Chemical Models of Peptide Formation in Translation

ARTICLE *in* BIOCHEMISTRY · FEBRUARY 2010

Impact Factor: 3.02 · DOI: 10.1021/bi1000273 · Source: PubMed

2 AUTHORS, INCLUDING:

Eddie Ray Watts
Stanford University
7 PUBLICATIONS **230** CITATIONS

SEE PROFILE

Published in final edited form as:
Biochemistry. 2010 March 16; 49(10): 2177–2185. doi:10.1021/bi1000273.

Chemical Models of Peptide Formation in Translation

R. Edward Watts and **Anthony C. Forster***
Department of Pharmacology and Vanderbilt Institute of Chemical Biology, Vanderbilt University Medical Center, 2222 Pierce Ave., Nashville, TN 37232, USA.

Abstract

Ribosomal incorporations of *N*-alkyl amino acids including proline are slower than incorporations of non-*N*-alkyl L-amino acids. The "chemical reactivity hypothesis" proposes that these results, and the exclusion of non-proline *N*-alkyl amino acids from the genetic code, are explained by intrinsic chemical reactivities of the amino acid nucleophiles. However, there is little data on the reactivities relevant to physiological conditions. Here, we use non-enzymatic, aqueous-based assays to model 11 amino acid nucleophiles in dipeptide formation. The relative rates in the non-enzymatic and translation systems correlate well, supporting the chemical reactivity hypothesis and arguing that peptide bond formation, not accommodation, is rate limiting for natural Pro-tRNA^{Pro} isoacceptors. The effects of *N*-substitution sterics, side chain sterics, induction and pKa were evaluated in the chemical model. The dominant factor affecting relative rates was found to be *N*-substitution sterics.

Ordinary natural substrates for translation encompass 19 amino acids and one *N*-alkyl amino acid, proline, esterified to a larger number of transfer RNAs that recognize 61 sense codons. Ribosomes thus utilize some 61^2 different codon pairs to template the formation of some 400 different dipeptide bonds (and many more dipeptides when supplied with unnatural amino acids). Despite this enormous diversity of natural chemical couplings, it is presumed that any codon pair can direct efficient coupling between cognate amino acids if the substrate concentrations are not limiting. The "uniform decoding rate hypothesis" (1) goes further by predicting that all natural couplings occur at the same rate at saturation. This hypothesis has been verified by measuring *in vitro* rates of dipeptide formation from fMet-tRNA$_i^{fMet}$ and nine amino acids (Phe, Ala, Gly, Ile, Val, Glu, Leu, Lys and His) delivered by 10 of the 46 aminoacyl-tRNAs of *E. coli* at 10 different codons at the ribosomal A site (1). The generally believed mechanism of uniformity (1) is evolutionary tuning and the masking of peptide bond formation rates by a much slower accommodation step (movement of the A-site substrate into the peptidyl-transferase center) (2). This "rate-limiting accommodation hypothesis" is based on the observation that accommodation is rate-limiting in dipeptide formation when the A site substrate is Phe-tRNA^{Phe} chemically labeled with a large fluorescent molecule (2). However, it should be noted that experimental challenges have prevented testing the rate-limiting accommodation hypothesis with other substrates.

Recently, we found that natural *E. coli* Pro-tRNA^{Pro} isoacceptors provide interesting exceptions to the uniform decoding hypothesis (3). These substrates couple three- to six-fold slower than natural Phe-tRNA^{Phe} to fMet-tRNA$_i^{fMet}$ *in vitro*, depending on the Pro codon (<15% standard errors). The uniform decoding hypothesis also does not explain our dramatically different incorporation efficiencies of unnatural *N*-alkyl amino acids (3,4); rates

*To whom correspondence should be addressed: Dept. of Pharmacology, Vanderbilt University Medical Center, 2222 Pierce Ave., Nashville, TN 37232, USA. Phone: (615) 936 3112. Fax (615) 936 5555. a.forster@vanderbil.edu.

SUPPORTING INFORMATION AVAILABLE Tables S1-S3 containing kinetic rate constants and pKa values associated with Figures 3 and 5. This material is available free of charge via the Internet at http://pubs.acs.org.

of dipeptide formation with natural fMet-$tRNA_i^{fMet}$ using a very efficient synthetic $tRNA^{Phe}$-based adaptor at the ribosomal A site were Phe > Ala > Pro > *N*-Me-Phe > *N*-Bu-Phe ((3); structures in Figure 1; summary in Figure 6 below). This order of rates was correlated with the theoretical intrinsic chemical reactivities of the amino nucleophiles, a "chemical reactivity hypothesis" (4). Consistent with this hypothesis, the rate-limiting step for all but our fastest coupling amino acid (Phe) occurred after GTP hydrolysis by the aminoacyl-tRNA carrier, elongation factor Tu (EFTu) (3). This narrowed down the steps responsible for the large difference in rates to (i) substrate release from EF-Tu.GDP, (ii) substrate accommodation, and/or (iii) peptide bond formation. It is unclear how our amino acid substitutions could cause such differences in accommodation, as might be predicted by the rate-limiting accommodation hypothesis. Alternatively, corresponding differences in peptide bond formation rates would be predicted by the chemical reactivity hypothesis. It therefore seems that there are at least some natural exceptions to the uniform decoding rate hypothesis and the rate-limiting accommodation hypothesis (the *E. coli* $ProtRNA^{Pro}$ isoacceptors), and that there are many unnatural exceptions (*N*-alkyl aminoacyl-tRNAs).

The chemical reactivity hypothesis could also explain a long-standing mystery regarding the origin of the genetic code (4) in the "RNA world." *N*-alkyl amino acids are major products of prebiotic synthetic experiments and meteorite analyses, so why did Nature select Pro as the only *N*-alkyl amino acid in the genetic code (5)? The chemical reactivity hypothesis proposes that Pro was selected because it is the most reactive nucleophile of the *N*-alkyl amino acids under physiological conditions.

Although the chemical reactivity hypothesis makes specific predictions about the relative reactivities of the amino acid nucleophiles on the ribosome and in the RNA world, there is little data on the reactivities relevant to physiological conditions. Prior comparisons of the nucleophilicities of various amino acids and *N*-alkyl amino acids in non-enzymatic peptide bond formation (aminolysis) were done mainly to optimize amino acid couplings in solid phase peptide synthesis (*e.g.* Gly = Pro > Ala > Phe in (6); Ala > Gly > Phe > Pro in (7)). These reactivity orders differ from our expected relative chemical reactivities of the amino nucleophiles under physiological conditions (Phe > Ala > Pro; (3)). However, such studies are of limited relevance to translation because they used activated intermediates and complex multistep mechanisms (*e.g.* carbodiimide and hexafluorophosphate 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium (HATU) chemistries) that are very different from aminoacyl-tRNA ester chemistry, and reactions have been performed almost exclusively in anhydrous organic solvents to prevent hydrolysis. Anhydrous solvents are problematic for modeling translation because pH is not well-controlled and pKa's can differ from aqueous solution by many units (8). pH is an important variable because the % protonation of the amine nucleophile varies considerably between different amino acids under physiological conditions ((9); Table S3), and deprotonation is necessary for nucleophilic attack. Though it is possible, if not likely, that the amine is not fully accessible to bulk solvent on the ribosome (10) and its pKa is shifted, even deprotonation of the ammonium form by a proton shuttle involving RNA (11) should be affected by amine pKa. Indeed, the rates of peptide formation by puromycin analogs at the ribosomal peptidyl transferase center under physiological conditions exhibit dependence on pH, and this pH dependence is altered by changing the pKa of the amine nucleophile (12). These data and considerations argue that aqueous buffers at near-neutral pH are preferable for modeling peptide bond formation (13) in translation. Yet, the only relevant such rate comparison that we are aware of showed Gly > Ala > Pro > Phe (14). This order also differs from our expected relative chemical reactivities, although this study was limited by using activated acetates rather than activated amino acids as electrophiles, by using phosphate as the ester leaving group (which is much less hindered sterically than ribose of tRNA), and by not testing *N*-alkyl amino acids other than Pro. Thus, we examine here alternative leaving

groups under physiological-like conditions in order to model the nucleophilicities of amino acids studied on the ribosome.

