Site-Specific Incorporation of Unnatural Amino Acids into Proteins

Anne Strømgaard,[b] Anders A. Jensen,[a] and Kristian Strømgaard*[a]

1. Introduction

Rational modification of protein structure and function is still a discipline in its infancy. Conventional mutagenesis is limited to the 20 natural amino acids as building blocks. The introduction of unnatural amino acid mutagenesis has expanded the repertoire of amino acids that can be incorporated into proteins considerably. Site-specific incorporation of unnatural amino acids has wide-ranging applications in the structure–function studies of proteins, including studies of protein stability and folding, protein–ligand and protein–protein interactions. The technique allows incorporation of biophysical probes, for example, fluorescent and photolabile probes, into proteins, as well as backbone modifications, such as the replacement of amide bonds with ester bonds. This has been used for a number of structure–function studies of proteins in prokaryotic and eukaryotic cells, both in vitro and in vivo. These studies include soluble proteins as well as membrane-bound proteins. In the following article we will give an introduction to protein synthesis and an overview of the techniques currently available for site-specific incorporation of unnatural amino acids into proteins. The advantages and limitations of the techniques will be discussed, and case studies of the modification of membrane-bound proteins will be presented.

2. Protein Synthesis

The genetic code is contained in deoxyribonucleic acid (DNA) and consists of 64 three-base codons that encode a total of 20 amino acids. Three of the 64 codons, UAG (amber), UGA (opal), and UAA (ochre), are termination codons, alternatively called stop or nonsense codons. The DNA molecule itself has not been associated with any specific cellular function but it is transcribed (copied) into messenger ribonucleic acid (mRNA), which subsequently is translated into proteins.

Protein synthesis (translation) requires transfer RNA (tRNA), which carries an amino acid to the ribosome, where protein synthesis takes place (Scheme 1). The tRNA is charged with an amino acid corresponding to its anticodon (three nucleobases complementary to the codon of the mRNA). The charging of each amino acid is carried out by specific aminoacyl-tRNA synthetases (aaRSs) forming aminoacyl-tRNA (aa-tRNA). Consequently, an aaRS exists for each amino acid.[1,2] Although anticodons of tRNAs are general for all organisms, aaRSs are often species-specific, that is, aaRSs from one species do not aminoacylate tRNA from another species.[3] This orthogonality is useful in relation to the incorporation of unnatural amino acids into proteins.

Translation of mRNA into proteins is carried out by the ribosome in four steps: initiation, elongation, termination, and recycling. In the first step, an initiation complex is formed between the ribosome, mRNA, initiator tRNA (a tRNA charged with methionine), initiation factors, and guanosine triphosphate (GTP; Scheme 1). In the second step, an aa-tRNA is transported by an elongation factor (EF) to the mRNA in the ribosome, where the anticodon of the aa-tRNA is matched against the codon of the mRNA. Cognate interaction between anticodon and codon leads to peptide bond formation between the newly arrived amino acid and the methionine in the ribosome, thereby generating a dipeptide. The elongation cycle proceeds until a stop codon in the mRNA is encountered, which triggers association of a release factor (RF) to the ribosome. The RF–ribosome interaction leads to hydrolysis of the ester bond between the C-terminal amino acid of the peptide chain and the last aa-tRNA encountered by the ribosome. Consequently, the polypeptide chain is released from the ribosome. During the final step of protein synthesis, the ribosome is recycled for another round of protein synthesis (Scheme 1).

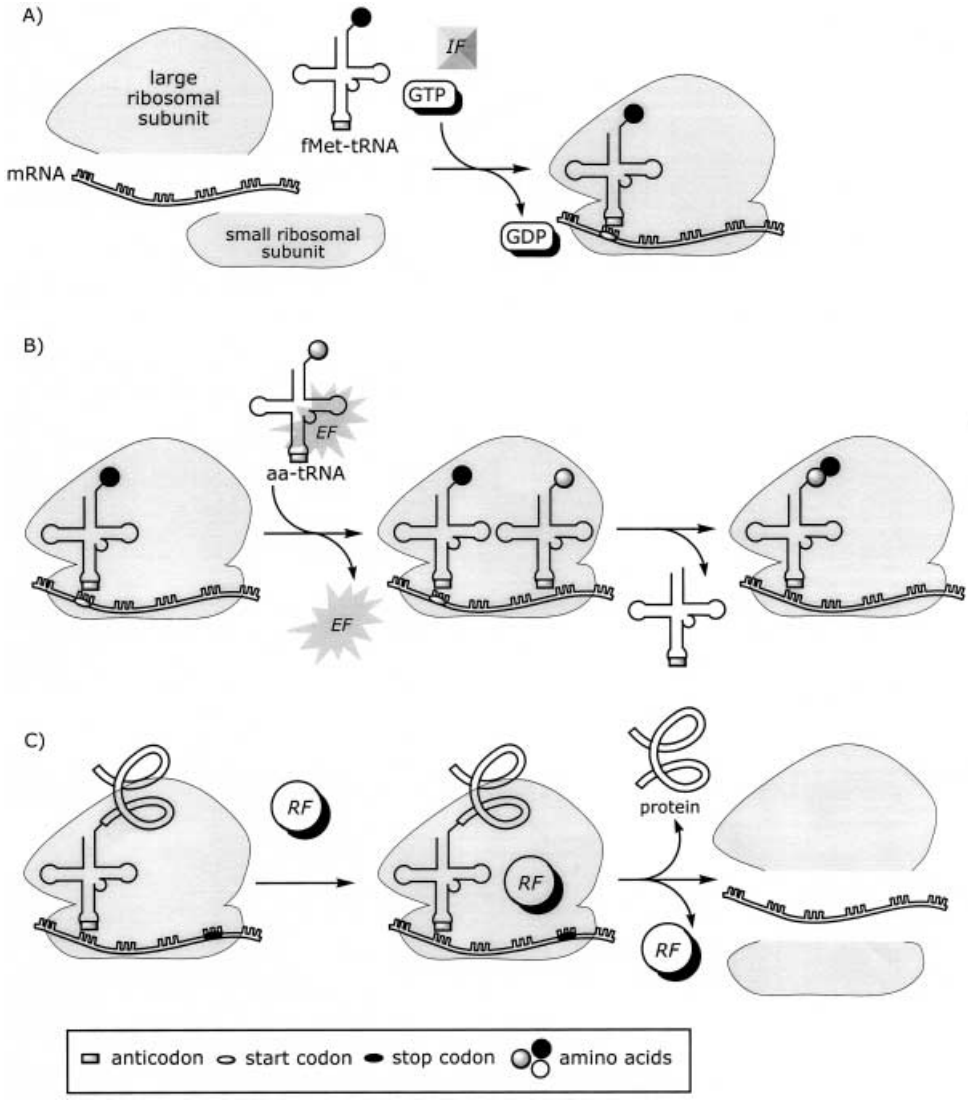
3. Strategies

Protein synthesis allows several approaches to site-specific incorporation of unnatural amino acids into proteins. Specific tRNAs have been charged with an unnatural amino acid, and unnatural base pairs that code for unnatural amino acids have been introduced.[4–8] Methods for incorporation of close analogues of natural amino acids have existed for many years but these methods are beyond the scope of this review, just as methods for residue-specific incorporation of unnatural amino acids[7,9,10] and chemical synthesis of proteins containing unnatural amino acids[11,12] will not be discussed in this context.

