RESEARCH ACCOMPLISHMENTS

Nature readily creates and utilizes chemical diversity for the evolution of potent natural products, enzymes with new functions, and even complex systems. Rather than compete with Nature, my laboratory looks to co-opt biological systems to synthesize and evolve chemical diversity by bringing together modern methods in chemical synthesis and DNA technology. The last century saw a revolution in our understanding of the reactivity of small molecules and ability to synthesize small molecules of defined molecular structure, realized as the modern drug industry. My research aims to bring this level of control and understanding to complex biological systems. Manipulation of these biological systems should not only allow us to make new and useful materials on a whole new scale, but also provide fundamental insight into the mechanism of these complex biological systems. Our long-term goal is to understand protein function at the molecular level, looking at isolated proteins in solution, large protein complexes, and finally protein function in biological networks in living cells.

Chemical Complementation. Advances in computation and directed evolution hold promise for being able to understand and design protein catalysts with tailor-made structures and functions. Such proteins could be used as materials, reagents, and even therapeutics. The question of how a protein's primary amino acid sequence dictates its three dimensional fold and function is not resolved and is of fundamental importance to our understanding of living systems and the design and synthesis of higher-order structures. Directed evolution attempts to recapitulate the natural evolution of proteins with new structures and functions, but on an experimentally accessible timescale. Genetic methods have the advantage of DNA encoding, but are limited to the repertoire of chemistry used by nature. Here we have sought to combine the advantages of genetic assays with the flexibility of synthetic chemistry by linking enzyme catalysis to traditional genetic assays for reporter gene transcription via small molecules. The genetics allows us to use DNA encoding, and the small molecule chemistry allows us to readily extend this assay to new chemical reactions. Currently, we are using directed evolution both to ask fundamental questions about the molecular basis for enzyme catalysis and to engineer enzymes with new and useful properties.

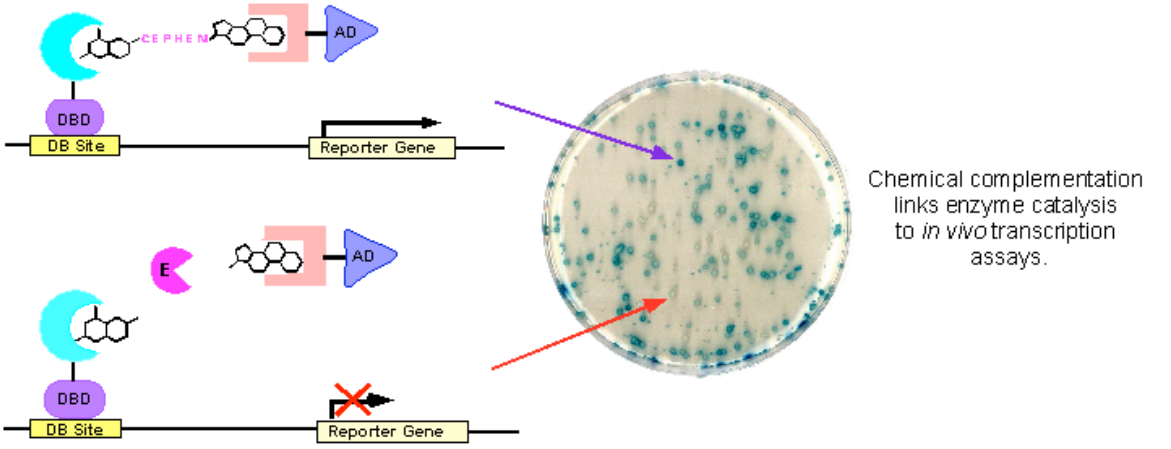
***Ribosome Chemistry*.** The ribosomal biosynthetic machinery, a large complex of protein and RNA, is among Nature's most sophisticated biosynthetic machineries. The ribosome in essence allows template-encoded synthesis of polymers of defined length and composition. Unlike in most biosynthetic machines, substrate recognition is separate from the catalytic center in the ribosome, suggesting it may be particular tolerant to substrate manipulation. Here, our goal is to extend efforts to use synthetic aminoacyl-tRNAs to expand the genetic code, instead to read-out the 64 natural codons with artificial substrates using a purified translation system. This project is in collaboration with Prof. Steve Blacklow and Dr. Tony Forster at Harvard Medical School. Currently, we are using this system for ribosome display of peptidomimetics and to test the adaptor hypothesis, one of the fundamental tenets of translation.

