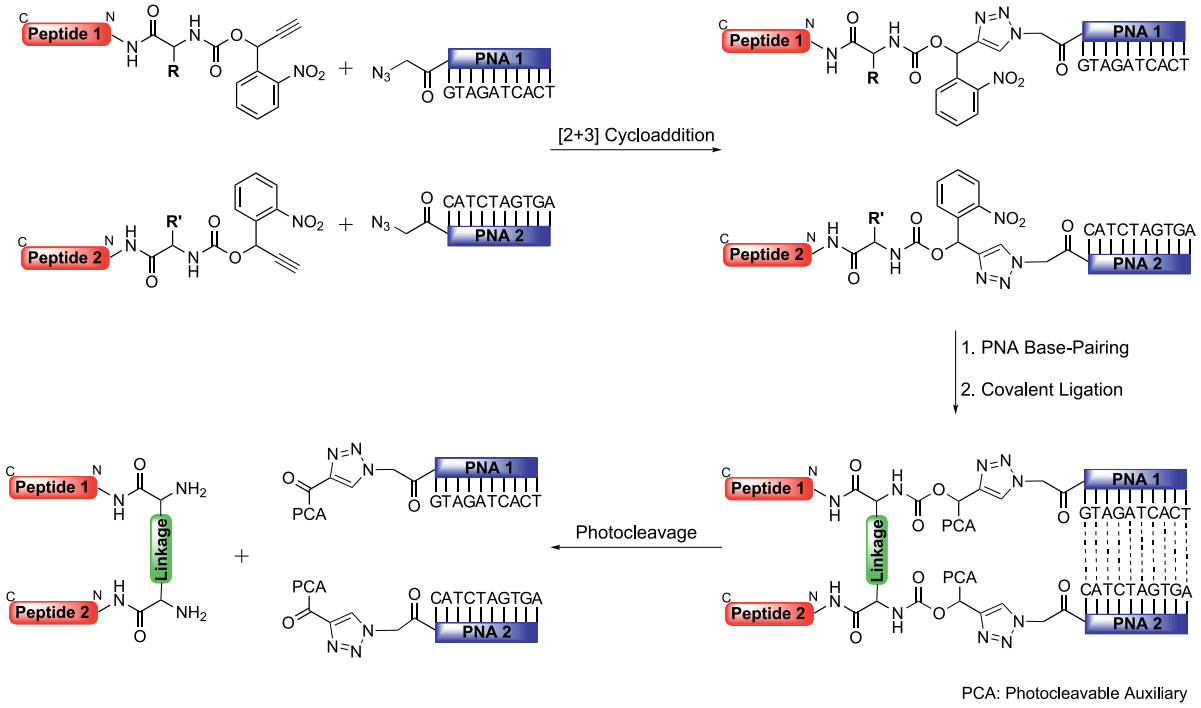
1 Introduction

Proteins are involved in countless cellular transformations. Owing to their diverse and significant role in biological systems, biologically relevant macromolecules have gained attention as target structures in order to elucidate their functions and to design novel bioactive materials with respect to medical applications.[1,2] Especially, site selective functionalizations of proteins allow the examination of the spacial and temporal localization, dynamics and function of biologically active samples in a cellular environment and more importantly even at the molecular level.[3] However, in many cases selectively functionalized proteins are difficult to access by applying recombinant techniques as these are limited to genetically encoded amino acids and site directed modifications are often impossible to control.[1]
The total or semi-synthesis of proteins utilizing chemical transformations provides a solution to overcome these limitations. Chemical protein synthesis is a powerful tool that allows to introduce biophysical probes, fluorophores or post translational modifications such as phosphorylation or glycosylation in a selective manner.[3–5] The methods of choice are to either utilize artificial amino acid derivatives or to manipulate the reactive moieties in the side chains of amino acid residues. Herein, orthogonal bioconjugation strategies often exploit the distinct reactivity of the sulfhydryl moiety of cysteine residues benefiting from its low abundance in native proteins. Modifications are achieved by alkylation or the formation of mixed disulfide bonds. Other approaches involve the transformation of the amine group or the hydroxy moiety of lysine or serine, respectively.[6] Modern chemical synthesis techniques do not only enable bioorthogonal labeling but furthermore, provide researchers with invaluable tools to prepare protein domains. The inherent limitations of the stepwise solid phase peptide synthesis have stimulated the development of synthetic methods to sequentially assemble unprotected peptide fragments in order to obtain active proteins.[4] In particular, the native chemical ligation (NCL) strategy which yields a native peptide bond at the ligation junction revolutionized the scope of chemical protein synthesis.[7,8] This most frequently applied technique allows to combine two unprotected peptide fragments by an irreversible S $\rightarrow$ N acyl shift between an *N*-terminal cysteine residue and a *C*-terminal thioester. The numerous applications emphasize its robustness and general implementation.
An example highlighting the success of NCL is the recently reported preparation of a prion protein containing both lipidated and phosphorylated modifications.[9] These glycosylphosphatidylinosityl (GPI) anchored prion proteins are involved in metabolic diseases or infections