[a] *Dr. A. A. Jensen, Dr. K. Strømgaard*
Department of Medicinal Chemistry
The Danish University of Pharmaceutical Sciences
Universitetsparken 2, 2100 Copenhagen (Denmark)
Fax: (+45) 3530-6040
E-mail: krst@dfuni.dk

[b] *A. Strømgaard*
Department of Molecular Biology, University of Aarhus
Gustav Wieds Vej 10C, 8000 Aarhus C (Denmark)

 DOI: 10.1002/cbic.200400060



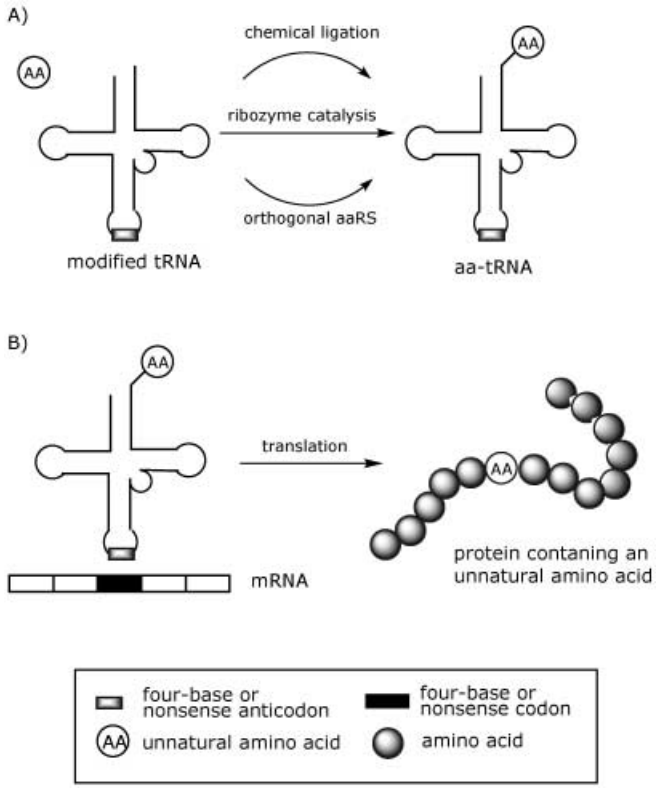
Scheme 1. *A simplified representation of protein synthesis (translation). A) Initiation: The two parts of the ribosome are assembled, and initiator tRNA (fMet-tRNA) is placed in the ribosome according to the start codon in mRNA. This process is driven by initiation factors (IF) and GTP. B) Elongation: An aminoacylated tRNA (aa-tRNA) is transported to the ribosome by an elongation factor (EF). Cognate interaction between the codon of mRNA and the anticodon of the tRNA leads to transfer of the polypeptide chain to the newly arrived tRNA, thus extending the polypeptide chain by one amino acid. The tRNA containing the polypeptide chain is then translocated and the ribosome is ready to start another cycle of elongation. This process continues until the synthesis of the protein is complete and a stop codon is reached. C) Termination: A stop codon triggers binding of a release factor (RF), which releases the completed protein from the ribosome, together with the tRNA. The ribosomal subunits and mRNA are hereafter ready for another round of protein synthesis. GTP = guanosine triphosphate, GDP = guanosine diphosphate.*

3.1. Nonsense suppression

In 1989 a novel biosynthetic in vitro method that allowed site-specific incorporation of unnatural amino acids into proteins was introduced independently by Chamberlin[13,14] and Schultz,[15,16] with their respective co-workers, based on earlier work on nonsense suppression.[8,17] The term "nonsense suppression" refers to the use of stop (nonsense) codons and suppressor tRNAs, which recognize stop codons. The method is based on the fact that only one of three stop codons in the genetic code is necessary for termination of protein synthesis. Hence, the two unused stop codons can be exploited for the introduction of unnatural amino acids. One or both of these stop codons can be introduced into DNA encoding for the target protein by conventional site-directed mutagenesis[14] or by synthesis of the mRNA.[7] The amber (UAG) stop codon has been most frequently used for the incorporation of unnatural amino acids but recently the use of the ochre (UAA) stop codon[18,19] and the potential use of the opal (UGA) stop codon have been reported.[20,21] A modified suppressor tRNA is chemically acetylated with an unnatural amino acid, and this aa-tRNA is recognized by the mRNA carrying the specific stop codon, whereby the unnatural amino acid is incorporated into the protein at the specific position. The suppressor aa-tRNA will inevitably be in competition with endogenous RFs for the additional stop codon that has been introduced into the mRNA. However, the target protein containing the unnatural amino acid can in most cases be easily separated from the truncated protein originating from the action of RFs.

The first step in nonsense suppression is the construction of the aminoacylated suppressor tRNA (Scheme 2).[15,22,23] Hecht and co-workers have acylated and ligated a dinucleotide to the 3′-terminus end of a truncated tRNA in the presence of T4 RNA ligase, thereby providing an aa-tRNA charged with an unnatural amino acid.[24] This methodology has been exploited[15,25,26] and improved by Schultz and co-workers.[22,27] Recently, a simplified version has been introduced by Sisido and co-workers, whereby the acylation of the dinucleotide is carried out in cationic micelles.[28] Most often, gel electrophoresis or radiolabeling has been used to examine formation of the aa-tRNA but recently Dougherty, Lester and co-workers have applied matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometry as an facile alternative.[29]

An interesting alternative is to use ribozymes as catalysts for the formation of aa-tRNAs, as shown by Suga and co-workers.[30–35] In vitro evolution of a ribozyme made it possible to charge a specific tRNA with a specific unnatural amino acid[30]



Scheme 2. *Principles for site-specific incorporation of unnatural amino acids into proteins. A) Synthesis of a tRNA containing the unnatural amino acid is currently carried out by using one of three methods: 1) a truncated, modified suppressor tRNA is ligated to a dinucleotide holding the unnatural amino acid, 2) a ribozyme can be engineered by in vitro evolution to catalytically aminoacylate the tRNA with an unnatural amino acid, or 3) an engineered aaRS can carry out the amioacylation with the unnatural amino acid in vivo. B) Site-specific incorporation of an unnatural amino acid into a protein is mediated by a matching of either a stop or a four-base anticodon of the tRNA that carries the unnatural amino acid and the corresponding stop or four-base mRNA codon.*

and provide multiple turnover activity. This approach has so far only been used in vitro but the technique has the potential to be applied to in vivo systems, and it could be an attractive alternative to the development of specific tRNA/aaRS pairs (see later). To achieve an efficient incorporation of the unnatural amino acid into the target protein, it is crucial that the aa-tRNA carrying the unnatural amino acid is not deacylated by native aaRSs. Therefore, the tRNA either has to be modified so that it is no longer a substrate for aaRSs or a tRNA from an organism unrelated to the transcription/translation system has to be used.[36–38] For example, Schultz and co-workers have applied a tRNA derived from yeast in an *Escherichia coli* transcription/translation system.[15,16]

The nonsense suppression method has been used to incorporate a large number of structurally diverse unnatural amino acids, representing a large variety of functionalities, into proteins. This has allowed studies of enzymatic activity, protein stability, biomolecular recognition, and protein–protein interactions, studies that would not have been possible with conventional mutagenesis.[39–49] In most cases the unnatural amino acids have been α-amino acids but non-α-amino acids[50] and, most notably, α-hydroxy acids have also been incorporated, with the latter introducing an amide-to-ester mutation in the protein backbone.[14,45,51] Attempts to use the method for incorporation of D-amino acids have not been successful, since modification of the ribosome was required.[52] However, these studies have shown that translation factors and the ribosome are compatible with many types of unnatural amino acids.

