Id3, Inhibitor of DNA Binding and Cell Differentiation: Synthesis and Conformational Analysis of the Full-Length Protein and its Truncated Analogues

Dissertation

zur Erlangung des Doktorgrades
der Naturwissenschaften (Dr. rer. nat.)
der Fakultät für Chemie und Pharmazie
der Universität Regensburg

vorgelegt von
Jaroslava Svobodová
aus Semily

Regensburg 2007

Die Arbeit wurde angeleitet von:	Prof. Dr. A. Buschauer
Promotionsgesuch eingereicht am:	8. März 2007
Promotionskolloquium am	20. April 2007
Prüfungsausschuß:	Vorsitzender: Prof. Dr. R.Winter 1. Gutachter: Prof. Dr. A. Buschauer 2. Gutachter: Prof. Ing. I. Stibor, CSc. 3. Prüfer: Prof. Dr. B. König

Die vorliegende Arbeit wurde in der Zeit von Oktober 2003 bis Oktober 2006 an der Fakultät für Chemie und Pharmazie der Universität Regensburg in der Arbeitsgruppe Prof. Dr. A. Buschauer unter der Leitung von Dr. C. Cabrele angefertigt.

Dr. C. Cabrele danke ich herzlich für die Überlassung des interessanten Themas und ihre stetige Unterstützung.
Prof. Dr. A. Buschauer danke ich für die Möglichkeit diese Arbeit an seinem Lehrstuhl durchführen zu dürfen.

To my parents

Table of Contents

List of Abbreviations

Amino acid residue	One-letter code	Three-letter code
Alanine	A	Ala
Arginine	R	Arg
Asparagine	N	Asn
Apartic acid	D	Asp
Cysteine	C	Cys
Glutamine	Q	Gln
Glutamic acid	E	Glu
Glycine	G	Gly
Histidine	H	His
Isoleucine	I	Ile
Leucine	L	Leu
Lysine	K	Lys
Methionine	M	Met
Phenylalanine	F	Phe
Proline	P	Pro
Serine	S	Ser
Threonine	T	Thr
Tryptophan	W	Trp
Tyrosine	Y	Tyr
Valine	V	Val
Aminoisobutyric acid	X	Aib
Norleucine	-	Nle

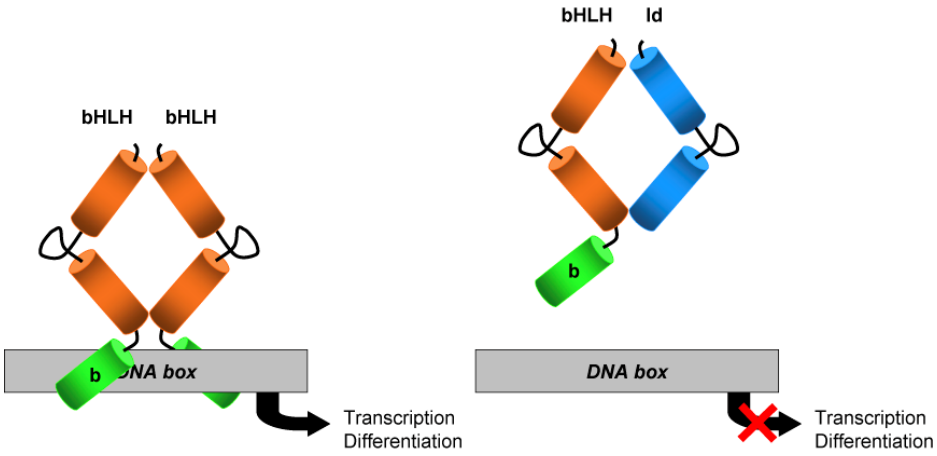
Ac	acetyl
ACN	acetonitrile
All	allyl
AM	aminomethyl
BAL	backbone amide linker
BMP	bone morphogenic protein
Boc	*tert*-butyloxycarbonyl
CD	circular dichroism
CDK	cyclin-dependent kinase
CDKI	cyclin-dependent kinase inhibitor
CEP	circulating endothelial precursor cell
CSN	constitutive photomorphogenesis signalosome
DBU	1,8-diazabicyclo[5.4.0]undec-7-ene
DCM	dichloromethane
DIC	*N,N*-diisopropylcarbodiimide
DIPEA	*N,N*-diisopropylethylamine
DKP	diketopiperazine
DMF	*N,N*-dimethylformamide
DMSO	dimethylsulfoxide
Dns	dansyl
DTT	dithiotreitol
DVB	divinylbenzyl
EPL	expressed protein ligation
eq	equivalent(s)
ERK	extracellular signal-regulated kinase
ESI	electrospray ionisation
FAM	carboxyfluorescein
Fmoc	9-fluorenylmethoxycarbonyl
FRET	fluorescence resonance energy transfer
GnHCl	guanidium chloride
HBTU	*O*-(1-benzotriazolyl)-1,1,3,3-tetramethyluroniumhexafluoro-phosphate
HIV	human immunodeficiency virus
HLH	helix-loop-helix
HOBt	*N*-hydroxybenzotriazole
*HS*Bn	benzylmercaptane
*HS*PAA	*p*-acetamidothiophenol
*HS*PAE	3-mercaptopropionic acid ethyl ester
MALDI-TOF	matrix-assisted laser desorption ionisation time-of-flight
MAPK	mitogen-activated protein kinase
MBHA	methylbenzhydrylamine
MCK	muscle-specific creatine kinase
MMP	matrix metalloproteinase
MS	mass spectrometry
Mtt	methyltrityl
MW	molecular weight
MyoD	myogenic determination factor
NCL	native chemical ligation
NES	nuclear export signal
NLS	nuclear localization signal
NMP	1-methyl-2-pyrrolidinone
PAGE	polyacrylamide gel electro-phoresis
PAM	phenylacetamidomethyl
Pbf	2,2,4,6,7-pentamethyldihydro-benzofuran-5-sulfonyl
PG	protecting group
PP	polyproline
pRb	retinoblastima protein
PS	polystyrene
PTTG	pituitary tumor transforming gene
PyBOP	benzotriazole-1-yl-oxy-tris-pyrrolidinophosphonium hexa-fluorophosphate
RMSD	root mean square deviation
RP-HPLC	reverse phase high performance liquid chromatograhy
SAB	sulfamyl butyryl
SDS	sodium dodecyl sulfate
SPPS	solid phase peptide synthesis
TAMRA	carboxytetramethylrhodamine
Tat	transactivator of transcription protein
t-Bu	*tert*-butyl
TCF	ternary complex factor
TCR	T-cell receptor
TFA	trifluoroacetic acid
TFE	2,2,2-trifluoroethanol
TGF	transforming growth factor
THF	tetrahydrofuran
TIS	triisopropylsilane
TMSBr	trimethylsilylbromide
t_R	retention time
TRIS	tris(hydroxymethylamino-methane)
trt	trityl
UV	ultraviolet
VEGF	vascular endothelial growth factor
VSMC	vascular smooth muscle cell

I. Biological Role of the Id3 Protein

I.1 Introduction

The family of the *Id* proteins [1, 2], *i*nhibitors of *D*NA binding and cell *d*ifferentiation (Id1-4), represents the class V of the large group of the helix-loop-helix (HLH) transcription factors [3], and is characterized by the fact that its members do not display a basic region N-terminal to the HLH domain. By contrast, the related bHLH proteins contain not only the HLH domain which is required for protein-protein interactions, but also the basic Lys-/Arg-rich sequence (indicated by the letter "b") which is important for protein-DNA interactions. Both regions are essential for their activity as DNA transcription regulators. Indeed, the bHLH proteins bind to DNA as homo- or heterodimers folded into a parallel four-helix bundle with the two basic extensions correctly juxtaposed to allow the recognition and binding to the so-called DNA E box (CA*NN*TG), N box (CAC*N*AG) or Ets site (GGAA/T) [4] present in promoter regions of the target genes (Scheme I.1, left)

Scheme I.1. Effects of the Id proteins on bHLH-mediated DNA transcription. Activation of transcription by formation of the DNA-bHLH dimer complex (*left*), and inhibition of transcription by sequestering the bHLH factors in inactive dimers with the Id proteins (*right*).



In developing cells, the bHLH proteins mostly regulate the expression of genes which are required for terminal cell differentiation and cell cycle arrest in the G1 phase. As the Id proteins only possess the HLH domain, but not the basic DNA-binding region, the dimers formed with bHLH proteins are unable to bind to DNA (Scheme I.1, right). This leads to an inhibition of the bHLH-mediated DNA transcription by the Id proteins. Dimerization partners of the Id proteins among the bHLH family are the ubiquitously expressed E-proteins (E12, E47, E2-2 and HEB) and the tissue-specific myogenin-regulating factors MyoD and

myogenin [5, 6]. Norton and co-workers have proposed the following mechanism to explain how the Id proteins can override the high energy of the bHLH-DNA complexes to sequester the bHLH factors: the nuclear Id proteins interact with the transiently free bHLH proteins and not with the chromatin-associated ones, and prevent the reassociation of the free-ones with the chromatin domains [7].

It is believed that the bHLH fold is stabilized only upon dimerization and binding to the DNA. In the dimeric form, the structure of the bHLH domain is very similar among all the HLH family members, as shown by a comparative study based on known crystal structures and homology models [8]. In the case of the HLH domain of Id3, a three-dimensional homology model based on the crystal structure of the DNA-bound homodimers of E47, MyoD and Max has shown that the highly favored packing of the hydrophobic core together with the absence of repulsive interactions within the N-termini explains the formation of an Id3 homodimer [9].

A splice variant of human and rat Id3 has been detected, which is characterized by different dimerization properties [10]. In both species, the isoform contains a modified C-terminal region [11] where the last 19 C-terminal residues are replaced by a new sequence of 60 (human, Id3L) or 29 (rat, Id3a) residues. These isoforms have different expression patterns and effects on cell growth and death control with respect to the shorter variant of Id3. For example, expression level of Id3 is low in normal blood vessels, whereas Id3a is totally absent. However, expression of both is induced upon injury [12]. Moreover, Id3L has been found in advanced human atherosclerotic plaques [13] and has impaired ability to form heterodimers with E47 [10]. Interestingly, gene delivery of Id3a was found to induce apoptosis and inhibit Id3 expression in vascular smooth muscle cells (VSMCs) [14].

I.2 Id3 Protein Expression

In general, Id proteins are overexpressed in proliferating, undifferentiated normal and cancer cells, while their expression level is low or absent in non-proliferating cells [15]. The expression pattern of Id3 overlaps with that of Id1 during embryogenesis [16]. Both proteins are required for neurogenesis, angiogenesis and vascularization of tumors [17]. The Id3 gene was identified as an immediate early response gene induced in the early phase after mitogen stimulation by serum growth factors or forskolin, or after stimulation of protein kinase C by tumor-promoting phorbol esters [18]. In contrast, Id3 gene silencing might be controlled by DNA methyltransferases, histone acetyl transferases and histone deacetylases [2]. Homologues of the Id3 protein were found in a range of non-mammalian animals,

such as in *Drosophila melanogaster extramacrochaetae* [19, 20], *Xenopus* [21], zebrafish [22] and chick [23]. Unlike Id1 and Id2, the Id3 protein in mammals and its non-mammalian homologues exhibit a highly conserved organization of the coding gene, which indicates that they all evolved from a common ancestor.

The human Id3 gene is mapped to chromosome 1p36 [24-26]. Deletions of the latter are often present in neuroblastoma and other tumors derived from the neural crest. Accordingly, tumor cell lines with N-myc protein overexpression have generally low Id3 gene expression, whereas cell lines with high Id3 gene levels do not overexpress N-myc. Mutually exclusive expression of these two genes, Id3 and N-myc, has also been found in developing mouse tissues, suggesting an inverse correlation between the two genes in human neuroblastomas and during normal development [24].

Id3 protein level and function can be regulated during transcription (alternative splicing), by modification of the protein stability (ubiquitination), by posttranslational modifications (phosphorylation), and also by the presence of binding partners. These aspects are discussed below.

I.3 The Role of the N- and C-Terminal Regions on the Id3 Protein Function

With 119 amino acid residues (13 kDa), Id3 is the smallest member of the Id protein family. The 41-residue long HLH motif is responsible for the interaction with the bHLH proteins. However, the N- and C-termini are also important for the Id3 function. The N-terminus contains a consensus site for cyclin-dependent kinase 2 (CDK2), in which serine at position 5 (Ser-5) is phosphorylated by cyclin-CDK2 complexes, with consequent modulation of the Id3 affinity to the E-proteins. The N-terminal amino group is important for ubiquitination. Deletion studies on the Id3 C-terminus have indicated that the 15 C-terminal residues are critical for the full inhibitory activity of Id3 on the muscle-specific creatine kinase (MCK) reporter gene, suggesting that the C-terminus of Id3 also mediates important biological functions [27, 28]. C-terminal mutants of Id3 and a chimeric protein containing the N-terminal part of Id3 and the C-terminus of Id2 did not inhibit the MyoD function; however, they could interact with E47. This suggests that the biological action of Id3 might require additional functions that depend on its C-terminal residues. Both the C-terminal region and HLH domain of Id3 are required for full inhibition of MCK reporter gene activity and for growth-suppressive activity in immature B lymphoma cells [29]. Id3 is probably not only a passive transcription repressor but also an active one. As in all active transcription

repressors, a high amount of proline, glutamine and alanine residues can be found in the C-terminal domain of Id3, particularly proline (29 % in the C-terminus) [27].

I.4 Id3 Protein Stability

Id3 is a short-lived protein with a half-life of 20 min [30]. It is difficult to express a large amount of Id proteins because of their low stability inside the cells. Their degradation is mediated by several regulation pathways: for example, Id3 is N-terminally degraded through the ubiquitin-proteasome pathway [31], similarly to Id1 and Id2. Constitutive photomorphogenesis 9 signalosome (CSN), which is thought to cooperate with the ubiquitin-proteasome system in the regulation of protein stability in all eukaryotes [32], is also involved in the Id3 protein stability pathways. Unlike Id2 and Id4, Id1 and Id3 bind to the CSN subunit CSN5 (and/or CSN7 for Id3) [33]. The kinases associated to CSN phosphorylate CSN-bound proteins, leading to inhibition of Id1 and Id3 ubiquitination, as supported by the fact that inhibitors of CSN-associated kinases significantly induce Id ubiquitination and proteasome-dependent degradation [33].

The stability of Id3 can also be affected by its dimerization partners. Protein p204, a member of the interferon-inducible p200 protein family, is a growth inhibitory factor and is involved in the differentiation of skeletal [34] and cardiac muscle myoblasts [35]. It induces differentiation by binding to the Id proteins (Id1, Id2 and Id3) and by accelerating their ubiquitination and proteasome degradation. Additionally, protein p204 has the ability to decrease the Id protein level in the nucleus, where Id enforces its inhibitory function, by driving its translocation from the nucleus to the cytoplasm. In contrast, the degradation of Id3 is slowed down when bound to the Gata4 protein [35] or to E47 [36]. After separation from Gata4, Id3 is translocated to the cytoplasm and degraded via the ubiquitin-proteasome pathway. Also the heterodimer Id3/E47 is less sensitive to degradation by the proteasome and the half-life of Id3 is thus extended.

I.5 Id3 Protein Localization and Phosphorylation

The Id proteins functions depend on their distribution between the nucleus and the cytoplasm; although their small size allows them to enter and exit the nucleus by passive diffusion, there is an indication that their subcellular localization is regulated also by other pathways. Id1 and Id2 proteins are predominantly localized in the cytoplasm, whereas Id3 and Id4 have a nucleo-cytoplasmic distribution [37]. In contrast to Id1 [38] and Id2 [37], Id3 possesses neither a nuclear export signal (NES) essential for cytoplasmic localization nor

a nuclear localization signal (NLS). Upon coexpression with E47, which contains a NLS region, Id3 accumulates in the nucleus. E47 acts as a nuclear chaperone for Id3 [36].

The subcellular localization of many proteins is regulated by protein modifications such as phosphorylation. Id2, Id3 and Id4 possess a conserved consensus CDK2 phosphorylation site at Ser-5 [39, 40]: phosphorylation at this position occurs at the late G1 state of the cell cycle, which corresponds to the time both for activation of cyclin A and cyclin E dependent CDK2 and for the formation of a specific E-box-DNA-binding complex. Phosphorylation of Id3 changes its specificity for abrogating E-box-dependent homo- or heterodimer complexes. Unphosphorylated Id3 inhibits an E12 homodimer complex but not the E12-MyoD heterodimer. Phosphorylated Id3 has greater ability to abolish the E12-MyoD complex but nullifies the inhibition of the E12 homodimer.

Phosphorylation of Id3 by CDK2 also negatively affects its ability to inhibit DNA binding of ternary complex factors (TCFs) with Ets-domain. Phosphorylation of Id3 decreases protein-protein interactions between the Id protein and TCFs and thereby reduces the DNA-binding inhibitory activity of the Id3 protein [41].

The Id proteins are required for the entry of the cells into the S phase; however, a CDK-dependent phosphorylation alters the ability of Id3 to promote the S phase entry in fibroblast cells (phosphorylated Id3 cannot promote cell growth), suggesting a G1 into S phase cell cycle regulatory function of the Id proteins [40]. The situation is, however, different in VSMCs [42].

