4 Concepts of a Peptide-Based Screening System

Most important for the evaluation of the properties and interactions of fluorine-containing amino acids in native protein environments is the access to fluoroalkyl substituted peptides. Therefore, peptide building blocks (fluorine-containing amino acids) have to be synthesized in an enantiomerically pure form and incorporated into the peptide chain.

4.1 Fluorine-Containing Amino Acids as Peptide Building Blocks

For most fluoroalkyl substituted amino acids, such as side chain-modified α-amino acids, standard protocols for synthesis and incorporation into peptides are available. However, some classes of these non-natural peptide building blocks, e.g., fluorinated $C^{\alpha,\alpha}$-dialkylated amino acids require special and sometimes advanced procedures and methodologies.

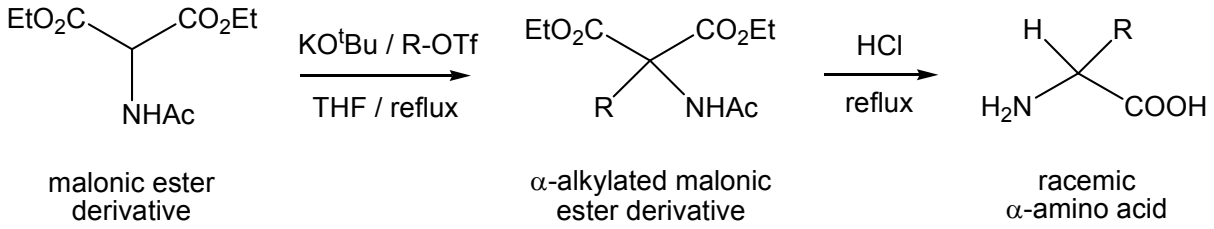
4.1.1 Synthesis of Fluorine-Containing Amino Acids

Since the main focus of this work was not on synthetical aspects of the amino acids that had to be investigated, only a brief summary about the synthetic routes towards side chain-modified and $C^{\alpha,\alpha}$-dialkylated fluorinated amino acids that have been studied in this work will be given in this chapter. In general, both the position as well as the extent of fluorine substitution within an amino acid dictates the methods that have to be applied for the chemical synthesis of these building blocks.

4.1.1.1 Side Chain-Modified Amino Acids

Many fluorinated amino acids have been successfully prepared by classical methods of amino acid chemistry, such as amination reactions, Strecker synthesis, hydantoin synthesis, Erlenmeyer azlactone synthesis, synthesis from N-substituted aminomalonic esters, and others.[140] This access is based on the general stability and chemical inertness of the C-F bond and on similar reactivities of several fluorinated and non-fluorinated reactants. One very popular synthesis route towards racemic amino acids starts from the diethyl ester of N-acetyl-aminomalonic acid (Scheme 4.1). After deprotonation of the α-carbon atom, the amino acid side chain is introduced, using the corresponding triflate or alkyl bromide. Ester cleavage, decarboxylation and deacetylation is carried out in one step with concentrated HCl

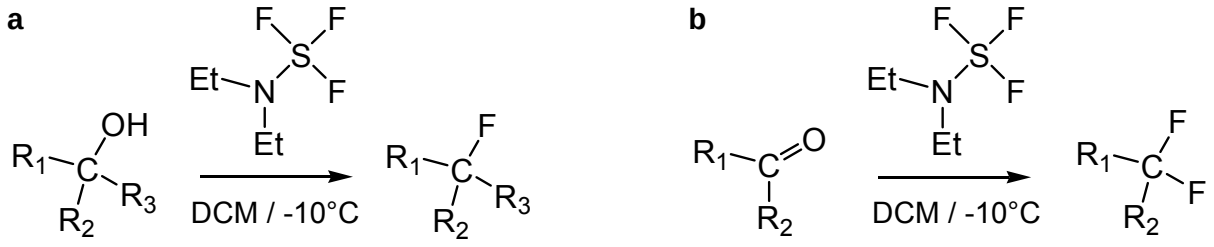
under reflux, resulting in the racemic α-aminocarbonic acid. This strategy was applied for the synthesis of the racemic precursor of *(S)*-4,4,4-trifluoroethylglycine, an amino acid that was studied in this work.



Scheme 4.1: *Amino acid synthesis starting from N-acetyl-aminomalonic ester.*

However, the applicability of these methods is limited, mainly caused by the change of polarity of the functional groups and bonds in the vicinity of fluorine atoms, the lower reactivity of some polyfluorinated compounds, and the considerable lability of the C-F bonds in the β- and, to some extent, also in the γ-position to carbonyl.

Alternative routes apply fluorination techniques of organofluorine chemistry for the introduction of fluorine into protected amino acids and their precursors, respectively.[141] Amongst other reagents, such as fluorine, trifluoromethyl hypofluorite, perchloryl fluoride, and sulfur tetrafluoride, different fluorides can be used. Diethylaminosulfur trifluoride (DAST) is a very prominent fluorination reagent for the replacement of a hydroxyl or carbonyl oxygen with fluorine (Scheme 4.2).



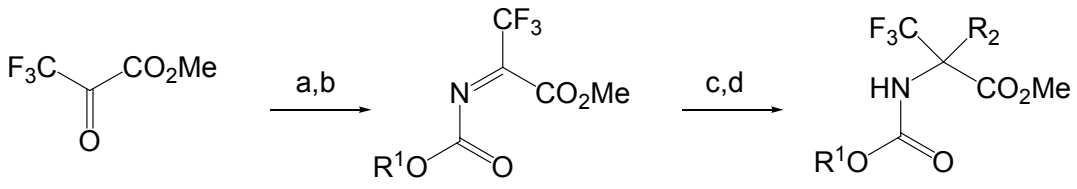
Scheme 4.2: *Fluorination of **a)** alcohols and **b)** carbonyls with DAST (according to Wilkinson.[142])*

The advantages of DAST are reactions under very mild conditions and the ease of application compared to others, such as sulfur tetrafluoride. Other techniques towards side chain-fluorinated amino acids use homogeneous catalysis[143] or enzymatic synthesis[144] for the incorporation of fluorine-containing groups into amino acid precursors. Stereoselective

synthesis routes using DAST fluorination were applied for the preparation of two amino acids that were studied in this work, *(S)*-4,4-difluoroethylglycine and *(S)*-4,4-difluoropropylglycine.

4.1.1.2 $C^{\alpha,\alpha}$-Dialkylated Amino Acids

Racemic C^{α}-trifluoromethyl substituted amino acids with side chains of high structural diversity can be produced via several synthesis routes that have been developed.[145,146] The most general approach is the amidoalkylation of carbon nucleophiles with alkyl-2-(alkoxycarbonylimino)-3,3,3-trifluoropropionates (Scheme 4.3).[147] α-Trifluoromethylalanine, an amino acid that was studied in this work, was prepared by this synthesis route.



R^1 = Z R^2 = alkyl, benzyl

a) R^1OCONH_2, b) $(CF_3CO)_2O$/pyridine, c) R^2MgX, d) H_3O^+

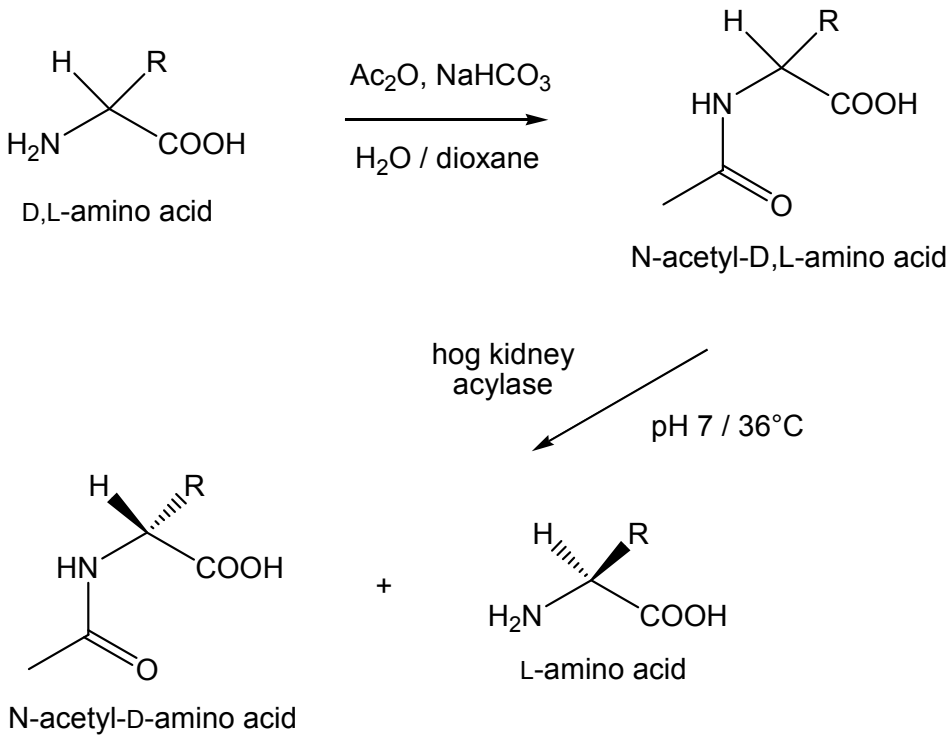
Scheme 4.3: *Synthetic route towards fully protected C^{α}-trifluoromethyl substituted amino acids (according to Koksch et al.[148])*