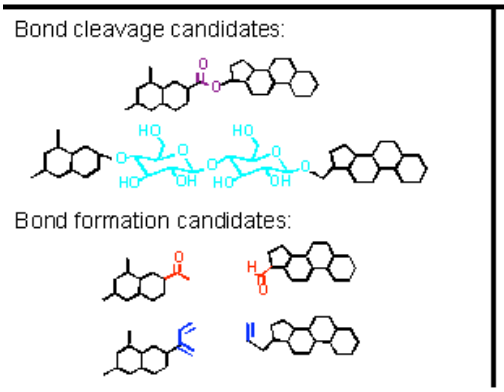
In Vivo Labeling. Finally, in collaboration with Prof. Mike Sheetz in Columbia's Biological Sciences Department, we are developing methods for selectively labeling proteins with small molecules inside the cell. The short-term goal of this project is to provide chemical surrogates to GFP for multi-color tagging and FRET applications. The long-term goal is to extend the power of synthetic chemistry to living systems.

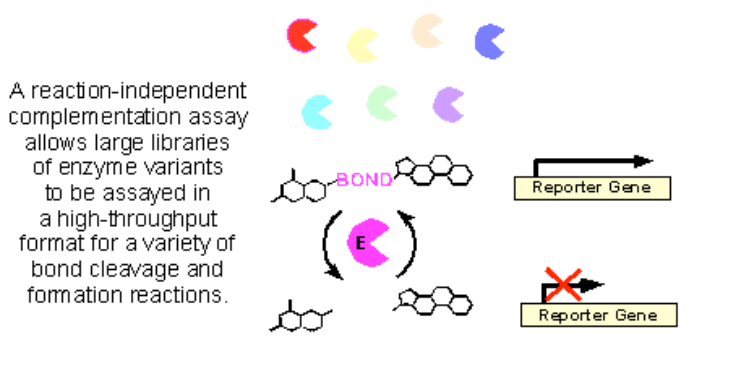
PROGRESS IN THREE CURRENT RESEARCH AREAS

I. CHEMICAL COMPLEMENTATION

Overview. Directed evolution offers a new approach for manipulating and understanding protein function. Directed evolution involves generating large pools of protein variants and then assaying these variants *en masse* for the desired function. While a powerful tool, directed evolution is restricted to enzymes that are inherently screenable or selectable--for example, enzymes where the product is fluorescent or an essential metabolite. Thus, at the start of our research program, my laboratory has focused on developing a general, high-throughput assay for enzyme catalysis based on the yeast three-hybrid assay ("Chemical Complementation") that should allow directed evolution to be applied to a broad range of chemical reactions. This assay detects enzyme catalysis of bond formation or bond cleavage reactions based on covalent coupling of two small molecule ligands *in vivo*. The heterodimeric ligand reconstitutes a transcriptional activator, turning on transcription of a reporter gene. Bond formation is detected as activation of an essential reporter gene; bond cleavage, repression of a toxic reporter gene. The assay is high-throughput because it can be run as a growth selection where only the cells containing functional enzyme survive. The assay can be readily extended to new chemistry simply by synthesizing dimeric ligands with different substrates as chemical linkers. Such a general assay should find broad use not only in directed evolution, but also proteomics, drug discovery, and enzymology.

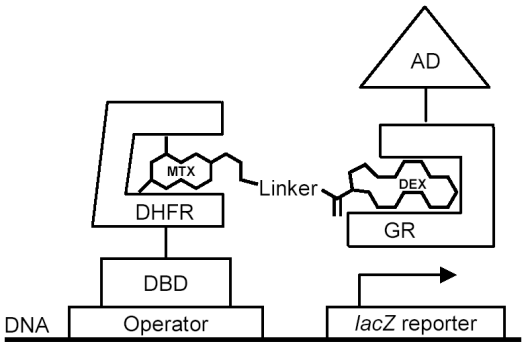






First, we reported the small molecule transcription assay that is the basis of chemical complementation. This work was reported in *J. Am. Chem. Soc.* and featured in *Chem. & Eng. News*. Then, we demonstrated chemical complementation using a well-studied enzyme-catalyzed reaction, β-lactam hydrolysis by a β-lactamase. This work was published in *Proc. Natl. Acad. Sci. USA* and featured in both *Proc. Natl. Acad. Sci. USA* and *Chem. & Eng. News*. The long-term goal is to apply this assay to directed evolution, enzymology, drug discovery, and proteomics. Currently we are focused on two applications. The first, a basic science application, to understand the molecular basis for the difference in chemical reactivity between two proteins believed to be evolutionarily related, a penicillin-binding protein and a β-lactamase. The second, to use directed evolution to engineer glycosynthase enzymes that can be used for the synthesis of carbohydrates. Recently, we have completed our first efforts in both areas.