affecting brain or neural tissue such as the CREUTZFELDT-JAKOB disease.[10] However, the recombinantly expressed GPI proteins do not contain their membrane anchor, thus limiting their exact study. The successful synthesis of the lipidated GPI protein utilizing chemical protein synthesis methods gave new insight into how this complex posttranslational modification influences the structure and function of prion proteins.[11]
Nevertheless, the advantage of utilizing the unique reactivity of cysteine residues to facilitate the NCL reaction is also its main drawback. The cysteine content in native proteins is with 1.7% extremely low and therefore engineered cysteine residues at the ligation junction are often required, which might have a strong influence with respect to the overall protein folding.[4] A range of desulfurization techniques of cysteine residues are available to circumvent this drawback.[4,12,13] Most commonly, cysteine residues are transformed to yield an alanine at the ligation site.[4] However, the main obstacles are that the thioether of methionine occasionally participates in an unwanted desulfurization reaction or that an epimerization of secondary alcohols during the reaction can occur.[4,14,15]
Nowadays, a number of chemoselective transformations have been identified to either ligate peptide fragments or to allow site selective functionalizations in a chemoselective manner. Prominent examples are the traceless STAUDINGER ligation,[16] the ketoacid-hydroxylamine (KAHA) ligation,[17] copper catalyzed [3+2] cycloaddition[4] or enzymatic methods.[18] However, with our growing understanding of fundamental biochemical processes there is a constant need to develop novel bioorthogonal transformations that do not only permit the selective ligation of peptides but in addition enable chemoselective functionalizations in a competing environment. This will allow to address underlying biochemical problems and provide access to various novel pharmaceutical applications.
In this thesis, a novel chemical ligation strategy is developed which will not only allow the use of a range of ligation techniques to form a permanent linkage between two peptide fragments but above all will provide the possibility of simultaneous functionalizations at different ligation sites in a competing environment (Scheme 1.1). The required chemoselectivity will be facilitated by peptide nucleic acid (PNA) guiding units which are attached to the *N*-terminus of the peptides. Based on the templating effect of two complementary PNA oligomers, derived from the hybridization of canonical nucleobases, the respective residues for ligation are preorganized in close proximity to each other. Moreover, this PNA/PNA mediated ligation strategy provides a general method to address distinct ligation sites which are defined just by their nucleobase sequence. In consequence, this novel ligation strategy offers access to nearly unlimited orthogonality. However, PNA oligomers are just a tool to enable directed alignment and need to be cleaved after the formation of a covalent linkage between the two peptide fragments. This is facilitated by applying a traceless photocleavable auxiliary as a caging group. Within this project, a suitable photocleavable auxiliary is developed in compliance with the requirements of the ligation strategy.



Scheme 1.1: Principle of novel PNA/PNA mediated ligation strategy. After successful formation of a covalent linkage between two peptide fragments the PNA guiding units are cleaved by applying UV light.

In a second project, peptides for specific DNA bending are developed. These peptides mimic the DNA binding and bending abilities of integration host factor (IHF), which is a small and well characterized protein consisting of an α- and a β-subunit and is responsible for inducing extreme conformational changes in double stranded DNA.[19] However, the important elements leading to this high selectivity need yet to be identified. The design of a small peptide which mimics the sequence specific DNA binding and bending characteristics of the α-subunit of the IHF protein has been implemented by the DIEDERICHSEN research group.[20,21] In order to elucidate the important elements in more detail, novel IHF mimicking peptides are presented comprising two DNA recognition units to resemble the native protein in function more closely. It is anticipated that these IHF mimics will expand our comprehension in identifying the origin of the high specificity and in addition will provide a tool which allows to address DNA sequences in a specific manner.

Furthermore, in a cooperation project with the SCHWAPPACH research group the synthesis of a literature known substrate-kinase crosslinker is shown.[22] This crosslinker may help to elucidate the temporal and spacial order of protein kinase A induced phosphorylation of the tandem pore domain acid sensitive potassium channel (TASK) which results in its cell surface localization.