While the racemic $C^{\alpha,\alpha}$-dialkylated amino acids, obtained by the synthetic route described above, have to be optically purified afterwards, a strategy for the diastereoselective synthesis of αTfm amino acid-containing dipeptides uses amidoalkylation of carbon nucleophiles via *in situ* formation of homochiral cyclic acyl imines.[149]

4.1.1.3 Enzymatic Resolution of Racemic Fluorinated Amino Acids

Considering the divergent biological activities of the enantiomers of fluoroalkyl substituted amino acids and their diastereomeric peptide derivatives, the availability of these compounds in enantiomerically pure form is highly desirable. The most popular strategy for the stereochemical resolution of canonical and non-natural amino acids uses hydrolytic enzymes that specifically accept one, mostly the L-enantiomer, of the peptide building block.

One approach is based on the enantioselective deacylation of N-acyl L-amino acid derivatives applying acylases.[150] After synthesis and N-acetylation of the racemic 4,4,4-trifluoroethylglycine, the *(S)*-enantiomer was isolated via stereoselective deacetylation by hog kidney acylase (Scheme 4.4). The pure L-amino acid was then used for investigations in this work.



Scheme 4.4: *Optical resolution of a racemic amino acid. After N-acylation of the racemic mixture, hog kidney acylase selectively deacetylates the L-form while leaving the D-derivative unconverted.*

There are several examples in literature for successful enzymatic enantioselective deacylation of fluorinated amino acids. While N-trifluoroacetylated aromatic fluorinated amino acids, such as several fluoro- and pentafluorophenylalanines as well as fluoro- and difluorotryptophans are good substrates for Carboxypeptidase A, acylase I is preferably applied for the enantioselective hydrolysis of N-acetylated, trifluoromethyl-substituted aliphatic L-amino acids. *R*(+)2-(Trifluoromethyl)alanine can be obtained from the racemic N-trifluoroacetyl derivative by partial hydrolysis with an enantioselectivity of 99,1% using hog kidney aminoacylase.[151] This enzyme is probably the most applied biocatalyst for the optical resolution of racemic amino acids.

The stereochemical resolution of fluorinated amino acid can also be accomplished by enantioselective hydrolysis of their ester derivatives. The proteases α-chymotrypsin and subtilisin Carlsberg have succesfully been applied for the optical resolution of fluorinated and

trifluoromethyl-substituted aromatic amino acids, such as several phenylalanine and tryptophan derivatives.[152] A broad substrate specificity and enantioselectivity in the hydrolysis of esters with halogenated ester leaving groups can be achieved with lipases. This could be applied for the optical resolution of different 2,2,2-trifluoroethyl or 2-chloroethyl esters of fluorophenylalanines.

Alternatively, $C^{\alpha,\alpha}$-dialkylated amino acid amides can be hydrolyzed enantioselectively with amides.[153] A technology that uses amidase preparations from microorganisms can be applied for the preparation of a variety of enantiopure $C^{\alpha,\alpha}$-fluoroalkyl substituted amino acids.[154] Amidase from *Mycobacterium neoaurum* could be shown to hydrolyze *R,S*-(αCF_3)Ala-NH_2, *R,S*-(αCF_2Cl)Ala-NH_2, and *R,S*-(αCF_2Br)Ala-NH_2 with high enantioselectivities to give pure *R*-amino acids. Amidase from *Ochrobactrum anthropi* accepts *R,S*-(αCF_2H)Phe-NH_2 as substrate.

4.1.2 Incorporation of Fluorine-Containing Amino Acids into Peptides

In order to benefit from the impact fluoroalkyl chains have on structure, stability, and bioactivity of peptides, routine methods have to be available for the site-directed incorporation of fluorine-substituted amino acids. Since relatively young techniques like native chemical ligation[155] and expressed protein ligation[156] enable the assembly of proteins from altered peptide fragments, the insertion of challenging building blocks into polypeptide chains becomes increasingly interesting for protein engineering as well.

4.1.2.1 Side Chain-Modified Amino Acids

Although the inductive effect of the highly electronegative fluorine substituents affects the basicity of the α-amino function of the side chain-substituted amino acid, N-terminal and C-terminal incorporation of these non-natural building blocks into peptides can be accomplished by standard coupling routines[155] in SPPS chemistry.[157] Furthermore, monofluorinated surrogates and alkyl-fluorinated analogues of aliphatic residues can be inserted into peptides and proteins by biotechnological means, using auxothrophic bacterial strains.[158] If a certain native amino acid is absent in the media, such bacteria use its altered building block for protein synthesis. However, only minimal variations in side chain structure, such as hydrogen by fluorine replacements are accepted. This method is not adaptive to the insertion of fluorinated analogues of non-natural amino acids. Consequently, the

incorporation of sterically and/or electronically demanding fluoroanalogues needs alternative routes.

More than 100 different non-natural amino acids, including several monofluorinated derivatives, have been biosynthetically incorporated into proteins by applying nonsense codon suppression methodologies.[156] This approach uses suppressor tRNAs that are aminoacylated with the non-natural amino acid of interest in cell-extract or cell-intact translation systems. The methodology is an excellent tool for the site-specific amino-acid incorporation, covering a wide range of structural diversity. However, a considerable drawback of this method is the complexity of molecular biological, biochemical, and organochemical techniques that have to be combined. Alternatively, rationally or genetically engineered tRNA-synthetases can be utilized that connect the non-natural amino acid to a specific tRNA.[157] Consequently, no chemical loading of the tRNA is required.