Dexamethasone-Methotrexate Yeast Three-Hybrid Transcription Assay. [Published in H. Lin, W. Abida, R.T. Sauer, **V.W. Cornish**. "Dexamethasone-Methotrexate: An Efficient Chemical Inducer of Protein Dimerization *In Vivo*." *J. Am. Chem. Soc.*, **122**, 4247-4248 (2000). Featured in *Chem. & Eng. News*, **78**, 52 (2000).].

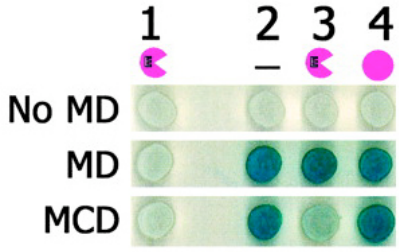


Yeast three-hybrid assay. An operator that binds the DBD is placed upstream of a *lacZ* reporter gene. The Dex-Mtx CID dimerizes the DBD-DHFR and AD-GR protein chimeras, effectively reconstituting the transcriptional activator (DBD-AD) and activating transcription of the *lacZ* reporter gene.

The first step to building the chemical complementation assay was to design a heterodimeric ligand that could efficiently reconstitute a transcriptional activator *in vivo*. Essentially what this required was that we develop ligand-receptor pairs other than FK506 or FK506 analogs that could be used as "Chemical Inducers of Dimerization" (CIDs). Methotrexate (Mtx) and dihydrofolate reductase (DHFR) were chosen because of the ease with which Mtx analogs can be synthesized and the high affinity of the Mtx-DHFR interaction (K_D = 10 pM). Dexamethasone (Dex) and the hormone-binding domain of the glucocorticoid receptor (GR) had been reported previously. We demonstrated the efficacy of the Dex-Mtx CID based on activation of a *lacZ* reporter gene in a yeast three-hybrid assay. The ease of synthesis of Dex-Mtx allows the linker between Dex and Mtx to be changed readily, and, thus, chemical complementation to be applied to a wide range of chemical reactions. Proof of the advantages of this CID, GPC Biotech just reported in *Chem. Biol.* the use of our Mtx yeast three-hybrid system to discover both known and novel protein targets of cyclin-dependent kinase inhibitors.

Chemical Complementation: Proof-of-Principle with a β-Lactamase Enzyme. [K. Baker, C. Bleczinski, H. Lin, G. Salazar-Jimenez, D. Sengupta, S. Krane, **V.W. Cornish**. "Chemical Complementation: A Reaction-Independent, High-Throughput Genetic Assay for Enzyme Catalysis." *Proc. Natl. Acad. Sci. USA,* **99**, 16537-16542 (2002). Featured in a commentary in *Proc. Natl. Acad. Sci. USA,* **99**, 16513-16515 (2002) and in *Chem. & Eng. News*, **81**, 24, (2003).].

The well-studied enzyme-catalyzed reaction of cephalosporin hydrolysis by the P99 cephalosporinase was used to develop and demonstrate the chemical complementation assay. A Dex-Cephem-Mtx CID with the cephem substrate as the linkage between Dex and Mtx was synthesized. Using standard *lacZ* transcription assays, transcription activation was shown to be turned off by expression of the cephalosporinase enzyme, presumably because the enzyme catalyzes hydrolysis of the cephem bond. To confirm that the change in transcription of the reporter gene was in fact caused by enzymatic turnover of the Dex-Cephem-Mtx substrate, *lacZ* transcription was shown to be



***LacZ* transcription assay of cephalosporinase-dependent Dex-Cephem-Mtx-induced *lacZ* transcription.** Yeast strains containing a *lacZ* reporter gene were grown on X-gal indicator plates with or without Dex-linker-Mtx molecules as indicated. Columns 1-4 correspond to yeast strains containing a DBD fusion protein, a AD fusion protein, and enzyme, as follows: 1, DBD-DHFR, AD, P99 cephalosporinase. 1 lacks GR and is used as a negative control. 2, DBD-DHFR, AD-GR, no enzyme; 3, DBD-DHFR, AD-GR, P99 cephalosporinase. 4, DBD-DHFR, AD-GR, P99:S64A cephalosporinase. The rows correspond to individual X-gal plates, which have different small molecules as indicated: No MD, No Dex-Mtx; MD, 1 μM Dex-Mtx; MCD, 10 μM Dex-Cephem-Mtx.