Dougherty, Lester and co-workers have used the nonsense suppression method for site-specific incorporation of unnatural amino acids into proteins expressed in *Xenopus* oocytes; this is an attractive system for electrophysiological studies of ion channels, receptors, and transporters.[53–55] The oocyte is coinjected with two RNA species: the modified mRNA encoding for the target protein and the aa-tRNA chemically acylated with an unnatural amino acid. This coinjection results in synthesis and surface expression of the target protein containing the unnatural amino acid. This has been used in numerous studies of integral membrane proteins (see below).[56,57] Recently the injection strategy has been extended to include incorporation of unnatural amino acids into proteins in a mammalian cell expression system; unnatural amino acids were incorporated into green fluorescent protein (GFP) and the nicotinic acetylcholine receptor in Chinese hamster ovary (CHO) cells and cultured hippocampal neurons by using microelectroporation.[58]

Schultz and co-workers have presented an alternative nonsense suppression approach to incorporate unnatural amino acids into proteins in vivo, based on an approach introduced by Furter.[59] In this methodology, a custom-made tRNA/aaRS pair is introduced, where an aaRS is engineered so that it only recognizes the unnatural amino acid and efficiently acylates the corresponding tRNA. This means that a specific aaRS has to be generated for each unnatural amino acid. The unnatural amino acid is added to the growth media, taken up by the host organism, and incorporated into the protein by the specific tRNA/aaRS pair.[60] The tRNA/aaRS pair has to be orthogonal to the host organism, that is, the aaRS must not recognize endogenous tRNAs and the suppressor tRNA cannot be a substrate for endogenous aaRSs. This methodology has been successfully utilized for the incorporation of several different unnatural amino acids into proteins in *E. coli*.[21,37,59–66] *O*-methyl-L-tyrosine has for example been site-specifically incorporated into β-lactamase in *E. coli* by using an orthogonal tRNA/aaRS pair originating from *Methanococcus jannaschii*.[7,67–69] The tRNA/aaRS pair from *M. jannaschii* has been used to incorporate various unnatural amino acids, such as photoactivatable amino acids[70–72] and keto-containing amino acids,[73–75] into proteins.[70–74,76–79] Recently, a fully autonomous bacterium encoding an unnatural amino acid was reported,[80] where an *E. coli* was engineered so it could synthesize *p*-amino-phenylalanine. This unnatural amino acid was incorporated into sperm whale myoglobin by using the orthogonal tRNA/aaRS pair from *M. jannaschii*.

Several orthogonal tRNA/aaRS pairs have been proposed to suppress nonsense codons in eukaryotic cells[37,81,82] and in mammalian cell lines.[83] Furthermore, site-specific incorporation of an unnatural amino acid into eukaryotic proteins in a

mammalian cell line by using a heterologous tRNA/aaRS pair has also been demonstrated.[84] Recently, the orthogonal tRNA/aaRS approach was used to incorporate unnatural amino acids into proteins in a eukaryotic cell for the first time; an orthogonal tRNA/aaRS pair originating from *E. coli* was used to add five novel amino acids into proteins in the yeast *Saccharomyces cerevisiae*.[85,86] Very recently, the same technique was used for site-specific incorporation of glycosylated amino acids to make glycoproteins, with myoglobin as an example, in an *E. coli* expression system.[87]

3.2. Extended codons

The major limitation of the nonsense suppression method is that the genetic code only contains three stop codons, which limits the theoretical numbers of different unnatural amino acids that can be incorporated in a single protein to two. To overcome this limitation Sisido and colleagues have explored an alternative strategy by using extended codons and frame shift suppression.[88–91] In this approach, an mRNA containing an extended codon consisting of four or five bases is read by a modified aa-tRNA (acylated with an unnatural amino acid and containing the corresponding extended anticodon), and a full-length protein containing an unnatural amino acid at the specific site is obtained. If the extended codon is read as a three-base codon by an endogenous tRNA, the reading frame will be shifted by one base. This will eventually result in a premature encounter with a stop codon and early termination of protein synthesis, thereby resulting in a truncated protein.

In certain species, some naturally occurring codons are rarely used and the amount of their corresponding tRNA is low. This has been used in the design of four-base codons, which are derived from these rarely used codons, to minimize the competition between the four-base anticodon tRNA and endogenous tRNA. Several such four-base codons have been studied along with four-base codons derived from stop codons. A number of four-base codons (AGGU, CGGU, CCCU, CUCU, and GGGU) were successfully read by corresponding tRNAs, whereas four-base codons derived from termination codons were not successfully decoded.[8,89]

The four-base codon technique has been used to incorporate a large number of unnatural amino acids into proteins in *E. coli*.[92–96] It has also been used to incorporate two different unnatural amino acids into two different sites of a single protein, thereby showing that four-base codons are not only orthogonal to their host organism but also to each other.[91,97,98] Recently, the use of five-base codons has been reported.[99] In this study, 16 different mRNAs, each containing one of the five-base codons $CGGN_1N_2$ (N_1 and N_2 indicate one of the four bases), were decoded by aa-tRNAs with complementary five-base anticodons to give rise to full-length proteins, each containing an unnatural amino acid. Moreover, it was shown that at least two of the five-base codons (CGGUA and CGGUG) were orthogonal to each other. Schultz and co-workers have also explored the multiple-base codon strategy and found that three-, four-, and five-base codons can be decoded efficiently.[100,101]

4. Advantages and Limitations

Site-specific incorporation of unnatural amino acids into proteins is a fairly new technique, and the focus has so far primarily been on development and improvement of the different techniques. In the following section, the advantages and limitations of these approaches will be described.

The major advantage of the nonsense suppression technique lies in its simplicity. The application of the method in vitro requires a relatively simple translational system and results in fast production and, if necessary, easy purification of proteins. The most important disadvantages of the nonsense suppression technique, however, are the often low yields of the mutated protein and the inherent limitations to the number of different unnatural amino acids that can be incorporated into the target protein. One of the reasons for the low yields of mutated protein is the competition between the suppressor tRNA and release factors.[102] In an *E. coli* transcription/translation system this problem can be addressed by the use of S-30 extracts, which are characterized by the lack of release factor 1 (RF1).[36,44] Alternatively, the competition can be avoided by using expression systems with purified components such as the recently introduced "protein synthesis using recombinant elements" (PURE) system.[103]

Another reason for the often low yields of mutated protein in the nonsense suppression studies is that the synthetic suppressor aa-tRNA is consumed during the translation process and is not regenerated. However, other factors may contribute, as it is claimed that repeated addition of suppressor aa-tRNA did not increase the yields of mutated protein.[8] Recently, problems associated with the synthetic suppressor aa-tRNA were circumvented by the introduction of an engineered tRNA/aaRS pair, which generated (and regenerated) the aa-tRNA in the expression system.[68] The main advantage of this system is that it is applicable to site-specific incorporation of unnatural amino acids into proteins in vivo and to prokaryotic as well as eukaryotic cells. The major limitation of the approach is that a specific tRNA/aaRS pair must be generated for each unnatural amino acid and engineering of specific aaRSs is particularly technically demanding.