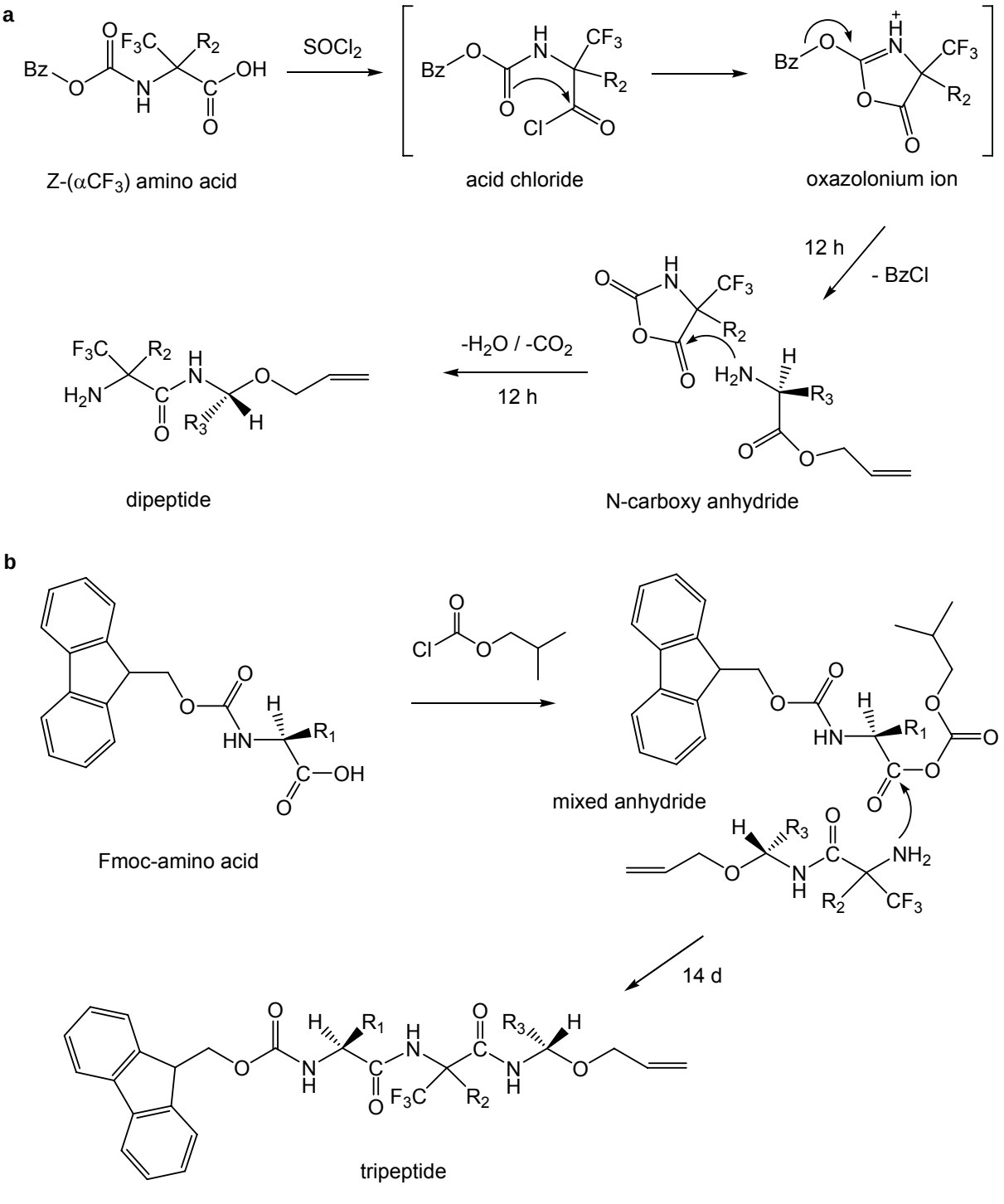
4.1.2.2 $C^{\alpha,\alpha}$-Dialkylated Amino Acids

The incorporation of $C^{\alpha,\alpha}$-dialkylated amino acids into peptides is much more challenging, compared to their C^{α}-monosubstituted analogues. Thereby, the structure as well as the position and the content of fluorine within the amino acid determine the incorporation strategy. Fluoromethylation at the C^{α}-atom leads to a situation in which the high steric demand in combination with the strong electron-withdrawing effect drastically lowers the reactivity of the amino function.[158]

Several (αCF_2H), (αCF_2Cl), and (αCF_2Br) amino acids have been successfully incorporated into peptides by applying *in situ* peptide coupling protocols.[159] The coupling of these amino acids to the peptide N-terminus can easily be accomplished with the reagent combination DIC/HOAt. The use of amino acid fluorides and chlorides as well as certain in situ coupling procedures, using newly developed guanidinium and phosphonium reagents, give excellent yields for the much more demanding incorporation of these building blocks at the C-terminus of the peptides.

In contrast, $C^{\alpha,\alpha}$-trifluoromethylated amino acids demand elaborated solution methods.[160] For the synthesis of longer peptides containing (αCF_3)amino acids via SPPS, it is often required to synthesize tripeptides with the fluorinated building block in the middle position that can then be inserted into the polypeptide chain by standard protocols (Scheme 4.5). This strategy was applied for the insertion of (αCF_3)alanine into peptides in this work. The tripeptide synthesis demands very strong activation procedures in both the reaction of the αTfm-amino acid as acyl donor and the coupling to its amino group. Therefore, anhydride formation as carboxyl group activation can be used in either of the coupling steps. N-carboxy anhydrides

(NCA) are prepared by heating the Z-protected (αCF_3)amino acid with thionyl chloride (Scheme 4.5a).



Scheme 4.5: *Synthesis of tripeptides with an (αCF_3) amino acid in the middle position;* ***a)*** *Dipeptide synthesis via N-carboxy anhydride (NCA) activation of the Z-(αCF_3) amino acid and* ***b)*** *Coupling of an Fmoc-amino acid to the α-amino group of the dipeptide, applying the strategy of mixed anhydride for carboxyl group activation.*

This amino acid coupling method can be used for (αCF_3)amino acids because the nucleophilicity of the α-amino group of the newly formed dipeptide is dramatically lowered due to the inductive effect of the trifluoromethyl group. Otherwise, this function would compete with the amino acid ester as a nucleophile, resulting in oligomerization. While the carboxyl group activation via mixed anhydride strategy proceeds very fast, the following coupling of the activated amino acid to the α-amino function of an (αCF_3)amino acid can take weeks (Scheme 4.5b). For comparison, amino acid coupling reactions in standard solid phase peptide chemistry are usually finished after 20 min. For (αCF_3)amino acids with bulky side chains, such as (αCF_3)Leu or (αCF_3)Phe, the anhydride activation procedure fails. In these cases, advanced organochemical methods have to be applied.[161]

The insertion of (αCF_3)amino acids into polypeptide chains via the described chemical solution methods demands time consuming reactions with often insufficient yields. An attractive alternative to classical peptide chemistry is presented by the combination of chemical and enzymatic methods. The valuable characteristics that enzymes provide are racemization-free catalysis under mild reaction conditions, high regio- and stereoselectivity, and the requirement of only minimal side chain protection.[162] Successful enzymatic fragment condensations of both (αCF_3)amino acid-containing acyl donors and –acyl acceptors with a variety of electrophilic or nucleophilic peptide fragments, respectively, have been achieved under catalysis with several proteases.[163]

The direct coupling of (αCF_3)Ala and (αCF_2H)Ala methyl esters as acyl donors to different amino acid amides can be performed using Carboxypeptidase Y (CPY), an enzyme with a broad substrate specificity.[164] Both enantiomers of either of the fluoroalkyl-substituted peptide building blocks are converted by the enzyme without any further activation. The direct enzymatic coupling of $C^{\alpha,\alpha}$-fluoroalkyl substituted amino acids, however, is still a challenging topic because of the distinct substrate specificity of proteases. The substrate mimetic concept helps to overcome these limitations.[165] In this strategy, a 4-guanidinophenyl ester (OGp) functionality within the enzyme substrate mediates the acceptance of non-specific amino acids for Arg-specific proteases, such as trypsin, thrombin, and clostripain, as well as of the Phe-specific protease α-chymotrypsin (Figure 4.1). The mimicking group can be introduced easily via *in situ* coupling procedures and is cleaved during enzymatic condensation reaction. Several (αCF_2H)- and (αCF_3)amino acid guanidinophenyl esters are accepted by trypsin as acyl donors and can be coupled directly to various nucleophiles of different length and sequence.[166]

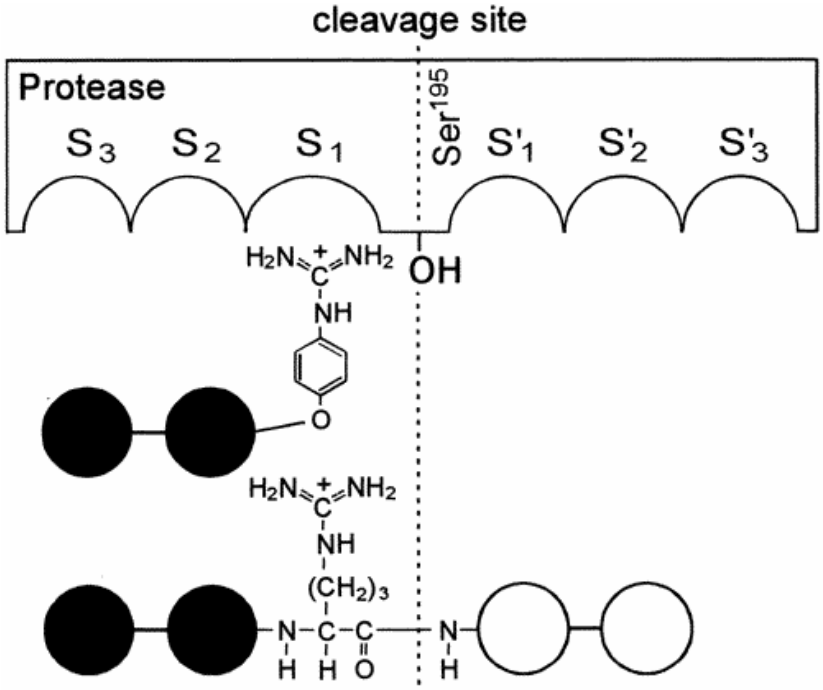


Figure 4.1: *Schematic representation of the binding of a substrate mimetic (OGp ester) to the active site of trypsin (according to Thormann et al.[167])*