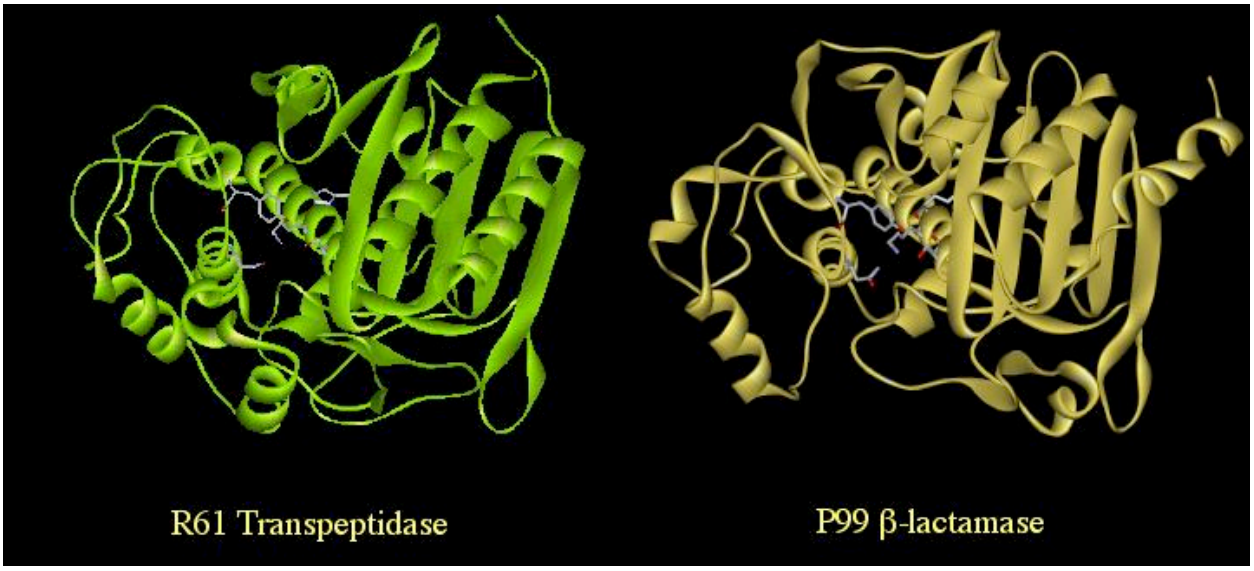
unaffected by expression of an inactive P99:S64A mutant. Finally, a *lacZ* screen was used to isolate the wild-type cephalosporinase from a pool of inactive variants.

Further, we have shown that the levels of *lacZ* transcription correlate with the catalytic efficiency of a series of cephalosporinase mutants [D. Sengupta *et al.* (V.W. Cornish), *Biochemistry,* **43**, 3570-3581 (2004)]. The *lacZ* screen has been used to investigate the mechanism of cephalosporinase resistance to third-generation cephalosporin antibiotics [B. Carter and V.W. Cornish), *unpublished results*].

Molecular Basis for the Difference in Reactivity Between Penicillin-Binding Proteins and β-Lactamases. [S. Goldberg, W. Iannuccilli, T. Nguyen, J. Ju, **V.W. Cornish**. "Identification of Residues Critical for Catalysis in a Class C β-Lactamase by Combinatorial Scanning Mutagenesis." *Protein Science,* **12**, 1633-1645 (2003). B.F. Gherman, S.D. Goldberg, **V.W. Cornish**, R.A. Friesner. "Mixed Quantum Mechanical/Molecular Mechanical (QM/MM) Study of the Deacylation Reaction in a Penicillin Binding Protein (PBP) Versus in a Class C β-Lactamase." *J. Am. Chem. Soc., accepted.* S. Goldberg, **V.W. Cornish**. "Reenacting the Evolution of a β-Lactamase from a Penicillin-Binding Protein." *in preparation.* Featured in *Chem. & Eng. News*, **81**, 35-36, 38-40 (2004).].