The theoretical number of different unnatural amino acids that can be incorporated into a protein using nonsense suppression is two, and recently RajBhandary and co-workers used two different nonsense codons (UAG and UAA) to incorporate two different unnatural amino acids into the same protein.[18] Hecht and co-workers used a combination of nonsense and frame-shift suppression to incorporate two different unnatural amino acids into a single protein.[104] Nonsense suppression can also be combined with other approaches such as the use of missing, rare, or unassigned codons as nonsense sites.[105] Finally, Tirrel and co-workers have used the "degeneracy" of the genetic code, that is, the fact that amino acids are encoded for by more than one codon, which potentially can be applied for site-specific incorporation of unnatural amino acids into the protein.[106]

The advantage of using frame-shift suppression instead of nonsense suppression is threefold: It should be easier to incor-

porate more than one unnatural amino acid into the same protein, there is no competition with RFs, which is important for in vivo incorporation, and finally it seems as if four-base codons are generally more efficient than stop codons in incorporating unnatural amino acids into proteins.

Incorporation of many types of unnatural amino acids into proteins has been reported, and probably more have been tested but not reported. This makes it difficult to get an overall idea of the limitations as to which unnatural amino acids can be incorporated into proteins. However, it is clear that the translation machinery tolerates a wide variety of diverse amino acid side chains. The few reports concerning the limiting factors have suggested that incorporation of unnatural amino acids into proteins is not limited so much by their size or hydrophobic character as by their geometry and the shape of their side groups.[8, 93, 107]

5. Structure–Function Studies

The number of studies in which site-specific incorporation of unnatural amino acids has been applied to investigate biological questions is limited. Schultz and co-workers have studied a few soluble proteins such as T4 lysozyme[108–110] and staphylococcal nuclease (SNase).[111–114] Pioneering work has been carried out by Dougherty, Lester and co-workers in studies of integral membrane proteins, particularly the nicotinic acetylcholine (nACh) receptor,[53, 115–125] and the 5-hydroxy-tryptamine (5-HT_3) receptor,[118, 120, 126] as well as of the voltage-gated potassium channels Kir2.1[127] and Shaker B.[125] In all of these studies the nonsense suppression method has been applied, and the functional consequences of the unnatural amino acid mutagenesis have been measured by electrophysiology in *Xenopus* oocytes. The cell-surface expression levels of the target protein are often low because of the low efficiency of the nonsense suppression method. Hence, the highly sensitive *Xenopus* oocytes system is excellent for such studies. Finally, it is essential that truncated proteins arising from the action of RFs are not functional, since this would clearly complicate the interpretation of the data generated.

The nACh receptors belong to a superfamily of ligand-gated ion channels (LGICs), which mediate fast chemical synaptic transmission in the central and peripheral nervous systems. LGICs are a pentameric assembly of subunits, each composed of an extracellular N-terminal domain and four transmembrane domains (M1–M4).[128] Agonist binding takes place at the interface between two subunits in the extracellular N-terminal domain of the receptor and induces a conformational change in the receptor protein, which triggers an opening of the ion channel and permits ions to cross the cell membrane (Figure 1).

In this review we will focus on studies of the muscle-type nACh receptor, a heteromeric LGIC composed of two α1 subunits and β, γ, and δ subunits. Initial investigations by using site-specific incorporation of unnatural amino acids were focused on the agonist-binding site of the receptor (Figure 1). This site had previously been examined by photoaffinity labeling, and it had been established that nine aromatic tryptophan and tyrosine residues were located near the agonist-binding site,[129] although the individual roles of these residues were unknown.[56] It had been speculated that these residues could bind the agonist, acetylcholine (ACh), through cation–π binding, where the quaternary amine of ACh and the negative electrostatic potential on the face of an aromatic ring generates a noncovalent binding force.[130] In the initial studies, three tyrosine residues residing in the α1 subunit of the muscle nACh receptor (αTyr93, αTyr190, and αTyr198) were examined[53, 115] by incorporation of 14 different unnatural derivatives of phenylalanine and tyrosine at the three sites. Although all three tyrosine residues were found to be important for agonist binding, none of them appeared to be involved in a direct cation–π binding. In a follow-up study, the importance of four tryptophan residues (αTrp86, αTrp149, αTrp184, and γTrp55/δTrp57), which had also been implicated in ACh binding, was studied.[116] Unnatural tryptophan derivatives were incorporated at the four sites, and αTrp149 was identified as a likely candidate for a cation–π interaction, as it was in direct contact with the quaternary amine of ACh. Analogously, it was shown that the corresponding tryptophan residue in the 5-HT_3 receptor, Trp183, also forms a cation–π interaction with the protonated amine of serotonin (5-HT).[118, 120]

The function of tyrosine residues αTyr93 and αTyr198 for agonist binding to the muscle nACh receptor have been explored by a "decaging" technique.[122] In this case, the phenolic hydroxy groups of tyrosine were protected with the photosensitive *o*-nitrobenzyl protecting group, which is readily removed by UV irradiation, thus restoring the original tyrosine residue (Figure 1). Whereas the "caged" αTyr93 and αTyr198 (*o*-nitrobenzyl-contaning) mutants of the receptor were not responsive to ACh, decaging of the residues restored the functional properties of the receptor. The decaging methodology has also been applied to study the M2 segment of the γ subunit of muscle-type nACh receptors,[124] where *o*-nitrobenzyl-protected cysteine and tyrosine residues were incorporated into a specific position in the M2 segment. Analogously, *o*-nitrobenzyl-protected tyrosine has been incorporated into a K^+ channel.[127] By using a similar principle, a (2-nitrophenyl)glycine group, which causes peptide-bond cleavage upon irradiation, was site-specifically incorporated into the intracellular region of a K^+ channel, as well as into the extracellular and transmembrane regions of the nACh receptor.[125]

The insight into the agonist-binding site of the nACh receptors obtained in studies with site-specific incorporation of unnatural amino acids was confirmed when an X-ray crystallographic structure of an ACh-binding protein was solved.[131] The ACh-binding protein (AChBP) is a structural and functional homologue of the agonist-binding domain of the nACh receptor α subunit.[131] The AChBP was crystallized with HEPES (*N*-2-hydroxyethylpiperazine-*N*^{9}-ethanesulphonic acid) rather than ACh in the agonist-binding site, and the crystal structures revealed that the positively charged amino group of HEPES makes a cation–π interaction with the equivalent of αTrp149.[131] Furthermore, the nine tyrosine and tryptophan residues corresponding to those studied in the muscle-type nACh receptor constituted the agonist-binding site as an aro-