4.2 Structural Requirements for a Peptide-Based Screening System

The systematic evaluation of the properties and molecular interactions that fluorinated moieties exert within native polypeptide environments demands a model polypeptide that contains the fluoroalkyl-substituted amino acids as well as methods for the detection and characterization of the investigated features. The properties of fluoroalkylated amino acids like steric demand, their hydrophobicity, polarity, and specific interactions with all kinds of native amino acids will affect the structural stability of the model polypeptide as well as its folding behavior. The aim is to develop a versatile screening system that allows the study of many different fluoroalkyl-substituted amino acids differing in size, content of fluorine, and position of fluorination within an appropriate time, with manageable costs, and moderate amount of work. Therefore, several important requirements for such a peptide-based screening system have to be fulfilled, in order to obtain reproducible and comparable results:

1) The model polypeptide must be easy to synthesize and the incorporation of fluorinated building blocks has to be accomplished by effective peptide coupling methods.

2) The model polypeptide has to be of manageable size and must fold into a predefined and well-known 3D-structure, forming specific interfaces, such as hydrophobic and polar interaction domains.

3) The model polypeptide must contain substitution positions for the fluorinated building blocks in different interaction domains. These residues have to have defined amino acid positions as their native interaction partners within a specific interfaces. Therefore, the amino acid substitutions have to be tolerated by the 3D-structure of the model polypeptide. That means that the overall fold must be stable towards all alterations to be made but sensitive enough to detect changes.

4) The methodologies that are used for the detection of the influences, the inserted fluoroalkyl moieties exert on structural stability and folding behavior of the model system, have to be sensitive enough to differentiate between amino acid side chains of minimal alterations. Since a systematic investigation requires the screening of multiple substituted amino acids, the experimental effort of these methods must be managable.

4.3 The α-Helical *Coiled Coil* Folding Motif

To accomplish points 1 to 3 of the above-mentioned requirements, the α-helical *coiled coil* motif can perfectly serve as such a model system. Approximately 3% of all protein residues in nature form *coiled coils*,[168] which shows that this structural motif is very common in biological systems. It is found in transcription factors,[169] tropomyosin,[170] chaperones,[171] and many more.[172]

4.3.1 Structure and Design Principles of the Folding Motif

Since its first description in 1953 by Crick,[173] the structural principles of the *coiled coil* motif have been analyzed in detail and have become very well understood.[174] Review articles that summarize the results appear frequently.[175,176,177] The *coiled coil* folding motif is known to possess an extraordinarily high structural stability even when composed of relatively short peptide strands. *Coiled coils* typically consist of two to five right-handed α-helices, which are wrapped around each other in either parallel or antiparallel orientation to form a left-handed superhelical twist (Figure 4.2).

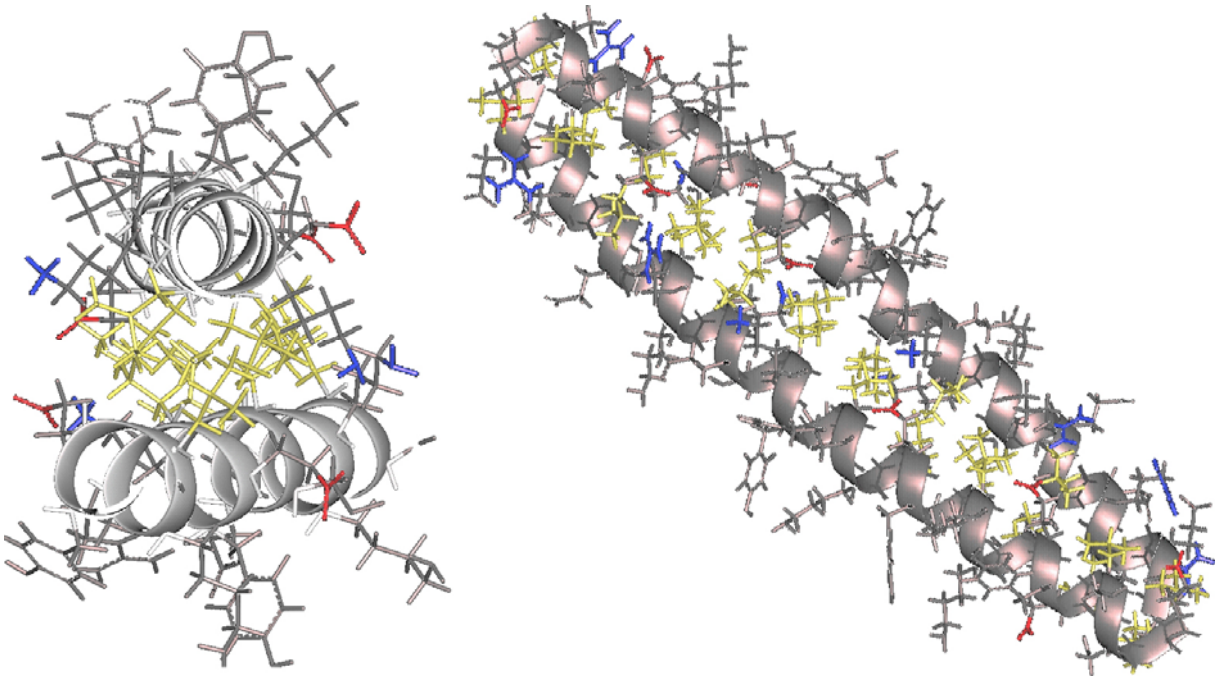


Figure 4.2: *View along the superhelix axis **(left)** and side view **(right)** of a model of a dimeric coiled coil peptide. The hydrophobic core is shown in yellow and functional groups of the amino acid side chains within the charged interaction domain are highlighted in blue (positively charged groups) and red (negatively charged groups).*

The primary structure of each helix is characterized by a periodicity of seven residues, the so-called 4-3 heptad repeat, which is commonly denoted as $(a\text{-}b\text{-}c\text{-}d\text{-}e\text{-}f\text{-}g)_n$. The positions **a** and **d** are typically occupied by nonpolar residues (Leu, Ile, Val) and form a special interaction surface at the interface of the helices by hydrophobic core packing ("knobs-into-holes"). In contrast, the positions **e** and **g** are solvent-exposed and often occupied by charged amino acids (Glu, Lys, Arg), forming inter-helical salt bridges in a charged interaction surface (Figure 4.3).

In the remaining heptad repeat positions **b**, **c**, and **f**, located solvent exposed at the outer sides of the motif, polar residues are mostly found, which can stabilize the conformation by intra-helical salt bridges.[178] The hydrophobic core provides the major contribution to the structural stability of the *coiled coil,* while the allocation of the two positions **a** and **d** of this interface by the different types of hydrophobic residues, β- and γ-branched, controls the order of aggregate fomation. In contrast, the inter-helical ionic pairing positions **e** and **g** mainly dictate the orientation specificity (parallel versus antiparallel) as well as the preference for homo- or heterotypic *coiled coil* formation. Amino acids in these positions provide an abated contribution to the overall stability of the structural motif and the determination of the oligomerization state, compared to that of the hydrophobic core residues.