MATERIALS AND METHODS

Materials

Formyl-methionine and *N*-Me-glycinamide hydrochloride were purchased from Bachem Americas, Inc. (Torrance, CA) and 3 mm NMR tubes from Bruker Biospin (Billerica, MA). All other compounds, unless noted below, were purchased from Sigma-Aldrich, Inc. (St. Louis, MO).

Synthesis of Reagents

Formyl-methionine-hydroxy-succinimidyl ester (fMet-NHS)—Formyl-methionine (1240.5 mg, 7mmol) was dissolved in 75 ml CH_2Cl_2. Dimethylaminopyridine (43 mg, 0.35 mmol, 0.05 equiv.), *N*-(3-dimethylaminopropyl)-*N*′-ethylcarbodiimide hydrochloride (2.7 g, 14 mmol, 2 equiv.), and *N*-hydroxysuccinimide (1.6 g, 14 mmol, 2 equiv.) were added sequentially and stirred overnight at room temp. The reaction mixture was concentrated under reduced pressure to ~25 ml, and loaded onto a silica flash column solvated with CH_2Cl_2. It was eluted with 0-30% acetone in CH_2Cl_2, concentrated under reduced pressure, dried with toluene azeotrope, and the resultant oil dried under high vacuum 48 hrs. Yield: 1.721 g, 6.27 mmol, 89.6%. ^{1}H NMR (400 MHz, $CDCl_3$): δ 2.122 (3H, s), 2.177 (2H, m), 2.647 (2H, m), 2.864 (4H, bs), 5.222 (1H, dddd, J = 5.2, 7.6 Hz), 6.604 (1H, d, J = 8.0 Hz), 8.267 (1H, s). ^{13}C NMR (100 MHz, $CDCl_3$): δ 15.32, 25.33, 29.41, 31.51, 48.38, 160.75, 167.30, 168.54. High resolution mass spectrometry (HRMS) *m/z* calculated for $C_{10}H_{14}N_2O_5S$, 274.0623, found, 297.0518 (m + Na^+). The pKa of the NHS leaving group is 5.9 (15).

N-methyl-phenylalaninamide—*N*-methyl phenylalanine (400 mg, 2.23 mmol) was dissolved in 17.1 ml of 2:1 dioxane:water. 9-Fluorenylmethyl chloroformate (577 mg, 2.23 mmol, 1 equiv.) and K_2CO_3 (616.4 mg, 4.46 mmol, 2 equiv.) were added. After 24 hrs, TLC in 7% acetone and 0.5% HOAc in CH_2Cl_2 showed very little 9-Fluorenylmethyl chloroformate present (Rf = 0.6), and a preponderance of product (Rf = 0.5) when visualized by UV. Staining the plate with ninhydrin showed no amine present in the reaction mixture, whereas a known sample of *N*-methylphenylalanine stained red on the baseline. Staining with bromocresol green showed the presence of a carboxylic acid with an Rf = 0.5, corresponding to the Rf of the product, but no carboxylic acid on the baseline. The reaction was acidified, and brine (15 ml) added, before extraction (4 × 100 ml) with EtOAc and solvent removal with reduced pressure. This reaction was repeated exactly to increase the amount of total product available. Without further purification, 1.0 g of the Fmoc-*N*-methyl-phenylalanine was dissolved in 25 ml dry THF, and triethylamine (382 ul, 2.74 mmol, 1.1 equiv) was added. This mixture was cooled in an ice/water bath, whereupon ethylchloroformate (261 ul, 2.74 mmol, 1.1 equiv) was added in aliquots over 2 mins. This was stirred in the ice/water bath for 1 hr, then ammonium hydroxide (1.18 ml, an excess) at room temp was added in one aliquot to the cold reaction mixture with vigorous stirring. The reaction was allowed to warm to room temp and was stirred overnight. The reaction was diluted with water and extracted with EtOAc (4 × 50 ml). These combined organic phases contained Fmoc-protected *N*-methyl phenylalaninamide. The combined aqueous phases were basified with K_2CO_3, and extracted with EtOAc (4 × 50 ml). These combined organic phases contained *N*-methyl phenylalaninamide that had been deprotected by the basic conditions of the reaction. The combined organic phases from both extractions were concentrated under reduced pressure and 10 ml of 20% piperidine in DMF were added and reacted for 24 hrs. The DMF/piperidine was removed under reduced pressure, and the material purified by flash chromatography in 0-10% methanol in 3% triethylamine/ CH_2Cl_2. Yield: 286 mg, 1.6 mmol, 64%. ^{1}H NMR (400 MHz, CD_3CN): δ 2.305 (3H, s), 2.751

(1H, dd, J = 8.4, 13.6 Hz), 2.85 (2H, bs), 3.027 (1H, dd, J = 5.2, 13.6 Hz), 3.162 (1H, dd, J = 5.2, 8.4 Hz), 6.344 (1H, bs), 7.063 (1H, bs), 7.223 (5H, m). ^{13}C NMR (100 MHz, CD_3CN): δ 35.98, 40.70, 67.53, 127.83, 129.80, 130.74, 140.22, 176.86. HRMS *m/z* calculated for $C_{10}H_{14}N_2O$, 178.1106, found, 179.1181 (m + H^+).

N-butyl-phenylalaninamide—Phenylalaninamide (492.6 mg, 3 mmol) and butanal (537 ul, 6 mmol) were added to 15 ml methanol at room temp and stirred 15 min. The reaction vessel was cooled in an ice/water bath for 15 mins, whereupon $NaBH_4$ (158.9 mg, 4.2 mmol) was added in 4 equal aliquots at 5 min intervals. The reaction was allowed to stir and warm to room temp overnight. TLC in 5% acetone, 5% triethylamine in dichloromethane showed starting material and a product with ninhydrin stain. The reaction mixture was cooled again in an ice/water bath, 537 ul more butanal was added, and stirred 15 mins, whereupon 158.9 mg more $NaBH_4$ was added in 4 aliquots at 5 min intervals. The reaction was allowed to stir and warm to room temp over 24 hrs. TLC showed greatly diminished starting material relative to the product. The crude was purified by flash chromatography with a 0-10% gradient of methanol in a mixture of 10% ethyl acetate, and 2% triethylamine in dichloromethane. Solvents were removed by reduced pressure. Yield: 287.3 mg, 1.3 mmol, 43.5%. ^{1}H NMR (400 MHz, D_6-acetone): δ 0.804 (3H, dd, J = 4.8 Hz), 1.198 (2H, m), 1.306 (2H, m), 2.390 (1H, ddd, J = 4.4, 8.8 Hz), 2.501 (1H, ddd, J = 4.8, 9.6 Hz), 2.715 (1H, dd, J = 5.6, 9.2 Hz), 3.009 (1H, dd, J = 3.2, 9.2 Hz), 3.206 (1H, dd, J = 3.6, 6.0 Hz), 6.25 (1H, bs), 7.05 (1H, bs), 7.189 (5H, m). ^{13}C NMR (100 MHz, D_6-acetone): δ 13.20, 19.89, 31.95, 39.20, 47.90, 64.01, 126.21, 128.17, 129.11, 138.62, 175.67. HRMS *m/z* calculated for $C_{13}H_{20}N_2O$, 220.3107, found, 221.1650 (m + H^+).

