs f-MX(F/Y)Y].

Because isoleucine, lysine, and arginine are potentially important for deformylase recognition but were eliminated from the library, peptides were synthesized with these amino acids incorporated into various positions to test their role in substrate binding/catalysis. Other substitutions were also made to confirm the importance of the selected amino acids at each position. Replacement of the N-terminal methionine in f-MAYY by isoleucine resulted in a marginally active substrate, with a k_{cat}/K_M value of 4.4×10^3 M^{-1} s^{-1} (Table 2, entry 8). Substitution of lysine or arginine at this position, however, rendered the peptides essentially inactive toward the *E. coli* deformylase (Table 2, entries 9 and 10). At the penultimate position, most of the amino acids are tolerated, with the exception of negatively charged aspartate and glutamate. Indeed, replacement of the asparagine of f-MNYL by an aspartyl residue resulted in a 6-fold reduction in activity (Table 2, entries 11 vs 12). At the P_3' position, substitution of lysine or arginine for F/Y in the consensus sequence actually increased the activity by 1.5−5.6-fold (Table 2, compare entries 1 and 2 vs 13 and 14; 15 vs 18). At the P_4' position, replacement of tyrosine by lysine or arginine slightly decreased the enzymatic activity (Table 2, compare entries 2 vs 16 and 17), whereas substitution of leucine or serine resulted in significantly poorer substrates (Table 2, compare entries 1 vs 19; 2 vs 18; 6 vs 20; 7 vs 21; 4 vs 22). Finally, when all three residues at the P_2'−P_4' positions were replaced by negatively charged aspartate, the resulting peptide, f-MDDD (Table 2, entry 23), was a very poor substrate, consistent with the absence of aspartate or glutamate among the selected sequences (Figure 2). Therefore, the consensus sequence for *N*-formylmethionyl peptides should be modified as f-MXZY (Z = K, R, Y, or F).

Table 3: Kinetic Constants for N-Acetyl Peptides

substrate	K_M (μM)	k_{cat} (s^{-1})	k_{cat}/K_M (M^{-1} s^{-1})
f-ML-pNA	20.3 ± 1.3	37.8 ± 1.8	2.9×10^6
Ac-ML-pNA[a]	−	−	42
FAc-ML-pNA	25.0 ± 2.4	0.016 ± 0.001	640
F_2Ac-ML-pNA	3.9 ± 0.3	0.30 ± 0.01	7.8×10^4
F_3Ac-ML-pNA[a]	−	−	136

[a] From Wei and Pei (1997) (*23*).

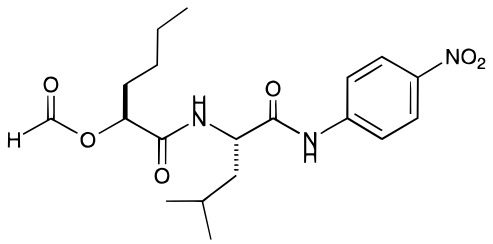

FIGURE 3: Structure of ester 2-formyloxycaproylleucyl-*p*-nitroanilide.

Acylase Activity of Deformylase. Peptide deformylase is highly selective for the formyl group vs acetyl group (4.5×10^4-fold) (*23*). To gain insight into the origin of this selectivity, we synthesized and tested peptides acylated with α-fluoroacetyl groups, FAc-ML-*p*NA, F_2Ac-ML-*p*NA, and F_3Ac-ML-*p*NA (*23*), as potential substrates. While peptide deformylase had very poor activity toward Ac-ML-*p*NA (k_{cat}/K_M = 42 M^{-1} s^{-1}) and nonsaturation kinetics up to 160 μM substrate, substitution of fluorine for the α-H's increased that activity significantly (Table 3). Moreover, the fluorinated substrates exhibited saturation kinetics, with k_{cat} values of 0.016 and 0.3 s^{-1} and K_M values of 25 and 3.9 μM for the mono- and difluorinated peptides, respectively. However, replacement of all three α-H's with fluorine resulted in a poor substrate, which showed no trend of saturation at 40 μM substrate (*23*).

Esterase Activity of Peptide Deformylase. The ester 2-formyloxycaproylleucyl-*p*-nitroanilide (Figure 3) was synthesized and used to examine the esterase activity of peptide deformylase (by assay method C). The *E. coli* enzyme efficiently hydrolyzes the ester substrate, with a k_{cat}/K_M value of 2.7×10^5 M^{-1} s^{-1}. This activity is approximately 10-fold lower than that of the corresponding amide, f-ML-*p*NA ($k_{cat}/K_M = 2.9 \times 10^6$ M^{-1} s^{-1}). Unfortunately, the limited solubility (~10 μM in neutral aqueous buffer) prevented the determination of the individual k_{cat} or K_M values.