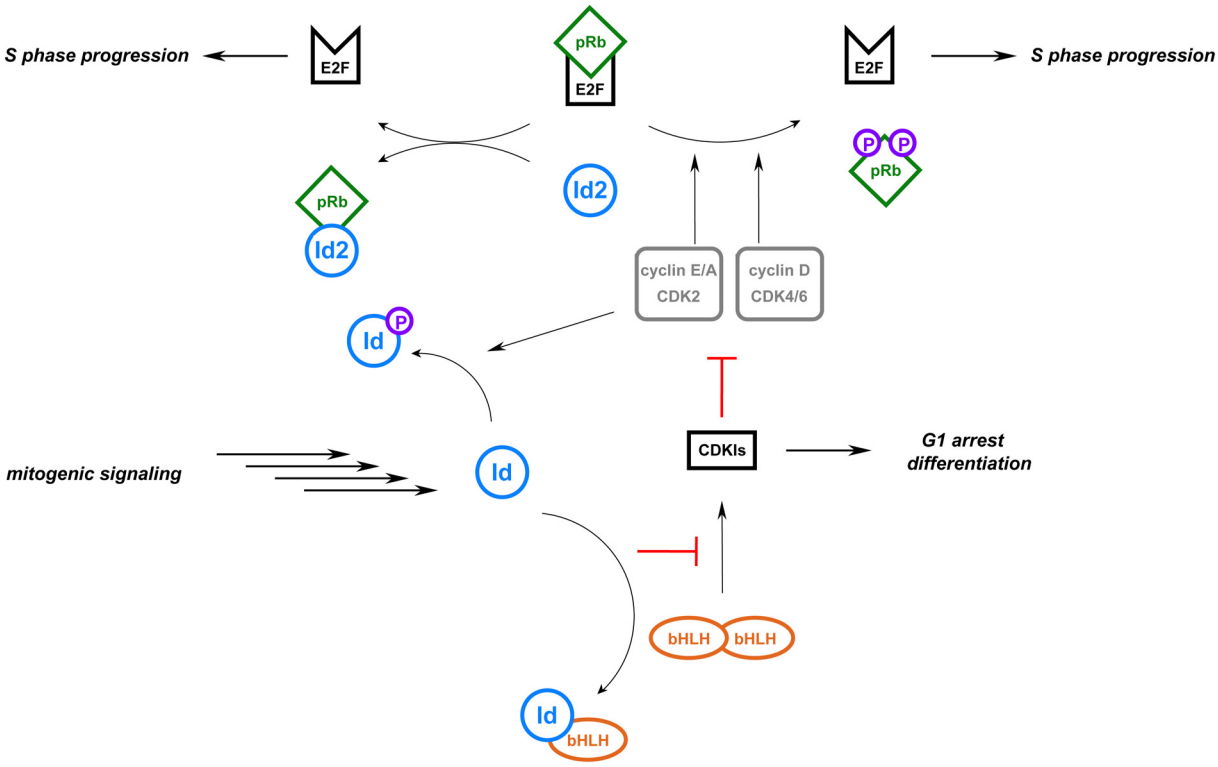
Overexpression of the Id proteins leads to apoptosis or cell proliferation, depending on the cell line. For example, overexpressed Id3 protein does not promote apoptosis in muscle cells (myoblasts) [43, 44], but leads to apoptosis in serum-deprived rat embryo fibroblasts. Id3 protein-induced apoptosis is mediated at least in part through p53-independent mechanisms and could be efficiently rescued by coexpression with E47 and antiapoptotic Bcl2 and $BclX_L$ [45]. Induction of apoptosis is correlated with the ability of Id3 to promote the cell cycle S phase entry. The close association between apoptosis and the pathways, through which Id3 promotes the G1 to S phase cell cycle transition, is based on functional properties of phosphorylated and unphosphorylated Id3 forms. Unphosphorylated Id3 (present throughout early G1), but not the phosphorylated form (persisting from late G1 throughout the S phase) drives apoptosis. A lowering of functional CDK2 activity and consequent hypophosphorylated Id protein in serum-deprived fibroblasts might therefore trigger apoptosis induced by overexpression of the Id gene.

In conclusion, phosphorylation plays an important role in regulating the activity (negative effect in fibroblasts), dimerization specificity and stability (inhibition of ubiquitin-mediated proteasomal degradation) of the Id3 protein.

I.6 Id3 in the Cell Cycle Control

The Id proteins tune the balance between cell growth and differentiation by negatively regulating the function of bHLH transcription factors. In mammals, the pocket proteins of the retinoblastoma (Rb) family control a crucial proliferation checkpoint [46]. The Rb pathway is negatively influenced either by direct interaction with Id2 or indirectly by inhibition of Ets-mediated transcription of $p16^{INK4a}$ [41] and Id protein heterodimerization with bHLH proteins (Scheme I.2).

Scheme I.2. Id proteins and cell cycle control.



The E-proteins and MyoD block G1 into S phase cell cycle transition mediated by Id proteins through the transcriptional activation of the $p21^{Cip1}$ and consequent inactivation of

CDKs, thus enhancing the activity of the Rb protein (pRb) [15]. CDKs in complex with cyclins sequentially phosphorylate pRb, thus releasing it from its transcriptionally repressive complex with E2F and facilitating the S phase progression. The Id proteins disturb the activity of the bHLH factors and thereby downregulate CDK inhibitors (CDKI: $p21^{Cip1}$, $p16^{INK4a}$) [44] and activate the CDKs for pRb phosphorylation and cell cycle progression into the S phase.

I.7 Id3 and its Cellular Pathways

I.7.1 Interaction with bHLH Proteins [47]

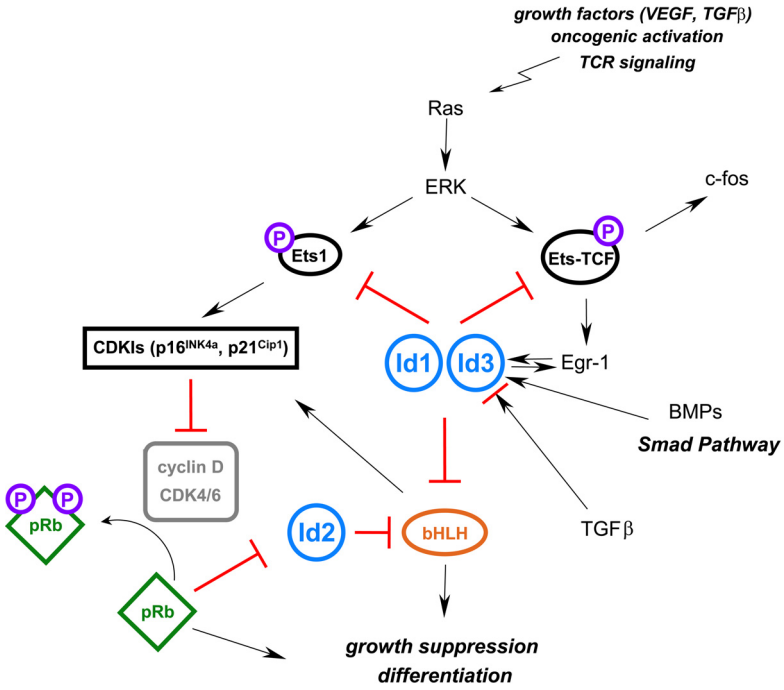
The Id3 protein is a general negative regulator of the bHLH transcription factors, especially the E-proteins and MyoD. It binds to the bHLH proteins, prevents DNA binding and thereby blocks transcription activation. Id3 interacts with both spliced forms of the E2A gene, E47 and E12, as well as with the related protein E2-2. E47 and E12 preferentially associate with Id3 rather than with any other Id protein *in vitro* and *in vivo* [48].

I.7.2 Ras-ERK-MAP Kinase Pathway, Ets Proteins

There is a link between extracellular signal-regulated kinase (ERK) and mitogen-activated protein kinase (MAPK) signaling, and the HLH transcriptional regulators E2A, HEB and Id3 during thymocyte maturation [49]. Id3 protein expression is regulated by the Ras-ERK-MAP kinase pathway which is activated by T-cell receptor (TCR) signaling, resulting in a significant reduction in E2A/HEB DNA binding activity. ERK in turn activates the Ets domain-containing transcription factor SAP-1. Activated SAP-1 then induces expression of the zinc finger-containing protein Egr-1 [50], which causes a direct activation of Id3 protein expression and a subsequent decrease in E-protein DNA binding activity, and, ultimately, promotion of thymocyte maturation (Scheme I.3).

Id3 together with Id1 shuts down the upstream Ets transcription factors by direct physical interaction. Inhibition of Ets1, a non-bHLH protein, by Id1 and Id3 proteins leads to inhibition of $p16^{INK4a}$ expression and, consequently, allows cyclin D-CDK4-mediated phosphorylation of pRb and the liberation of the Id2 protein from a hypophosphorylated pRb-Id2 complex [41, 51].

Scheme I.3. Id proteins pathways: Ras-ERK-MAPK kinase pathway.



I.7.3 The Pax Pathway

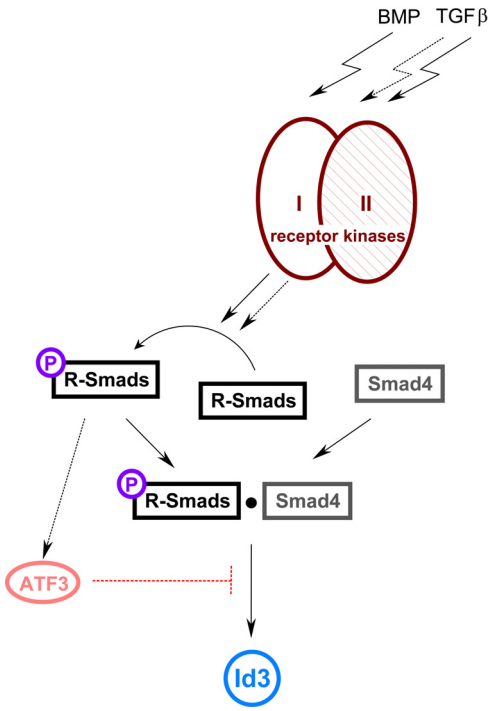
Id1-3 also bind and repress Pax transcription factors (Pax-2/5/8) which play important roles in cellular differentiation and organism development [52]. Binding occurs through the paired-DNA-domain of the Pax proteins and results in the disruption of DNA-bound complexes. Interaction of the Id proteins with the Pax proteins leads to the inhibition of promoters regulated by Pax complexes, such as the B-cell-specific mb-1 promoter. A subset of Ets-domain transcription factors, including Elk-1, can also form ternary complexes with the paired-domain transcription factor Pax-5. The Id proteins cause dissociation of both Pax and Ets-domain proteins from DNA in a manner similar to their effect on bHLH proteins [53]. Nevertheless, both Pax and Ets-domain proteins bind DNA as monomers. Thus, removal of an obligatory dimerization partner, as it is necessary for the inhibition of the bHLH protein heterodimers, is not a universal mode of action for the Id proteins.

I.7.4 TGFβ and BMP Regulation of Id Expression [54, 55]

The transforming growth factor (TGF) β family possesses a broad biological potential, such as the regulation of cell growth, differentiation and fate determination in various cell types including epithelial cells, endothelial cells, hematopoietic cells and mesenchymal cell types including fibroblasts, chondrocytes and osteoblasts. Members of the TGFβ family,

including TGFβs, activins, bone morphogenic proteins (BMPs) and growth/differentiation factors transmit their pleiotropic effects through transmembrane serine/threonine kinase type I and type II receptors. Upon ligand-induced heteromeric complex formation between type I and type II receptors, the type I receptor becomes phosphorylated and activated by the type II receptor kinase, and mediates specific intracellular signals. Smad proteins are the central signal mediators of the TGFβ superfamily activated by type I receptor kinases. Smad proteins are subdivided into three subtypes: receptor-regulated Smads (R-Smads), common-mediator Smad (Smad4) and inhibitory Smads (Smad6/7). R-Smads are further subdivided into two subtypes: Smad2/3 for TGF-β and activin signaling pathways, Smad1/5/8 for BMP signaling pathways. Inhibitory Smad proteins repress the action of BMPs by inhibiting the receptor-mediated phosphorylation of Smad1/5/8 or by competing with Smad4 for the binding to Smad1/5/8. Among the three different classes of Smads, only R-Smads interact directly with activated type I receptors and become activated through phosphorylation. R-Smads then form heteromeric complexes with Smad4, followed by nuclear translocation and regulation of gene transcription in collaboration with other transcription factors and transcription regulators (Scheme I.4).

Scheme I.4. Regulation of Id3 expression by TGFβ and BMP proteins.



There are two pathways in the signaling cascade of the BMP family: the first is Smad-dependent via Smad phosphorylation, and the other is a Smad-independent pathway involving MAPK. Stimulation of epithelial or endothelial cells with BMP factors upregulates the expression of the Id genes via phosphorylation of BMP-responsive Smad1/5 which binds to Smad-responsive elements in the Id promoter and thus activates its transcription. In contrast, TGFβ inhibits the expression of the Id genes, as TGFβ-responsive Smad3 activates the transcription repressor ATF3 which inhibits the Smad-initiated transcription. ATF3 co-operates with TGFβ-responsive Smad3, but not with BMP-specific Smad1, which allows cells to distinguish between TGFβ and BMP inputs. Unlike in epithelial cells, TGFβ is able to induce the synthesis of the Id proteins in endothelial cells by binding the type I receptor serine-threonine kinase ALK-1 that then activates the BMP R-Smads [56]. Also in immune cells, TGFβ upregulates Id2 and Id3 protein expression leading to the sequestration of the E2A gene products [51].

I.7.5 The VEGF Pathway

Vascular endothelial growth factor (VEGF), a homodimeric heparin-binding glycoprotein, exerts its biological effects by binding to the tyrosine kinase receptors VEGFR1 and VEGFR2. VEGF is produced by almost every cell type, and is required for the angiogenic switch. VEGFR1 is important for vascular modeling and regulation of endothelial cell-matrix interactions. VEGFR2 is the critical receptor which controls proliferation, differentiation and migration of endothelial cells and is expressed in some non-endothelial cells, too. Pituitary tumor transforming gene (PTTG) is an oncogene known to promote thyroid cell proliferation. High PTTG expression observed in thyroid cancers upregulates VEGFR2 expression and VEGF secretion, and there is evidence that PTTG takes part in thyroid tumorigenesis by promoting a VEGF-VEGFR2-Id3 mitogenic autocrine mechanism in human thyroid cells. VEGF is involved in VEGF-VEGFR2-Id3 pathway by activating VEGFR2 which in turn upregulates Id3 expression. Id3 is an important gene in VEGF-dependent angiogenesis. Expression of VEGFR2 and/or VEGFR1 in non-endothelial cells including ovarian, prostate and pancreatic carcinoma cells suggests the existence of a potential autocrine stimulatory pathway. VEGF and VEGFR inhibitors have been used against cancer progression and metastasis in the last decade [57].

VEGF leads to a pronounced upregulation of Id1 and Id3 in the bone marrow, presumably in the to-be-mobilized progenitor cells. Activation of protein kinase (MAPK) pathway by VEGFR stimulation could concern the Id1 and Id3 promoters at the Egr-1 site, as demonstr-

ated for other cell types [54, 58, 59]. In the bone marrow, the Id proteins are required for the mobilization and proliferation of endothelial progenitor cells which enter the bloodstream as circulating endothelial progenitors and migrate to the tumor site.

VEGF and TGFβ induce the expression of Id1 and Id3 through the MAPK cascade in human umbilical vein endothelial cells. Forced expression of the Id proteins induces proliferative activity (angiogenesis) accompanied by downregulation of $p16^{INK4a}$ caused by direct inhibition of Ets1/2 transcription factors [41]. Overexpression of the Id proteins induces pro-angiogenic processes such as transmigration and expression of the matrix metalloproteinases MMP-2 and MMP-9.

Inhibition of expression and/or function of Id1 and Id3 has a therapeutic value for conditions associated with pathological angiogenesis, not only of cancer but also of rheumatoid arthritis. Indeed, Id1 and Id3 are overexpressed in the endothelial cells of the synovial tissue in rheumatoid arthritis, a chronic inflammatory disease characterized by aggressive proliferation [60].

I.8 Id Proteins as Targets in Cancer

The Id proteins are potential targets in cancer therapy [61], because they are heavily involved in pathways regulating proliferation, differentiation, angiogenesis, migration, invasion and cell-cell interaction [4, 62-64]. They are mainly overexpressed during embryogenesis and development as well as in tumor cells, whereas their expression significantly decreases in normal mature adult tissues [17, 63].

Id1 and Id3 are important for tumor angiogenesis, and their loss causes vascular defects in both xenografts and the more physiologically relevant genetic tumor models. The angiogenic defect in Id-deficient mice may be attributed to combined downregulation of MMP-2 and fibroblast growth factor receptor 1 and $\alpha_6\beta_4$ integrin. A partial decrease of Id levels significantly inhibits tumor invasion and metastasis because of reduced expression of MMPs and vascularization with consequent reduction of oxygen and nutrients supply to the tumor cells [57].

There are at least four essential hallmarks for malignancy (anaplasia, uncontrolled proliferation, sustained neoangiogenesis and invasion of neighboring tissues) that are connected by dysregulated Id expression during tumorigenesis [46]. The Id proteins are required for vascularization of tumors. Nevertheless, some tumors, e.g. the Lewis lung carcinoma, can grow despite the lack of Id expression and of a functional vasculature [50].