Synthesis and characterization of dipeptide products—One ml-scale reactions using conditions identical to the kinetic assays (with the highest concentration of nucleophile) were performed and allowed to run > 10 hrs without removal of aliquots or quenching. These samples were purified by prep HPLC, dissolved in 30% $CD_{13}CN/D_2O$ and analyzed by ^{1}H NMR, ^{13}C NMR and High Resolution Mass Spectroscopy. Aliquots of the purified, characterized compounds were added to quenched timepoint samples saved from the kinetic assays. In all cases, the putative product peaks in the timepoint samples increased relative to the hydrolysis by-product and hydroxylamine-quenched starting material upon addition of the known standard compound. NMR data for the primary rotamers are given.

Formyl-methionine-glycinamide—^{1}H NMR (600 MHz, 30% CD_3CN/70% D_2O): δ 2.175 (2H, m), 2.351 (3H, s), 2.694 (2H, m), 3.988 (2H, dd, J=16.8 Hz), 4.677 (~2H, m) (this signal obscured by the water signal), 8.290 (1H, s). ^{13}C NMR (150 MHz, 30% CD_3CN/70% D_2O): δ 11.49, 26.52, 27.65, 39.42, 48.81, 159.69, 161.66, 170.84. HRMS *m/z* calculated for $C_8H_{15}N_3O_3S$, 233.0834, found, 256.0725 (m + Na^+).

Formyl-methionine-alaninamide—^{1}H NMR (600 MHz, 30% CD_3CN/70% D_2O): δ 1.619 (3H, d, J = 7.2 Hz), 2.209 (2H, m), 2.356 (3H, s), 2.791 (2H, m), 4.522 (1H, ddd, J = 7.2 Hz), 4.754 (~1H, m) (this signal was obscured by the water signal), 8.377 (1H, s). ^{13}C NMR (150 MHz, 30% CD_3CN/70% D_2O): δ 11.59, 14.14, 46.74, 48.61, 161.46, 169.88, 174.51. HRMS *m/z* calculated for $C_9H_{17}N_3O_3S$, 247.0991, found, 270.0883 (m + Na^+).

Formyl-methionine-phenylalaninamide—^{1}H NMR (600 MHz, 30% CD_3CN/70% D_2O): δ 1.908 (2H, ddd, J = 7.2 Hz), 2.224 (3H, s), 2.304 (2H, m), 3.093 (1H, dd, J = 11.4, 14.4 Hz), 3.549 (1H, dd, J = 4.8, 14.4 Hz), 4.579 (1H, dd, J = 7.2 Hz), 4.887 (1H, dd, J = 4.8, 11.4 Hz), 7.530 (5H, m), 8.282 (1H, s). ^{13}C NMR (150 MHz, 30% CD_3CN/70% D_2O): δ 11.46, 25.99, 28.01, 34.24, 48.92, 51.79124.55, 126.16, 126.48, 134.31, 161.19, 169.98, 172.89. HRMS *m/z* calculated for $C_{15}H_{21}N_3O_3S$, 323.1304, found, 346.1196 (m + Na^+).

Formyl-methionine-prolinamide—^{1}H NMR (600 MHz, 30% CD_3CN/70% D_2O): δ 2.163 (2H, m), 2.232 (1H, m), 2.306 (2H, m), 2.371 (3H, s), 2.806 (2H, m), 3.930 (1H. m), 4.059 (1H, m), 4.611 (1H, dd, J = 6.0, 8.4 Hz), 5.119 (1H, dd, J = 4.8, 9.0 Hz), 8.338 (1H, s). ^{13}C NMR (150 MHz, 30% CD_3CN/70% D_2O): δ 11.76, 22.06, 26.64, 26.93, 27.61, 45.29, 46.30, 57.70, 160.95, 168.59, 173.78. HRMS *m/z* calculated for $C_{11}H_{19}N_3O_3S$, 273.1147, found, 296.1043 (m + Na^+).

Formyl-methionine-N-Me-glycinamide—^{1}H NMR (600 MHz, 30% CD_3CN/70% D_2O): δ 2.134 (2H, m), 2.351 (1H, s), 2.742 (2H, m), 3.448 (3H, s), 4.24 (2H, dd, J=16.8, 43.2 Hz), 5.298 (1H, dd, J = 4.2, 9.0 Hz), 8.329 (1H, s). ^{13}C NMR (150 MHz, 30% CD_3CN/70% D_2O): δ 11.72, 26.64, 27.68, 34.14, 44.80, 48.54, 160.81, 169.65, 170.24. HRMS *m/z* calculated for $C_9H_{17}N_3O_3S$, 247.0991, found, 270.0884 (m + Na^+).

Dipeptide formation kinetics

An electrophile-containing solution was added to an excess of nucleophile-containing solution with vigorous stirring. Aliquots of the reaction mixture were removed and added to a quench buffer at recorded intervals. The quenched samples were diluted and analyzed by ^{1}H NMR (*e.g.* Figure 2A). The electrophilic solution contained fMet-NHS and pyrazine (as an internal NMR integration standard) in 80% deuterated DMSO and 20% deuterated acetonitrile (ACN). The DMSO was necessary to effect rapid dissolution of the fMet-NHS in the aqueous nucleophilic buffer. The ACN was added to the DMSO solution immediately prior to use to keep the electrophilic solution from freezing at the reaction temp of 4°C. The final fraction of organic solvent in D_2O was 7.5%. The nucleophile-containing buffer was comprised of varying concentrations of amino acid amide hydrochloride salts in D_2O and potassium phosphate buffer. The nucleophilic buffer was adjusted to a reading of 7.0 on the pH meter by addition of KOD. The pD of a solution with a reading of 7.0 on the pH meter was calculated to be 7.4 (16). The ionic strength of the reactions was kept constant across all concentrations of all nucleophiles by adjusting the final [Cl^-] with added NaCl. The starting concentrations of species in the reaction mixture were: fMet-NHS (6.7 mM); pyrazine (1.25 mM); $KHPO_4$ buffer (40 mM); NaCl (785 mM); the concentrations of amino acid amides varied. High ionic strength was used because it substantially improved assay reproducibility by greatly reducing hydrolysis upon quenching. Changes in rates of aminolysis due to increased ionic strength did not change the relative order of rates of the nucleophiles. For each reaction, 13-16 time-points of 100 ul were collected and quenched by adding them to 35 ul quench buffer containing 300 mM hydroxylamine and 100 mM $KHPO_4$ buffer in deuterium oxide at pD 7.4 (Figure 2A).

Samples were analyzed with 600 MHz ^{1}H NMR (Bruker) at 298K with T_1 = 10s. At least 16 transients were collected for each time-point. Sufficient resolution of the formyl proton resonances of the different species present required use of 3 mm NMR tubes and dilution with D_2O to help overcome tuning and shimming problems arising from the high concentrations of salt. It was also necessary to add deuterated ACN to the quenched samples to improve resolution. Resolution may have been improved by lowering the viscosity of the solvent to allow better molecular tumbling of analytes and/or by a chemical shift effect from changes in pH upon addition of the organic solvent. However, all of the 135 ul quenched samples were diluted to 420 ul with D_2O and ACN. Achieving sufficient resolution in the alaninamide and phenylalaninamide reaction samples required adding ACN to quenched samples to give a final concentration of 30% deuterated ACN in D_2O, whereas, the prolinamide, glycinamide and *N*-Me-glycinamide reactions required 20% ACN. Results were quantified by integrating resonances of the formyl protons on the dipeptide product, free acid (hydrolysis) by-product, and hydroxylamine (quenched) adduct of the starting material with reference to the integration of the pyrazine internal standard. The data collected were fitted using equations for competing reactions (*e.g.* Figure 2B).