2 Design and Synthesis of IHF Mimicking Peptides Containing Two DNA Recognition Units

2.1 The Integration Host Factor

The integration host factor (IHF) of Escherichia coli (E. coli) is an architectural protein which binds and bends double stranded DNA in a sequence specific manner.[19] Especially, the ability of IHF to provide extreme conformational changes in the structure of DNA by inducing a sharp U-turn of about 160° made it an interesting target for research.[23–29] Initially, IHF was identified as a host factor required for the integration of bacteriophage λ genome into its host chromosome but it became soon apparent that IHF is involved in a range of processes in the cellular machinery.[23,30] Kinking of double stranded DNA is not only important for chromosomal compaction but also for the regulation of diverse cellular functions which requires the assembly of higher-ordered nucleoprotein complexes like replication or translation of DNA.[30,31]

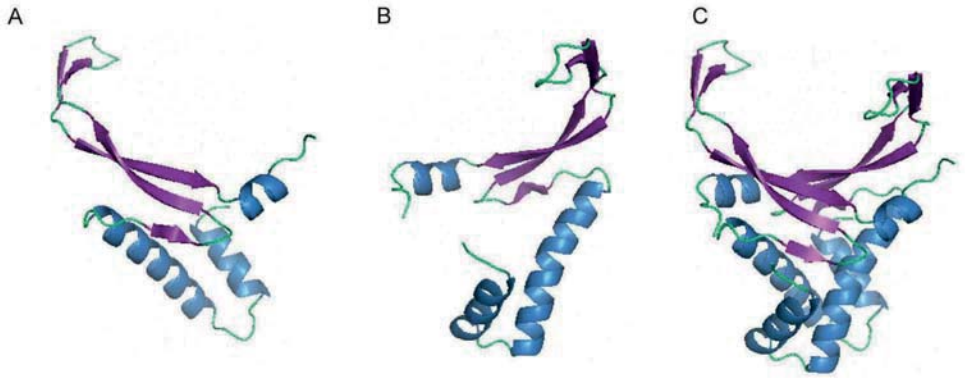


Figure 2.1: Molecular structure of the β- (A) and α-subunits (B) of the IHF protein (C) (PDB code: 1IHF).[19]

Thus, IHF belongs to the group of histone-like proteins which are all involved in the assembly of higher ordered DNA structures. Histone-like proteins are named after the eukaryotic histone proteins, however, they only resemble them in function.[31–34] Moreover, IHF displays high structural similarities to the other members of the DNABII protein family especially to the heat-unstable protein (HU). Nevertheless, whereas IHF binds sequence specifically to DNA, HU is a non-specific DNA binding and bending protein.[35,36]
IHF is a small heterodimeric protein consisting of an α- and a β-subunit (each ∼ 10 kDa) which are 30% identical in sequence and are encoded by the E. coli genes himA and hip/himD, respectively.[19,23] Each subunit is composed of three α-helices and five β-sheets (Figure 2.1). Furthermore, the subunits are intertwined forming a compact central body from which two flexible β-ribbon arms extend. Additionally, the basic amino acids in the helices are oriented to the outside and thereby facilitate the formation of a positively charged surface (Figure 2.2).[37,38]

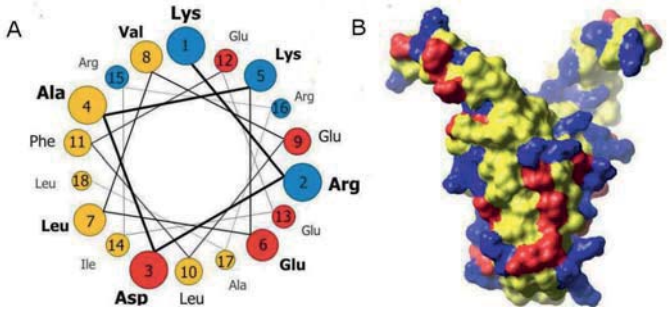


Figure 2.2: Schematic representation of amino acid contribution within the IHF.[20]

2.2 Recognition and Bending of Double Stranded DNA

The IHF-DNA co-crystal structure reveals that IHF exclusively binds to the minor groove of double stranded DNA, thereby displaying no contact to the major groove.[19] The recognition proceeds through an indirect readout of sequence dependent structural parameters and not by recognizing individual nucleobases.[19,37,39] IHF binds with high selectivity to the three distinct segments H′, H1 and H2 which are contained within the attachment site of bacteriophage λ (*attP*) having the highest affinity towards the H′ element.[24] All three binding sites contain the following 13 base pair (bp) core-consensus sequence: 5′-**WATCAR**NNNN**TTR**-3′ (wherein W = dA or dT, R = purine base dA or dG, and N = any nucleotide) towards which IHF has a 10^3-10^4 times higher affinity compared to other sequences.[35,40] Mutations in this region are always associated with a loss in affinity. Even a total affinity loss is observed when the middle **T** in the **TTR** element is exchanged.[25,41] Furthermore, a dA/dT-rich element is often located at the 5′ end of the core-consensus sequence. Common characteristics of such dA/dT elements are small and narrow grooves[42] which might in turn increase the affinity of IHF for the core-consensus sequence (Figure 2.3).[43,44]