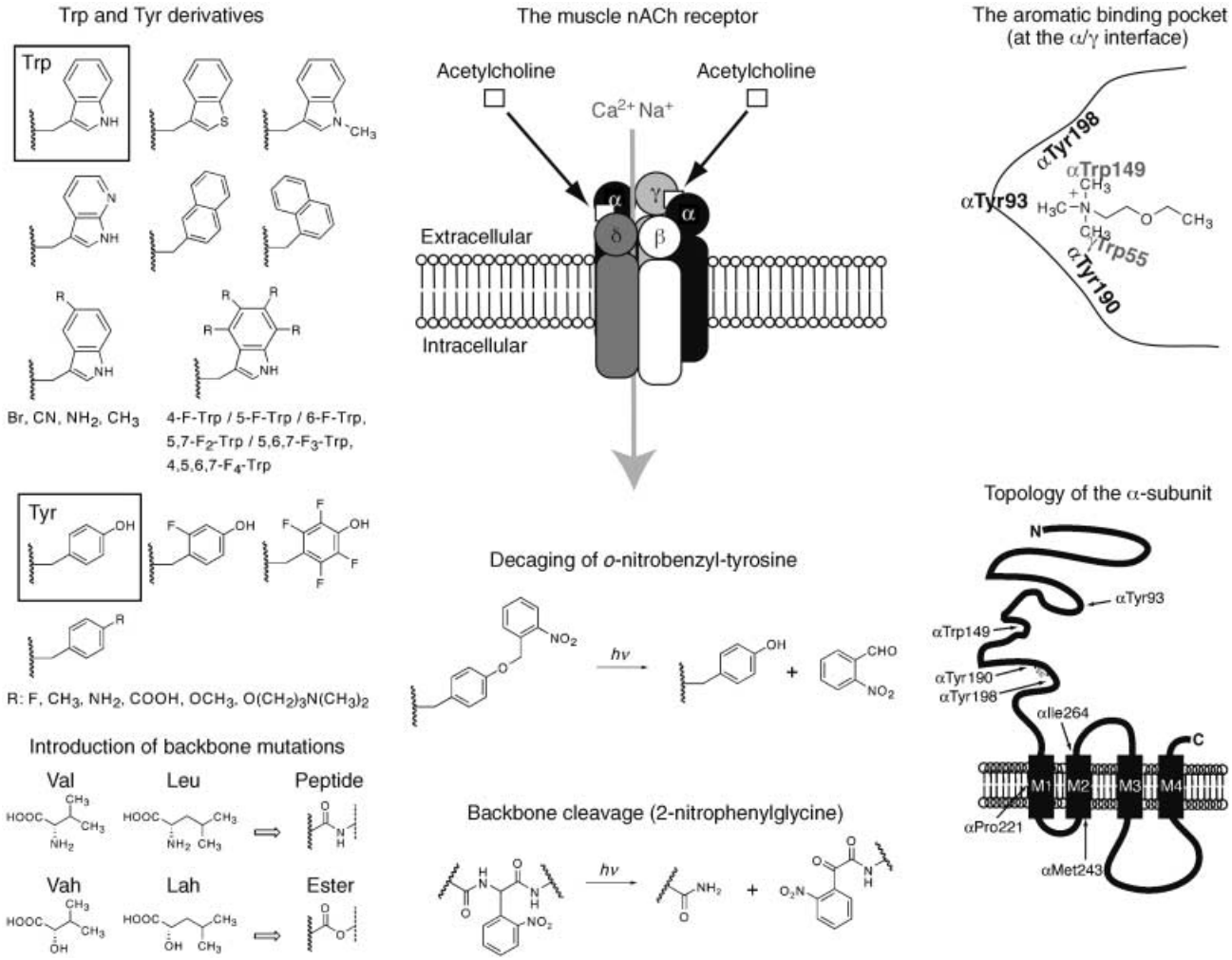


Figure 1. *Probing nACh receptor function by incorporation of unnatural amino acids. The pentameric muscle nACh receptor is activated by acetylcholine binding to the α/γ and α/δ subunit interfaces; this triggers the opening of the ion channel, thereby enabling the influx of Na^+ and Ca^{2+} ions into the cell. The topology of the α subunit of the receptor and the five aromatic amino acids demonstrated to constitute the aromatic box surrounding the quaternized amino group of acetylcholine is depicted to the right of the receptor. Examples of the different unnatural amino acids incorporated into the nACh receptor and the applications of these by the groups of Lester and Dougherty are given to the left and below the receptor. Lah = leucic acid, Vah = α-hydroxy valine.*

matic cage.[131,132] Thus, the AChBP structure verified the overall conclusions of the studies by Dougherty, Lester and coworkers.

The incorporation of unnatural amino acids allows the introduction of backbone mutations in the target protein or more specifically the introduction of an ester linkage instead of the amide bond. The substitution of an amide to an ester bond affects the backbone hydrogen-bonding interactions and the secondary structure of the protein dramatically, whereas the size, shape, and chemical nature of the side chains are retained.[39] This fact was exploited in a study of the functional importance of a proline (αPro221) in the M1 domain of the α subunit of nACh receptor (Figure 1).[123] αPro221 is a highly conserved residue in the LGIC family, and previous conventional mutation studies had shown this residue to be important for gating of the nACh receptor, whereas it did not appear to be crucial for folding, assembly, and surface expression of the receptor.[123] Furthermore, it had been speculated that the hydrogen-bonding properties of proline, which can only act as a hydrogen acceptor, were important for the receptor function. This was evaluated by incorporation of α-hydroxy acids, such as lactic acid and leucic acid, α-hydroxy analogues of alanine and leucine, respectively, into this site of the nACh receptor. These mutated receptors displayed functional properties similar to wild-type nACh receptors, thus emphasizing the importance of the hydrogen-bond-accepting properties of αPro221.[123]

Amide-to-ester bond mutations were also used to examine conformational changes in the M2 domain of the activated nACh receptor (residues αMet243 to αIle264, Figure 1).[123] The M2 domains of the five subunits in the nACh receptor constitute the ion-channel pore of the receptor, and thus the region plays a key role in receptor gating. Introduction of leucic acid and α-hydroxy valine at various positions of the M2 domain of the α subunit (Figure 1) resulted in significantly changed agonist potencies at the receptor. This suggested significant structural changes, rather than just a reorientation, in the M2 backbone upon agonist binding and opening of the ion channel.[123]

Similarly, a backbone amide-to ester mutation has also been used for studying the Kir2.1 K^+ channel.[133]

Finally, Turcatti et al. have used site-specific incorporation of unnatural amino acids to carry out fluorescence resonance energy transfer (FRET) studies at the neurokinin-2 (NK2) receptor, which is the only example of the technique being applied in a study of a G-protein-coupled receptor.[134] A fluorescent donor group, 3-*N*-(7-nitrobenz-2-oxa-1,3-diazol-4-yl)-2,3-diaminopropionic acid, was incorporated into either position 104 or 248 of the NK2 receptor. Incorporation of the unnatural amino acid resulted in fully functional receptors, and the authors were able to estimate the distance between a fluorescent ligand and the fluorescent donor group of the unnatural amino acid by using FRET techniques.[134]

Conclusion and Outlook

The ability to introduce amino acids not encoded in the genetic code into proteins constitutes a principal step forward in our ability to investigate and understand structure–function relationships of proteins. Considering the structural diversity of the unnatural amino acids that have been incorporated into proteins so far, it will be possible to probe ligand–protein and protein–protein interactions in a much more sophisticated manner than by conventional mutagenesis. Until now much effort has been devoted to the exploration and improvement of the strategies and techniques used in the unnatural amino acid mutagenesis. The techniques are now at a point where a more widespread use of the techniques should be possible.