Equations

For A + B + H_2O → P + D, where P is the aminolysis product of starting material ester A and amine B, and D is the hydrolysis product of A and water,

$$d[\mathrm{A}]/dt = -k_1[\mathrm{A}][\mathrm{B}] - k_2[\mathrm{A}] \text{ (disappearance of starting material A)}$$
$$d[\mathrm{P}]/dt = k_1[\mathrm{A}][\mathrm{B}] \text{ (appearance of aminolysis product)}$$
$$d[\mathrm{D}]/dt = k_2[\mathrm{A}] \text{ (appearance of hydrolysis product)}$$

Since the rate of disappearance of starting ester is equal to the sum of the rates of appearance of the products of the aminolysis and hydrolysis reactions, then:

$$k_{Total} = k_1 + k_2$$

Integrating the aminolysis product rate equation gives:

$$[\mathrm{P}] = (-k_1/(k_1+k_2))^*[\mathrm{A}]_0{}^*[\mathrm{B}]_0{}^*\left(1 - e^{-(k1+k2)t}\right)$$

where $[A]_0$ is the initial concentration of starting material ester, and $[B]_0$ is the initial concentration of starting material amine. Removing $[B]_0$ from this equation gives:

$$[\mathrm{P}] = \left(-k_{apparent}/(k_1+k_2)\right)^*[\mathrm{A}]_0{}^*\left(1 - e^{-(k1+k2)t}\right) \quad \text{Eq. 1}$$

where $k_{apparent}$ is the rate constant for the aminolysis reaction that is dependent on the initial amine concentration $[B]_0$, and which gives k_1 as the slope of the line from the relation:

$$k_{apparent} = k_1[\mathrm{B}]_0 \quad \text{Eq. 2}$$

Fitting Equation 1 to the data collected in kinetics assays performed at varying concentrations of amine gave a set of values for $k_{apparent}$ that were plotted against the concentration of amine using Equations 2 and 3.

$$k_{apparent} = k_3[\mathrm{B}]_0 + k_4[\mathrm{B}]_0{}^2 \quad \text{Eq. 3}$$

Equation 2 above assumes a negligible amount of base catalysis occurs under the conditions of the assays. Equation 3 governs the reaction where the amino acid amide acts as a general base catalyst (right of Tables S1 and S2), and where the reaction proceeds by both catalyzed and uncatalyzed mechanisms. In Equation 3, k_3 represents the rate of the uncatalyzed reaction pathway and k_4 represents the rate of the catalyzed reaction pathway. All data were fit with regressions using GraphPad Prism 4. The concentration of unprotonated amine was calculated with the Henderson-Hasselbalch equation.

Relative rates of dipeptide formations in competition assays

To a D_2O buffer containing 30 mM $KHPO_4$, pD 7.4, 100 mM NaCl, *N*-Me-glycinamide and an amino acid nucleophile was added a solution of fMet-NHS in deuterated DMSO to make a final fMet-NHS concentration of 30 mM and a final concentration of DMSO of 10% v/v. The ratio of *N*-Me-glycinamide to the amino acid nucleophiles varied depending on the solubility

of the amino acid nucleophile, but typically ranged from a four-fold excess of *N*-Me-glycinamide to an equimolar ratio for each nucleophile. The initial concentration of the amino acid nucleophile was never below 3x that of fMet-NHS to ensure reaction rates would remain relatively constant with respect to both nucleophiles. Reactions were run for 24 hrs at room temp, diluted with D_2O and titrated with deuterated ACN. NMR scans were taken between ACN additions until adequate resolution of the fMet-AA dipeptide product and fMet-OH hydrolysis product was achieved. Typically, the final concentration of ACN fell between 15-30%. NMR resonances were assigned with the help of standards, and pyrazine in the fMet-NHS sample served as an integration standard. Samples were spun on a Bruker cryoprobe 600 MHz NMR at 410K with 16 transients collected and a relaxation delay of 10 seconds. Relative rates of nucleophile reactivity were taken to equal the relative concentrations of dipeptide products at equilibrium adjusted by the relative initial concentrations of nucleophiles in accordance with the following equation:

$$\text{Relative Rate} = \left(\left[\text{fMet} - \text{AA}_{\text{eq}}\right]^{*} [N - \text{Me} - \text{Gly}_0]\right) / \left(\left[\text{fMet} - N - \text{Me} - \text{Gly}_{\text{eq}}\right]^{*} [\text{AA}_0]\right)$$

where [fMet-AA_{eq}] is the equilibrium concentration of the dipeptide containing fMet and the test amino acid nucleophile, [fMet-*N*-Me-Gly_{eq}] is the equilibrium concentration of the dipeptide containing fMet and *N*-Me-glycinamide, [*N*-Me-Gly_0] is the initial concentration of *N*-Me-glycinamide and [AA_0] is the initial concentration of the test amino acid nucleophile. Relative rates were constant across the concentration range for all amino acid nucleophiles. As expected, the concentration of fMet-OH hydrolysis product was constant across all experiments with a given amino acid nucleophile since the concentration of deuteroxide was held constant by the buffer.

Relative rates of dipeptide formation from N-Me-Phe-NH_2 versus Phe-NH_2

fMet-NHS hydrolyzes much faster than it reacts with *N*-Me-Phe-NH_2, preventing detection of fMet-*N*-Me-Phe-NH_2 product by NMR or UV absorbance. However, in competition reactions between *N*-Me-Phe-NH_2 and Phe-NH_2 for coupling to fMet-NHS, both dipeptide products gave detectable mass spec extracted ion peaks, enabling estimation of the relative rates (assuming their similar structures give similar relative ionization efficiencies). Five reactions containing 2.5 mM fMet-NHS, 7.5 mM Phe-NH_2 and amounts of *N*-Me-Phe-NH_2 that were equimolar or 5-fold, 10-fold, 50-fold and 100-fold in excess over Phe-NH_2 were conducted in ~400 mM H_2O in DMSO. The reactions were run for 20 min at 90C, then 16 hrs at room temp. Five hundred ul 1M phosphate buffer, pH 4.2, and 100 ul brine were added, followed by 4 ml of EtOAc. After vigorous stirring, the organic layer was removed, dried with sodium sulfate, pushed through a silica plug and evaporated under reduced pressure. Products were resuspended in 50 ul ACN and then diluted to 150 ul with water. Samples were analyzed by ultra-high pressure liquid chromatography-mass spectrometry (UPLC-MS), the identities of the dipeptide products in the reaction were confirmed by HRMS, and the extracted ion peaks of the dipeptide product masses were integrated. Plots of product ratios to starting material ratios gave straight lines. HRMS *m/z* calculated for fMet-Phe-NH_2, $C_{15}H_{21}N_3O_3S$, 323.1304 found, 346.1201 ($m+Na^+$). HRMS *m/z* calculated for fMet-*N*-Me-Phe-NH, $C_{16}H_{23}N_3O_3S$, 337.1460, found, 360.1379 ($m+Na^+$).

Determination of amino acid amide pKa's

Calculation of the concentration of unprotonated amine in the reaction buffer required measurement of the pKa's of the amino acid amide nucleophiles under the solvent conditions of the study: [NaCl] = 585 mM, [amino acid amide HCl] = 185 mM, 7% DMSO in H_2O. 200 mM KOH was added in measured aliquots and the pH measured with a Mettler Toledo SevenEasy pH meter calibrated with standard buffer solutions. For each titration > 40

observations were made. Inflection points at the beginning and end of the titrations were found by plotting the instantaneous slope of the pH observations vs. the amount of titrant added (Figure 3). The addition of NaCl and DMSO had a minimal effect on pKa based on comparisons with the literature (Table S3 and (9)). The inflection points appear as peaks and the pH at the mid-point between these peaks was taken to be the pKa of the compound. Equation 4 (17) was used to calculate pKa's in D_2O (Table S3).

$$\text{pKa in } D_2O = (\text{pKa in } H_2O - 0.42)/0.929 \quad \text{Eq. 4}$$

The relative pKa's of the amino acid amides correlate with the relative pKa's of the amino groups of the amino acids (pKa of Gly = 9.58, Phe = 9.09, Ala 9.71, Pro 10.47, *N*-Me-Gly = 9.97) reported in the CRC Handbook online.