Tumor progression depends on a multistage angiogenic process; the initial quiet phase is connected with quiescent endothelial cells. The second stage is associated with an angiogenic switch, including increased tumor growth and significantly increased endothelial proliferation, migration, vascular dilation and microhemorrhages. Id gene expression stimulates angiogenesis of endothelial cells through the inhibition of the angiogenesis inhibitor thrombospondin-1 and through the activation of VEGFR-mediated functions [4]. Although Id1 and Id3 are not expressed in adult bone marrow, VEGF-upregulated Id expression in bone marrow cells suggests that VEGF-activated Id1 and Id3 protein expression is essential for the mobilization and recruitment of circulating endothelial precursor cells (CEPs) and $VEGFR1^+$ hematopoietic cells to the tumor vasculature. Therefore, new therapies target $Id1^{+/+}Id3^{+/+}VEGFR2^+$ CEPs and $VEGFR1^+$ hematopoietic precursor cells to inhibit tumor angiogenesis [65].

I.9 Conclusion and Perspectives

The Id proteins subtly, but decisively determine the cell fate by a complex interplay of regulatory pathways, as seen with Id3 protein which is required for neurogenesis, angiogenesis and vascularization.

Id3 activity is regulated at different levels, e.g. by gene expression or at the protein level. Id3 phosphorylation and ubiquitination, alternative longer spliced variants and expression of Id3 binding partners may also play a role: for example, a modulation of Id3 activity is provided by a switch from the expression of Id3 to that of Id3L, as the latter cannot bind to E-proteins, or by overexpression of E47 which sequesters Id3 and thus inhibits its function. Phosphorylation of Id3 also affects its role in the cell cycle progression and proliferation. Additionally, transfer of Id3 from the nucleus to the cytoplasm represents another mechanism to control the Id functions.

Thus far, the interest in Id proteins mainly focused on their biological role in physiological and pathological conditions, while information on their biophysical and structural features was scarce. Purpose of the presented PhD thesis was the investigation of the smallest of the Id proteins, Id3, from a structural point of view, using a chemical and spectroscopic approach. The major challenge was the total chemical synthesis of the full-length Id3 protein and of related analogues whose conformation has been successively examined.

I.10 Literature

[1] Norton, J. D., Deed, R. W., Craggs, G., and Sablitzky, F. (1998) Id helix-loop-helix proteins in cell growth and differentiation. *Trends Cell Biol 8*, 58-65.

[2] Lim, R. W., and Wu, J. M. (2005) Molecular mechanisms regulating expression and function of transcription regulator inhibitor of differentiation 3. *Acta Pharmacol Sin 26*, 1409-20.

[3] Massari, M. E., and Murre, C. (2000) Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol 20*, 429-40.

[4] Sikder, H. A., Devlin, M. K., Dunlap, S., Ryu, B., and Alani, R. M. (2003) Id proteins in cell growth and tumorigenesis. *Cancer Cell 3*, 525-30.

[5] Langlands, K., Yin, X., Anand, G., and Prochownik, E. V. (1997) Differential interactions of Id proteins with basic-helix-loop-helix transcription factors. *J Biol Chem 272*, 19785-93.

[6] Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L., and Weintraub, H. (1990) The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell 61*, 49-59.

[7] O'Toole, P. J., Inoue, T., Emerson, L., Morrison, I. E., Mackie, A. R., Cherry, R. J., and Norton, J. D. (2003) Id proteins negatively regulate basic helix-loop-helix transcription factor function by disrupting subnuclear compartmentalization. *J Biol Chem 278*, 45770-6.

[8] Chavali, G. B., Vijayalakshmi, C., and Salunke, D. M. (2001) Analysis of sequence signature defining functional specificity and structural stability in helix-loop-helix proteins. *Proteins 42*, 471-80.

[9] Wibley, J., Deed, R., Jasiok, M., Douglas, K., and Norton, J. (1996) A homology model of the Id-3 helix-loop-helix domain as a basis for structure-function predictions. *Biochim Biophys Acta 1294*, 138-46.

[10] Deed, R. W., Jasiok, M., and Norton, J. D. (1996) Attenuated function of a variant form of the helix-loop-helix protein, Id-3, generated by an alternative splicing mechanism. *FEBS Lett 393*, 113-6.

[11] Yeh, K., and Lim, R. W. (2000) Genomic organization and promoter analysis of the murine Id3 gene. *Gene 254*, 163-71.

[12] Forrest, S., and McNamara, C. (2004) Id family of transcription factors and vascular lesion formation. *Arterioscl Throm Vas 24*, 2014-20.

[13] Matsumura, M. E., Li, F., Berthoux, L., Wei, B., Lobe, D. R., Jeon, C., Hammarskjold, M. L., and McNamara, C. A. (2001) Vascular injury induces posttranscriptional regulation of the Id3 gene: cloning of a novel Id3 isoform expressed during vascular lesion formation in rat and human atherosclerosis. *Arterioscl Throm Vas 21*, 752-8.

[14] Forrest, S. T., Barringhaus, K. G., Perlegas, D., Hammarskjold, M. L., and McNamara, C. A. (2004) Intron retention generates a novel Id3 isoform that inhibits vascular lesion formation. *J Biol Chem 279*, 32897-903.

[15] Zebedee, Z., and Hara, E. (2001) Id proteins in cell cycle control and cellular senescence. *Oncogene 20*, 8317-25.

[16] Yokota, Y. (2001) Id and development. *Oncogene 20*, 8290-8.

[17] Lyden, D., Young, A. Z., Zagzag, D., Yan, W., Gerald, W., O'Reilly, R., Bader, B. L., Hynes, R. O., Zhuang, Y., Manova, K., and Benezra, R. (1999) Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature 401*, 670-7.

[18] Christy, B. A., Sanders, L. K., Lau, L. F., Copeland, N. G., Jenkins, N. A., and Nathans, D. (1991) An Id-related helix-loop-helix protein encoded by a growth factor-inducible gene. *Proc Natl Acad Sci USA 88*, 1815-9.

[19] Ellis, H. M. (1994) Embryonic expression and function of the Drosophila helix-loop-helix gene, extramacrochaetae. *Mech Dev 47*, 65-72.

[20] Garrell, J., and Modolell, J. (1990) The Drosophila extramacrochaetae locus, an antagonist of proneural genes that, like these genes, encodes a helix-loop-helix protein. *Cell 61*, 39-48.

[21] Zhang, H., Reynaud, S., Kloc, M., Etkin, L. D., and Spohr, G. (1995) Id gene activity during Xenopus embryogenesis. *Mech Dev 50*, 119-30.

[22] Sawai, S., and Campos-Ortega, J. A. (1997) A zebrafish Id homologue and its pattern of expression during embryogenesis. *Mech Dev 65*, 175-85.

[23] Kee, Y., and Bronner-Fraser, M. (2001) The transcriptional regulator Id3 is expressed in cranial sensory placodes during early avian embryonic development. *Mech Dev 109*, 337-40.

[24] Ellmeier, W., Aguzzi, A., Kleiner, E., Kurzbauer, R., and Weith, A. (1992) Mutually exclusive expression of a helix-loop-helix gene and N-myc in human neuroblastomas and in normal development. *Embo J 11*, 2563-71.

[25] Deed, R. W., Hirose, T., Mitchell, E. L., Santibanez-Koref, M. F., and Norton, J. D. (1994) Structural organisation and chromosomal mapping of the human Id-3 gene. *Gene 151*, 309-14.

[26] Arnold, J. M., Mok, S. C., Purdie, D., and Chenevix-Trench, G. (2001) Decreased expression of the Id3 gene at 1p36.1 in ovarian adenocarcinomas. *Br J Cancer 84*, 352-9.

[27] Melnikova, I. N., and Christy, B. A. (1996) Muscle cell differentiation is inhibited by the helix-loop-helix protein Id3. *Cell Growth Differ 7*, 1067-79.

[28] Chen, B., Han, B. H., Sun, X. H., and Lim, R. W. (1997) Inhibition of muscle-specific gene expression by Id3: requirement of the C-terminal region of the protein for stable expression and function. *Nucleic Acids Res 25*, 423-30.

[29] Hata, K., Yoshimoto, T., and Mizuguchi, J. (2004) CD40 ligand rescues inhibitor of differentiation 3-mediated G1 arrest induced by anti-IgM in WEHI-231 B lymphoma cells. *J Immunol 173*, 2453-61.

[30] Bounpheng, M. A., Dimas, J. J., Dodds, S. G., and Christy, B. A. (1999) Degradation of Id proteins by the ubiquitin-proteasome pathway. *Faseb J 13*, 2257-64.

[31] Light, W., Vernon, A. E., Lasorella, A., Iavarone, A., and LaBonne, C. (2005) Xenopus Id3 is required downstream of Myc for the formation of multipotent neural crest progenitor cells. *Development 132*, 1831-41.

[32] Deng, X. W., Dubiel, W., Wei, N., Hofmann, K., Mundt, K., Colicelli, J., Kato, J., Naumann, M., Segal, D., Seeger, M., Carr, A., Glickman, M., and Chamovitz, D. A. (2000) Unified nomenclature for the COP9 signalosome and its subunits: an essential regulator of development. *Trends Genet 16*, 202-3.

[33] Berse, M., Bounpheng, M., Huang, X., Christy, B., Pollmann, C., and Dubiel, W. (2004) Ubiquitin-dependent degradation of Id1 and Id3 is mediated by the COP9 signalosome. *J Mol Biol 343*, 361-70.

[34] Liu, C. J., Ding, B., Wang, H., and Lengyel, P. (2002) The MyoD-inducible p204 protein overcomes the inhibition of myoblast differentiation by Id proteins. *Mol Cell Biol 22*, 2893-905.

[35] Ding, B., Liu, C. J., Huang, Y., Yu, J., Kong, W., and Lengyel, P. (2006) p204 protein overcomes the inhibition of the differentiation of P19 murine embryonal carcinoma cells to beating cardiac myocytes by Id proteins. *J Biol Chem 281*, 14893-906.

[36] Deed, R. W., Armitage, S., and Norton, J. D. (1996) Nuclear localization and regulation of Id protein through an E protein-mediated chaperone mechanism. *J Biol Chem 271*, 23603-6.

[37] Kurooka, H., and Yokota, Y. (2005) Nucleo-cytoplasmic shuttling of Id2, a negative regulator of basic helix-loop-helix transcription factors. *J Biol Chem 280*, 4313-20.

[38] Makita, J., Kurooka, H., Mori, K., Akagi, Y., and Yokota, Y. (2006) Identification of the nuclear export signal in the helix-loop-helix inhibitor Id1. *FEBS Lett 580*, 1812-6.

[39] Hara, E., Hall, M., and Peters, G. (1997) Cdk2-dependent phosphorylation of Id2 modulates activity of E2A-related transcription factors. *Embo J 16*, 332-42.

[40] Deed, R. W., Hara, E., Atherton, G. T., Peters, G., and Norton, J. D. (1997) Regulation of Id3 cell cycle function by Cdk-2-dependent phosphorylation. *Mol Cell Biol 17*, 6815-21.

[41] Ohtani, N., Zebedee, Z., Huot, T. J., Stinson, J. A., Sugimoto, M., Ohashi, Y., Sharrocks, A. D., Peters, G., and Hara, E. (2001) Opposing effects of Ets and Id proteins on p16INK4a expression during cellular senescence. *Nature 409*, 1067-70.

[42] Forrest, S. T., Taylor, A. M., Sarembock, I. J., Perlegas, D., and McNamara, C. A. (2004) Phosphorylation regulates Id3 function in vascular smooth muscle cells. *Circ Res 95*, 557-9.

[43] Atherton, G. T., Travers, H., Deed, R., and Norton, J. D. (1996) Regulation of cell differentiation in C2C12 myoblasts by the Id3 helix-loop-helix protein. *Cell Growth Differ 7*, 1059-66.

[44] Yokota, Y., and Mori, S. (2002) Role of Id family proteins in growth control. *J Cell Physiol 190*, 21-8.

[45] Norton, J. D., and Atherton, G. T. (1998) Coupling of cell growth control and apoptosis functions of Id proteins. *Mol Cell Biol 18*, 2371-81.

[46] Lasorella, A., Uo, T., and Iavarone, A. (2001) Id proteins at the cross-road of development and cancer. *Oncogene 20*, 8326-33.

[47] Loveys, D. A., Streiff, M. B., and Kato, G. J. (1996) E2A basic-helix-loop-helix transcription factors are negatively regulated by serum growth factors and by the Id3 protein. *Nucleic Acids Res 24*, 2813-20.

[48] Deed, R. W., Jasiok, M., and Norton, J. D. (1998) Lymphoid-specific expression of the Id3 gene in hematopoietic cells. Selective antagonism of E2A basic helix-loop-helix protein associated with Id3-induced differentiation of erythroleukemia cells. *J Biol Chem 273*, 8278-86.

[49] Bain, G., Cravatt, C. B., Loomans, C., Alberola-Ila, J., Hedrick, S. M., and Murre, C. (2001) Regulation of the helix-loop-helix proteins, E2A and Id3, by the Ras-ERK MAPK cascade. *Nat Immunol 2*, 165-71.

[50] Benezra, R. (2001) The Id proteins: targets for inhibiting tumor cells and their blood supply. *Biochim Biophys Acta 1551*, F39-47.

[51] Ruzinova, M. B., and Benezra, R. (2003) Id proteins in development, cell cycle and cancer. *Trends Cell Biol 13*, 410-8.

[52] Roberts, E. C., Deed, R. W., Inoue, T., Norton, J. D., and Sharrocks, A. D. (2001) Id helix-loop-helix proteins antagonize pax transcription factor activity by inhibiting DNA binding. *Mol Cell Biol 21*, 524-33.

[53] Yates, P. R., Atherton, G. T., Deed, R. W., Norton, J. D., and Sharrocks, A. D. (1999) Id helix-loop-helix proteins inhibit nucleoprotein complex formation by the TCF ETS-domain transcription factors. *Embo J 18*, 968-76.

[54] Perk, J., Iavarone, A., and Benezra, R. (2005) Id family of helix-loop-helix proteins in cancer. *Nat Rev Cancer 5*, 603-14.

[55] Hollnagel, A., Oehlmann, V., Heymer, J., Ruther, U., and Nordheim, A. (1999) Id genes are direct targets of bone morphogenetic protein induction in embryonic stem cells. *J Biol Chem 274*, 19838-45.

[56] Miyazono, K., and Miyazawa, K. (2002) Id: a target of BMP signaling. *Sci STKE 2002*, PE40.

[57] Kim, D. S., Franklyn, J. A., Boelaert, K., Eggo, M. C., Watkinson, J. C., and McCabe, C. J. (2006) Pituitary tumor transforming gene (PTTG) stimulates thyroid cell proliferation via a vascular endothelial growth factor/kinase insert domain receptor/inhibitor of DNA binding-3 autocrine pathway. *J Clin Endocr Metab 91*, 4603-11.

[58] Sakurai, D., Tsuchiya, N., Yamaguchi, A., Okaji, Y., Tsuno, N. H., Kobata, T., Takahashi, K., and Tokunaga, K. (2004) Crucial role of inhibitor of DNA binding/differentiation in the vascular endothelial growth factor-induced activation and angiogenic processes of human endothelial cells. *J Immunol 173*, 5801-9.

[59] Pammer, J., Reinisch, C., Kaun, C., Tschachler, E., and Wojta, J. (2004) Inhibitors of differentiation/DNA binding proteins Id1 and Id3 are regulated by statins in endothelial cells. *Endothelium 11*, 175-80.

[60] Sakurai, D., Yamaguchi, A., Tsuchiya, N., Yamamoto, K., and Tokunaga, K. (2001) Expression of ID family genes in the synovia from patients with rheumatoid arthritis. *Biochem Biophys Res Commun 284*, 436-42.

[61] Fong, S., Debs, R. J., and Desprez, P. Y. (2004) Id genes and proteins as promising targets in cancer therapy. *Trends Mol Med 10*, 387-92.

[62] Norton, J. D. (2000) ID helix-loop-helix proteins in cell growth, differentiation and tumorigenesis. *J Cell Sci 113 (Pt 22)*, 3897-905.

[63] Coppe, J. P., Smith, A. P., and Desprez, P. Y. (2003) Id proteins in epithelial cells. *Exp Cell Res 285*, 131-45.

[64] Desprez, P. Y., Sumida, T., and Coppe, J. P. (2003) Helix-loop-helix proteins in mammary gland development and breast cancer. *J Mammary Gland Biol 8*, 225-39.

[65] Benezra, R., Rafii, S., and Lyden, D. (2001) The Id proteins and angiogenesis. *Oncogene 20*, 8334-41.

II. Full-Length Id3 Protein: A Target for Peptide Chemists [1]

II.1 Introduction

The synthesis of small proteins by the chemical approach is an attractive alternative to the expression in bacteria as well as cell-free systems, as it allows to carry out a variety of chemical manipulations to the native target by the use of backbone surrogates, unnatural amino acids, conformational constraints or site-specific labels. All these synthetic tools are helpful for the analysis of the protein structure and folding. This chapter describes the total chemical synthesis and conformational study of the Id3 protein, and shows that the product obtained by chemical methods spontaneously undergoes self-association and forms disulfide-bonded dimers.