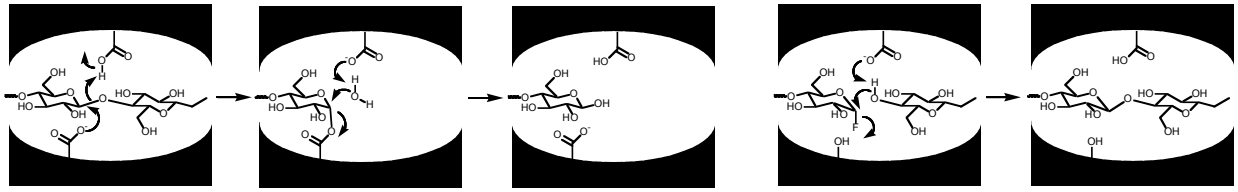
Directed evolution offers a new tool for understanding protein structure and function by allowing us to watch proteins acquire new function through mutation of the amino acid sequence. Here, in collaboration with the Friesner group at Columbia, we are combining directed evolution with computational studies to understand the difference in reactivity between two proteins believed to be evolutionarily related. Penicillin-binding proteins (PBPs) are the targets of β-lactam antibiotics, and β-lactamases are the bacterial resistance enzymes that hydrolyze and inactivate these antibiotics. β-Lactamases are believed to have evolved from an ancestral PBP. Interestingly, the two proteins have conserved three-dimensional structures and active-site residues. Both are serine-protease type enzymes. PBPs are inactivated by β-lactams, while β-lactamases turn them over, because PBPs are inefficient catalysts of acyl-enzyme hydrolysis. The rate constant for acyl-enzyme hydrolysis differs by six-orders of magnitude for these two proteins with typical substrates. QM/MM calculations of the

ground states and transition states for catalysis of this step by the R61 transpeptidase PBP and the P99 β-lactamase suggest that the difference in reactivity stems from the ability of the active-site Tyr general base to interact with a stabilizing hydrogen-bonding network in the acyl-enzyme complex. "Isostere Scanning" of the P99 β-lactamase and directed evolution of the PBP support this hypothesis.



Engineering Glycosynthases for Use in Carbohydrate Synthesis. [H. Lin, H. Tao, **V.W. Cornish**. "Directed Evolution of a Glycosynthase Via Chemical Complementation." *submitted.*].

Despite their fundamental role in biological processes and potential use as therapeutics, it still remains difficult to synthesize carbohydrates. Enzymes, with their control of both regio- and stereo-chemistry, provide an obvious alternative to traditional small molecule chemistry for carbohydrate synthesis. Recently, Withers and co-workers demonstrated that retaining glycosidases can be re-engineered to glycosynthases simply by mutating the nucleophilic Glu residue at the base of the active site to a small hydrophobic residue and using an α-fluoro donor. While several retaining glycosidases have now been reported, these glycosynthase variants are not sufficiently active for use on a



Retaining Glycosidase **Glycosynthase**

Conversion of a glycosidase to a glycosynthase. A glycosidase has two acidic residues in the active site, one acting as the general acid/base, and the other acting as the nucleophile to attack the anomeric position where cleavage of the glycosidic bond is going to occur. If the nucleophilic acidic residue is changed to a small hydrophobic residue, such as a Ser or Ala, the protein will not be able to hydrolyze glycosidic bonds since it lacks the nucleophile. But it can act as a glycosynthase, accepting a glycosyl fluoride as the glycosyl donor and forming glycosidic bonds with suitable glycosyl acceptors.