Acknowledgements

We are grateful to Dr. Leif Christensen, Novo Nordisk A/S (Denmark), for critically reading the manuscript. We thank the Alfred Benzon Foundation (to K.S.) and the Lundbeck Foundation (to A.A.J.) for financial support.

Keywords: membrane proteins · mutagenesis · protein engineering · tRNA · unnatural amino acids

[1] M. Ibba, D. Söll, *EMBO reports* **2001**, *2*, 382.
[2] M. Ibba, D. Söll, *Annu. Rev. Biochem.* **2000**, *69*, 617.
[3] K. Wakasugi, C. L. Quinn, N. Tao, P. Schimmel, *EMBO J.* **1998**, *17*, 297.
[4] I. Hirao, T. Ohtsuki, T. Fujiwara, T. Mitsui, T. Yokogawa, T. Okuni, H. Nakayama, K. Takio, T. Yabuki, T. Kigawa, K. Kodama, T. Yokogawa, K. Nishikawa, S. Yokoyama, *Nat. Biotechnol.* **2002**, *20*, 177.
[5] S. Matsuda, A. A. Henry, P. G. Schultz, F. E. Romesberg, *J. Am. Chem. Soc.* **2003**, *125*, 6134.
[6] H. Liu, J. Gao, S. R. Lynch, D. Saito, L. Maynard, E. T. Kool, *Science* **2003**, *302*, 868.
[7] L. Wang, P. G. Schultz, *Chem. Commun.* **2002**, 1.
[8] M. Sisido, *Biopolymers* **2003**, *7*, 25.
[9] A. J. Link, M. L. Mock, D. A. Tirrell, *Curr. Opin. Biotechnol.* **2003**, *14*, 603.
[10] B. Bilgicer, K. Kumar, *J. Chem. Edu.* **2003**, *80*, 1275.
[11] J. A. Borgia, G. B. Fields, *Trends Biochem. Sci.* **2000**, *18*, 243.
[12] G. Casi, D. Hilvert, *Curr. Opin. Struc. Biol.* **2003**, *13*, 589.
[13] J. D. Bain, C. G. Glabe, T. A. Dix, A. R. Chamberlin, *J. Am. Chem. Soc.* **1989**, *111*, 8013.
[14] J. D. Bain, S. D. Edward, C. G. Glabe, D. A. Wacker, M. H. Lyttle, T. A. Dix, A. R. Chamberlin, *Biochemistry* **1991**, *30*, 5411.
[15] C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Science* **1989**, *244*, 182.
[16] J. A. Ellman, D. Mendel, S. J. Anthony-Cahill, C. J. Noren, P. G. Schultz, *Methods Enzymol.* **1991**, *202*, 301.
[17] T. Hendrickson, V. de Crécy-Lagard, P. Schimmel, *Annu. Rev. Biochem.* **2004**, *73*, in press.
[18] C. Köhrer, J. Yoo, M. Bennett, J. Schaack, U. L. RajBhandary, *Chem. Biol.* **2003**, *10*, 1095.
[19] C. Köhrer, J. Xie, S. Kellerer, U. Varshney, U. L. RajBhandary, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 14310.
[20] J. P. Capone, J. M. Sedivy, P. A. Sharp, U. L. RajBhandary, *Mol. Cell. Biol.* **1986**, *6*, 3059.
[21] J. C. Anderson, P. G. Schultz, *Biochemistry* **2003**, *42*, 9598.
[22] S. A. Robertson, C. J. Noren, S. J. Anthony-Cahill, M. C. Griffith, P. G. Schultz, *Nucleic Acids Res.* **1989**, *17*, 9649.
[23] J. D. Bain, D. A. Wacker, E. E. Kuo, M. H. Lyttle, A. R. Chamberlin, *J. Org. Chem.* **1991**, *56*, 4615.
[24] T. G. Heckler, L.-H. Chang, Y. Zama, T. Naka, M. S. Chorghade, S. M. Hecht, *Biochemistry* **1984**, *23*, 1468.
[25] J. R. Roesser, C. Xu, R. C. Payne, C. K. Surratt, S. M. Hecht, *Biochemistry* **1989**, *28*, 5185.
[26] G. Baldini, B. Martoglio, A. Schachenmann, C. Zugliani, J. Brunner, *Biochemistry* **1988**, *27*, 7951.
[27] S. A. Robertson, J. A. Ellman, P. G. Schultz, *J. Am. Chem. Soc.* **1991**, *113*, 2722.
[28] K. Ninomiya, T. Kurita, T. Hohsaka, M. Sisido, *Chem. Commun.* **2003**, 2242.
[29] E. J. Petersson, M. Shahgholi, H. A. Lester, D. A. Dougherty, *RNA* **2002**, *8*, 542.
[30] Y. Bessho, D. R. W. Hodgson, H. Suga, *Nat. Biotechnol.* **2002**, *20*, 723.
[31] H. Murakami, N. J. Bonzagni, H. Suga, *J. Am. Chem. Soc.* **2002**, *124*, 6834.
[32] H. Murakami, D. Kourouklis, H. Suga, *Chem. Biol.* **2003**, *10*, 1077.
[33] H. Murakami, H. Saito, H. Suga, *Chem. Biol.* **2003**, *10*, 655.
[34] H. Saito, H. Suga, *J. Am. Chem. Soc.* **2001**, *123*, 7178.
[35] H. Saito, K. Watanabe, H. Suga, *RNA* **2001**, *7*, 1867.
[36] M. A. Gilmore, L. E. Steward, A. R. Chamberlin, *Top. Curr. Chem.* **1999**, *202*, 78.
[37] A. K. Kowal, C. Köhrer, U. L. RajBhandary, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 2268.
[38] S. T. Cload, D. R. Liu, W. A. Froland, P. G. Schultz, *Chem. Biol.* **1996**, *3*, 1033.
[39] J. A. Ellman, D. Mendel, P. G. Schultz, *Science* **1992**, *255*, 197.
[40] V. W. Cornish, D. R. Benson, C. A. Altenbach, K. Hideg, W. L. Hubbell, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1994**, *91*, 2910.
[41] J. K. Judice, T. R. Gamble, E. C. Murphy, A. M. Vos, P. G. Schultz, *Science* **1993**, *261*, 1578.
[42] H. Chung, D. R. Benson, P. G. Schultz, *Science* **1993**, *259*, 806.
[43] D. Mendel, J. A. Ellman, Z. Chang, D. L. Veenstra, P. A. Kollman, P. G. Schultz, *Science* **1992**, *256*, 1798.
[44] S. J. Anthony-Cahill, M. C. Griffith, C. J. Noren, D. J. Suich, P. G. Schultz, *Trends Biochem. Sci.* **1989**, *14*.
[45] J. T. Koh, V. W. Cornish, P. G. Schultz, *Biochemistry* **1997**, *36*, 11314.
[46] B. Wang, K. C. Brown, M. Lodder, C. S. Craik, S. M. Hecht, *Biochemistry* **2002**, *41*, 2805.
[47] J. D. Bain, C. Switzer, A. R. Chamberlin, S. A. Benner, *Nature* **1992**, *356*, 537.
[48] V. W. Cornish, P. G. Schultz, *Curr. Opin. Struc. Biol.* **1994**, *4*, 601.
[49] L. E. Steward, C. S. Collins, M. A. Gilmore, J. E. Carlson, J. B. A. Ross, A. R. Chamberlin, *J. Am. Chem. Soc.* **1997**, *119*, 6.
[50] B. M. Eisenhauer, S. M. Hecht, *Biochemistry* **2002**, *41*, 11472.
[51] I. Shin, A. Y. Ting, P. G. Schultz, *J. Am. Chem. Soc.* **1997**, *119*, 12667.
[52] L. M. Dedkova, N. E. Fahmi, S. Y. Golovine, S. M. Hecht, *J. Am. Chem. Soc.* **2003**, *125*, 6616.
[53] M. V. Nowak, P. C. Kearney, J. R. Sampson, M. E. Saks, C. G. Labarca, S. K. Silverman, W. Zhong, J. Thorson, J. N. Abelson, N. Davidson, P. G. Schultz, D. A. Dougherty, H. A. Lester, *Science* **1995**, *268*, 439.
[54] M. V. Nowak, J. P. Gallivan, S. K. Silverman, C. G. Labarca, D. A. Dougherty, H. A. Lester, *Methods Enzymol.* **1998**, *293*, 504.
[55] M. E. Saks, J. R. Sampson, M. V. Nowak, P. C. Kearney, F. Du, J. N. Abelson, H. A. Lester, D. A. Dougherty, *J. Biol. Chem.* **1996**, *271*, 23169.