DISCUSSION

Peptide deformylase has evolved to act as a broad-specificity enzyme that deformylates polypeptides bearing diverse N-terminal sequences. However, it is clear from this work as well as our earlier work (*16*) that it does so at drastically different rates. Our library screening indicates that an optimal substrate should have the sequence f-MXZY (X

= any amino acid except for Asp and Glu; Z = K, R, Y, or F) or f-FY(F/Y)X. The most prominent feature of the selected sequences is the preference for amino acids with hydrophobic side chains at all four positions, consistent with the fact that the deformylase active/binding site is comprised of largely hydrophobic residues (*27*−*29*). The observed sequence specificity can be explained by our structural studies of the deformylase bound with an inhibitor, (*S*)-2-*O*-(H-phosphonoxy)caproyl-L-leucyl-*p*-nitroanilide, which mimics the tetrahedral intermediate of the deformylase-catalyzed reaction (*30*). In this enzyme−inhibitor complex, the side chain of the P_1' residue (*n*-butyl) is deeply buried in a hydrophobic pocket formed by both protein side chains and the aromatic ring of the *p*-nitroanilide group. The intimate hydrophobic interactions between the protein and the P_1' side chain likely make key contributions to the overall affinity and anchor the formamide moiety for nucleophilic attack by a metal-bound water molecule/hydroxide ion. This limits the types of amino acid side chains that can fit into the pocket. Some side chains may be able to fit into the pocket with high affinity, but may place the scissile formamide group in a nonoptimal position or orientation so that the hydrolysis rate (k_{cat}) is slow. Note that several peptides in Table 2 have relatively high affinity (as indicated by their low K_M values) to the deformylase but have very low k_{cat} values (e.g., f-GAFL, f-KAYY, and f-HVYL). The side chain of the P_2' residue, L-leucine, is engaged to hydrophobic interactions with the solvent-exposed Leu-91 and the protein backbone on one face, and the other face, is completely exposed to the solvent. The lack of specific interactions suggests that many amino acids would be tolerated at this position, and this is indeed the case. Negatively charged residues are disfavored, presumably because of their unfavorable interactions with the hydrophobic Leu-91 and/or the protein backbone. At the P_3'/P_4' positions, the *p*-nitroanilide group sits in a shallow hydrophobic pocket formed by hydrophobic side chains of the protein and the side chain of the P_1' residue of the inhibitor. The aromatic side chains of the P_3' and P_4' residues in the selected peptides likely occupy the same pocket. The shallow pocket permits many other hydrophobic side chains to bind, albeit with less optimal interactions. This feature is important, for it allows the deformylase to accept a diverse array of sequences as substrates. Indeed, besides phenylalanine and tyrosine, the selected peptides also contain histidine, leucine, norleucine, and glutamine at the P_3' and P_4' positions (Figure 2). Lysine and arginine appear to be even more favorable than phenylalanine or tyrosine at the P_3' position. This may be explained by the amphipathic nature of these two side chains. They have hydrophobic methylene groups to interact with the hydrophobic patch on the protein surface and have a positively charged terminal group to interact with the hydrophilic solvent molecules. The f-FY-(F/Y)X peptides may bind to the deformylase in a different manner.

The selection of the *N*-formylglycyl and *N*-formylhistidyl peptides from the library was not expected, and the reasons for their selection are not yet clear. They represent 11 and 4 out of the total of 58 selected sequences, respectively, and yet have very low activities toward the *E. coli* deformylase ($k_{cat}/K_M < 200\ M^{-1}\ s^{-1}$). The fact that these peptides have similar P_3' and P_4' residues to the selected class I peptides argues that they are not due to background hydrolysis of the formyl group or any other random events. Possibly, the f-Gly and f-His peptides were selected because of their relatively high affinity to the deformylase [relative to many f-methionyl peptides which failed to saturate the deformylase at ≥2 mM (*16*)]. Resin beads that carry high-affinity peptides might recruit more enzyme molecules to their surface so that the total amount of deformylated peptides was high enough to render the beads positive ($V_{max} = k_{cat}[E]$) even though the k_{cat} value (determined in solution phase) was low.

In addition to the sequence specificity, peptide deformylase is highly specific for the formyl group; its activity toward *N*-acetyl peptides is orders of magnitude lower (*4*, *23*). This could be due to either steric hindrance of the larger acetyl group or the inherently lower electrophilicity of acetamide compared to formamide. To determine which factor or whether both factors are causing the observed rate difference, we synthesized and tested a series of α-fluoro-substituted acetyl peptides as substrates. Our results show that substitution of one and two fluorines progressively increased the catalytic rate (Table 3). This indicates that the lower rate for acetyl peptides is due, at least in part, to the lower electrophilicity of the acetyl group. The 6-fold decrease in the K_M value upon difluorination is probably due to the introduction of specific interactions between the fluorine atoms and the protein (e.g., H bonding). However, the decreasing trend in activity going from difluoro- to trifluoroacetyl peptides and the lack of saturation kinetics for acetyl and trifluoroacetyl peptides suggest that the steric factor also plays a role. The trifluoromethyl group is significantly larger in size than a methyl group (*31*, *32*).

In summary, screening of resin-bound combinatorial peptide libraries coupled with sequencing by MALDI mass spectrometry is shown to be a rapid method for the identification of optimal substrates for peptide deformylase. The advantage of this method is that individual sequences can be obtained, thus allowing for the identification of any specific ligands regardless of their abundance in the library. It should be applicable to other acylases and peptidases. The sequence specificity information obtained in this work will serve as a starting point for investigating the catalytic mechanism and in vivo function of peptide deformylase as well as designing selective deformylase inhibitors.