II.2 Solid-Phase Synthesis of the Id3 Protein

Stepwise solid-phase methodology is widely applied for the synthesis of peptides with chain length of up to ca. 50 residues. Above this limit, the accumulation of byproducts related to the lack of quantitative coupling/deprotection reactions represents the major problem [2]. Nevertheless, since the introduction of the solid-phase peptide synthesis (SPPS) by Merrifield [3], a number of stepwise syntheses of polypeptides and small proteins both with the Boc (*tert*-butyloxycarbonyl) and the Fmoc (9-fluorenylmethoxycarbonyl) strategies have been reported [4-14]. Thus, we decided to investigate whether the latter approach could be applied to the Id3 protein, too. The polypeptide chain was automatically synthesized on Wang resin which had been manually preloaded by treatment with Fmoc-His(trityl)-OH, HOBt (*N*-hydroxybenzotriazole), DIC (*N*,*N'*-diisopropylcarbodiimide), and DIPEA (*N*,*N*-diisopropylethylamine); the resulting loading was 0.6 mmol/g. Double couplings of N^{α}-Fmoc-protected amino acids were accomplished with *in situ* activation by HBTU (*O*-(1-benzotriazolyl)-1,1,3,3-tetramethyluronium hexafluorophosphate), HOBt and DIPEA in DMF/NMP (*N*,*N*-dimethylformamide/1-methyl-2-pyrrolidinone); the Fmoc group was removed by piperidine/DMF/NMP. Neither capping nor monitoring of the acylation yields was carried out. However, to control the quality of the growing chain, small amounts of peptide were detached from the resin after the coupling of Gln-66, Gly-57, Leu-41 and Ser-21. The cleavage mixture consisted of trifluoroacetic acid (TFA) containing 10% (v/v) scavengers: water, triisopropylsilane (TIS) and ethanedithiol (EDT). The good homogeneity of the four intermediate products (**II.1-4**) allowed us to continue the synthesis until Met-1 was

coupled (**II.5**). All five crude sequences, corresponding to the C-terminus 82-119 elongated with helix-2 (**II.1**), helix-2 and the loop (**II.2**), the entire HLH motif (**II.3**), the HLH motif and part of the N-terminus (**II.4**) and, finally, the entire Id3 protein (**II.5**) were then purified and characterized by mass spectrometry (Table II.1) and analytical HPLC (Figure II.1).

Table II.1. Amino acid sequence of the human Id3 protein and analytical data of the synthetic Id3 polypeptides **II.1**-**6**.

Human Id3[a]				
M[1]KALSPVRGCYEAVCCLSERSLAIARGRGKGPAAEEPLSL[40] LDDMNHCYSRLRELVP[56]GVPRGTQLS[65]QVEILQRVIDYILDLQ[81] VVLAEPAPGPPDGPHLPIQTAELAPELVISNDKRSFCH[119]				
Synthetic Id3 polypeptides				
Product	**Sequence**	**$MW_{calcd.}$ (Da)**	**MW_{found} (Da)**	**t_R (min)**[b]
II.1	Ac-(66-119)-OH	5999.0	5999.8	23.3
II.2	Ac-(57-119)-OH	6895.0	6895.4	21.4
II.3	Ac-(41-119)-OH	8838.2	8838.5	21.1
II.4	Ac-(21-119)-OH	10813.5	10813.4	19.8
II.5	H-(1-119)-OH	12969.1	12972.3	22.4
II.6	H-[Q-59, P-63]-(1-119)-OH	12969.1	12971.6	21.3

[a] Underlined sequences: helix-1 (41-56) and helix-2 (66-81); double underlined sequence: loop (57-65).
[b] Gradient: 10% ACN for 3 min, 10-70% ACN in 40 min.

The pure product **II.5** was subjected to tryptic digestion which cleaved the peptide bonds formed by the carboxyl groups of Lys and Arg. Additional molecular masses could be detected and assigned to disulfide-bonded fragments. This indicates that some oxidation processes of the five cysteine thiol groups (at positions 10, 15, 16, 47 and 118) occurred spontaneously.

The above-described synthetic procedure was further applied to the preparation of the full-length Id3 analogue **II.6**. This analogue contains two amino acid substitutions in the loop region: Pro-59→Gln and Gln-63→Pro. Previous site-directed mutagenesis studies of the Id1 protein carried out by Pesce and Benezra [15] showed that the loop of Id1 is critical for the protein function. Replacement of the loop regions of the four Id HLH motifs by the 6-amino-hexanoyl moiety also compromised the helical character [16], suggesting that the Id loops

play a role in the folding of the Id dimerization domains. One of the fully-conserved positions in the loops is occupied by proline, a residue which is known to influence the secondary structure of peptides and proteins and to induce a kink in the backbone due to the geometric constraints imposed by the five-membered ring. Therefore, we wanted to see whether shifting the position of proline within the loop would affect the protein folding. We chose to invert the native residues Pro-59 and Gln-63 located exactly two positions after the N-terminus and before the C-terminus of the loop, respectively. Synthetically, these two exchanges were not significant, as the analytical HPLC profiles of the crude products of **II.5** and **II.6** were comparable (Figure IX.8 in the Experimental Part). The product was confirmed by mass spectrometry with and without tryptic digestion after purification by preparative RP-HPLC.

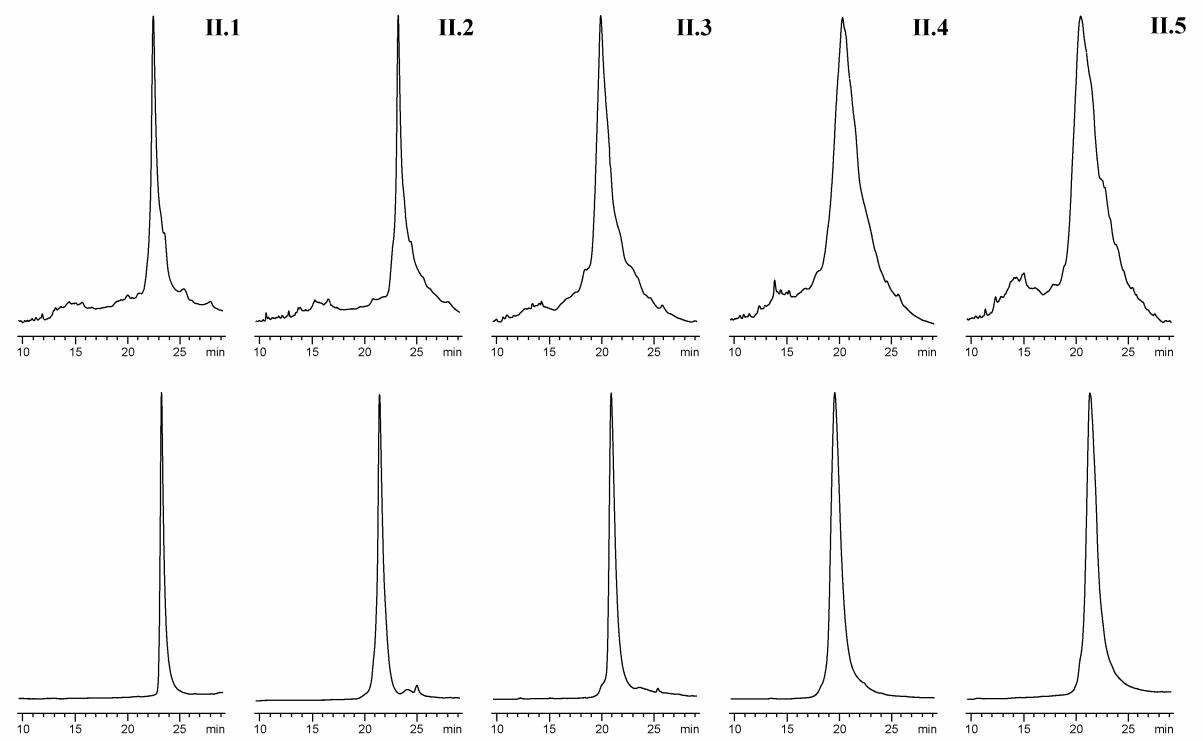


Figure II.1. Analytical RP-HPLC profiles of the Id3 polypeptides **II.1-5** before (*top*) and after (*bottom*) purification by preparative RP-HPLC.

II.3 Homodimerization of the Id3 Protein

Expressed Id3 protein has been shown to form homodimers [17]. Therefore, we investigated the self-association properties of the chemically synthesized protein. A solution of protein **II.5** (200 μM in 50 mM Tris buffer, pH 6, or in 50 mM phosphate buffer, pH 7.25) was flushed with nitrogen and incubated overnight to allow the association process. Then, one aliquot was exposed to air to allow the formation of disulfide bridges, whereas another aliquot

was crosslinked by glutaraldehyde. Finally, the untreated (fresh or overnight incubated), oxidized and crosslinked samples were analyzed by SDS-PAGE under reducing (in the presence of dithiothreitol, DTT) and non-reducing conditions. In the absence of DTT, the untreated Id3 samples showed one band for the monomer and one band for the dimer (Figure II.2, lanes c and e), whereas only the monomer was present upon treatment with DTT (Figure II.2, lanes b and d). These results suggest that the Id3 protein **II.5** spontaneously forms homodimers which can be stabilized by disulfide bridges. The presence of Id3 homodimers was also confirmed in samples oxidized by air (Figure II.2, lane f) and chemical crosslinking (Figure IX.18 in the Experimental Part). Mutant **II.6** also forms homodimers in non-reducing conditions (Figure II.2, lane h). This suggests that the two substitutions in the loop region do not influence the self-association process.

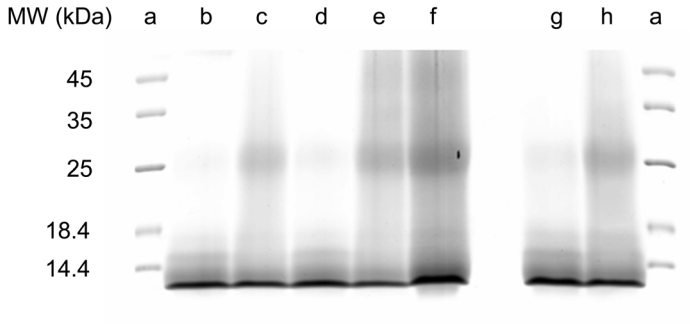


Figure II.2. SDS-PAGE characterization of the homodimers formed by the Id3 protein **II.5** or mutant **II.6**. The loaded samples correspond to protein **II.5**, freshly prepared (lanes b and c) or incubated overnight (lanes d and e) which was denatured in the presence (lanes b and d) or absence (lanes c and e) of DTT, and to protein **II.5** incubated overnight, oxidized by air and denatured in the absence of DTT (lane f). Lanes g and h correspond to a fresh sample of mutant **II.6** in the presence or absence of DTT, respectively. The molecular weight marker was loaded twice (lanes a).

II.4 Conformation of the Id3 Protein

The conformation of the synthetic Id3 protein **II.5** was studied by circular dichroism (CD) spectroscopy. The spectrum of the sample in phosphate buffer (0.1 M, pH 7.25) displayed an intensive negative band close to 206 nm (attributed to the $\pi \rightarrow \pi^*$ amide transition), a negative shoulder near 222 nm (attributed to the $n \rightarrow \pi^*$ amide transition) and a positive band below 200 nm (Figure II.3, right).

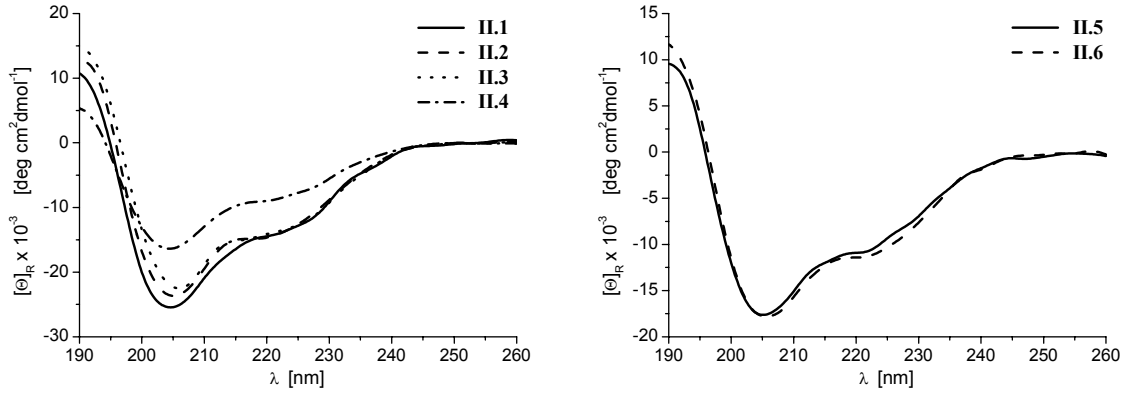


Figure II.3. CD spectra of the N-terminally truncated Id3 analogues **II.1-4** (*left*) and of the synthetic full-length Id3 protein **II.5** and its double mutant **II.6** (*right*). All samples (30 μM) were dissolved in phosphate buffer (0.1 M, pH 7.25).

Such a CD shape is similar to the one attributed to the α-helix conformation. However, the π→π* band is blue-shifted (by about 2 nm) and remarkably more intense than the n→π* band. A different type of helix, the 3_{10}-helix, is characterized by a strong negative Cotton effect in the region 202-206 nm and by a weak one near 222 nm, with a ratio $R = [\Theta_{222}]_R / [\Theta_{202-206}]_R$ equal to 0.3-0.4 [18]. This value was ~ 0.6 for the Id3 protein, an intermediate value between the typical values found for α-helices (R ~ 1) and 3_{10}-helices. This might suggest that both helix types coexist. However, the R value alone is not sufficient to assess the presence of a 3_{10}-helix [19], as other structural elements could play a role in determining the CD shape below 210 nm. In particular, unordered and extended conformations should be considered, because they both show a strong negative band close to 200 nm [20]. Moreover, we know from a previous study that the N- and C-terminal regions of the Id3 protein adopt no preferred conformation, at least as isolated domains [16], and it is probable that they remain partly flexible even in the entire protein. Accordingly, secondary structure analysis of the CD spectrum of Id3 using the CONTIN algorithm [21, 22] indicated that at least 30% of the sequence is unstructured (Table II.2). A similar overall conformation was observed for the double mutant **II.6** (Figure II.3, right panel, and Table II.2), which suggests that altering the position of a potential proline-induced kink of the backbone within the loop does not cause significant structural changes and that this region of the Id3 protein is flexible enough to tolerate such amino acid substitutions.

The CD properties of the full-length Id3 analogues were further compared to those of the N-terminally truncated Id3 fragments **II.1-3** (Figure II.3, left panel), which were obtained as intermediates during the synthesis of the full-length protein **II.5**. It was observed that the number of helical residues increased with the number of residues belonging to the HLH motif (Table II.2). As the C-terminal domain 82-119 is likely to be mainly unstructured [16], the helical fraction found for the peptides **II.1-3** (ca. 40% each) should be mostly attributed to the complete HLH motif (peptide **II.3**) or portions of it (peptides **II.1** and **II.2**). Moreover, the fact that the full-length protein **II.5** and the analogue **II.3** lacking the N-terminus (residues 1-40) displayed a similar number of helical residues indicates that the N-terminal part of the protein does not significantly contribute to the helix content of the whole protein.

Table II.2. Conformational properties of the synthetic Id3 polypeptides **II.1-4**, full-length Id3 protein **II.5** and its double mutant **II.6** at 30 μM in phosphate buffer (0.1 M, pH 7.25).

Product	No. of residues	% Helix (no. of residues)[a]	% Secondary structure elements by CONTIN[b] (no. of residues)			
			helix	β-strand	turns	unordered
II.1	54	39 (21)	40 (22)	5 (3)	23 (12)	32 (17)
II.2	63	39 (25)	43 (27)	5 (3)	22 (14)	30 (19)
II.3	79	38 (30)	44 (35)	5 (4)	21 (16)	30 (24)
II.4	99	24 (23)	20 (20)	24 (24)	20 (20)	36 (36)
II.5	119	29 (35)	34 (40)	12 (14)	24 (29)	30 (36)
II.6	119	31 (37)	38 (45)	7 (8)	24 (29)	31 (37)

[a] The helix percentage was estimated by the ellipticity value at 222 nm using the procedure of Chen and Yang [23] (for details see paragraph IX.7 in the Experimental Part).

[b] The CD spectra were analyzed by the CONTIN algorithm.

Surprisingly, the analogue **II.4** lacking the segment 1-20 was characterized by a moderate amount of β-strand structure and reduced helix character in comparison to the shorter analogue **II.3** and the full-length sequence **II.5**. Indeed, the ratio between the helix and the β-strand fractions was < 1 for the N-terminally truncated analogue **II.4**, and ~ 3 for the full-length protein **II.5**. However, in the presence of 2,2,2-trifluoroethanol (TFE, 40%, v/v, in phosphate buffer), the ratio α/β increased to ~ 3 also for the analogue **II.4** (Figure II.4). These

results indicate that a hydrophobic environment stabilized the helix structure, whereas a hydrophilic environment favored self-assembly and β-strand structure.