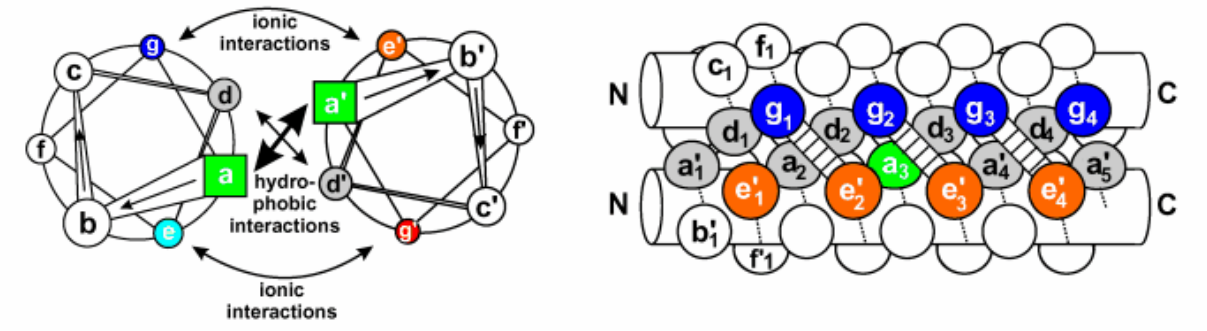


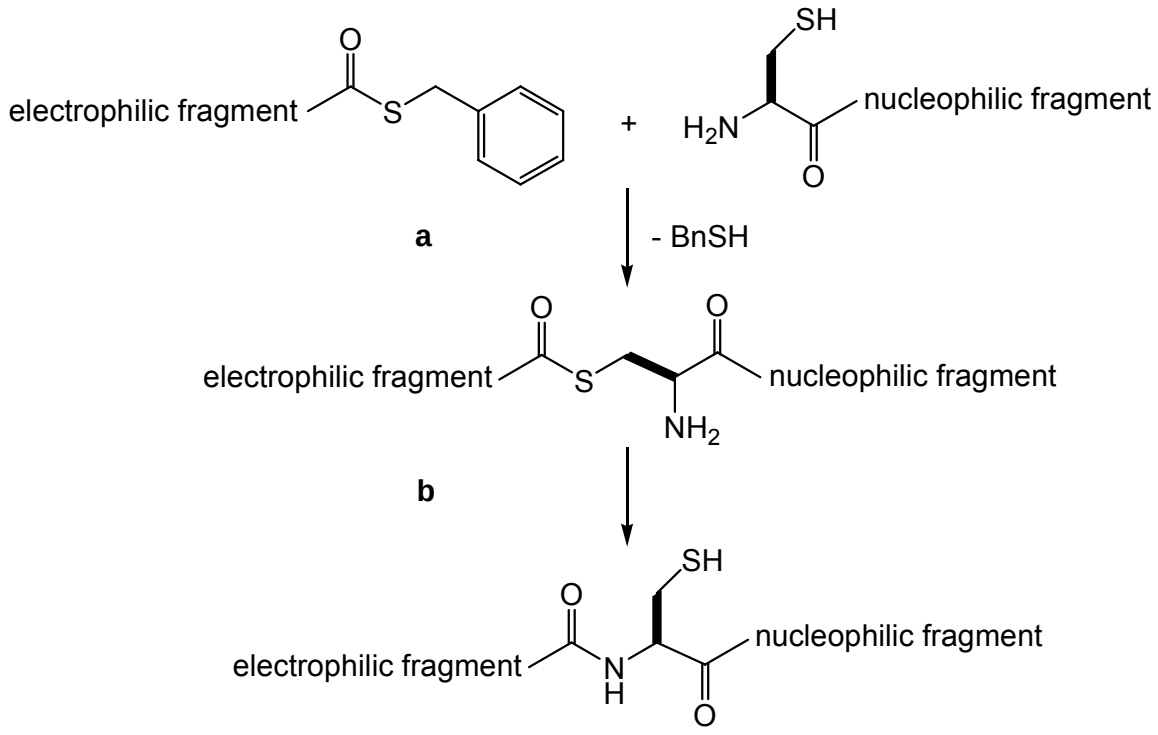
Figure 4.3: *Helical-wheel representation **(left)** and schematic side view **(right)** of a parallel coiled coil dimer, illustrating the specific interaction surfaces of this folding motif (according to Mason et al.)*

These features can, therefore, be used for the *de novo* design of *coiled coil* peptides. Furthermore, structure and functional properties of *coiled coils* can be improved via substitution of the hydrophobic residues in positions **a** and **d** to form alanine-,[179] phenylalanine-,[180] or tryptophan-based cores.[181]

4.3.2 Self-Replication Properties of *Coiled Coil* Peptides

As described above, the structural stability of *coiled coil* peptides and proteins as well as their defined and highly ordered hydrophobic and polar interaction surfaces makes this folding motif a perfect model for the investigation of the interaction properties of fluorinated residues with native amino acids within a native polypeptide environment. Additionally, their ability to self-replicate offers the possibility to evaluate the influence these non-natural building blocks exert on protein folding processes.

Chemically, this self-replication process is based on native peptide ligation (Kent ligation). Two peptide framents, an electrophilic peptide with a C-terminal thioester and a nucleophilic peptide with an N-terminal cysteine residue, ligate to a full length peptide. Thereby, a peptide-like amide bond is formed (Scheme 4.6). In a first step, a transesterification occurs between the C-terminal thioester moiety of the electrophilic peptide fragment and the thiol group of the N-terminal cysteine side chain of the nucleophilic peptide fragment. This initial thioester ligation product undergoes an intramolecular reaction by a nucleophilic attack of the α-amino group, forming a native peptide bond. Since this reaction is highly chemo- and regioselective, peptides can be used for ligation without protection of their functional backbone and side chain groups.



Scheme 4.6: *The native ligation reaction based on Kent strategy.* ***a)*** *initial transesterification reaction between the C-terminus of the electrophilic peptide and the N-terminal cysteine side chain of the nucleophilic peptide, followed by* ***b)*** *intramolecular rearrangement forming the peptide bond.*

If the two peptide fragments that are modified for Kent ligation can fold into *coiled coil* structure by annealing to a monomeric full-length *coiled coil* peptide strand, the reactive groups for native ligation are brought into close proximity, which accelerates the reaction. The full-length peptide strand acts as a ligase catalizing the ligation of the shorter peptide fragments by tertiary structure formation.[182] In the case of *coiled coil* homo-oligomerization, the ligation product represents the ligase for its own reproduction. The peptides undergo a self-replication (Figure 4.4).[183] The products formed in this auto-catalyzed reproduction cycle are identical copies of the template itself. The initial step is an un-catalyzed peptide bond formation between both ligation fragments, forming a full-length *coiled coil* strand (the ligase) for catalysis of the self-replication rounds.

Protein folding processes in this replication cycle, the association of the peptide fragments with the template and the dissociation of the ligated full-length *coiled coil*, are influenced by amino acid side chain interactions within the specific interaction domains as well as between the unfolded monomeric peptides. Therefore, the impact of fluorine-containing non-natural peptide building blocks on protein folding can be studied by measuring the turnover rate of the self-replication of substituted *coiled coil* peptides or proteins.

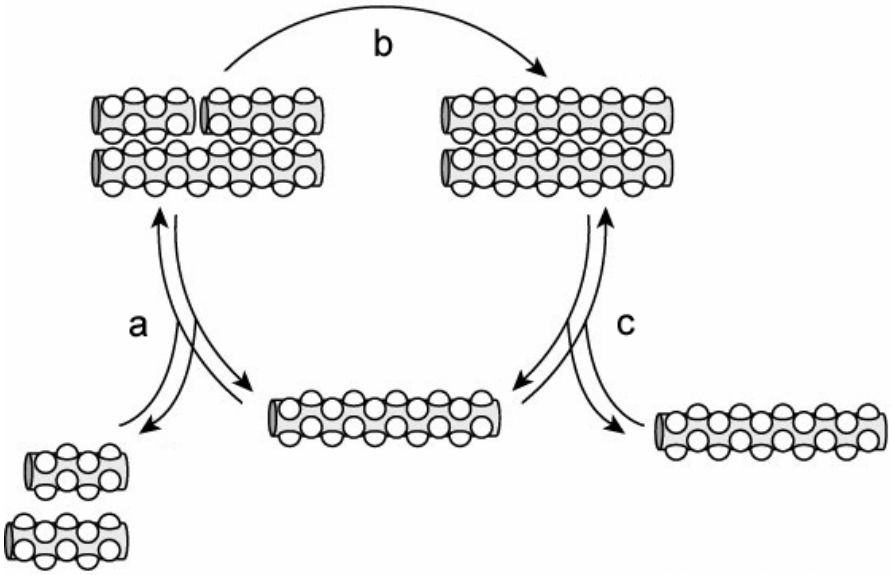


Figure 4.4: *Self-replication cycle of coiled coil peptides;* ***a)*** *two peptid fragments form a coiled coil with a full-length template;* ***b)*** *structure formation catalizes native ligation;* ***c)*** *dissoziation of the peptide strands provide new monomeric templates for further replication rounds (adapted from Saghatelian et al.[184])*

4.4 Phage Display Technology[185]

A further feature of the peptide-based screening system that had to be established by this work is that it can be used to select the preferred interaction partners for fluoroalkyl-substituted amino acids out of a pool of twenty canonical amino acids. One possible concept for the production of peptide libraries comprise multiple synthesis methods using combinatorial chemistry.[186] The big advantage of this strategy is the possibility to incorporate a wide variety of non-amino acid entities to add both constrained and chemical variation to the libraries.