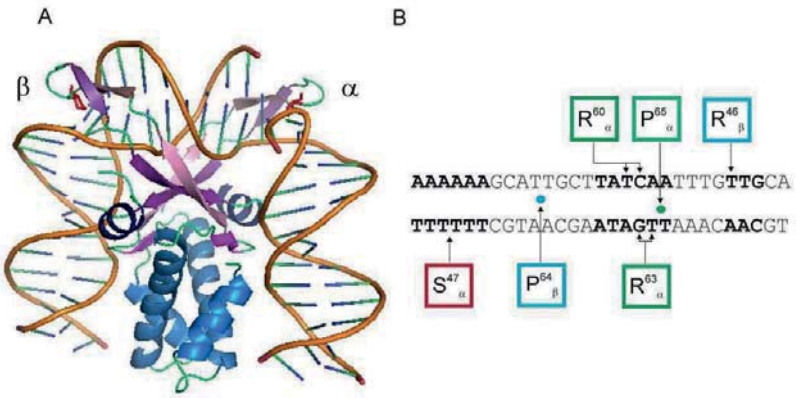
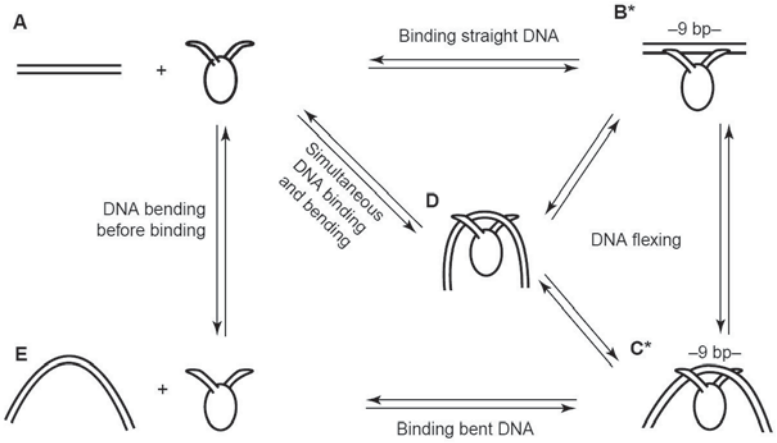


Figure 2.3: A Representation of the IHF-DNA co-crystal structure with a 35 bp DNA sequence, wherein the intercalating proline residues are highlighted in red. The β-sheets are indicated in violet and the α-helices are depicted in blue (PDB code: 1IHF). **B** The H′-binding site containing the core-consensus sequence and the dA/dT-rich 5′ element (both marked in bold). The locations of IHF-DNA interactions are indicated.

Surprisingly, only the β-ribbon of the α-subunit displays a sequence specificity for selective DNA binding while interacting with the **WATCAR** element and is, therefore, of particular importance. However, the basis for the sole specificity of the α-subunit is not yet clarified. Furthermore, the proline residues Pro^{61} and Pro^{72} are of conformational significance and contribute to sequence dependent recognition.[19] Both ribbons are embedded in the minor groove of the DNA with a proline residue located at the tip of each ribbon (Pro^{65} of the α-subunit and Pro^{64} of the β-subunit). These residues intercalate between the base pairs, disrupt the π-π stacking interactions and induce two significant kinks in the DNA conformation 9 bp apart

from each other.[19,27,35] Additionally, the β-ribbon of the α-subunit forms hydrogen bonds with the nucleobases via two arginine residues (Arg^{60} and Arg^{63}), whereas at corresponding positions two arginine residues of the β-subunit do not exhibit these interactions. This difference could be attributed to the fact that the β-arm of the α-subunit extends further into the minor groove in comparison to that of the β-subunit.[19] Almost all members of the DNABII protein class exhibit arginine residues in similar positions, nonetheless, they display usually no sequence specificity.[35] Further interactions are facilitated via the *N*-termini of the six helices making contact with the phosphate backbone of the DNA.[19] All these interactions together render the IHF protein with a capacity to induce a bend of > 160° into the DNA structure.[19,37] In order to facilitate and stabilize this extreme conformational change in the DNA structure, IHF utilizes two established principles, frequently found for DNA bending proteins. Firstly, the hydrophobic proline residues, located at the tip of each ribbon, intercalate between the base pairs causing a kinked DNA conformation. Secondly, the curved DNA is wrapped around the IHF protein by electrostatic binding. In addition, the bent is stabilized as the compact body of IHF is located at the center of the curved DNA and neutralizes the negatively charged phosphates.[37]
There are several feasible mechanisms regarding how DNA bending is facilitated by the IHF protein. Either IHF binds to an already pre-bent DNA or IHF binding occurs to a linear DNA followed by bending the DNA double strand. Alternatively, the DNA binding and bending follows a concerted mechanism (Figure 2.1).