ACKNOWLEDGMENT

We thank Kirk Beebe and Stephane Grimme of this laboratory for the synthesis and assay of some of the peptides, and Dr. Yen-Ho Chu and his group members for helpful suggestions during the course of this work.

REFERENCES

1. Meinnel, T., Mechulam, Y., and Blanquet, S. (1993) *Biochimie 75*, 1061−1075.
2. Adams, J. M., and Capecchi, M. (1966) *Proc. Natl. Acad. Sci. U.S.A. 55*, 147−155.
3. Webster, R. E., Engelhardt, D. L., and Zinder, N. D. (1966) *Proc. Natl. Acad. Sci. U.S.A.* 55, 155−161.
4. Adams, J. M. (1968) *J. Mol. Biol. 33*, 571−589.
5. Takeda, M., and Webster, R. E. (1968) *Proc. Natl. Acad. Sci. U.S.A. 60*, 1487−1494.
6. Livingston, D. M., and Leder, P. (1968) *Biochemistry 8*, 435−443.
7. Ball, L. A., and Kaesberg, P. (1973) *J. Mol. Biol. 79*, 531−537.

8. Mazel, D., Pochet, S., and Marliere, P. (1994) *EMBO J. 13*, 914−923.
9. Hauschild-Rogat, P. (1968) *Mol. Gen. Genet. 102*, 95−101.
10. Marasco, W. A., Phan, S. H., Krutzsch, H., Showell, H. J., Feltner, D. E., Nairn, R., Becker, E. L., and Ward, P. A. (1984) *J. Biol. Chem. 259*, 5430−5439.
11. Milligan, D. L., and Koshland, D. E., Jr. (1990) *J. Biol. Chem. 265*, 4455−4460.
12. Warren, W. C., Bentle, K. A., Schlittler, M. R., Schwane, A. C., O′Neil, J. P., and Bogosian, G. (1996) *Gene 174*, 235−238.
13. Bogosian, G., Violand, B. N., Dorward-King, E. J., Workman, W. E., Jung, P. E., and Kane, J. F. (1989) *J. Biol. Chem. 264*, 531−539.
14. Hoffert, S. P. (1998) *Scientist 12 (No. 8)*, 1−6.
15. Meinnel, T., and Blanquet, S. (1995) *J. Bacteriol. 177*, 1883−1887.
16. Rajagopalan, P. T. R., Datta, A., and Pei, D. (1997) *Biochemistry 36*, 13910−13918.
17. Rajagopalan, P. T. R., Yu, X. C., and Pei, D. (1997) *J. Am. Chem. Soc. 119*, 12418− 12419.
18. Rajagopalan, P. T. R., and Pei, D. (1998) *J. Biol. Chem. 273*, 22305−22310.
19. Bodanszky, M. (1993) *Principles of Peptide Synthesis*, 2nd ed., Springer-Verlag, Germany.
20. Yu, Z., and Chu, Y.-H. (1997) *Bioorg. Med. Chem. Lett. 7*, 95−98.
21. Lam, K. S., Salmon, S. E., Hersh, E. M., Hruby, V. J., Kazmierski, W. M., and Knapp, R. J. (1991) *Nature 354*, 82−84.
22. Furka, A., Sebestyen, F., Asgedom, M., and Dibo, G. (1991) *Int. J. Pept. Protein Res. 37*, 487−493.
23. Wei, Y., and Pei, D. (1997) *Anal. Biochem 250*, 29−34.
24. Youngquist, R. S., Fuentes, G. R., Lacey, M. P., and Keough, T. (1995) *J. Am. Chem. Soc. 117*, 3900−3906.
25. Vagner, J., Barany, G., Lam, K. S., Krchnak, V., Sepetov, N. F., Ostrem, J. A., Strop, P., and Lebl, M. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 8194−8199.
26. Cheung, Y. W., Abell, C., and Balasubramanian, S. (1997) *J. Am. Chem. Soc. 119*, 9568−9569.
27. Meinnel, T., Blanquet, S., and Dardel, F. (1996) *J. Mol. Biol. 262*, 375−386.
28. Chan, M. K., Gong, W., Rajagopalan, P. T. R., Bing, H., Tsai, C. M., and Pei, D. (1997) *Biochemistry 36*, 13904−13909.
29. Becker, A., Schlichting, I., Kabsch, W., Schultz, S., and Wagner, A. F. V. (1998) *J. Biol. Chem. 273*, 11413−11416.
30. Hao, B., Gong, W., Rajagopalan, P. T. R., Zhou, Y., Pei, D., and Chan, M. K. (1998) *Biochemistry* (submitted for publication).
31. White, D., and Coville, N. J. (1994) *Adv. Organomet. Chem. 36*, 95−158.
32. Datta, D., and Majumdar, D. (1991) *J. Phys. Org. Chem. 4*, 611−617.

BI9820412