The major drawback of screening such chemistry-derived synthetic libraries is the large complexity of the synthetic lab work. Furthermore, the measured binding properties of all of the library members have to be evaluated and characterized in order to select the best binders. This means there is a data set of n^{20} when screening a library with n randomized amino acid positions. Using iterative screening strategies, new libraries with reduced complexity are synthesized and tested based on the outcomes of the former selections. The key to controlling complex libraries is to link protein recognition and DNA replication. This way, the library is introduced on a DNA level and screened on a peptide or protein level. The main idea behind this concept is to capture only the best binders in a library screen and the

following amplification of the genes encoding for them. The selected genom can then be used for further (and stronger) selection or to characterize the peptide library members.

A common methodology that follows this concept is phage display technology. The peptide or protein that contains the library is displayed on the surface of bacteriophage particles, which are viruses that infect bacteria. The genes encoding it are encased by the particle. Individual phage can then be rescued from libraries by interacting of the displayed protein with the cognate ligand, that contains the interaction partners for the randomized amino acid positions. DNA can be amplified by phage infection of bacteria. Different types of bacteriophage have been used for library selection. Since only M13 was used in this work, the following sections will exclusively deal with this phage.

4.4.1 The Filamentous Phage M13

The filamentous bacteriophage M13 is a member of the group of viruses that contain a circular single-stranded DNA genome, encased in a long protein capsid cylinder. The Ff class of the filamentous phage, which includes M13, uses the tip of the F conjugative pilus as a receptor to facilitate the infection of Escherichia coli, containing the F plasmid. M13 does not kill its host during productive infection. The single-stranded viral DNA is replicated in the cytoplasm of the bacteria and the capsid proteins are all synthesized as integral (cytoplasmic) membrane proteins. After assembly of these proteins around the DNA, about 1000 phage in the first generation and about 100-200 particles in all further ones are released, resulting in titers of 10^{11} to 10^{12} phage per ml. Since the design of a phage-displayed peptide library and the development of a corresponding screening method requires detailed knowledge about structure and life cycle of the bacteriophage, these topics will be detailed below.

4.4.1.1 Structure and Genome of Bacteriophage M13

The M13 particle is approximately 6.5 nm in diameter and 930 nm in length. The genome is a single-stranded, covalently closed DNA molecule of about 6400 nucleotides. It encodes 11 genes which are grouped in the genome, according to their function in the life cycle of the phage (Figure 4.5). One group of them, genes *II*, *V*, and *X*, encode proteins that are required for the replication of the phage genome. A second group that comprises genes I, XI, and IV, encode proteins that are involved in the membrane-associated assembly of the bacteriophage. The final group encodes the minor capsid proteins pIII, pVI, pVII, and pIX, as

well as the major coat protein pIX. Two strong terminators of transcription divide the genome into two transcription regions: a frequently transcribed region containing genes *II* through *VIII*, where transcription initiates from three strong promotors, and an infrequently transcribed region, containing genes III through IV, where transcription initiates from two promoters (P).

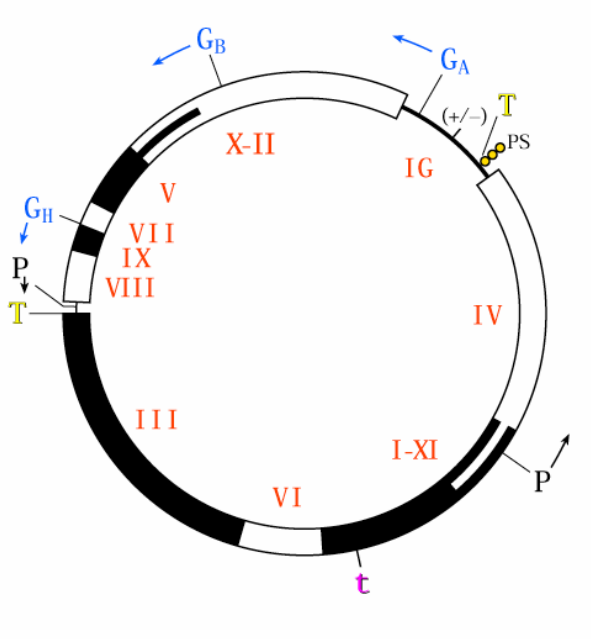


Figure 4.5: *Genome of the bacteriophage M13. Roman numbers denote the genes. IG refers to the intergenic region; T, the two strong terminators; t, the weak terminator in pI; G_A, G_B, and G_H, the promoters of the frequently transcribed region; P, the promoters for the infrequently transcribed regions; PS, the packaging signal; and (+/-), the relative position of the origins of replication for the viral (+) and complemetary (-) DNA strands (according to Barbas et al.)*

The phage capsid consists of the five capsid proteins. The length of the cylinder is made of approximately 2700 molecules of the 50-amino acid major coat protein pVIII (Figure 4.6). The pVIII monomers form uninterrupted α-helices (except for the amino-terminal five residues) and are tightly packed by gently wrapping around the long axis of the virus in a right-handed way. Thereby, the middle parts of the proteins interact with each other and, thus, form the stable inner core of the phage cylinder. Four positively charged lysine residues at the inside-wall-forming carboxy terminus of each pVIII molecule interact with the sugar phosphate backbone of the DNA. At one end of the phage particle, there are about five copies of each of the minor coat proteins pVII (33 residues) and pIX (32 residues). The other end contains approximately five copies each of the minor coat proteins pIII (406 rsidues) and pVI (112 residues).

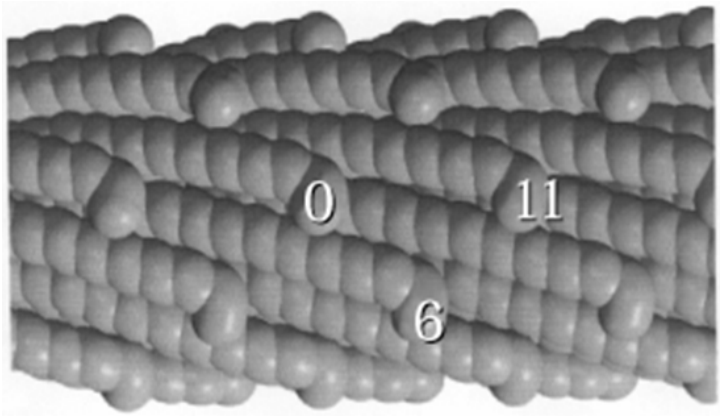


Figure 4.6: *Representation of the orientation of the pVIII molecules along the cylinder part of the phage. The three nearest neighbors of one pVIII molecule are indexed as 0, 6, and 11 (according to Barbas et al.)*

4.4.1.2 Infection and Reproduction of Bacteriophage M13

Besides the F pilus a protein tube that is assembled from pilin subunits and extends from the cytoplasmic membrane out into the medium, three proteins (TolQ, TolR, and TolA) that are anchored in the cytoplasmic membrane with the bulk of their residues exposed in the periplasm are indispensable tools in phage infection. Infection is initiated by the binding of the tip of the bacterial F pilus to the pIII protein of the phage coat. After binding, the pilus retracts, bringing the pIII end of the phage particle to the periplasm. As a result, TolA can act as a coreceptor for the attachment of the N-terminal part of the pIII protein. During the infection process, major coat proteins pVIII disassemble into the cytoplasmic membrane as the phage DNA is translocated into the cytoplasm. It is suggested that TolQ and TolR are involved in the formation of a channel, which is required for the DNA to traverse the membrane.

Once the viral DNA has entered the cytoplasm, bacterial enzymes synthesize the complementary (-) strand and convert it into negative supercoils, the so-called parental replicative form RF (Figure 4.7). The (-) strand of this RF is the template for transcription, finally resulting in newly synthesized phage proteins. Two of them, pII and pX, are involved in the replication of the viral RF DNA. Since DNA replication causes enhanced production of viral protein, pV accumulates in the cytoplasm. When the amount of pV reaches a critical concentration, dimers of this protein bind newly synthesized single-stranded DNA and prevent its conversion to RF DNA. Thus, DNA replication is switched to a synthesis of single-stranded pV-DNA complexes, which can be packed into new phage particles.