Scheme 2.1: Explaining of feasible routes for DNA binding and bending by IHF.[35]

Estimating the rate constants faces the problem of distinguishing the uni-molecular rearrangement (DNA binding and bending) from the bi-molecular rearrangement (DNA-IHF complex association).[28] The kinetics of IHF induced DNA bending were elucidated including temperature-jump and stopped-flow experiments in combination with fluorescence resonance energy transfer (FRET) measurements. The chosen conditions permitted nanosecond resolutions, which were required to examine profound conformational alterations in protein-DNA-com-

plexes. It was indicated that in the case of high IHF concentrations it is a two step mechanism. IHF binds first to a linear DNA followed by rate-determining bending of DNA (Figure 2.4). Nevertheless, in the case of low IHF concentrations a concerted mechanism is observed. [27,28,39,45–47]

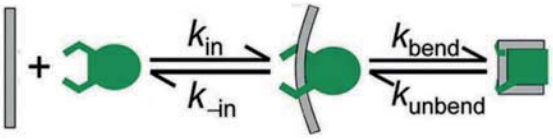


Figure 2.4: Mechanism of IHF induced bending of DNA. In a bi-molecular step IHF binds to DNA followed by a uni-molecular bending step. [27]

2.3 General Design of a Small Peptide Mimicking IHF-like DNA Binding and Bending

The native integration host factor is a well characterized protein which is responsible for inducing extreme conformational changes in double stranded DNA. Together with its small molecular weight and its ability to recognize double stranded DNA sequences with a high specificity, IHF is an eminently suitable target structure to design peptides which imitate the specific DNA binding and bending characteristics. Moreover, IHF mimicking peptides would not only help to elucidate the important elements responsible for the high selectivity for a target DNA but would additionally provide a tool to address DNA sequences in a sequence specific manner. The design and synthesis of a small peptide, which mimics the sequence specific DNA binding and bending characteristics of the native IHF protein, was reported by LIEBLER *et al.* (Figure 2.5).[20]