Effect of pD on reactivity

In a related chemical model for dipeptide formation, the rate of aminolysis was shown to be dependent on pD in a manner dictated by the concentration of free amine, as expected (18). We used a simpler approach to confirm the expected effect of unprotonated amine concentration in our model by performing assays identical to our kinetics assays except that the reactions were allowed to reach equilibrium. One series of reactions varied concentrations of total amine at fixed pD 7.4. The other series of reactions used a fixed concentration of total amine and varied pD by addition of small aliquots (< 0.5% of total reaction volume) of 3 M HCl in 90% D_2O/water or 3 M KOD in 99% D_2O. In the fixed pD assays, the median range in the concentration of unprotonated amine was 3.25-fold. In the varying pD assays, the mean range in the concentration of unprotonated amine was 3.5-fold, and the mean range in concentration of deuteroxide ion was ~10-fold, from pD 7.19 to pD 8.22. The reactions were allowed to run 18 hrs ($\sim t = \infty$), and equilibrium concentrations of aminolysis and hydrolysis products quantified by NMR as per the kinetics assays. The fraction of fMet-NHS converted to dipeptide via aminolysis correlated with concentration of unprotonated amine. The fractions of ester converted by aminolysis were essentially the same regardless of whether unprotonated amine concentrations were varied by changing concentration of total amine at constant pD or by changing pD at constant concentration of total amine (Figure 4).

RESULTS AND DISCUSSION

Our initial experiments were based upon a promising chemical model of dipeptide formation that used aqueous buffer at physiological pH. The electrophile was formyl-Phetrifluoroethyl ester and the nucleophile was glycinamide (Gly-NH_2; the amide neutralizes the unphysiological negative charge of the carboxylic acid; (18)). However, we found that this system could not be adapted readily to other amino acid amide nucleophiles because peptide bond formation was much slower than ester hydrolysis, even at the highest possible concentrations. Presumably Gly-NH_2 was the most reactive of the amino acid amides we attempted because it is the least hindered sterically. Using the trifluoroethyl ester of fMet instead of fPhe resulted in even faster hydrolysis. Schemes for continuous *in situ* activation of fMet were also deemed impractical due to mechanistic complexity.

Next, we turned to carboxylic acid activation by *N*-hydroxysuccinimidyl (NHS) ester formation because of its utility in labeling proteinogenic lysines in aqueous solution and because its steric bulk is more comparable to ribose. fMet-NHS (Figure 1, top left) proved sufficiently soluble in aqueous buffer, sufficiently activated as an electrophile, and sufficiently resistant to hydrolysis (provided that reactions were cooled to 4°C) to enable differentiation of its reaction rate with four of the five nucleophiles we had used on the ribosome. Because *N*-Me-Phe-NH_2 and *N*-Bu-Phe-NH_2 had very low reactivities, the least sterically hindered amino and *N*-

alkyl amino acids possible, Gly-NH_2 and *N*-Me-Gly-NH_2 were added to our set to model the effect of *N*-methylation (Figure 1).

The aqueous component of the buffer was 99.9% D_2O to enable quantification of reaction components by ^{1}H-NMR. Though the kinetic isotope effect may reduce reaction rates, it is expected to be too small (~10%) to alter the order of reactivity across the set of nucleophiles studied here. A small amount of organic solvent (7.5%) was also necessary for solubility and to prevent freezing; this had little effect on amine pKa's (Table S3), and even living cells contain a significant concentration of organics, such as lipids.

The apparent rates of dipeptide formation were plotted against the concentration of total amine (protonated and unprotonated) to incorporate all effects without data manipulation, as is done for ribosomal rates (Figure 5A). There was some curvature in the plots due to base catalysis (detailed in the methods; see left side of Table S1). The order of reactivity was Gly-NH_2 > Phe-NH_2 > Ala-NH_2 > Pro-NH_2 > *N*-Me-Gly-NH_2 > undetectable *N*-Me-Phe-NH_2 and *N*-Bu-Phe-NH_2. This order correlates well with the order of our ribosome-catalyzed reactions when controlling for the effect of the tRNA and codon by using the same tRNA/codon combination (Figure 6; $P < 0.05\%$). Though *N*-Me-Phe-NH_2 incorporation was undetectable by our NMR assay, a mass spectrometry assay enabled an approximation of its incorporation rate as 500 fold slower than Phe-NH_2 (Figure 6; see Materials and Methods). This confirmed the effect of *N*-methylation observed by comparing Gly-NH_2 with *N*-Me-Gly-NH_2 (630 fold; Figure 5A and Table S1). However, the extents of rate changes with different amino acids in the presence and absence of ribosomes did not correlate perfectly, especially for *N*-methylation (Figure 6). Possible explanations include nucleophile orientation preferences when bound to tRNA and the peptidyl transferase center that are not experienced in free solution and/or different chemical and steric effects of the tRNA ester versus our NHS ester substituent. Our chemical model rates in Figure 5A also matched fairly well ribosomal rates with different tRNAs and codons (cognate tRNAs and their various codons: within experimental errors, those ribosomal dipeptide rates are Gly-$tRNA^{Gly}$ = Phe-$tRNA^{Phe}$ = Ala-$tRNA^{Ala}$ > Pro-$tRNA^{Pro}$ (1,3)). The more uniform rates with cognate tRNAs on the ribosome may be due to tuning by the tRNAs to decrease differences in the reactivity of the amino acids (as shown for Pro (3)) and/or a rate-limiting conformational step for non-Pro-$tRNA^{Pro}$ substrates.

We also investigated the chemical principles responsible for the differences in non-enzymatic rates. In physiological buffer, the rate of dipeptide formation is expected to be governed by sterics and also the concentration of unprotonated amine as calculated by the Henderson-Hasselbalch equation (18). The latter expectation was investigated in our system as follows. First, we measured the pKa's of the amino acid amides under the salt and solvent conditions of the dipeptide assay (Figure 3; Table S3). Then, we found that the change in aminolysis yield was the same whether it was produced by changing the concentration of total amine (at constant pH) or by changing the concentration of free amine (at constant total amine concentration) through pH changes calculated from our measured pKa's and the Henderson-Hasselbalch equation (Figure 4). This indicated that pD changes affected aminolysis much more than hydrolysis, and supported the expectation that aminolysis rates are dictated by the concentation of unprotonated amine. Replotting the apparent rates vs concentration of unprotonated amine calculated from our pKa measurements gives much larger rate constants in all cases (Figure 5B; Table S2), indicating that the chemical model rates are dramatically affected by pH under physiological conditions. However, the different pKa's of the amino and *N*-alkyl amino acid amides do not correlate well with the relative rates (Table S3: amino acids are listed in decreasing order of rates from top to bottom, which does not match the order of pKa's; Table S2: rates are non-identical, which should not be the case if pKa was the dominant determinant of the relative rates). When the order of reactivities is examined by neglecting pH/pKa effects (Figure 5B and Table S2), the order of Phe and Ala changes so that the whole order is precisely

that predicted on the basis of nucleophile sterics. With respect to *N*-alkylation effects on reactivity, it is clear that the inhibitory steric effects completely dominate over potentially stimulatory inductive effects. These conclusions are in line with chemical theory that amino nucleophiles are more hindered sterically by direct substituents than by more distal substituents.