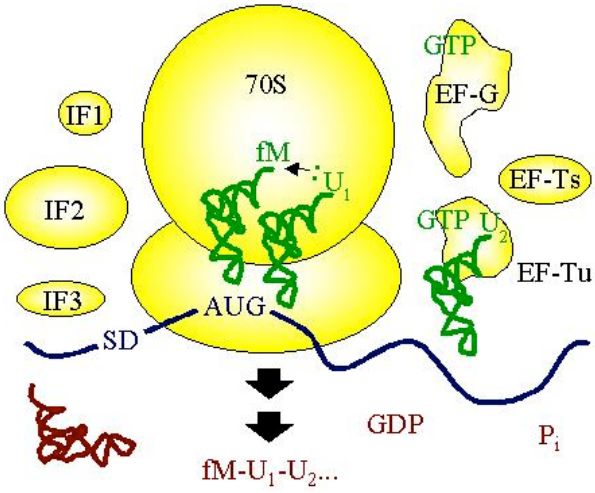
preparative scale. Directed evolution should offer a powerful approach for improving the activity of these enzymes and even modifying their substrate specificity; however, there are no natural selections for glycosynthase activity. Chemical complementation should open up directed evolution to carbohydrate chemistry. In this first paper, we showed that chemical complementation can link Cel7B:E197A glycosynthase activity to *LEU2* transcription activation *in vivo* using Dex disaccharide acceptor and Mtx disaccharide α-fluoro donor substrates. The *LEU2* selection was then used for an E197 saturation library, yielding mutations that increased the activity of the glycosynthase and that increased the expression levels of the protein.

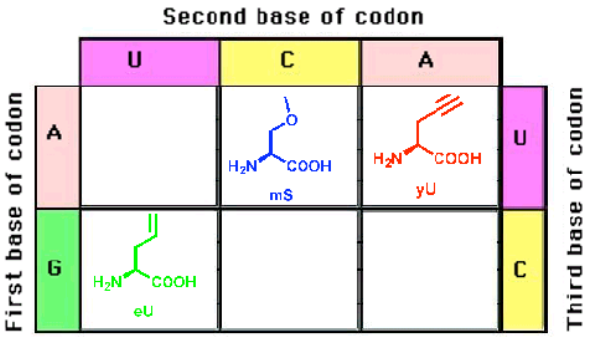
Ligand-Receptor Binding. [K. deFelipe, E. Althoff, B. Carter, **V.W. Cornish**. "Correlation Between the Dissociation Constant and the Transcription Read-Out in the Yeast Three-Hybrid Assay." *Biochemistry, accepted.*].

In addition to catalysis, we are exploiting the yeast three-hybrid assay to study ligand-receptor binding. It was proposed that the yeast three-hybrid assay could be used to identify the protein targets of small molecule drugs. The high-affinity of the Mtx/DHFR CID has improved the sensitivity of this assay and allowed us to go beyond initial proof-of-principle experiments. We have shown that the Mtx yeast three-hybrid system can detect small molecule-protein interactions with dissociation constants up to ca. 100 nM. In collaboration with Prof. Young-Tae Chang at New York University, we have applied the Mtx yeast three-hybrid system to an important pharmaceutical target, cyclin-dependent kinases (*unpublished data*).

II. RIBOSOME CHEMISTRY

Overview. The ribosome is a large complex of RNA and protein that catalyzes the translation of RNA into protein. The ribosome is impressive in its ability to convert between nucleic acid and peptide, two structurally unrelated polymers. In addition, unlike most natural enzymes, the ribosome shows broad substrate specificity and uses amino acids with diverse chemical side chains. The adaptor hypothesis suggests that the amino acid specificity of the ribosome should be entirely determined by the tRNA anticodon and independent of amino acid structure. Several laboratories have now shown in fact that suppressor tRNAs and unnatural base pairs can be used to add a "21st amino acid" to the genetic code. Rather than adding on to the genetic code, we set out to show that the genetic code could instead be "rewritten" using synthetic aminoacyl-tRNAs and a purified *in vitro* translation system. This project is in collaboration with Prof. Steve Blacklow and Dr. Tony Forster at Harvard Medical School. We have now completed a proof-of-principle for this strategy, demonstrating that we can synthesize a peptide containing three different unnatural amino acids in a row, each coded by a different, natural RNA codon. This research was published this year in *Proc. Natl. Acad. Sci. USA* and featured in *Chem. Biol.* The long-term goal of this project is to study the mechanism of translation and to manipulate the ribosome to catalyze polymers other than polypeptides. Currently, we are focused on adapting the purified translation system for ribosome display of peptidomimetics and testing the adaptor hypothesis.