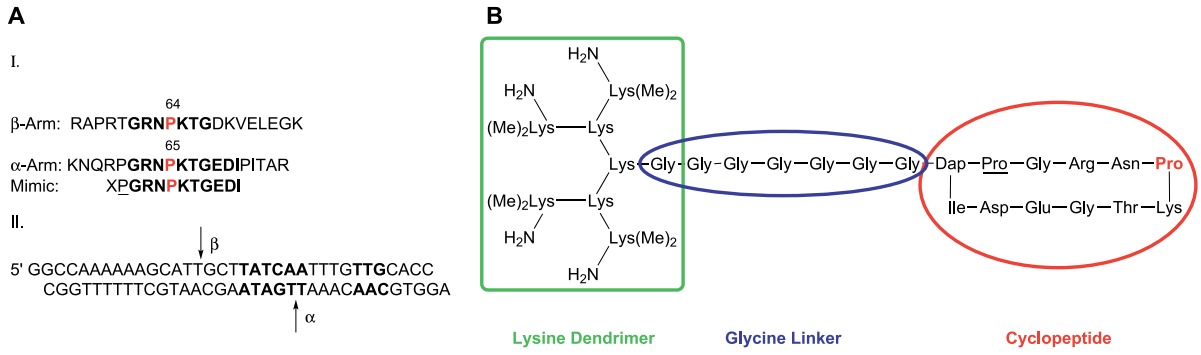


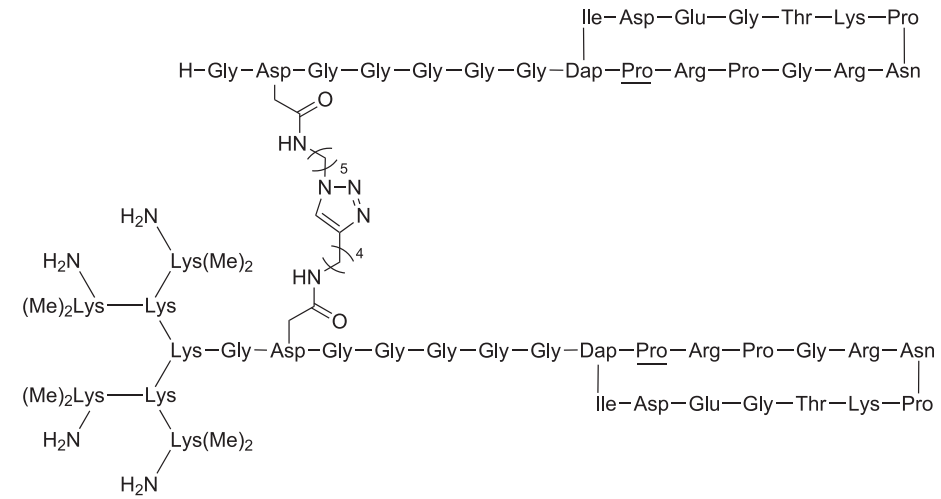
Figure 2.5: A (I) Excerpt of the amino acid sequence of the α- and β-subunits and the sequence of the first reported IHF mimicking peptide facilitating selective DNA binding,[20] wherein the conserved parts of the sequence around the intercalating proline residue (highlighted in red) are marked in bold. In the mimic sequence, X represents the unnatural diaminopropionic acid (Dap). (II) Applied DNA oligomers for the IHF-DNA co-crystallization.[19] The preserved consensus sequence for specific IHF binding is depicted in bold. Furthermore, the arrows indicate the position where the proline residue of the respective subunit intercalates between the base pairs. **B** Design of the first peptide mimicking selective IHF-like DNA binding and bending reported by LIEBLER *et al.*[20] The structure is composed of a three generation lysine dendrimer (green-framed), a heptaglycine linker (blue-framed) and a cyclopeptide (edged in red) containing the sequence around the intercalating proline (highlighted in red).

Herein, the design of IHF mimicking peptides consists of three parts: A cyclopeptide containing the DNA recognition sequence around the intercalating proline residue is selected for imitation of the β-ribbon of the α-subunit being responsible for sequence specific binding. The sequence is enlarged by two additional amino acids: The diaminopropionic acid for attachment of the glycine linker and for cyclization as well as the unnatural amino acid D-proline for improving the ability of the cyclopeptide to form a β-turn and also for facilitating close proximity of the respective moieties for the cyclization. The compact, positively charged body of the protein is imitated by a three generation lysine dendrimer which forms a basic surface under physiological conditions. For a better solubility in organic solvents during synthesis a

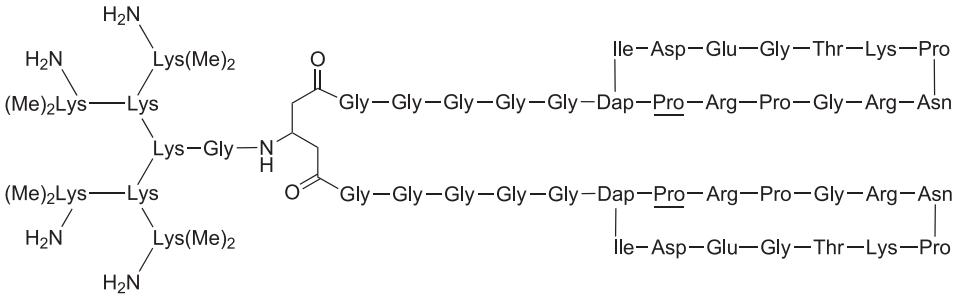
bismethylated lysine is used for the third generation. A heptaglycine linker connects the dendrimer with the cyclopeptide and adjusts both of them in the correct distance to each other for optimal DNA interaction.
In additional studies, variations in the design have been introduced in order to examine their ability to sequence specifically bind and bend DNA. Amino acid residues in the cyclopeptide have been exchanged with cysteine or histidine residues to facilitate a more rigid conformation and an improved preorganization by disulfide bond formation or metal ion complexation, respectively.[48] Furthermore, in IHF analogs the conserved intercalating proline residue was replaced by proline derivatives in order to elucidate the influence of the size of the side chain, the conformation of the five-membered ring and the aromaticity for bending double stranded DNA.[21] Moreover, owing to the structure of IHF and its recognition of DNA, it was described an attractive target for the design of therapeutics.[49]