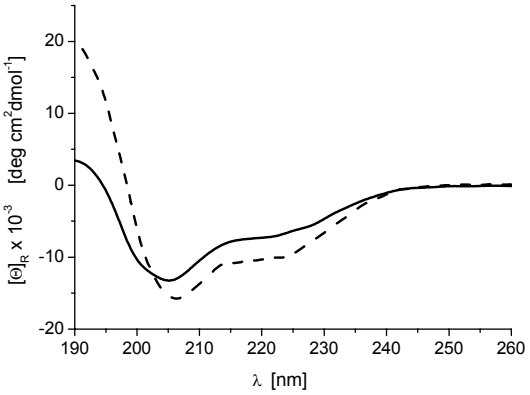


Figure II.4. Effect of TFE on the conformation of **II.4**. The solid line corresponds to the CD spectrum of a 91 μM sample in phosphate buffer (0.1 M, pH 7.25), whereas the dashed line to that of a 83 μM sample in phosphate buffer mixed with 40% TFE. The secondary structure element composition of peptide **II.4** in 40% TFE was estimated by the CONTIN algorithm and turned out to be as follows: 36% helix, 13% β-strand, 20% turn and 31% unordered structure.

The fact that increased β-strand content was not related to the presence (**II.5** and **II.6**) or absence (**II.3**) of the complete N-terminus (residues 1-40), but rather to the shortening of the N-terminus suggests that in the latter case the flanking region N-terminal to the HLH motif is crucial for secondary structure formation and/or aggregation. One possible explanation might be that residues 21-40 exert a destabilizing effect on the HLH fold – for example by pointing toward it – upon formation of a loop induced by the proline residues just proximal to the N-terminus of helix-1. This does not probably occur in the full-length protein due to the larger N-terminus. Moreover, as three out of the five cysteine residues of Id3 are located within the last 20 residues of the N-terminus, this region might be involved in additional intra-/inter-molecular interactions. Further studies will be necessary to understand the importance of the N-terminus in stabilizing the Id3 structure better.

II.5 Folding Stability of the Id3 Protein

The full-length Id3 protein **II.5** and its double-mutant **II.6** were subjected to chemical denaturation by guanidinium chloride. The titration curve, obtained by recording the change in the CD intensity at 222 nm of a solution of **II.5** in phosphate buffer (0.1 M, pH 7.25) at

room temperature upon denaturant addition, is shown in Figure II.5. High concentrations of the chaotropic agent prevented recording of the spectra at wavelengths below 210 nm; however, the change of the band at 222 nm could be observed, and exhibited a sigmoidal decrease in intensity with increasing concentration of the denaturant.

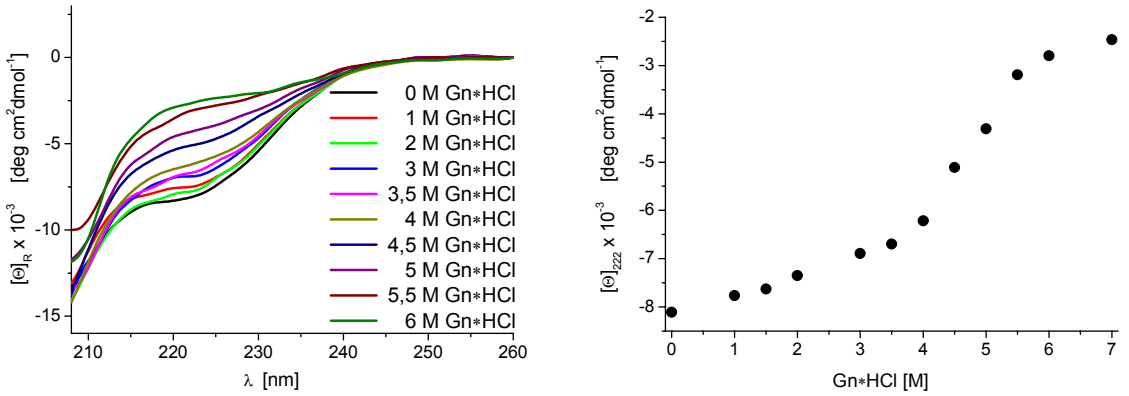


Figure II.5. Effect of increasing guanidinium chloride concentration on the helix content of protein **II.5** (100 μM in 0.1 M phosphate buffer, pH 7.25) at room temperature: CD spectra (*left*) and curve of the value of $[\Theta]_{R222}$ as a function of the guanidinium chloride concentration (*right*).

In order to compare the stability of the full-length proteins with those of the N-terminally truncated analogues **II.3** and **II.4**, the denaturation curves were expressed as the folded fraction (%) versus denaturant concentration (Figure II.6). The native and double-mutated full-length protein behaved similarly to the polypeptide **II.3**, lacking the N-terminus. They displayed a pre-unfolding phase characterized by a linear reduction of the folded fraction upon denaturant concentrations below 3.5 M, followed by a sharp decrease between 4 M and 5.5 M concentration of the denaturant. Assuming a two-state model transition that involves only the native (*N*) state and unfolded (*U*) state, all data sets but those of **II.4** were fitted using a non-linear least-square regression [24], which yielded the Gibbs free energy change for unfolding in aqueous solution in the absence of denaturant, $\Delta G^0_{H_2O}$ (for details see Equation IX.6 in the Experimental Part). This value was 28 kJ/mol and 26 kJ/mol for the full-length products **II.5** and **II.6**, respectively, whereas the values corresponding to the unfolding of the analogue **II.3** lacking the complete N-terminus and of the isolated Id3 HLH motif turned out to be of 17 kJ/mol and 22 kJ/mol, respectively. This suggests that the full-length protein was more stable than the shorter analogues lacking the N-terminal domain or both the N- and C-terminal domains, and therefore that both flanking regions play a stabilizing role in the Id3

fold. The $\Delta G^0_{H_2O}$ value found for the Id3 HLH domain was smaller than the one reported in the literature for the bHLH motifs of the Id3 dimerization partners E47 $\Delta G^0_{H_2O}$ = 44 kJ/mol) and MyoD ($\Delta G^0_{H_2O}$ = 33 kJ/mol) [25]: this is in agreement with the observation that the Id3 protein forms less stable homodimers than the corresponding binding partners [17].

In contrast, the polypeptide **II.4** lacking only the N-terminal segment 1-20 and containing a significant amount of β-strand structure exhibited a different behavior: there was an immediate loss of folded fraction already in 0.5 M denaturant, which, however, slowed down until the denaturant concentration reached 3.5 M. After that point, a sharp decrease of the curve was observed. More than one transition probably occurred, the first at submolar denaturant concentration, another one at ~ 3 M denaturant concentration and the last one above 4 M denaturant concentration.

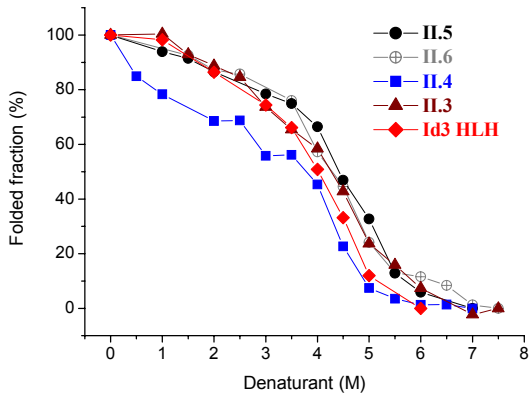


Figure II.6. Denaturation curves of **II.3-6** and of the Id3 HLH peptide (residues 41-81) obtained by titration of the peptide solutions in phosphate buffer (0.1 M, pH 7.25) by guanidinium chloride (*left*). The peptide concentration was in the range of 50-100 μM.

II.6 Conclusion

The Id3 protein has been shown to be a valuable target for solid-phase synthesis. This encourages the preparation of further chemically modified analogues of Id3 which will enable the investigation of its biophysical features. The results presented above have already shown that the synthetic Id3 protein adopts a stable α-helical conformation in the core region, and self-associates to form disulfide-bonded dimers.

II.7 Literature

[1] Svobodova, J., and Cabrele, C. (2006) Stepwise solid-phase synthesis and spontaneous homodimerization of the helix-loop-helix protein Id3. *Chembiochem 7*, 1164-8.

[2] Kent, S. B. (1988) Chemical synthesis of peptides and proteins. *Annu Rev Biochem 57*, 957-89.

[3] Merrifield, R. B. (1963) Solid Phase Peptide Synthesis. I. The Synthesis of a Tetrapeptide. *J Am Chem Soc 85*, 2149-2154.

[4] Clark-Lewis, I., Moser, B., Walz, A., Baggiolini, M., Scott, G. J., and Aebersold, R. (1991) Chemical synthesis, purification, and characterization of two inflammatory proteins, neutrophil activating peptide 1 (interleukin-8) and neutrophil activating peptide. *Biochemistry 30*, 3128-35.

[5] Dawson, P. E., and Kent, S. B. (2000) Synthesis of native proteins by chemical ligation. *Annu Rev Biochem 69*, 923-60.

[6] Goetz, M., Geoffre, S., Busetta, B., Manigand, C., Nespoulous, C., Londos-Gagliardi, D., Guillemain, B., and Hospital, M. (1997) Synthesis and CD studies of an 88-residue peptide containing the main receptor binding site of HTLV-I SU-glycoprotein. *J Pept Sci 3*, 347-53.

[7] Goud, N. A., McKee, R. L., Sardana, M. K., DeHaven, P. A., Huelar, E., Syed, M. M., Goud, R. A., Gibbons, S. W., Fisher, J. E., Levy, J. J., and et al. (1991) Solid-phase synthesis and biologic activity of human parathyroid hormone (1-84). *J Bone Miner Res 6*, 781-9.

[8] Gutte, B., and Merrifield, R. B. (1971) The synthesis of ribonuclease A. *J Biol Chem 246*, 1922-41.

[9] Abiko, T., and Sekino, H. (1993) Synthesis of prothymosin alpha deduced from nucleotide sequence of the murine cDNA and its effect on the impaired T lymphocytes of uremic patients. *Biotechnol Ther 4*, 213-20.

[10] Bonetto, V., Massignan, T., Chiesa, R., Morbin, M., Mazzoleni, G., Diomede, L., Angeretti, N., Colombo, L., Forloni, G., Tagliavini, F., and Salmona, M. (2002) Synthetic miniprion PrP106. *J Biol Chem 277*, 31327-34.

[11] Ferrer, M., Woodward, C., and Barany, G. (1992) Solid-phase synthesis of bovine pancreatic trypsin inhibitor (BPTI) and two analogues. A chemical approach for evaluating the role of disulfide bridges in protein folding and stability. *Int J Pept Protein Res 40*, 194-207.

[12] Fukuda, H., Irie, K., Nakahara, A., Ohigashi, H., and Wender, P. A. (1999) Solid-phase synthesis, mass spectrometric analysis of the zinc-folding, and phorbol ester-binding studies of the 116-mer peptide containing the tandem cysteine-rich C1 domains of protein kinase C gamma. *Bioorg Med Chem 7*, 1213-21.

[13] Kaiser, T., Luppa, P., and Voelter, W. (1999) Solid-phase synthesis, isolation and analysis of a mouse protein, the macrophage migration inhibitory factor. *J Chromatogr A 852*, 189-95.

[14] Dong, C. Z., Romieu, A., Mounier, C. M., Heymans, F., Roques, B. P., and Godfroid, J. J. (2002) Total direct chemical synthesis and biological activities of human group IIA secretory phospholipase A2. *Biochem J 365*, 505-11.

[15] Pesce, S., and Benezra, R. (1993) The loop region of the helix-loop-helix protein Id1 is critical for its dominant negative activity. *Mol Cell Biol 13*, 7874-80.

[16] Kiewitz, S. D., and Cabrele, C. (2005) Synthesis and conformational properties of protein fragments based on the Id family of DNA-binding and cell-differentiation inhibitors. *Biopolymers 80*, 762-74.

[17] Wibley, J., Deed, R., Jasiok, M., Douglas, K., and Norton, J. (1996) A homology model of the Id-3 helix-loop-helix domain as a basis for structure-function predictions. *Biochim Biophys Acta 1294*, 138-46.

[18] Toniolo, C., Crisma, M., Formaggio, F., Peggion, C., Broxterman, Q. B., and Kaptein, B. (2004) Molecular spacers for physicochemical investigations based on novel helical and extended peptide structures. *Biopolymers 76*, 162-76.

[19] Formaggio, F., Peggion, C., Crisma, M., Kaptein, B., Broxterman, Q. B., Mazaleyrat, J. P., Wakselman, M., and Toniolo, C. (2004) Recent contributions of electronic circular dichroism to the investigation of oligopeptide conformations. *Chirality 16*, 388-97.

[20] Sreerama, N., and Woody, R. W. (1994) Poly(pro)II helices in globular proteins: identification and circular dichroic analysis. *Biochemistry 33*, 10022-5.

[21] Provencher, S. W., and Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry 20*, 33-7.

[22] van Stokkum, I. H., Spoelder, H. J., Bloemendal, M., van Grondelle, R., and Groen, F. C. (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal Biochem 191*, 110-8.

[23] Chen, Y. H., Yang, J. T., and Chau, K. H. (1974) Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry 13*, 3350-9.

[24] Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry 27*, 8063-8.

[25] Wendt, H., Thomas, R. M., and Ellenberger, T. (1998) DNA-mediated folding and assembly of MyoD-E47 heterodimers. *J Biol Chem 273*, 5735-43.

III. Conformational Studies of the Id3 Helix-Loop-Helix Motif

III.1 Introduction

The Id3 protein forms heterodimers with bHLH transcription factors, in particular with the ubiquitously expressed members of the E family (E2A, E2-2, HEB), and thus inhibits the formation of DNA-binding bHLH homo- or heterodimers [1, 2]. The HLH motif spanning the Id3 residues 41-81 (peptide **III.1**) is required for dimerization. The latter, its analogues, and sequences corresponding to some of its binding partners were prepared by standard SPPS using the Fmoc strategy (Table III.1). Peptides **III.2** and **III.3** containing the substitution Phe→Tyr at positions 48 and 76, respectively, were prepared to see whether the hydroxyl group is involved in potential hydrogen bonding and consequent conformational stability. The importance of the cysteine residue within the first helix was evaluated using peptide **III.4**, which contains the replacement Gly→Cys. The secondary structure of all these HLH analogues was analyzed by CD spectroscopy and compared to that of the native Id3 HLH motif **III.1**. The HLH motifs of the bHLH proteins E47, E12, and MyoD (**III.5**, **III.6** and **III.7**) were prepared to compare their helical properties to those of the Id3 HLH peptide. Both helical segments of the Id3 HLH domain (**III.8** and **III.9**) were also synthesized separately, and their conformation in solution was studied

III.2 Solid-Phase Synthesis of the HLH Motifs of Id3, E47, E12, and MyoD

The peptides were synthesized on Rink amide resin using a double coupling of N^{α}-Fmoc-protected amino acids and activation *in situ* by HBTU/HOBt/DIPEA. Once the synthesis was complete, the N-terminus was acetylated and the peptides were cleaved from the resin, purified by preparative RP-HPLC, and characterized by analytical RP-HPLC and MALDI-TOF mass spectrometry (Figure IX.4 and IX.5 in the Experimental Part). The amino acid sequences of the native HLH motifs of human Id3, E47, E12 and MyoD are reported in Table III.1, and correspond to the results of sequence alignment studies [3]. Unlike the helical parts, the loops are completely non-conserved.

The HLH motif of Id3 (peptide **III.1**) contains two tyrosine residues at positions 48 and 76, which are conserved within the Id family, but not in the HLH family. Indeed, the Id3 Tyr-48 is replaced by a phenylalanine residue in E47 (**III.5**), E12 (**III.6**) and MyoD (**III.7**), whereas the Id3 Tyr-76 is replaced by a valine residue in E47 and E12, but conserved in MyoD. To see whether the presence of the tyrosine residues in Id3 is critical or not, the two

analogues **III.2** and **III.3**, in which either position was modified with phenylalanine, were prepared.

The Id3 cysteine residue is also conserved within the Id family, whereas it is replaced by an alanine residue in E47, E12 and MyoD. Again, to study the role of this position, the Id3 Cys-47 residue was replaced by a glycine residue in peptide **III.4**.

The peptides **III.5-7** were found to be much more difficult sequences than peptide **III.1**. In fact, analysis of the aggregation potential of the four sequences by using the program Peptide Companion predicted that the aggregation risk was high for the chains corresponding to E47, E12 and MyoD, and medium for the chain corresponding to Id3. After TFA cleavage, peptides **III.5-6** had to be additionally treated with trimethylsilylbromide (TMSBr)/EDT/TFA to reduce the oxidized methionine residues. This treatment was not necessary for peptide **III.1**.

Table III.1. Analytical data of the HLH motifs **III.1-7** and of helix-1 (**III.8**) and helix-2 (**III.9**) of Id3.