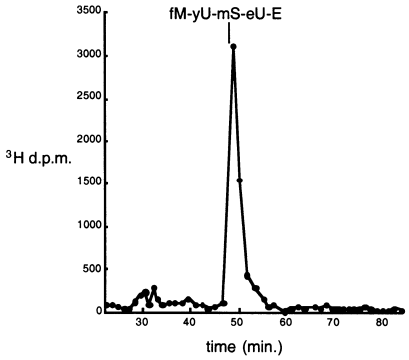




Miniature Genetic Code. [Published in A. Forster, Z. Tan, M. N. L. Nalam, H. Lin, H. Qu, **V.W. Cornish**, S. Blacklow. "Programming peptidomimetic syntheses by translating genetic codes designed *de novo*." *Proc. Natl. Acad. Sci. USA,* **100**, 6353-6357 (2003). Featured in *Chem. Biol.,* **10**, 586-587 (2003) and in *Chem. & Eng. News*, **82**, 64-68 (2004).]

A purified translation system should allow all 64 codons to be recognized by synthetic aminoacyl-tRNAs because there are no contaminating aminoacyl-tRNA synthetases that would otherwise proofread and recharge the tRNAs with the natural amino acids. To test this hypothesis, the purified translation system was used to synthesize a short peptide with three unnatural amino acids in a row--each coded by a different natural codon (Asn, Thr, and Val). Translation reactions showing that product formation was dependent on all three synthetic aminoacyl-tRNAs established that there was no cross-

reactivity. Co-migration of radiolabeled translation product with authentic peptide prepared by solid-phase synthesis on C18 reverse-phase HPLC further confirmed these results. This experiment demonstrates the feasibility of using different synthetic aminoacyl-tRNAs as substrates for all 64 codons in the genetic code.



Sequential incorporation of multiple different unnatural amino acids by redesigning multiple codons. Arbitrarily chosen codons (N, T, and V) were redefined in a rational manner to encode unnatural amino acids of arbitrary choosing by mutating the anticodon of $tRNA^{AsnB}$. The template normally encoding fMNTVE was then synthesized to test for the adjacent incorporation of three different unnatural amino acids using the appropriate synthetic adaptors. HPLC analysis of the radiolabeled translation products mixed with authentic unlabeled marker peptide (fM-yU-mS-eU-E synthesized on an Applied Biosystems synthesizer from commercial reagents) and analyzed by C18-reverse phase HPLC.

Ribosome Display. [A.C. Forster, **V.W. Cornish**, S.C. Blacklow. "Pure Translation Display." *submitted.*]

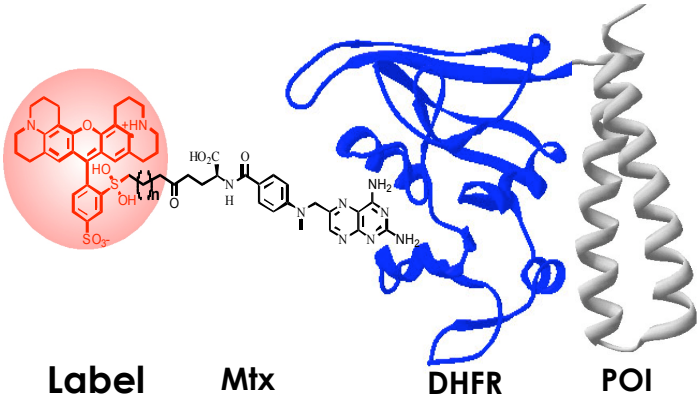
Ribosome display allows very large libraries (10^{12}-10^{15}) of peptides to be synthesized and selected for function because the peptides can be linked to their unique RNA sequences *in vitro*. Here we use the pure translation system for ribosome display, selecting for a biotinyl-amino acid via avidin beads.

Mechanistic Investigations. [Z. Tan, A. Forster, S. Blacklow, **V.W. Cornish**. "Backbone Specificity of the *E. coli* Ribosomal Biosynthetic Machinery." *submitted.*]

The adaptor hypothesis predicts that aa-tRNA decoding on the ribosome depends solely on the tRNA anticodon and is independent of the amino acid structure. Here we look at the ability of the *E. coli* ribosomal biosynthetic machinery to accept aa-tRNA substrates with different amino acid analogs, but the same tRNA body and anticodon as a test of this hypothesis.