[56] D. A. Dougherty, *Curr. Opin. Chem. Biol.* **2000**, *4*, 645.
[57] D. L. Beene, D. A. Dougherty, H. A. Lester, *Curr. Opin. Neurobiol.* **2003**, *13*, 264.
[58] S. L. Monahan, H. A. Lester, D. A. Dougherty, *Chem. Biol.* **2003**, *10*, 573.
[59] R. Furter, *Protein Sci.* **1998**, *7*, 419.
[60] D. R. Liu, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1999**, *96*, 4780.
[61] L. Wang, P. G. Schultz, *Chem. Biol.* **2001**, *8*, 883.
[62] D. R. Liu, T. J. Magliery, M. Pastrnak, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 10092.
[63] L. Wang, T. J. Magliery, D. R. Liu, P. G. Schultz, *J. Am. Chem. Soc.* **2000**, *122*, 5010.
[64] M. Pastrnak, P. G. Schultz, *Bioorg. Med. Chem.* **2001**, *9*, 2373.
[65] M. Pastrnak, T. J. Magliery, P. G. Schultz, *Helv. Chim. Acta* **2000**, *83*, 2277.
[66] S. Ohno, T. Yokogawa, I. Fujii, H. Asahara, H. Inokuchi, K. Nishikawa, *J. Biochem.* **1998**, *124*, 1065.
[67] L. Wang, A. Brock, B. Herberich, P. G. Schultz, *Science* **2001**, *292*, 498.
[68] L. Wang, *Science* **2003**, *302*, 584.
[69] T. Kobayashi, O. Nureki, R. Ishitani, A. Yaremchuk, M. Tukalo, S. Cusack, K. Sakamoto, S. Yokoyama, *Nat. Struct. Biol.* **2003**, *10*, 425.
[70] J. W. Chin, P. G. Schultz, *ChemBioChem* **2002**, *11*, 1135.
[71] J. W. Chin, S. W. Santoro, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 9026.
[72] J. W. Chin, A. B. Martin, D. S. King, L. Wang, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 11020.
[73] L. Wang, Z. Zhang, A. Brock, P. G. Schultz, *Proc. Natl. Acad. Sci. USA* **2003**, *100*, 56.
[74] Z. Zhang, B. A. C. Smith, L. Wang, A. Brock, C. Cho, P. G. Schultz, *Biochemistry* **2003**, *42*, 6735.
[75] V. W. Cornish, K. M. Hahn, P. G. Schultz, *J. Am. Chem. Soc.* **1996**, *118*, 8150.
[76] L. Wang, A. Brock, P. G. Schultz, *J. Am. Chem. Soc.* **2002**, *124*, 1836.
[77] L. Wang, J. Xie, A. A. Deniz, P. G. Schultz, *J. Org. Chem.* **2003**, *68*, 174.
[78] S. W. Santoro, L. Wang, B. Herberich, D. S. King, P. G. Schultz, *Nat. Biotechnol.* **2002**, *20*, 1044.
[79] L. Alfonta, Z. Zhang, S. Uryu, J. A. Loo, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, 14662.
[80] R. A. Mehl, J. C. Anderson, S. W. Santoro, L. Wang, A. B. Martin, D. S. King, D. M. Horn, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 935.
[81] J. W. Chin, A. Cropp, S. Chu, E. Meggers, P. G. Schultz, *Chem. Biol.* **2003**, *10*, 511.
[82] D. Kiga, K. Sakamoto, K. Kodama, T. Kigawa, T. Matsuda, T. Yabuki, M. Shirouzu, Y. Harada, H. Nakayama, K. Takio, Y. Hasegawa, Y. Endo, I. Hirao, S. Yokoyama, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 9715.
[83] H. J. Drabkin, H. Park, U. L. RajBhandary, *Mol. Cell. Biol.* **1996**, *16*, 907.
[84] K. Sakamoto, A. Hayashi, A. Sakamoto, D. Kiga, H. Nakayama, A. Soma, T. Kobayashi, M. Kitabatake, K. Takio, K. Saito, M. Shirouzu, I. Hirao, S. Yokoyama, *Nucleic Acids Res.* **2002**, *30*, 4692.
[85] J. W. Chin, T. A. Cropp, J. C. Anderson, M. Mukherji, Z. Zhang, P. G. Schultz, *Science* **2003**, *301*, 964.
[86] A. Deiters, A. Cropp, M. Mukherji, J. W. Chin, J. C. Anderson, P. G. Schultz, *J. Am. Chem. Soc.* **2003**, *125*, 11782.
[87] Z. Zhang, J. Gildersleeve, Y.-Y. Yang, R. Xu, J. A. Loo, S. Uryu, C.-H. Wong, P. G. Schultz, *Science* **2004**, *303*, 371.
[88] T. Hohsaka, M. Sisido, *Curr. Opin. Chem. Biol.* **2002**, *6*, 809.
[89] M. Sisido, T. Hohsaka, *Appl. Microbiol. Biotechnol.* **2001**, *57*, 274.
[90] T. Hohsaka, Y. Ashizuka, H. Murakami, M. Sisido, *J. Am. Chem. Soc.* **1996**, *118*, 9778.
[91] T. Hohsaka, Y. Ashizuka, H. Taira, H. Murakami, M. Sisido, *Biochemistry* **2001**, *40*, 11060.
[92] H. Murakami, T. Hohsaka, Y. Ashizuka, M. Sisido, *J. Am. Chem. Soc.* **1998**, *120*, 7520.
[93] T. Hohsaka, D. Kajihara, Y. Ashizuka, H. Murakami, M. Sisido, *J. Am. Chem. Soc.* **1999**, *121*, 34.
[94] H. Murakami, T. Hohsaka, Y. Ashizuka, K. Hashimoto, M. Sisido, *Biomacromolecules* **2000**, *1*, 118.
[95] M. Taki, T. Hohsaka, H. Murakami, K. Taira, M. Sisido, *FEBS lett.* **2001**, *507*, 35.