2.4 IHF Mimicking Peptides Containing Two DNA Recognition Units

In the presented approach, IHF mimicking peptides containing two recognition units in order to enhance the sequence specific DNA binding and bending characteristics, based on the structural conformation of previously reported IHF mimicking peptides, were synthesized. The native IHF protein is a heterodimer comprised of two β-ribbons with the ability to sequence specifically bind and bend double stranded DNA. Therefore, novel dimeric IHF mimics shall possess two cyclopeptides imitating the β-ribbons as DNA binding and bending units and thereby might resemble the native protein in function more closely.
The design and synthesis of IHF mimicking peptides bearing two cyclopeptides involved two distinct synthetic approaches. In the first approach, an IHF mimicking peptide **1** modified with two cyclopeptides is derived from copper mediated alkyne azide cycloaddition of respective peptide fragments, which is a frequently utilized method to ligate unprotected biomolecules (Figure 2.6). Herein, novel artificial amino acid derivatives modified with terminal alkyne or azide moieties contain long and flexible side chains and thus circumventing sterical hindrance during the ligation reaction. These derivatives will be introduced into the glycine linker region in close proximity of the lysine dendrimer in order to increase the flexibility of the DNA recognition units and hence enabling enhanced DNA interaction. In addition, the dimeric IHF mimic shall contain only one lysine dendrimer to avoid electrostatic repulsion of two positively charged bodies in close proximity of each other and thereby impeding the synthesis.
The design of the dimeric IHF mimic **2** for the second strategy relies entirely on natural peptide bonds (Figure 2.6). Herein, a branching amino acid derivative containing two equidistant acid moieties, like the β^3-aspartic acid, is to be incorporated in the glycine linker in close proximity of the lysine dendrimer for attachment of two cyclopeptide units via amide bond formation. Thereby, both DNA recognition units originate from the lysine dendrimer and resemble the native protein with regard to its structure more closely in comparison to the dimer formed by the first approach. By comparing the DNA bending abilities of both peptides it is expected to gain insight into the role of the flexibility of the recognition units.
Both dimers were synthesized according to Fmoc based SPPS protocols. In this study, only the homodimers containing two identical cyclopeptides were prepared to simplify the access of dimeric structures. Herein, the cyclopeptide contains the sequence around the intercalating proline of the α-subunit of the IHF protein which exhibits a better sequence specific binding to DNA in contrast to the β-subunit. Furthermore, the cyclopeptide was enlarged by the dipeptide Arg-Gly to allow comparison to previously synthesized IHF mimicking peptides investigating the importance of the DNA recognition sequence's length. Nevertheless, the future objective is to prepare dimeric IHF mimics comprising the DNA recognition units of both the IHF subunits and hence enhancing the resemblance in structure and function to the native IHF even more.



1

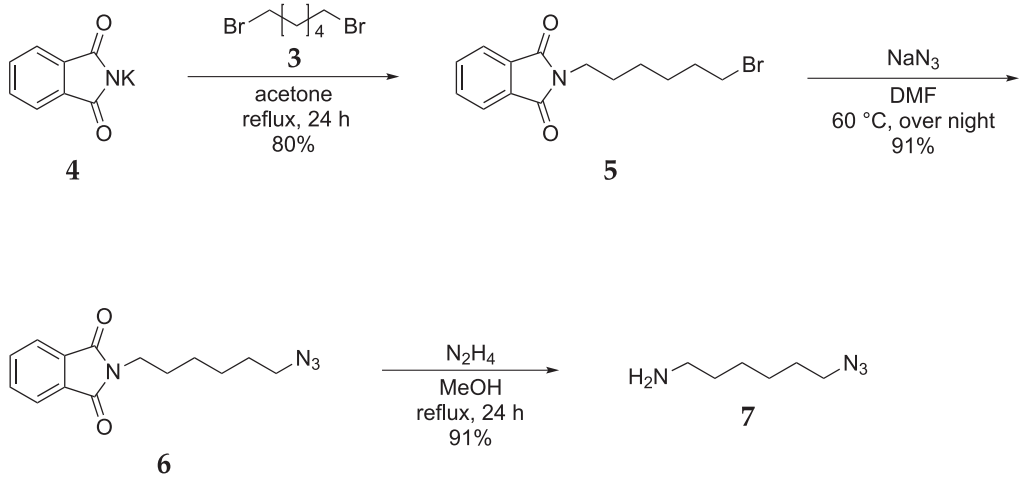


2

Figure 2.6: Structural conformation of two possible homodimeric IHF mimicking peptides based on Click chemistry **1** or derived entirely from peptide bonds **2**.