Finally, we consider the reactivities of the other six amino acids (Ile, Val, Glu, Leu, Lys and His) shown to have uniform decoding rates (within a two-fold error range) on their cognate tRNAs at the ribosomal A site (1). Of these six amino acids, Lys and His have side chain nucleophiles that are problematic for our fMet-NHS assay, so we selected the other four amino acids to test (Figure 7). The assay was modified to increase throughput by measuring the relative rates of each individual free amino acid in competition with an internal standard nucleophile, *N*-Me-Gly-NH_2. This internal standard was chosen because of its relatively slow reaction rate and the uncomplicated NMR spectrum of its product. As a further control, Phe was also competed with *N*-Me-Gly-NH_2 to allow comparison of results in Figures 5A and 8. The measured chemical reactivities of the natural amino acid nucleophiles are quite similar (Figure 8), consistent with both the uniform decoding hypothesis and the chemical reactivity hypothesis. As with the amino acid amide set in Figure 3, the relative rates do not correlate with the pKa's of the nucleophiles (pKa's are listed in Figure 7 legend, with *N*-Me-Gly-NH_2 and Phe having the *lowest* pKa's). Different steric and inductive properties of the side chains also did not substantially affect reaction rates, at least for the side chains examined together with the fMet-NHS electrophile. However, as with the amino acid amide set in Figure 3, *N*-alkylation sterics was again the main rate-determining factor.

In conclusion, dipeptide formation rates on the ribosome with natural and unnatural amino acid nucleophiles correlate well with chemical reactivities of the nucleophiles under physiological-like conditions in our chemical model. This correlation argues against extending the rate-limiting accommodation hypothesis to natural Pro-$tRNA^{Pro}$ isoacceptors. Rather, our correlation supports the chemical reactivity hypothesis as the explanation for slow translation rates with natural Pro-$tRNA^{Pro}$ isoacceptors, dramatically slower incorporations of other *N*-alkyl amino acids, and the lack of *N*-alkyl amino acids other than Pro in the genetic code. Pro and unnatural amino acids aside, the coupling rates of other amino acids to fMet measured in the chemical model are similar (as they are on the ribosome), which is compatible with both the uniform decoding hypothesis and the chemical reactivity hypothesis. Evaluation of the effects of *N*-substitution sterics, side chain sterics, induction and pKa in the chemical model showed that the dominant factor was *N*-substitution sterics.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Donald Stec for help with NMR, Måns Ehrenberg for discussions and comments on the manuscript, and Brian Bachmann, Craig Lindsley and Charles Sanders for comments on the manuscript.

This research was supported by the National Institutes of Health and the American Cancer Society.

ABBREVIATIONS

AA	amino acid
ACN	acetonitrile

HATU	hexafluorophosphate 2-(7-Aza-1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium
HRMS	high resolution mass spectrometry
NHS	N-hydroxysuccinimide
UPLC	ultra-high pressure liquid chromatography-mass spectrometry.

REFERENCES

1. Ledoux S, Uhlenbeck OC. Different aa-tRNAs are selected uniformly on the ribosome. Mol. Cell 2008;31:114–123. [PubMed: 18614050]
2. Pape T, Wintermeyer W, Rodnina MV. Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the E.coli ribosome. EMBO J 1998;17:7490–7497. [PubMed: 9857203]
3. Pavlov MY, Watts RE, Tan Z, Cornish VW, Ehrenberg M, Forster AC. Slow peptide bond formation by proline and other N-alkylamino acids in translation. Proc. Natl. Acad. Sci. U.S.A 2009;106:50–54. [PubMed: 19104062]
4. Zhang B, Tan Z, Dickson LG, Nalam MN, Cornish VW, Forster AC. Specificity of translation for N-alkyl amino acids. J. Am. Chem. Soc 2007;129:11316–11317. [PubMed: 17718568]
5. Weber AL, Miller SL. Reasons for the occurrence of the twenty coded protein amino acids. J. Mol. Evol 1981;17:273–284. [PubMed: 7277510]
6. Eichler J, Houghten RA. Identification of substrate-analog trypsin inhibitors through the screening of synthetic peptide combinatorial libraries. Biochemistry 1993;32:11035–11041. [PubMed: 8218169]
7. Herman LW, Tarr G, Kates SA. Optimization of the synthesis of peptide combinatorial libraries using a one-pot method. Mol. Diversity 1996;2:147–155.
8. Kim H-S, Chung TD, Kim H. Voltammetric determination of the pKa of various acids in polar aprotic solvents using 1,4-benzoquinone. J. Electroanalytical Chem 2001;498:209–215.
9. Chambers RW, Carpenter FH. On the preparation and properties of some amino acid amides. J. Am. Chem. Soc 1955;77:1523–1526.
10. Schmeing TM, Huang KS, Kitchen DE, Strobel SA, Steitz TA. Structural insights into the roles of water and the 2' hydroxyl of the P site tRNA in the peptidyl transferase reaction. Mol. Cell 2005;20:437–448. [PubMed: 16285925]
11. Trobro S, Åqvist J. Mechanism of peptide bond synthesis on the ribosome. Proc. Natl. Acad. Sci. USA 2005;102:12395–12400. [PubMed: 16116099]
12. Okuda K, Seila AC, Strobel SA. Uncovering the enzymatic pKa of the ribosomal peptidyl transferase reaction utilizing a fluorinated puromycin derivative. Biochemistry 2005;44:6675–6684. [PubMed: 15850401]
13. Satterthwait AC, Jencks WP. The mechanism of aminolysis of acetate esters. J. Am. Chem. Soc 1974;96:7018–7031. [PubMed: 4436508]
14. Koshland DE. Kinetics of peptide bond formation. J. Am. Chem. Soc 1951;73:4103–4108.
15. Notari RE. Hydroxamic acids II: Kinetics and mechanisms of hydroxyaminolysis of succinimide. J. Pharm. Sci 1969;58:1064–1068. [PubMed: 5346066]
16. Salomaa P, Schaleger LL, Long FA. Solvent deuterium (D) isotope effects on acid-base equilibriums. J. Am. Chem. Soc 1964;86:1–7.
17. Krezel A, Bal W. A formula for correlating pKa values determined in D_2O and H_2O. J. Inorg. Biochem 2004;98:161–166. [PubMed: 14659645]
18. Schroeder GK, Wolfenden R. The rate enhancement produced by the ribosome: an improved model. Biochemistry 2007;46:4037–4044. [PubMed: 17352494]

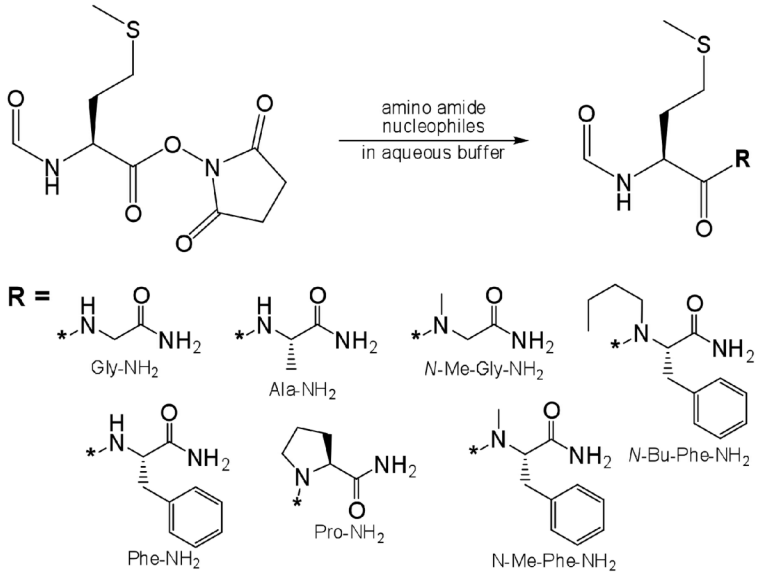

Figure 1.
Amino acid amide nucleophiles reacted with fMet-NHS in physiological-like buffer.