HLH sequences[a]				
Ac-LDDMNH**CY**SRLRELVP GVPRGTQLS QVEILQRVID**Y**ILDLQ-NH_2				hId3
Ac-VRDINEAFRELGRMCQ MHLKSDKAQT KLLILQQAVQVILGL-NH_2				hE47
Ac-VRDINEAFKELGRMCQ LHLNSEKPQT KLLILHQAVSVILNL-NH_2				hE12
Ac-LSKVNEAFETLKRCTS SNPNQRLP KVEILRNAIRYIEGL-NH_2				hMyoD
Synthesized peptides				
Product	**Sequence**[b]	**$MW_{calcd.}$(Da)**	**MW_{found} (Da)**	**t_R (min)**[c]
III.1	Ac-(41-81)-NH_2	4838.6	4839.6	22.1
III.2	Ac-[F-48]-(41-81)-NH_2	4822.6	4822.4	21.9
III.3	Ac-[F-76]-(41-81)-NH_2	4822.6	4829.0	22.6
III.4	Ac-[G-47]-(41-81)-NH_2	4792.5	4788.6	21.3
III.5	Ac-(559-599)-NH_2	4749.7	4751.0	21.2
III.6	Ac-(562-602)-NH_2	4754.7	4753.0	21.5
III.7	Ac-(122-160)-NH_2	4545.3	4541.4	19.0
III.8	Ac-(41-57)-NH_2	2058.4	2058.4	15.2
III.9	Ac-(65-81)-NH_2	2086.4	2085.5	26.6

[a] The underlined sequence corresponds to the loop region.

[b] Id3 peptides, with the exception of peptide **III.5** (hE47), peptide **III.6** (hE12) and peptide **III.7** (hMyoD).

[c] Gradient: 10% ACN for 3 min, 10-70% ACN in 40 min.

III.3 CD Spectroscopy

Secondary structure investigations of the synthetic peptides reported in Table III.1 were carried out by CD spectroscopy in phosphate buffer (0.1 M, pH 7.25), with and without 30% TFE, which stabilizes secondary structure elements by promoting the formation of intramolecular hydrogen bonds. As peptides **III.5** and **III.6** were not soluble in pure phosphate buffer, 30% TFE was added to increase their solubility. For comparison, a solution of peptide **III.1** in buffer with and without TFE was examined as well. The following stock solutions were prepared: 100 μM for **III.1**, **III.5-9**, 83 μM for **III.2**, 96 μM for **III.3** and 108 μM for **III.4**.

In phosphate buffer, the CD spectrum of **III.1** was characterized by a negative band close to 208 nm, attributed to the $\pi \rightarrow \pi^*$ transition, followed by a second negative band located near 222 nm, attributed to the $n \rightarrow \pi^*$ transition. A positive band was present below 200 nm (Figure III.1). This CD spectrum is typical for an α-helix, including the ratio R of ~ 1 between two minima ($R = [\Theta_{222}]_R / [\Theta_{202-206}]_R$). Composition of secondary structure elements of peptide **III.1** was extrapolated by the dichroic data using the CONTIN algorithm [4, 5]. Beside a high helical fraction (65%), β-strand (6%), turn (7%) and unordered (21%) elements have been found (Table III.2). Interestingly, the R value decreased from ~ 1 to 0.77 in the presence of TFE, and a significant blue shift of the whole CD spectrum was observed (Figure III.1). Such dichroic differences were interpreted by the CONTIN algorithm as an increase in turn and unordered structure at the expense of the helical content (Table III.2). This was unexpected, due to the well-known ability of TFE to induce ordered conformations; however, it is also known that TFE tends to stabilize single helices rather than packed helices or coiled coils. This is also supported by the decrease in the R value, as values close to and higher than 1 usually indicate helix packing or coiled coils [6]. Therefore, it is likely that TFE disturbs the highly compact helical packing of the Id3 HLH motif, and weakens its helical character.

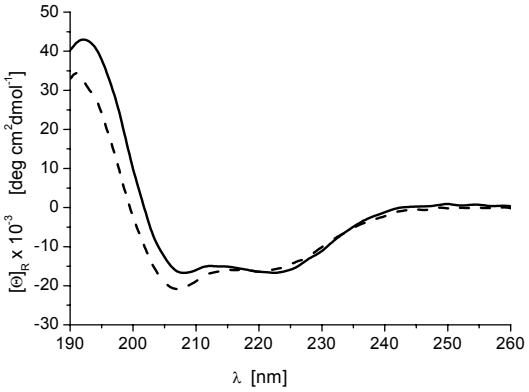


Figure III.1. CD spectra of the synthetic HLH motif of Id3 (**III.1**, 100 μM) in phosphate buffer (0.1 M, pH 7.25) (solid line) and in the presence of 30% TFE (dashed line).

The importance of the helix packing for a stable conformation of the Id3 HLH motif is supported by the observation that the isolated Id3 helix-1 and helix-2, peptides **III.8** and **III.9**, respectively, showed a different propensity to adopt helical conformation, with helix-1 being much less helical than helix-2 (Figure III.2). This suggests that helix-2 possesses intrinsic helicity, whereas the conformation of helix-1 may depend on the environment. Accordingly, the presence of 30% TFE stabilized helix-1, and a significant transition from unordered to helical conformation was observed: the negative band at 199.8 nm shifted to 206.4 nm, whereas a positive band below 198 nm appeared. The number of helical residues rose up to 10. Helix-2 was further stabilized by the presence of 30% TFE, too, and the number of helical residues increased from 10 up to 16 (Table III.2).

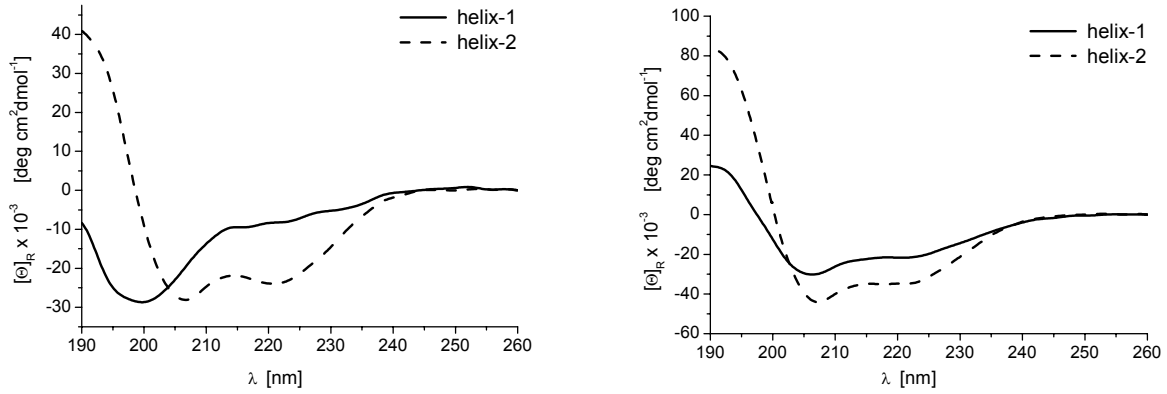


Figure III.2. CD spectra of helix-1 (**III.8**) and helix-2 (**III.9**) at 100 μM concentration in phosphate buffer (0.1 M, pH 7.25) (*left*) and in the presence of 30% TFE (*right*).

The analogues of the Id3 HLH motif **III.2**, **III.3** and **III.4** are helical in phosphate buffer, too (Figure III.3). However, unlike peptide **III.2** (Phe→Tyr-48), peptide **III.3** (Phe→Tyr-76) contains less helical residues than peptide **III.1** (17 vs. 27, Table III.2). This indicates that the Tyr-76 is more strongly involved in hydrogen bonding than Tyr-48, stabilizing the helical conformation.

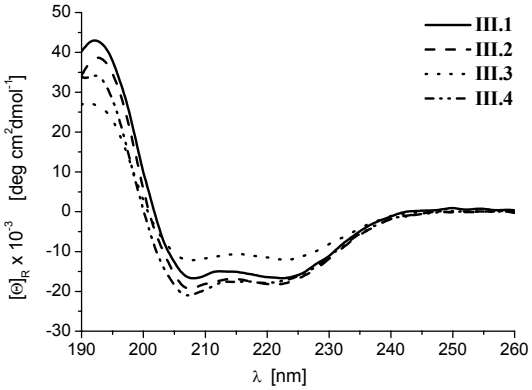


Figure III.3. CD spectra of the native Id3 HLH motif **III.1** and its analogues **III.2-4** at ~ 100 μM concentration in phosphate buffer (0.1 M, pH 7.25).

Peptide **III.4** (Gly→Cys-47), measured in phosphate buffer, also contains less helical residues than peptide **III.1**, which suggests that Cys-47 plays a role in the stabilization of the HLH folding.

Peptides **III.5-7**, representing the HLH motifs of the bHLH proteins E47, E12 and MyoD, were generally less helical than the Id3 HLH motif. Peptide **III.7** was the only one which was soluble in phosphate buffer (0.1 M, pH 7.25), whereas the others required the presence of 30% TFE (Figure III.4). The latter peptides have strong tendency to aggregate and the α-helical contribution to the CD spectra was low. Accordingly, a significant amount of β-strand fraction was estimated by the CONTIN algorithm, while the number of helical residues was only 7 or 8 (for peptides **III.5** or **III.6**, respectively) vs. 21 for peptide **III.1** (Table III.2). In contrast, peptide **III.7**, which could be measured in pure phosphate buffer, showed a different CD shape characterized by a blue-shifted $\pi\rightarrow\pi^*$ band at 202 nm, followed by the $n\rightarrow\pi^*$ band near 222 nm, with *R* value of 0.43. These dichroic properties resemble the ones of 3_{10}-helix. Obviously, peptide **III.7** also contains less helical residues than peptide **III.1** (5 vs. 27), and the composition of secondary structure elements is comparable to that of the peptides **III.5-6**; however, the ratio between regular and irregular α-helical elements was 1 for **III.5-6** and 0.5 for **III.7**, which would might reflect the presence of 3_{10}-helical turns in **III.7**.

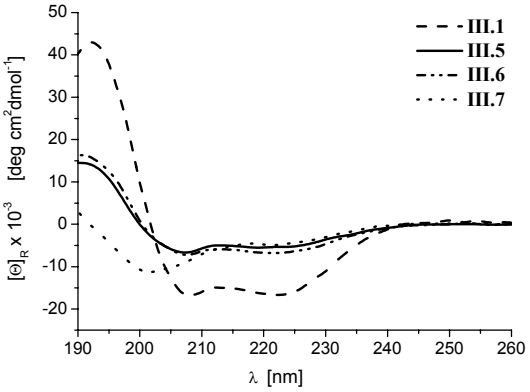


Figure III.4. CD spectra of the native Id3 HLH motif **III.1** and its bHLH binding partners **III.5-7**. The samples (100 μM) were dissolved in phosphate buffer (0.1 M, pH 7.25; peptides **III.1** and **III.7**) or in a mixture of phosphate buffer (0.1 M, pH 7.25) and TFE (70:30, v/v, peptides **III.5** and **III.6**).

Peptide **III.1** was characterized by a ratio α/β much higher than 1 (7 in the presence of TFE and 11 in pure phosphate buffer), whereas peptides **III.5-7** showed a α/β ratio smaller than 1 (0.6 for peptide **III.5**, 0.7 for peptide **III.6**, and 0.4 for peptide **III.7**). Interestingly, peptide **III.5** underwent a significant β→α transition upon incubation at 10 °C for 7 days (Figure III.5 and Table III.2). The α/β ratio increased from 0.6 to 2.7. This phenomenon was observed only for the HLH motif of E47, but not for E12 and MyoD, which suggests that the HLH motif of E47 has a superior helix propensity allowing for a slow conformational change towards the helix formation.

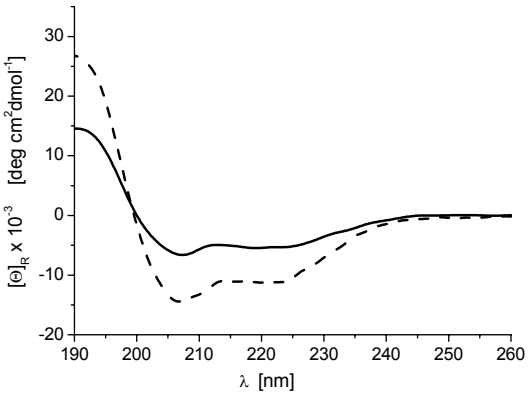


Figure III.5. CD spectra of a 100 μM solution of the HLH motif of E47 (peptide **III.5**) in a mixture of phosphate buffer (0.1 M, pH 7.25) and TFE (70:30, v/v), recorded immediately after preparation (solid line) and after incubation at 10 °C for 7 days (dashed line).

Table III.2. Conformational properties of the synthetic peptides **III.1**-**9** in phosphate buffer (0.1 M, pH 7.25) and/or in the presence of 30% TFE.

Product	Concentration (μM)	% Helix (no. of residues)[a]	% Secondary structure elements by CONTIN[b] (no. of residues)			
			helix	β-strand	turns	unordered
III.1	100	46 (19)	65 (27)	6 (2)	7 (3)	21 (9)
III.1[c]	100	46 (19)	52 (21)	7 (3)	16 (6)	26 (11)
III.2	83	51 (21)	60 (25)	4 (2)	15 (6)	21 (8)
III.3	96	34 (14)	42 (17)	12 (5)	17 (7)	29 (12)
III.4	108	49 (20)	54 (22)	6 (2)	16 (7)	24 (10)
III.5[c]	100	15 (6)	17 (7)	29 (12)	20 (8)	34 (14)
	100[d]	32 (13)	39 (16)	14 (6)	18 (7)	29 (12)
III.6[c]	100	20 (8)	19 (8)	26 (11)	20 (8)	35 (14)
	100[d]	15 (6)	17 (7)	29 (12)	20 (8)	35 (14)
III.7	100	13 (5)	12 (5)	31 (12)	21 (8)	36 (14)
	100[d]	15 (6)	15 (6)	29 (11)	22 (9)	34 (13)
III.8	100	29 (5)	9 (2)	7 (1)	30 (5)	54 (9)
III.8[c]	100	65 (11)	60 (10)	4 (1)	26 (4)	10 (2)
III.9	100	76 (13)	59 (10)	3 (1)	14 (2)	24 (4)
III.9[c]	100	100 (17)	95 (16)	0 (0)	0 (0)	5 (1)

[a] The helix percentage was estimated by the ellipticity value at 222 nm using the procedure of Chen and Yang [7]. The peptide length was the same for all samples (41 residues), with the exception of peptides **III.7** (39 residues), **III.8** and **III.9** (both 17 residues).

[b] The CD spectra were analyzed by the CONTIN algorithm.

[c] The peptide was measured in the presence of 30% TFE.

[d] The sample was incubated at 10 °C for 7 days.

III.4 Discussion and Conclusion

The Id3 helix-2 is intrinsically helical, whereas the conformation of the Id3 helix-1 may depend on the environment: indeed, it becomes helical after connection to helix-2 through the loop region, as observed for the HLH peptide **III.1**, or upon addition of TFE. This behavior is in agreement with the results reported in Chapter II, where the helical content of the Id3 sequence 57-119 (peptide **II.2**) increased after elongation with helix-1 (peptide **II.3**).

The hydroxyl group of the tyrosine residue in the helix-1 does not seem to be important for the conformation, as the Phe-48 mutant did not show any changes in the secondary structure composition. Tyr-48 is probably participating in the π-π interactions and not in the hydrogen bonding. In contrast, Tyr-76 must be involved in the hydrogen bonding which stabilizes the Id3 HLH motif, because after its replacement with the phenylalanine residue, there was a significant decrease in the number of helical residues. Accordingly, a homology model of the Id3 HLH dimer proposed by Norton and co-workers [8] shows an interchain contact between the side chain of Tyr-76 and the side chain of Gln-81 (Figure III.6). Therefore, the role of Tyr-76 seems to be important for the stabilization of the dimer.

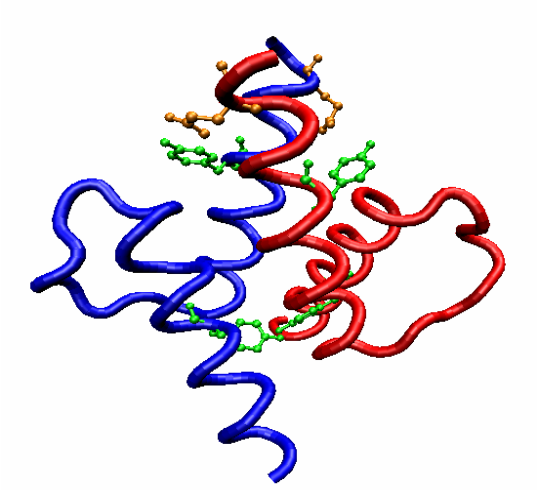


Figure III.6. The crystal structure of the DNA-bound dimer of the MyoD bHLH motif can serve as a model for the Id3 HLH motif because of the highest sequence identity and conserved residue similarity to the Id3 sequence over helix-1 and helix-2 [8]. The side chains of Tyr-76 and Gln-81 are indicated in green and orange, respectively. The crystal structure of MyoD was taken from the PDB data bank (PDB ID: 1MDY) [9], the sequence substitutions were done by using the spdbv program and the model structure was created with the VMD program.

Cys-47 is also important for the conformation: indeed, the glycine mutation negatively affected the helix character leading to an increase in the number of turn elements.

The HLH motifs of the bHLH proteins E47, E12 and MyoD were generally less helical than the Id3 HLH motif; this suggests that the HLH domains of the bHLH transcription factors require the binding to DNA or at least the presence of the N-terminal DNA binding domain to fold in a stable manner [10].