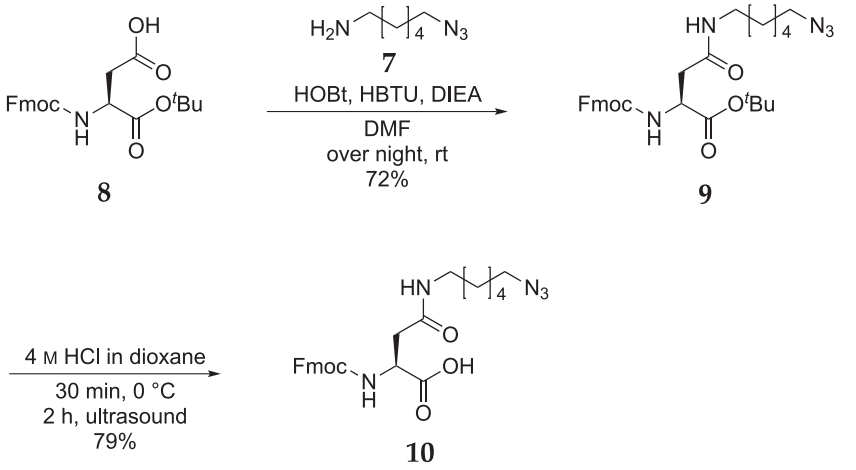
2.5 Synthesis of Artificial Amino Acids to Participate in the Ligation

Artificial amino acids were required to participate in the copper mediated alkyne azide cycloaddition for the formation of the dimeric IHF mimicking peptide **1** containing two DNA recognition units. Therefore, unnatural amino acids bearing terminal alkyne or azide moieties in their side chains suitable for Fmoc SPPS protocols were synthesized. The side chain of aspartic acid was elongated by amide bond formation with a primary amine bearing the respective terminal moiety for cycloadditon. The GABRIEL synthesis offers few synthetic steps under mild conditions to obtain primary amines from corresponding alkyl halides or alkyl sulfonates.[50] Herein, 1,6-dibromohexane (**3**) was alkylated with potassium phthalimide (**4**) furnishing 6-bromo-*N*-phthalimide (**5**)[51] which was subsequently converted into the corresponding azide **6** by treatment with NaN_3.[52] Afterwards, 6-azido-hexylamine (**7**) was generated by hydrazinolysis (Scheme 2.2).[52]



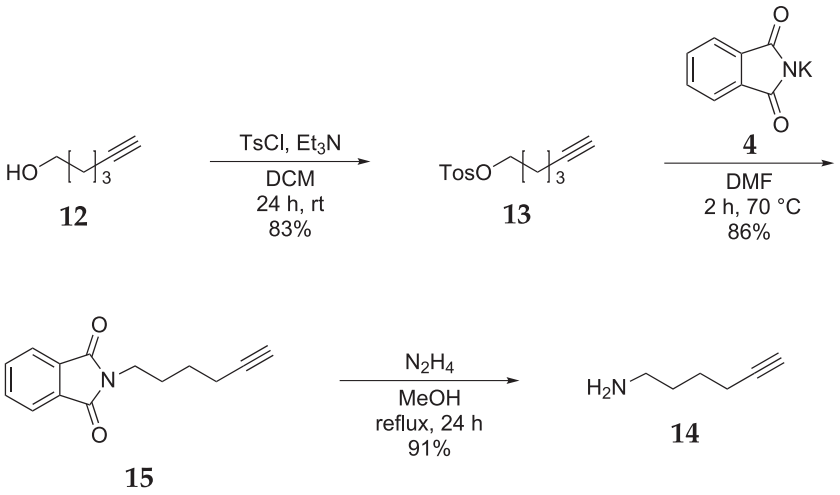
Scheme 2.2: Synthetic route to obtain 6-azido-hexylamine (**7**).[51,52]

Thereafter, 6-azido-hexylamine (**7**) was coupled to the carboxylic acid of Fmoc-Asp-O^{t}Bu (**8**) with standard conditions using HBTU and HOBt as activators and DIEA as activator base furnishing protected amino acid **9**. Removal of the ^{t}Bu moiety of **9** with TFA yielded the product **10** only in very low yields. However, milder cleavage conditions with 4 M HCl in dioxane gave the free acid **10** in 79% yield (Scheme 2.3).



Scheme 2.3: Synthesis of the artificial amino acid **10** bearing a terminal azide moiety in the side chain.

In the same manner, the artificial amino acid **11** bearing a terminal alkyne residue was synthesized (Scheme 2.6). Herein, 5-hexynol (**12**) was first tosylated yielding **13** and subsequently reacted with potassium phthalimide (**4**) by nucleophilic substitution (Scheme 2.4). Afterwards, hydrazinolysis yielded 5-hexynylamine (**14**). In an alternative reaction procedure, 5-hexynyl-*N*-phthalimide (**15**) was derived in a single conversion utilizing MITSUNOBU conditions, wherein 5-hexynol (**12**) was treated with phthalimide (**16**) in the presence of diethyl azodicarboxylate (DEAD) and PPh_3 (Scheme 2.5).[53]



Scheme 2.4: Synthesis of 5-hexynylamine (**14**).