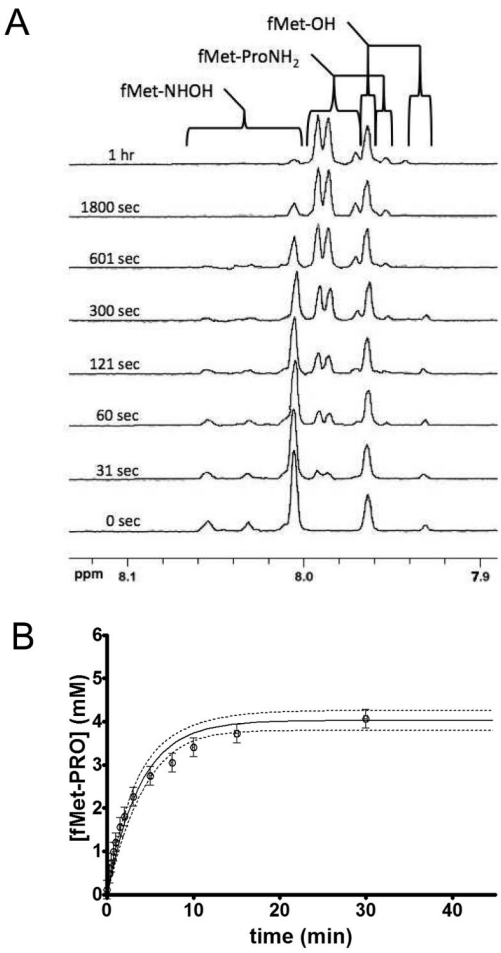

Figure 2.

Representative time course of dipeptide synthesis. **A.** NMR chemical shifts of products from a reaction between 762 mM total prolinamide and 6.7 mM fMet-NHS over time. Signals due to the formyl proton of fMet-Pro (all rotamers) are at 7.99, 7.98, 7.97 and 7.95 ppm. Residual fMet-NHS starting material was quenched with hydroxylamine, yielding fMet-NHOH (formyl protons at and above 8.00 ppm) and fMet-OH (resonances at 7.96 and 7.93). Concentrations of species were determined by integrating resonances with respect to the resonance of pyrazine, which was added to the electrophilic solution as an internal standard (resonance of pyrazine at 8.55 ppm is not shown). Sixteen time points were collected at each concentration, but only every other time point is shown for clarity. **B.** Representative fit of Equation 1 to the time course in **A**. Concentration error bars were determined by measuring differences in the ratio of the internal standard resonance area to the sum of all analyte resonance areas (this ratio should be invariant between time-points).

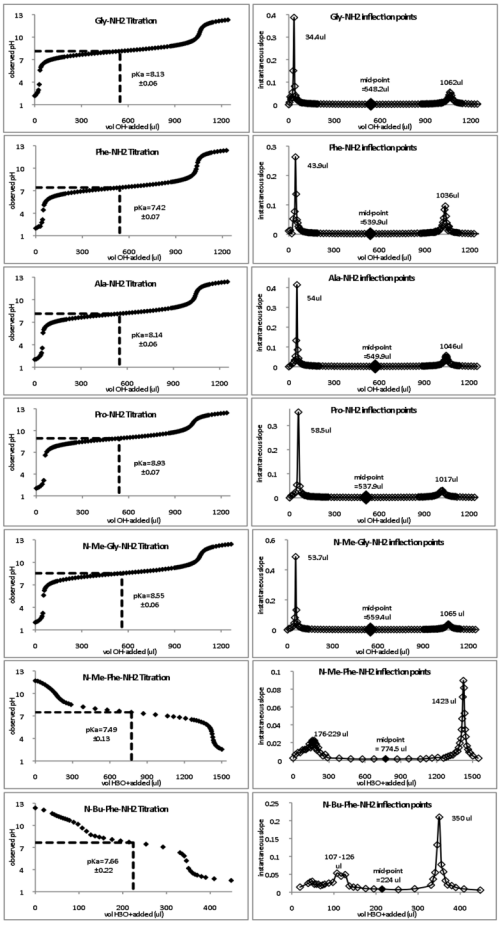

Figure 3.

Measurements of pKa's of amino acid amide nucleophiles. **Left.** pH Titration curves. The pKa was taken to be the pH at the midpoint of the titration. **Right.** Plots of instantaneous slope. The midpoint was determined by finding the inflection points of the titration.

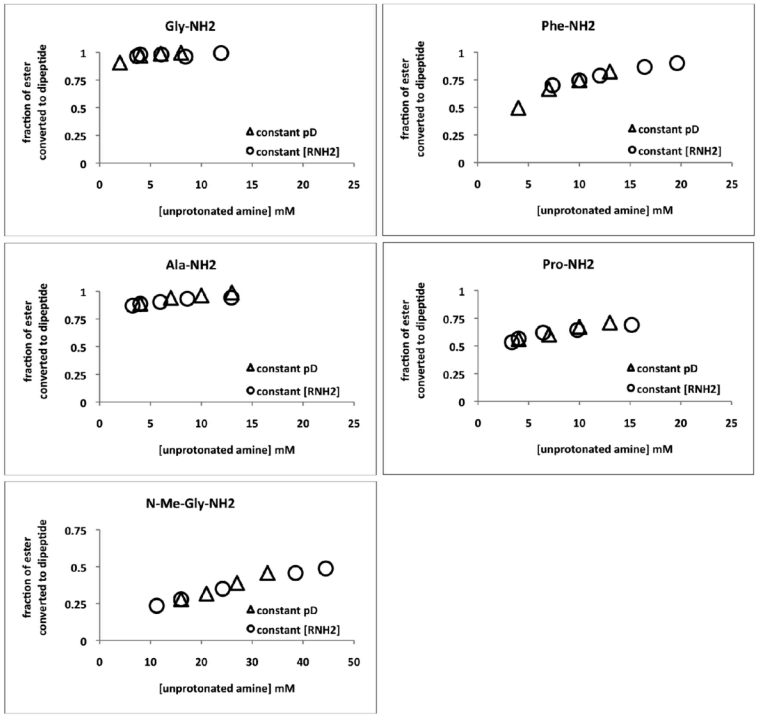

Figure 4.

Fractions of ester converted by aminolysis are unchanged regardless of whether the concentration of unprotonated amine is varied by changing concentration of total amine or by changing pD. Triangles show the effect of varying the concentration of unprotonated amine at a fixed pD. Circles show the effect of varying the concentration of unprotonated amine by changing pD at a fixed concentration of total amine.