III.5 Literature

[1] Benezra, R., Davis, R. L., Lockshon, D., Turner, D. L., and Weintraub, H. (1990) The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell 61*, 49-59.

[2] Norton, J. D., Deed, R. W., Craggs, G., and Sablitzky, F. (1998) Id helix-loop-helix proteins in cell growth and differentiation. *Trends Cell Biol 8*, 58-65.

[3] Chavali, G. B., Vijayalakshmi, C., and Salunke, D. M. (2001) Analysis of sequence signature defining functional specificity and structural stability in helix-loop-helix proteins. *Proteins 42*, 471-80.

[4] Provencher, S. W., and Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry 20*, 33-7.

[5] van Stokkum, I. H., Spoelder, H. J., Bloemendal, M., van Grondelle, R., and Groen, F. C. (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal Biochem 191*, 110-8.

[6] Su, J. Y., Hodges, R. S., and Kay, C. M. (1994) Effect of chain length on the formation and stability of synthetic alpha-helical coiled coils. *Biochemistry 33*, 15501-10.

[7] Chen, Y. H., Yang, J. T., and Chau, K. H. (1974) Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry 13*, 3350-9.

[8] Wibley, J., Deed, R., Jasiok, M., Douglas, K., and Norton, J. (1996) A homology model of the Id-3 helix-loop-helix domain as a basis for structure-function predictions. *Biochim Biophys Acta 1294*, 138-46.

[9] Ma, P. C., Rould, M. A., Weintraub, H., and Pabo, C. O. (1994) Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell 77*, 451-9.

[10] Wendt, H., Thomas, R. M., and Ellenberger, T. (1998) DNA-mediated folding and assembly of MyoD-E47 heterodimers. *J Biol Chem 273*, 5735-43.

IV. Conformational Fragment Analysis of the Id3 Protein

IV.1 Introduction

Previous chapter showed that the Id3 helix-1 is helical either in the presence of TFE or when connected to helix-2 which possesses intrinsic helicity. Accordingly, the entire HLH motif of Id3 is highly helical. In contrast, the regions preceding and following it do not prefer any conformation as isolated fragments [1]. In order to have some insights into the conformational interplay of the single subdomains, conformation of a series of peptides which span adjacent subdomains was analyzed (peptides **IV.2-5** and **IV.7-14** in Figure IV.1). For example, peptide **IV.2** reproduces the C-terminal part of the N-terminus and helix-1, and peptide **IV.3** reproduces helix-2 preceded by the loop and followed by the N-terminal part of the C-terminus. Comparison of these two peptides with the ones spanning only helix-1 (**III.8**) or helix-2 (**III.9**) gives an idea about the role of the adjacent regions on the conformation of the two helices. Similarly, peptides **IV.5** and **IV.7** suggest the influence of the N- and/or C-terminal elongations on the central HLH motif

Beside the Id3 fragments containing the native amino acid sequence, additional analogues displaying some amino acid substitutions have been synthesized. For example, peptide **IV.4** represents a mutant of peptide **IV.3**, in which four natural amino acids within helix-2 have been replaced with the achiral α-aminoisobutyryl (Aib) moieties, C^{α}-tetrasubstituted residues, which should change the helical character of the peptide segment. It is known that peptide sequences rich in Aib residues tend to fold into a 3_{10}-helical conformation which is stabilized by *i*, *i+3* interactions [2-7]. 3_{10}-Helices are usually very short and often located at the termini of α-helical segments and have been proposed as intermediates in the folding-unfolding process of α-helices. The entropic penalty for the necessary loop closure for the formation of *i*, *i+3* versus *i*, *i+4* hydrogen bonds is lower.

A phosphorylation site corresponding to Ser-5 of Id3—also present in Id2 and Id4—modulates the protein activity during the cell cycle and alters the bHLH protein binding specificity [8-11]. Thus, the phosphorylated peptide **IV.10** has been prepared and compared to the unphosphorylated peptide **IV.9** to see the effect of the posttranslational modification. Moreover, peptide **IV.11**, in which the consensus sequence for phosphorylation has been modified by replacing Pro-6 with glutamine residue, should suggest whether the N-end of the protein is sensitive to point mutations. Similarly, peptide **IV.8** representing the N-terminally truncated analogue of peptide **IV.9** should suggest the role of the flanking region 1-20. The latter

contains three cysteine residues (Cys-10, Cys-15 and Cys-16) which might form intra- and intermolecular disulfide bridges involving also Cys-47 of helix-1. The importance of the latter residue was evaluated by using the corresponding Gly-47 mutant of peptide **IV.9** (**IV.12**) or of peptide **II.3** (**IV.14**). Also, a double mutant of the native sequence 57-119 (peptide **IV.13**) containing two substitutions in the loop region was prepared, in which Pro-59 was replaced by glutamine residue and Gln-63 by proline residue to see if the presence of the cyclic amino acid preceding the N-end of helix-2 might influence the latter. The proline residue is a frequent helix terminator due to its rigid conformation and the inability of the N^{α}-atom to participate in a hydrogen bond, but it can also precede the beginning of a α-helix; therefore, shifting the proline position from 59 to 63 might potentially change the length of the helix-2.

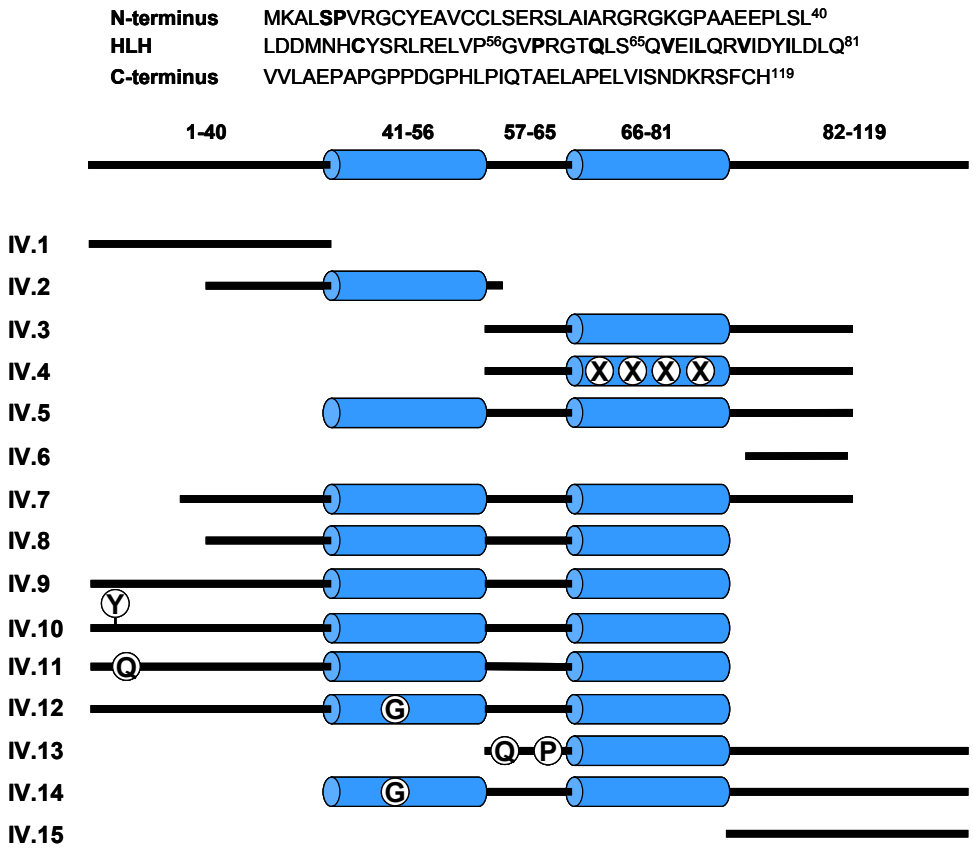


Figure IV.1. Sequence and schematic representation of the Id3 protein domains (N-terminus, HLH motif and C-terminus) and of the Id3 peptide sequences used for the conformational fragment analysis (the mutated positions are indicated in bold, X = Aib, Y = PO_3H_2).

Finally, peptide **IV.6** has been prepared to study an interesting region of the Id3 C-terminus which is particularly rich in proline residues. In proline-rich polypeptides, the constraints of the pyrrolidine ring stabilize a left-handed polyproline II (PPII) helical conformation which

plays an important role in many vital processes [12-14]. Thus far, nothing is known about the proline-rich region of Id3; with peptide **IV.6** it should be possible to evaluate the tendency of this Id3 segment towards a PPII structure.

IV.2 Solid-Phase Synthesis of the N-Terminal and C-Terminal Id3 Fragments and Related Analogues

Peptides **IV.1-12** were synthesized on Rink amide resin and cleaved as C-terminal peptide amides, whereas peptides **IV.13-15** were synthesized on Fmoc-His(trityl)-Wang resin and cleaved as free peptide acids. Peptide chain assembly was performed under standard conditions using double coupling of the N^{α}-Fmoc-protected amino acids and activation *in situ* by HBTU/HOBt/DIPEA. Fmoc-Aib-OH was inserted at positions 67, 70, 73 and 77 of the native Id3 sequence in peptide **IV.4** (in place of the hydrophobic residues Val, Leu, Val, and Ile) (Table IV.1). In peptide **IV.10**, the Fmoc-*O*-benzylphosphoserine was introduced at position 5 by standard procedure with activation *in situ* by HBTU/HOBt/DIPEA. After peptide removal from the resin, cleavage of the benzyl protecting group and reduction of the oxidized Met residues were accomplished simultaneously by treatment with a TFA mixture containing 5% (v/v) of TMSBr/EDT (2:3, v/v) for 1 hour.

Once the synthesis was completed, the N-terminus of peptides **IV.1-8** and **IV.13-14** was acetylated and the peptides were cleaved from the resin by TFA with 10% (v/v) scavengers. Peptides **IV.2** and **IV.7-12** were treated with TFA/TMSBr/EDT, as described above. The crude peptides were purified by preparative RP-HPLC and then characterized by analytical RP-HPLC (Figures IX.6-7 in the Experimental Part) and MALDI-TOF mass spectrometry.

Table IV.1. Analytical data of the synthetic Id3 fragments **IV.1**-**15**.

Synthetic Id3 fragments				
Product	**Sequence**	**$MW_{calcd.}$ (Da)**	**MW_{found} (Da)**	**t_R (min)**[a]
IV.1	Ac-(1-40)-NH_2	4189.9	4188.8	23.5
IV.2	Ac-(21-58)-NH_2	4133.8	4130.9	17.8
IV.3	Ac-(57-101)-NH_2	4882.7	4878.9	23.5
IV.4	Ac-[Aib-67,70,73,77]-(57-101)-NH_2	4798.5	4799.5	21.0
IV.5	Ac-(41-101)-NH_2	6825.9	6824.2	21.3
IV.6	Ac-(85-100)-NH_2	1633.8	1634.9	10.8
IV.7	Ac-(17-101)-NH_2	9286.8	9291.7	22.5
IV.8	Ac-(21-81)-NH_2	6813.9	6814.8	21.3
IV.9	H-(1-81)-NH_2	8969.6	8969.1	20.5
IV.10	H-[S(PO_3H_2)-5]-(1-81)-NH_2	9049.5	9071.8[b]	21.4
IV.11	H-[Q-6]-(1-81)-NH_2	9000.6	9002.3	21.3
IV.12	H-[G-47]-(1-81)-NH_2	8923.5	8925.3	20.3
IV.13	Ac-[Q-59, P-63]-(57-119)-OH	6895.0	6892.6	21.5
IV.14	Ac-[G-47]-(41-119)-OH	8792.1	8791.4	22.0
IV.15	H-(82-119)-OH	4016.6	4017.7	11.5

[a] Gradient: 10% ACN for 3 min, 10-70% ACN in 40 min for all peptides, with the exception of peptide **IV.15** (10-95% ACN in 40 min).
[b] $(M+Na)^+$.

IV.3 CD Spectroscopy

The secondary structure of the Id3 fragments and their analogues was investigated by CD spectroscopy in phosphate buffer (0.1 M, pH 7.25) or, in the case of the N-terminal Id3 peptides (**IV.9**, **IV.11** and **IV.12**), both in phosphate buffer and water, as their solubility in the buffer was limited. All peptides, with the exception of peptide **IV.15**, were measured at two different concentrations, ~ 100 μM and ~ 30 μM; however, no difference was observed.
Whereas the Id3 HLH motif was helical (peptide **III.1**), both the N-terminal (peptide **IV.1**) and C-terminal (peptide **IV.15**) domains were mostly unordered, as shown by the CD spectra characterized by an intense negative band around 199 nm (Figure IV.2).

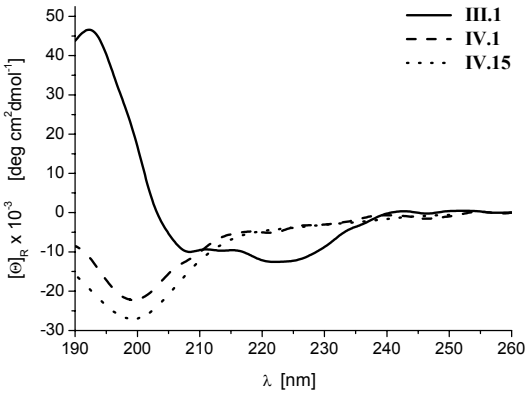


Figure IV.2. CD spectra of the Id3 HLH motif (**III.1**, 30 μM), N-terminus (**IV.1**, 30 μM) and C-terminus (**IV.15**, 43 μM) in phosphate buffer (0.1 M, pH 7.25).

Peptide **IV.2**, which contains part of the N-terminus (21-40) and helix-1, was also found to be mostly unordered (Figure IV.3, left). However, in contrast to the single helix-1 (peptide **III.8**), a moderate fraction of β-strand structure (28% vs. 7% of **III.8**) was estimated by the CONTIN algorithm. This suggests that the 20 N-terminal residues preceding helix-1 are not important for the helicity of helix-1, but rather favor a β-strand conformation (Table IV.2).

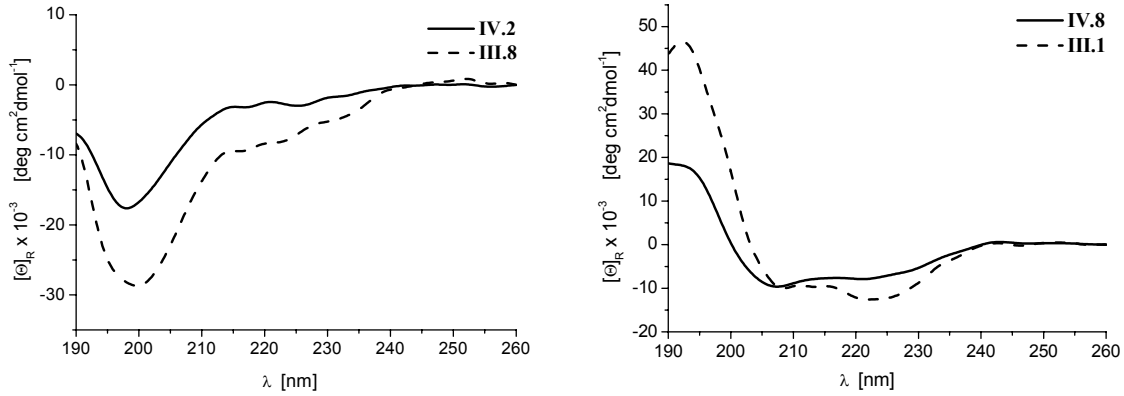


Figure IV.3. CD spectra of peptide **IV.2** and helix-1 (**III.8**) in phosphate buffer (0.1 M, pH 7.25) at 100 μM concentration (*left*). CD spectra of peptide **IV.8** and of the HLH motif (**III.1**) in phosphate buffer (0.1 M, pH 7.25) at 30 μM concentration (*right*).

The CD spectrum of peptide **IV.8**, which contains the N-terminal tail 21-40, similar to peptide **IV.2**, and the entire HLH motif, was characterized by two negative bands at 207.4 nm, attributed to the $\pi \rightarrow \pi^*$ transition, and 222 nm, attributed to the $n \rightarrow \pi^*$ transition, and by a positive

one below 200 nm (Figure IV.3, right). Comparison of the dichroic properties of this peptide and of the HLH motif alone (**III.1**) reveals a reduced helix stability for **IV.8** (decreased intensity of the bands at 222 nm and below 200 nm), probably due to a different helix packing (the ratio *R* was 0.80 for **IV.8** and 1.26 for **III.1**). In fact, the number of α-helical residues decreased from 24 for **III.1** to 18 for **IV.8** (Table IV.2). In contrast, the number of β-strand residues increased, thus leading to a α/β ratio of ~ 1 for **IV.8** against the value of ~ 5 for **III.1**. These observations have been already reported for the larger polypeptide **II.4** reproducing the Id3 residues 21-119 which contains an increased amount of β-strands with respect to the peptide **II.3** reproducing the Id3 residues 41-119. It can be therefore concluded that in all three cases studied (peptides **IV.2**, **IV.8** and **II.4**), the N-terminal tail 21-40 favored the formation of β-strand elements while disfavoring the helix conformation.