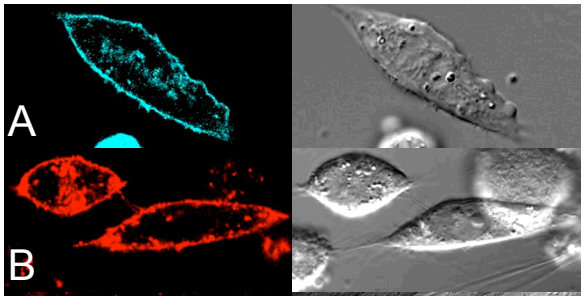
III. *IN VIVO* LABELING

Overview. Green fluorescent protein (GFP) and its derivatives have emerged as invaluable tools for tagging and visualizing proteins *in vivo*. GFP can be appended to the protein of interest simply by engineering the GFP fusion at the DNA level using standard molecular biology techniques. The localization, interaction and fate of the labeled protein can then be monitored in real time in intact cells using live cell imaging. While in theory the fluorescent properties of GFP can be modulated by mutating the protein sequence, in practice only limited changes and improvements to GFP have been possible because the chromophore is limited by the protein sequence. The difficulty in engineering well-behaved GFP variants with distinct absorption and emission maxima has hindered multi-colored tagging and FRET applications. Long before GFP was employed for protein labeling, proteins were labeled *in vitro* with small molecule fluorophores. For example, a unique Cys residue could be engineered on the surface of a protein and then labeled with a thiol-reactive small molecule. The advantage of labeling a protein with a small molecule is that the fluorescent properties of the small molecule can be readily varied and different types of labels, such as a photoaffinity label, can be employed. The problem, however, in extending this small molecule approach to *in vivo* labeling is that it is difficult to selectively label the protein of interest in the sea of proteins and other reactive species present in the cell. What is needed are approaches that combine the ability to genetically encode the label as for GFP with the flexibility of small molecule labels. In collaboration with the Sheetz laboratory in the Biological Sciences Department at Columbia, we are exploiting the high-affinity interaction between Mtx and DHFR to label proteins *in vivo* by fusing the protein of interest to DHFR and then labeling the protein with small-molecule Mtx conjugates. Recently, we have demonstrated the feasibility of this approach, labeling both a plasma membrane and a nuclear protein fused to DHFR with Mtx-Texas Red in Chinese Hamster Ovary (CHO) cells. This work has been published in *Angew. Chem.* Currently, we are focused on engineering orthogonal Mtx/DHFR variants for selectively labeling multiple proteins in the same cell and using the Mtx conjugates for applications beyond simple fluorescent labeling.



Methotrexate-Texas Red as an RFP Surrogate. [Published in L. Miller, J. Sable, P. Goelet, M. Sheetz, **V.W. Cornish**. "Methotrexate Conjugates: A Molecular *In Vivo* Protein Tag." *Angew. Chem. Int. Ed.,* **43**, 1672-1675 (2004).]

The goal of this work was to show that a fluorescent conjugate of Mtx could be used to label an *E. coli* DHFR fusion protein in cultured mammalian cells. DHFR (-/-) Chinese Hamster Ovary (CHO) cells were grown on coverslips, transfected with DNA encoding either plasma membrane (PM) or nuclear-localized DHFR, incubated in a medium containing Mtx-Texas Red (Mtx-TR), washed and imaged with a confocal fluorescent microscope. Mtx-TR efficiently labeled the fusion proteins, and the labeling could be competed out with an excess of free Mtx. This work establishes the feasibility of using the noncovalent interaction between Mtx and DHFR to label proteins *in vivo* with a wide variety of Mtx small molecule conjugates.



Non-covalent labeling of a DHFR tagged plasma membrane protein in DHFR-deficient CHO cells. (A-B) Confocal micrographs show fluorescence in left column, DIC images in right column. (A) excitation, 457nm. (B) excitation, 565 nm. Confocal micrograph (A) shows PM fluorescence in cells transiently expressing cyan fluorescent protein fused to PM-targeting sequence. (B) DHFR-deficient CHO cells transiently expressing PM-targeted DHFR. Cells were incubated in media containing 2 μM Texas Red™-conjugated methotrexate for 20 hrs., washed with PBS, mounted in media without Mtx-TR and imaged.

Orthogonal Methotrexate Tag. [L. Miller, M. Sheetz, **V.W. Cornish**, *unpublished data*.]

To allow labeling in a wide variety of cell types, a methotrexate-DHFR variant is needed that has no cross-reactivity with the natural cellular machinery. Toward this end, we are using a “bump-hole” strategy to engineer an orthogonal methotrexate-DHFR pair.

Virginia W. Cornish
Assistant Professor
Department of Chemistry
Columbia University
June 2, 2004