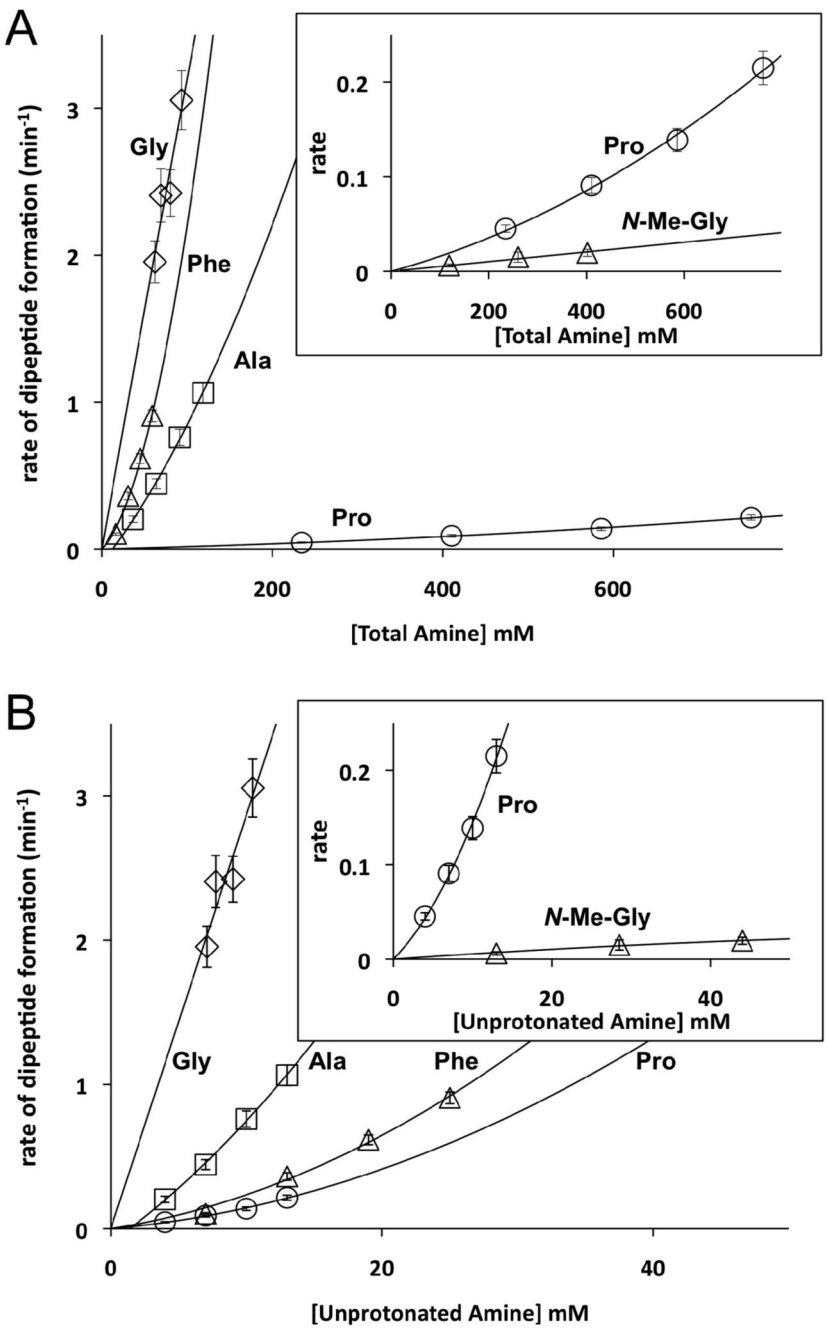

Figure 5.

Rates of dipeptide bond formation between fMet-NHS and amino acid amides in the chemical model of physiological aqueous pH 7.4. **A.** Plots of the apparent rates (min^{-1}) vs concentration of total amine. **B.** Plots of apparent rates vs concentration of unprotonated amine calculated from pKa measurements. Each point represents a time course at that particular concentration, with each error bar indicating a goodness-of-fit of each time course curve. Curvature due to base catalysis is detailed in Materials and Methods and Tables S1 and S2.

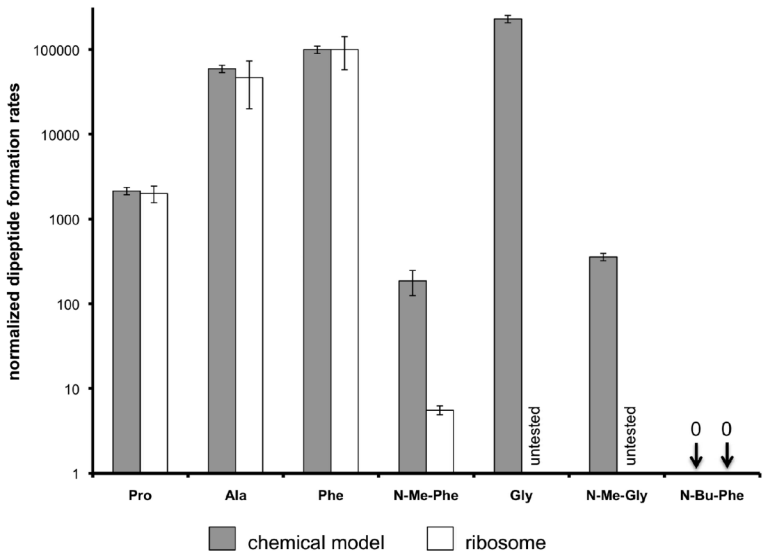

Figure 6.

Comparison between the relative rates of reaction in the chemical model system with those using $tRNA^{PheB}$ adaptor in the ribosomal system. The ribosomal reaction rates ($k_{acc,pep}$) do not include the GTP hydrolysis rates of EF-Tu (k_{GTP}) (3). Phe rates in both systems are normalized to 100,000 and the other rates are multiplied by the same factors and plotted on a log scale. Error bars indicate the 95% confidence interval (2 standard deviations).

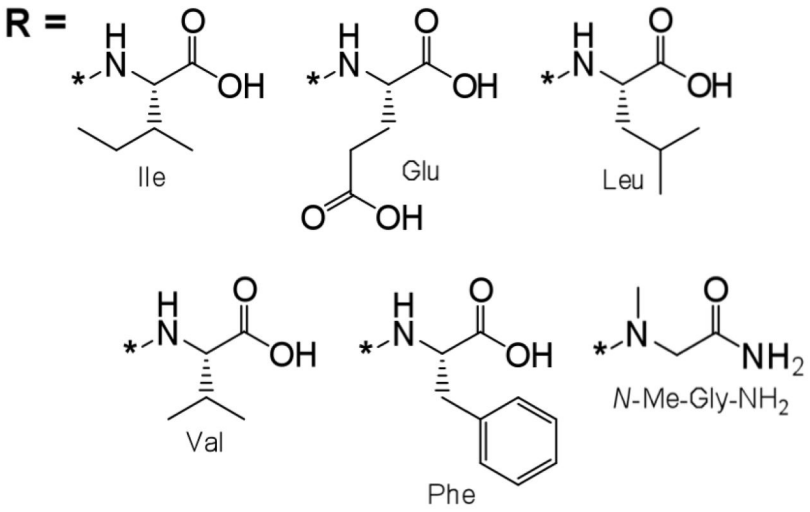

Figure 7.
Amino acid nucleophiles reacted with fMet-NHS in competition reactions with *N*-Me-glycinamide. The pKa's of the nucleophiles are Ile 9.60, Val 9.52, Glu 9.58, Phe 9.09, Leu 9.58 (CRC Handbook online) and *N*-Me-Gly-NH_2 8.55 (from Figure 3).

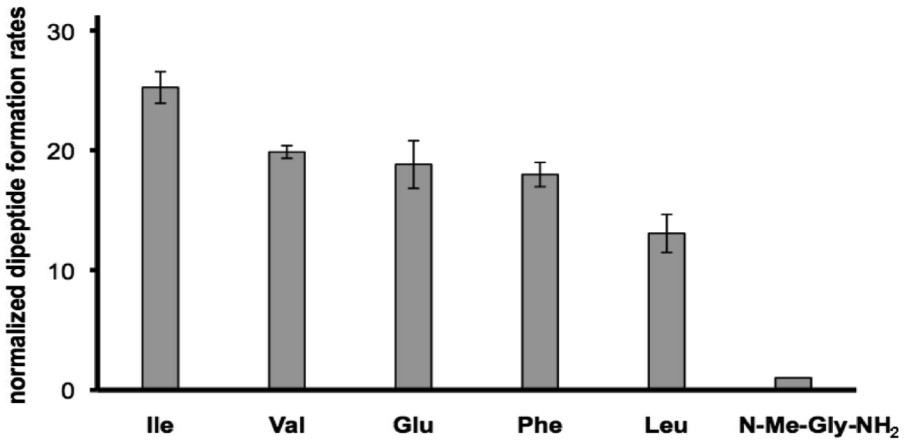

Figure 8.

Relative reaction rates of six nucleophiles. The rate for the standard competitor in all assays, *N*-Me-glycinamide, is normalized to 1.0 and the rates for the five natural amino acids (free acid forms) are multiplied by the same factor. Error bars indicate the 95% confidence interval (2 standard deviations).