[96] N. Muranaka, T. Hohsaka, M. Sisido, *FEBS lett.* **2002**, *510*, 10.
[97] T. Hohsaka, Y. Ashizuka, H. Sasaki, H. Murakami, M. Sisido, *J. Am. Chem. Soc.* **1999**, *121*, 12194.
[98] M. Taki, T. Hohsaka, H. Murakami, K. Taira, M. Sisido, *J. Am. Chem. Soc* **2002**, *124*, 14586.
[99] T. Hohsaka, Y. Ashizuka, H. Murakami, M. Sisido, *Nucleic Acids Res.* **2001**, *29*, 3646.
[100] T. J. Magliery, J. C. Anderson, P. G. Schultz, *J. Mol. Biol.* **2001**, *307*, 755.
[101] J. C. Anderson, T. J. Magliery, P. G. Schultz, *Chem. Biol.* **2002**, *9*, 237.
[102] G. F. Short III, S. Y. Golovine, S. M. Hecht, *Biochemistry* **1999**, *38*, 8808.
[103] Y. Shimizu, A. Inoue, Y. Tomari, T. Suzuki, T. Yokogawa, K. Nishikawa, T. Ueda, *Nat. Biotechnol.* **2001**, *19*, 751.
[104] R. D. Anderson, J. Zhou, S. M. Hecht, *J. Am. Chem. Soc.* **2002**, *124*, 9674.
[105] A. K. Kowal, J. S. Oliver, *Nucleic Acids Res.* **1997**, *25*, 4685.
[106] I. Kwon, K. Kirshenbaum, D. A. Tirrell, *J. Am. Chem. Soc* **2003**, *125*, 7512.
[107] T. Hohsaka, K. Sato, K. Takai, S. Yokoyama, *FEBS lett.* **1993**, *335*, 47.
[108] D. Mendel, J. A. Ellman, P. G. Schultz, *J. Am. Chem. Soc* **1993**, *115*, 4359.
[109] J. A. Ellman, B. F. Volkman, D. Mendel, P. G. Schultz, D. E. Wemmer, *J. Am. Chem. Soc* **1992**, *114*, 7959.
[110] D. Mendel, J. Ellman, P. G. Schultz, *J. Am. Chem. Soc* **1991**, *113*, 2758.
[111] A. Y. Ting, I. Shin, C. Lucero, P. G. Schultz, *J. Am. Chem. Soc* **1998**, *120*, 7135.
[112] J. Thorson, E. Chapman, E. C. Murphy, P. G. Schultz, *J. Am. Chem. Soc* **1995**, *117*, 1157.
[113] J. Thorson, E. Chapman, P. G. Schultz, *J. Am. Chem. Soc* **1995**, *117*, 9361.
[114] E. Chapman, J. Thorson, P. G. Schultz, *J. Am. Chem. Soc* **1997**, *119*, 7151.
[115] P. C. Kearney, M. W. Nowak, W. Zhong, S. K. Silverman, H. A. Lester, D. A. Dougherty, *Mol. Pharm.* **1996**, *50*, 1401.
[116] W. Zhong, J. P. Gallivan, Z. Zhang, L. Li, H. A. Lester, D. A. Dougherty, *Proc. Natl. Acad. Sci. USA* **1998**, *95*, 12088.
[117] L. Li, W. Zhong, N. Zacharias, C. Gibbs, H. A. Lester, D. A. Dougherty, *Chem. Biol.* **2001**, *8*, 47.
[118] D. L. Beene, G. S. Brandt, W. Zhong, N. M. Zacharias, H. A. Lester, D. A. Dougherty, *Biochemistry* **2002**, *41*, 10262.
[119] E. J. Petersson, A. Choi, D. S. Dahan, H. A. Lester, D. A. Dougherty, *J. Am. Chem. Soc.* **2002**.
[120] T. Mu, H. A. Lester, D. A. Dougherty, *J. Am. Chem. Soc.* **2003**, *125*, 6850.
[121] P. C. Kearney, H. Zhang, W. Zhong, D. A. Dougherty, H. A. Lester, *Neuron* **1996**, *17*, 1221.
[122] J. C. Miller, S. K. Silverman, P. M. England, D. A. Dougherty, H. A. Lester, *Neuron* **1998**, *20*, 619.
[123] P. M. England, Y. Zhang, D. A. Dougherty, H. A. Lester, *Cell* **1999**, *96*, 89.
[124] K. D. Philipson, J. P. Gallivan, G. S. Brandt, D. A. Dougherty, H. A. Lester, *Am. J. Physiol. Cell Physiol.* **2001**, *281*, C195.
[125] P. M. England, H. A. Lester, N. Davidson, D. A. Dougherty, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 11025.
[126] H. Dang, P. M. England, S. S. Farivar, D. A. Dougherty, H. A. Lester, *Mol. Pharm.* **2000**, *57*, 1114.
[127] Y. Tong, G. S. Brandt, M. Li, G. Shapovalov, E. Slimko, A. Karschin, D. A. Dougherty, H. A. Lester, *J. Gen. Physiol.* **2001**, *117*, 103.
[128] J. F. Leite, M. Cascio, *Mol. Cell. Neurosci.* **2001**, *17*, 777.
[129] P. Corringer, N. Le Novère, J. Changeux, *Annu. Rev. Pharmacol. Toxicol.* **2000**, *40*, 431.
[130] J. C. Ma, D. A. Dougherty, *Chem. Rev.* **1997**, *97*, 1303.
[131] K. Brejc, W. J. van Dijk, R. V. Klaassen, M. Schuurmans, J. van der Oost, A. B. Smit, T. K. Sixma, *Nature* **2001**, *411*, 269.
[132] T. K. Sixma, A. B. Smit, *Annu. Rev. Biophys. Biomol. Struct.* **2003**, *32*, 311.
[133] T. Lu, A. Y. Ting, J. Mainland, L. Y. Jan, P. G. Schultz, J. Yang, *Nat. Neurosci.* **2001**, *4*, 239.
[134] G. Turcatti, K. Nemeth, M. D. Edgerton, U. Meseth, F. Talabot, M. Peitsch, J. Knowles, H. Vogel, A. Chollet, *J. Biol. Chem.* **1996**, *271*, 19991.

Received: March 1, 2004