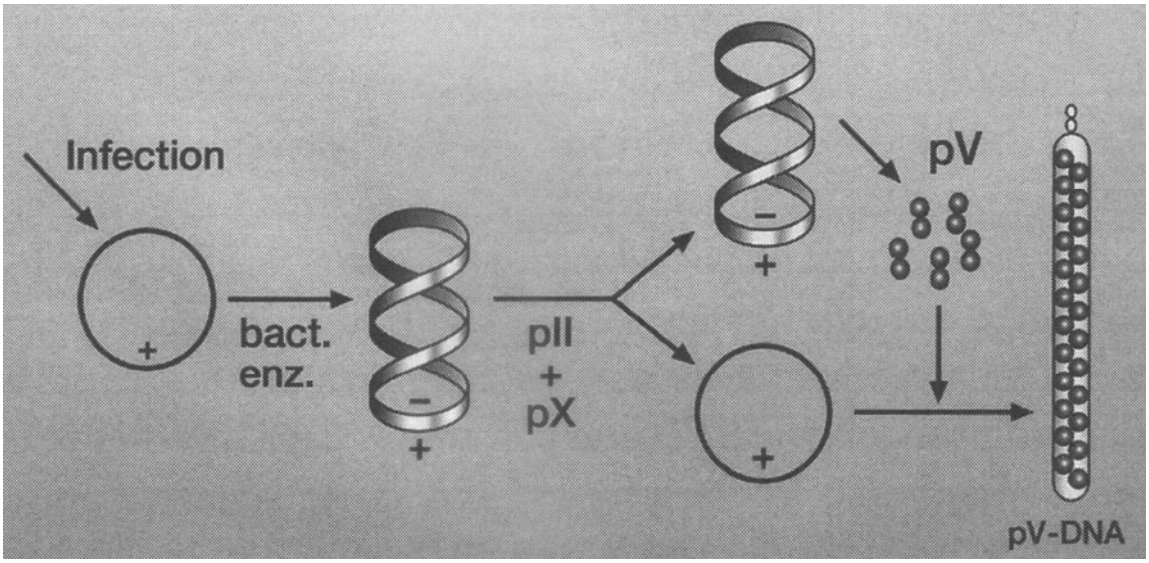


Figure 4.7: *DNA replication in the filamentous bacteriaphage M13. (+) refers to the viral strand and (-) the complimentary strand (according to Barbas et al.)*

The newly synthesized phage-coat proteins pVIII, pIII, pVI, pVII, and pIX accumulate in the cytoplasmic membrane for the assembly of new particles. Thereby, the C-terminal region of pIII constitutes the membrane-spanning part, while the major portion of this 406-amino acid protein is located in the periplasmatic space. The phage proteins pI and pXI are integrated in the cytoplasmic membrane and pIV in the outer membrane, respectively, to form some type of channel for the release of the emerging phage particle. During the phage production process that takes place within these cylinders, the pV dimers are removed and the capsid proteins are assembled around the single-stranded viral DNA as it is extruded through the assembly site (Figure 4.8). Thereby, the end of the particle containing the coat proteins pVII and pIX interacts with the DNA packing signal (PS) and emerges first from the bacterium. During the packing of the pVIII molecules around the DNA, the positively charged carboxy-terminal portions interact with the DNA, while the transmembrane portions interact with each other to form the capsid tube. When the end of the DNA is reached, assembly is terminated by the addition of the coat proteins pVI and pIII. The process of phage production and release additionally requires ATP and a proton motive force.

For peptide library display on a phage surface, the DNA encoding for the randomized peptide or protein has to be cloned into the gene of a phage coat protein. The altered protein will then be produced in bacteria and incorporated into the capsid of the emerging bacteriophage particle.

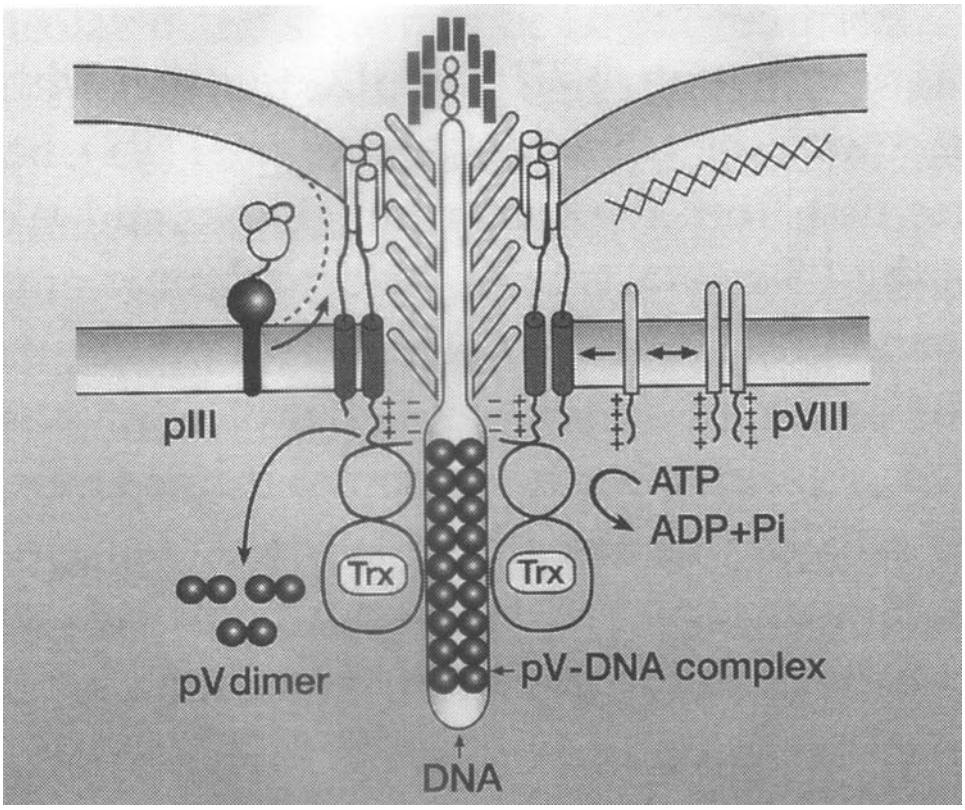


Figure 4.8: *Schematic representation of the elongation process on the assembly of the bacteriophage M13. Trx refers to thioredoxin (according to Barbas et al.)*

4.4.2 Display of Peptides and Proteins on Phage Particles

In general, any kind of protein or peptide can be displayed when fused to a capsid protein. A possible limitation is given by the fact that the artificial protein has to be connected to the periplasmic portion of the phage capsid protein in regard to its orientation when accumulated in the inner membrane before phage assembly (Figure 4.8). As a consequence, only chimeric proteins have a chance to be inserted into phage particle which are translocated efficiently through this bacterial membrane. Furthermore, they must not interfere with the processes that occur at the assembly site and have to fold properly. Whether a desired protein is displayed as fusion with a certain capsid protein depends on the bacterial physiology and is hardly predictable.

4.4.2.1 General Strategies for Phage Display

Although successful attempts have been made to display proteins at the carboxy-terminal portion of pVI as well as at the amino terminus of the minor coat proteins pVII and pIX, the

most common strategies are to produce chimeric pVIII and pIII proteins. Here, the artificial part is connected to the amino-terminal portion of the phage coat proteins. Therefore, the DNA sequence encoding the peptide or protein to be displayed is generally inserted between the sequence of the pVIII / pIII signal peptide and the amino-terminal coding region for the mature capsid protein. Display on the 50-amino acid major coat protein pVIII leads to phage particles with the peptide exposed along the surface of the phage (Figure 4.9).

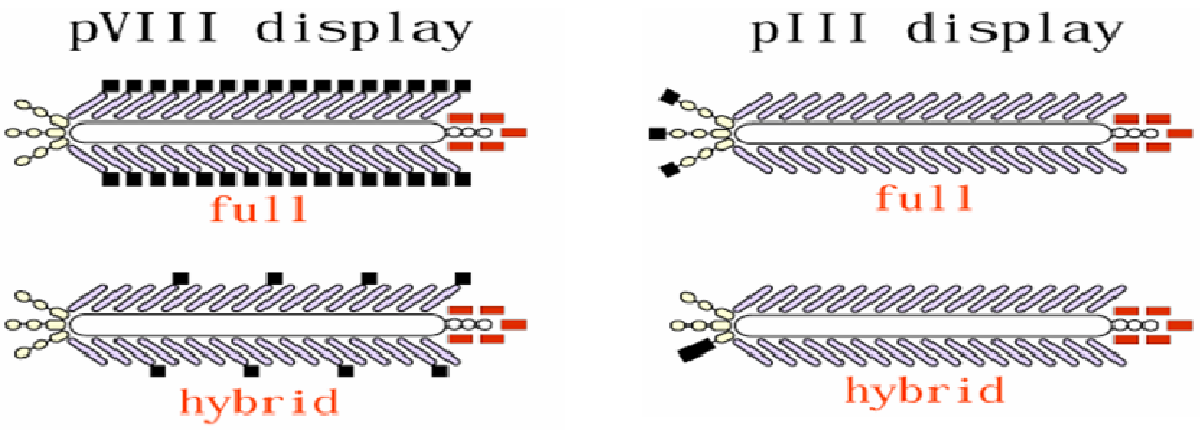


Figure 4.9: *Display of proteins on the filamentous bacteriophage M13. The dark solid symbols attached to pVIII or pIII, respectively, represent the foreign peptides or proteins (according to Barbas et al.)*

Since the phage cylinder is made of thousands of pVIII molecules and has to pass through the 7-nm pIV exit pore during particle assembly and release, the phage production process only accepts pVIII chimera with a fused peptide, which is limited to 6 or 8 amino acids. In contrast, chimeric molecules of the 406-amino acid minor coat protein pIII, containing large inserts, appear to be packed into the phage reasonably well. Since the insertion places the foreign protein at the very end of the packaged particle, it creates less steric hindrance when passing through the pIV exit pore during phage production. The disadvantage of this method is the limited display of only up to five molecules per phage particle (Figure 4.9). The fact that the pIII protein initiates phage infection constitutes a further problem of this strategy. Large inserts tend to lower phage infectivity or even make the phage non-infectious. A general solution to overcome the limitations of both fusion strategies to some extent is the production of smaller filamentous particles called phagemids. On the surface of such hybrid phages only a certain fraction of the pIII or pVIII molecules, respectively, are presented as fusion with the peptide or protein to be displayed (Figure 4.9).

While phage with completely chimeric pVIII or pIII proteins, respectively, is produced by cloning the foreign peptide-encoding DNA into a wild-type phage vector (called type 8 for pVIII display and type 3 for pIII display and containing the full phage genome), hybrid phage

is produced by the additional use of phagemid vectors (type 8+8 or type 3+3). Since only hybrid pIII display was used in this work, the next subchapter will exclusively cover this strategy.

4.4.2.2 Phagemid Display on Minor Coat Protein pIII

The phagemid genome contains the filamentous phage intergenic region with its origin of replication for viral and complementary strand synthesis as well as the hairpin packaging signal. It also contains a plasmid origin of replication and a gene encoding resistance to a specific antibiotic. In contrast to wild-type phage genomes, the phagemid vector does not contain the genes for all phage proteins but the one for the capsid protein used for chimeric fusion. Thus, the library to be displayed is cloned to the amino-terminal portion of the coat protein in the phagemid vector. The phagemid vector can maintain itself as a plasmid, which can direct the expression of the protein in bacteria, if desired.

The infection of the phagemid vector-containing bacteria with wild type phage, the so-called helper phage, activates the phage origin of replication and provides all of the phage derived proteins and enzymes required for phage replication and encapsulation of the genome. As a result, phage particles are produced in bacteria that contain both the chimeric and the wild type of the coat protein encoded on the phagemid vector. Two types of infectious particles are produced from cells: those containing the phagemid genome and those containing the wild type genome. Since phage are needed that bear the gene encoding for the displayed library peptide, helper phage with defective origin of replication or packaging signal are used, which allows the preferential packaging of the phagemid genome over the helper-phage genome.

Since pVIII display is limited to shorter peptides even when phagemid strategy is applied, the affinity selection from libraries, generated from larger polypeptides or proteins, has to be performed using pIII display. This methodology is greatly affected by pIII structure and function (Figure 4.10a). The amino-terminal domain of pIII is required for phage infectivity. The first part, N1, contains the amino-terminal 68 amino acids and mediates transloctaion of the DNA into the cytoplasm and the insertion of the coat proteins into the inner membrane during infection. The second part, N2, is made up of residues 87-217 and is responsible for binding to the F pilus. The carboxyterminal 150 residues make up the third domain of pIII, CT, and are essential for forming stable phage particles. All three parts of pIII are connected by glycine-rich linkers. A further important function of the amino-terminal domain of pIII is its ability to provide immunity to a cell already infected with phage. As a result, such bacteria are prevented from superinfection by other filamentous phage.

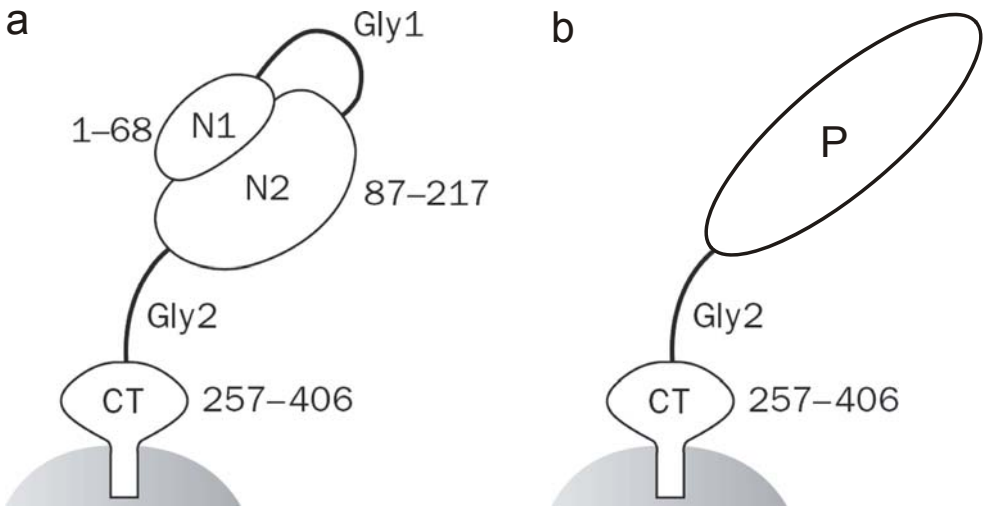


Figure 4.10: *Schematic representation of the structure of pIII, showing* ***a)*** *wild type coat protein containing domains N1, N2, and CT as well as the glycine-rich linkers Gly1 and Gly2 and* ***b)*** *phagemid encoded chimeric pIII with domains N1-N2 substituted by the library peptide or protein (P).*

Consequently, the chimeric pIII protein encoded in the phagemid vector has to be truncated by both amino-terminal domains, N1 and N2, otherwise infection by helper phage would be inhibited (Figure 4.10b). Since both types of pIII molecules, wild type encoded on helper phage genome and chimeric encoded on phagemid vector, will be present on the surface of the particle, wild type pIII will mediate infection via N1 and N2 domains and chimeric pIII bears the protein library. This strategy circumvents the decrease in phage infectivity that occurs when the foreign protein is displayed on all full-length pIII molecules (type 3 display) and enables the successful display of large proteins, such as antibody fragments on phage surface. Since both types of pIII compete for incorporation into the five pIII position within the virion, the average valency of the fusion protein on the phage particle can be modulated by controlling the expression of the phagemid cassette. A phagemid vector that can be applied for type 3+3 phage display is pComb3H, a derivative of pComb3[187] that contains the gene for amino acids 230-406 of pIII, which constitute the CT domain and the major part of the flanking glycine-rich linker.[188] The DNA for the protein to be displayed is ligated to the 5'-end of this gene resulting in the desired fusion protein.

Once the fusion protein is displayed on phage, binding of the foreign protein to the cognate ligand has to be proven before affinity selection from libraries can be performed. Since several strategies for such binding assays have been tested and applied in this work, the principles of these methodologies will be covered by the sections that describe the procedures used in detail.