Peptides **IV.5** and **IV.7** contain the HLH motif, the proline-rich region of the C-terminus, and, in the case of **IV.7**, the N-terminal tail 17-40. Both peptides are characterized by a negative CD band around 206 nm, attributed to the $\pi \rightarrow \pi^*$ transition, followed by a less intense one centered near 222 nm, attributed to the $n \rightarrow \pi^*$ transition, and a positive band below 195 nm (Figure IV.4, left). However, the CD curve of peptide **IV.7** is less intense than that of peptide **IV.5**. The ratio *R* was 0.58 for peptide **IV.5** and 0.50 for peptide **IV.7**. Peptide **IV.5** was characterized by a much higher α/β ratio (10) than peptide **IV.7** (0.7), which suggests that the presence of the N-terminal tail in the latter peptide favored a β-strand at the expense of α-helical conformation. Accordingly, the difference spectrum between **IV.7** and **IV.5** shows a positive component with a maximum near 205 nm and a shoulder near 222 nm, and a negative band below 200 nm, a shape which has been reported for distorted β-strands [15] (Figure IV.4, right). Such β-strand-inducing effect of the N-terminus has been already observed for the peptides **IV.2**, **IV.8** and **II.4**, as described above.

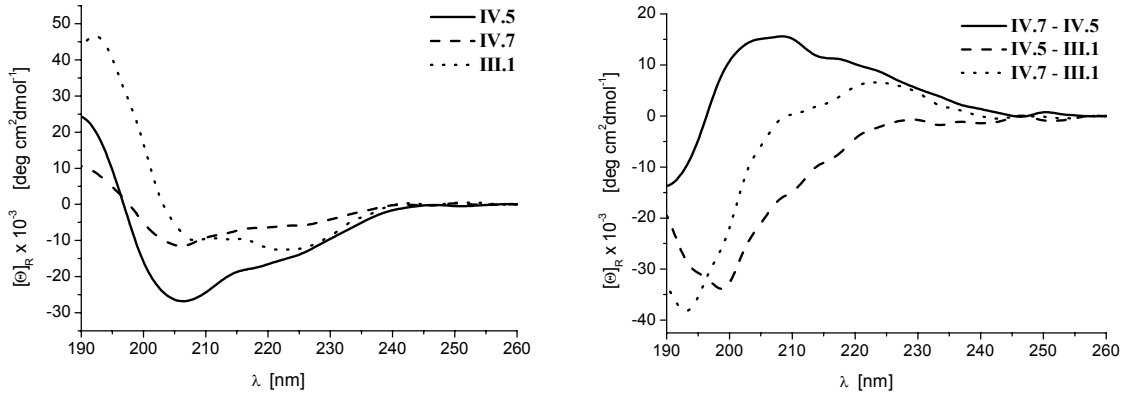


Figure IV.4. Comparison of the CD spectra of peptides **IV.5** and **IV.7** with the HLH motif **III.1** (all samples were at 30 μM concentration in 0.1 M phosphate buffer, pH 7.25) (*left*). CD spectra differences underlying the impact of the N-terminal part (residues 17-40) on **IV.5** (solid line), of the C-terminal part (residues 82-101) on **III.1** (dashed line) and of both terminal parts (dotted line) on **III.1** (*right*).

The HLH motif **III.1** was compared to peptide **IV.5** (displaying 20 additional C-terminal residues) and to peptide **IV.8** (displaying 20 additional N-terminal residues). Although the C-terminus is mostly unordered, it stabilizes and elongates helix-2 in its direction: indeed, the number of the helical residues increased from 24 to 30 upon connection the C-terminal part to the HLH motif. This result is in agreement with finding described in Chapter II, which concerned the comparison of the HLH motif with peptide **II.3** (residues 41-119): accordingly, in the presence of the entire C-terminus, the number of helical residues increased from 24 to 35. In contrast, the 20 N-terminal residues induce a transition from helical to β-strand structure, and the number of helical residues decreases from 24 to 18 upon connection the N-terminal part to the HLH motif, while the number of β-strand residues increased from 5 to 13. As already reported in Chapter II, these 20 residues could even abrogate the helicity gained by addition of the C-terminal part in peptide **II.4** which contains only 20 helical residues against the 35 ones of peptide **II.3** and the 24 ones of the HLH motif **III.1**.

Peptide **IV.3** contains the loop sequence and helix-2 with a portion of the C-terminus. Its CD spectrum is characterized by a negative band around 207 nm, attributed to the $\pi \rightarrow \pi^*$ transition, followed by a less intense one centered near 222 nm, attributed to the $n \rightarrow \pi^*$ transition. A positive band is present below 200 nm (Figure IV.5). Compared to peptide **II.2** (residues 57-119) a lower number of helical residues was found in peptide **IV.3**. The number was however larger than that of helix-2 alone. Thus, the entire C-terminus seems to stabilize the helix-2 in a more efficient manner than only the residues 82-101, which would support a

close spatial arrangement between helix-2 and the C-end. The number of helical residues significantly increased (from 17 to 30) when helix-1 was connected to the peptide **IV.3** to yield peptide **IV.5**. The same tendency was already observed in the presence of the entire C-terminus (comparison of peptides **II.2** and **II.3**, residues 57-119 and 41-119, respectively).
Peptide **IV.4**, containing four Aib residues within helix-2, is an analogue of peptide **IV.3**. Its CD spectrum shows a negative band at 201.2 nm and a second less intense negative band around 220 nm. No positive band below 200 nm was found (Figure IV.5). A dichroic parameter that is used to distinguish a 3_{10}-helix from a α-helix is the ratio *R* which is in the range of 0.3-0.4 for the 3_{10}-helix and close to 1 for the α-helix [3, 16]. In addition, the positive band near 195 nm is much less intense for a 3_{10}-helix than for a α-helix. The CD spectrum of peptide **IV.4** can be assigned to a 3_{10}-helix, with a ratio *R* of 0.26. This confirms the ability of the Aib residues to induce a 3_{10}- over a α-helix type.

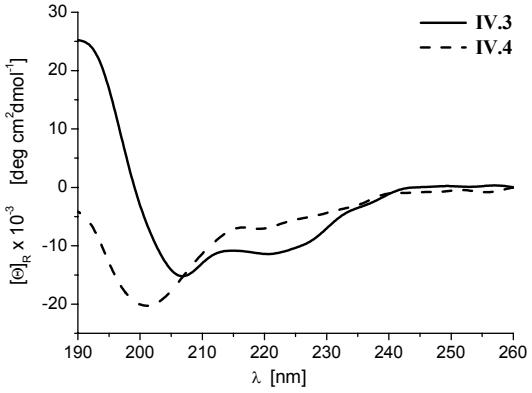


Figure IV.5. CD spectra of peptides **IV.3** and **IV.4** at 30 μM concentration in phosphate buffer (0.1 M, pH 7.25).

Peptide **IV.6** represents the proline-rich region in the Id3 C-terminus, also present in peptides **IV.3**-**5** and **IV.7**. Its CD spectrum is characterized by a single negative band centered at 199 nm (Figure IV.6). A typical CD curve for the polyproline helix II (PPII) in aqueous solution shows a weak positive band at 218-228 nm, attributed to the n→π* transition, and a strong negative band below 205 nm [14, 17, 18]. The PPII helix is the only known secondary structure which has a positive band in the wavelength range 218-228 nm. However, the absence of a positive band does not necessarily indicate the absence of PPII content [18], because other local structures which negatively contribute to the CD signal in this region might be included as well.

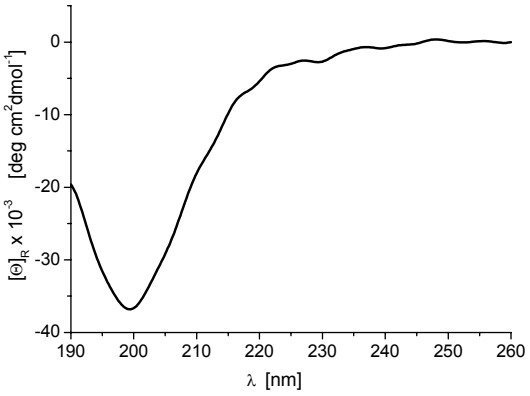


Figure IV.6. CD spectrum of peptide **IV.6** at 100 μM concentration in phosphate buffer (0.1 M, pH 7.25).

Proline residues can occur exclusively at the initial two or three positions of 3_{10}- and α-helices. Any following proline residue would act as a helix breaker, or at least to a kink of the helix backbone. PPII helices are a major component of unordered polypeptides [19, 20]. The PPII conformation is favored at low temperatures or upon addition of denaturants, such as urea or guanidinium chloride [14]. These denaturation agents interact with the backbone and force the polypeptide to extend. This can favor the formation of a PPII helix, which is indicated by an increase in the intensity of the positive band over 218-228 nm in the CD spectra. Denaturants clearly disfavor a compact native structure, but it is not yet clear whether the resulting backbone conformations are disordered. The denatured states probably possess significant amounts of locally ordered backbone structures [18]. Protein denaturation using urea and guanidinium chloride can be considered as a transition from a globally structured state to a dynamic ensemble of locally structured states. The fraction of PPII helix can be calculated from the CD band intensity of the local maximum, $[\Theta]_{max}$, using the following equation [14, 21]:

$$PPII = \frac{\Theta_{max} + 6100}{13700} \times 100\% \qquad \text{Eq. IV.1}$$

Temperature- and urea-dependent CD measurements of peptide **IV.6** have been carried out in order to see whether the PPII helix could be stabilized. The CD region of 218-228 nm has been monitored, as the intensity of the positive band in this region is considered to be proportional to the PPII helical content. For the temperature-scan experiment, peptide **IV.6** was cooled to a minimum of 5 °C and heated to a maximum of 45 °C; a positive band around

224 nm was observed at 5, 10 and 15 °C, whereas at higher temperatures, the CD signal became negative (Figure IV.7). Using the aforementioned Eq. IV.1, peptide **IV.6** was found to contain 59% PPII helix at 5 °C.

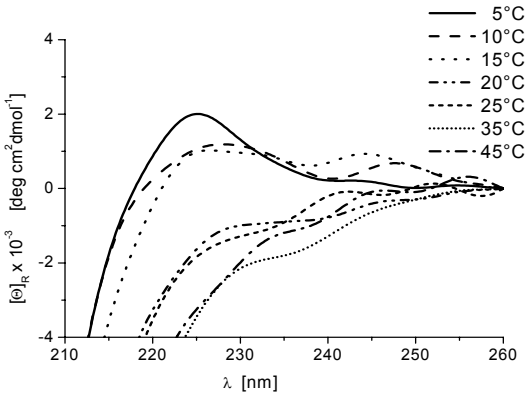


Figure IV.7. Effect of the temperature on peptide **IV.6** (100 μM) in 0.1 M phosphate buffer (pH 7.25).

Further, peptide **IV.6** was subjected to urea-titration. Increasing intensity of the positive band around 224 nm with increasing urea concentration is shown in Figure IV.8. Peptide **IV.6** was found to adopt a 63% PPII helix in the presence of 6 M urea. The ellipticity at 224 nm was plotted as a function of the urea concentration (Figure IV.8, right). The CD intensity increased in a linear manner over the concentration range of 0-3 M urea. At higher urea concentrations, the increment abruptly lowered. At low denaturant concentrations, a transition from a broad ensemble of conformations to an ensemble dominated by the PPII conformation probably occurs, whereas at higher denaturant concentrations, the ensemble is further restricted to the PPII conformation.

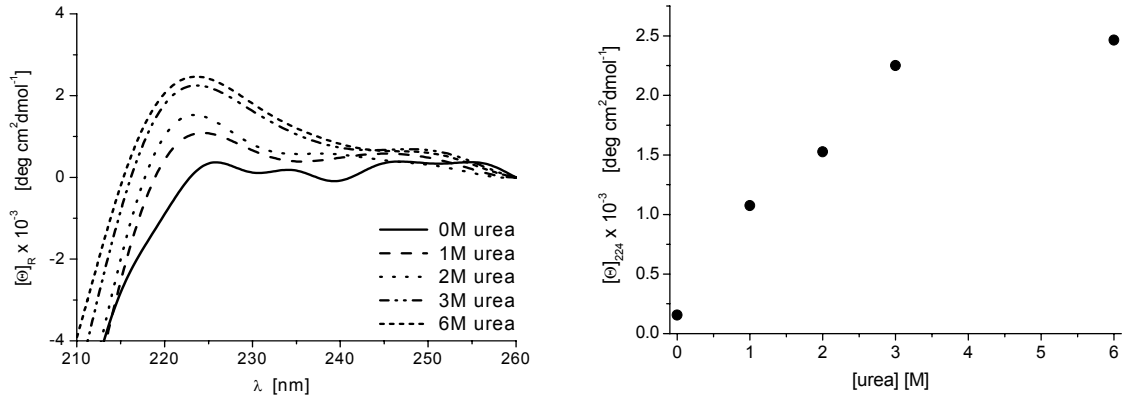


Figure IV.8. Effect of increasing urea concentration on peptide **IV.6** (100 μM) in phosphate buffer (0.1 M, pH 7.25). CD spectra (*left*) (at high concentrations of urea the CD spectra could be recorded only above 210 nm). The value of $[\Theta]_{224}$ is plotted as a function of the urea concentration (*right*).

The CD spectra of the N-terminal Id3 analogues **IV.9-12** are shown in Figure IV.9. The native sequence 1-81 (peptide **IV.9**) adopted a helical conformation both in phosphate buffer (0.1 M, pH 7.25) and in water. However, the CD intensity at 200-230 nm was higher in water, suggesting a superior helix character. Accordingly, the CONTIN analysis of the CD curves in buffer and in water indicated α/β ratio of 2 and 9, respectively (Table IV.2). The CD spectrum of the native sequence 1-81 phosphorylated at position 5 (peptide **IV.10**) was less intense than the unphosphorylated analogue, which was attributed to a decrease in helicity and an increase in β-strand structure (Table IV.2). Thus, the phosphorylation of Ser-5 seems to have a significant impact on the conformation of the Id3 sequence 1-81, which is likely to hold true for the full-length Id3 protein. In contrast, the CD spectrum of the mutant **IV.11**, in which the consensus sequence for the Cdk-2-dependent phosphorylation of Ser-5 was modified by substitution of Pro-6 by glutamine, was almost super-imposable to that of the unmodified sequence, suggesting that the mutation had no effect on the conformation. A deleterious mutation was that of Cys-47 with glycine (peptide **IV.12**), which dramatically decreased the helical content while favoring the formation of distorted β-strands under physiologic conditions (Figure IV.9). In water, the helical character could be only partially restored; accordingly, the difference CD spectrum between the Gly→Cys-47 mutant and the native sequence, which shows a positive component near 222 nm and a negative one below 200 nm (inset in Figure IV.9), indicates the presence of additional disordered elements and distorted β-strands [15]. Therefore, the sulfhydryl group of Cys-47 seems to be important for the stabilization of the conformation, probably by participating in hydrogen bonds.

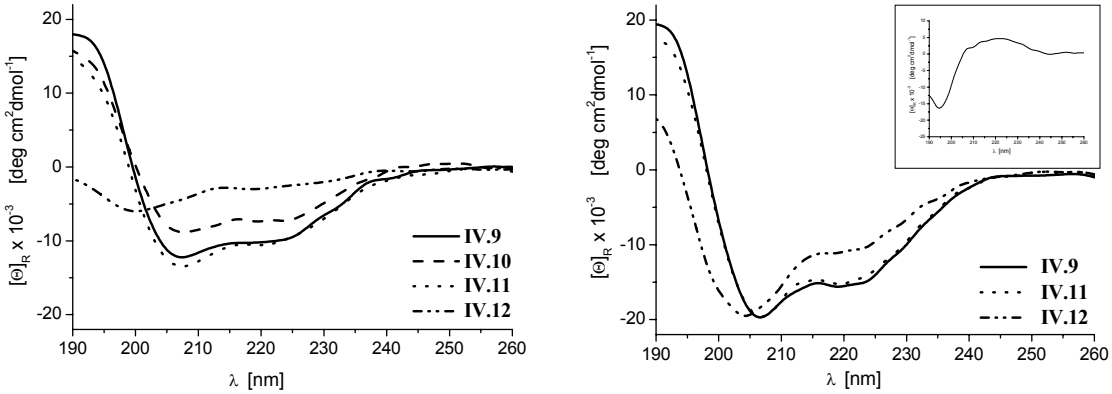


Figure IV.9. CD spectra of the N-terminal fragments **IV.9-12** at ~ 30 μM concentration in phosphate buffer (0.1 M, pH 7.25) (*left*) and in water (*right*, inset: difference CD spectrum between **IV.12** and **IV.9**).

One possible helix-stabilizing interaction might be an *i*, *i+4* hydrogen bond between the β-carboxylic group of Asp-43 and the sulfhydryl group of Cys-47. Further stabilization of the helix at pH below 7 could result from the formation of the ion pair between His-46 and Asp-42 or Asp-43. The presence of an electrostatic interaction would explain why also the native peptide **IV.9** was more helical in water than in phosphate buffer. In water, both peptides may be in fact stabilized by a salt bridge involving His-46, whereas in phosphate buffer, where the imidazole ring is mostly not ionized, only the native peptide may be stabilized by hydrogen bonds involving Cys-47.

Comparison of peptide **IV.9** with peptide **II.3** allows to evaluate the effect of the N- or C-terminal truncation of the Id3 protein (Figure IV.10). In fact, the analogue lacking the entire C-terminus was characterized by a reduced α/β ratio with respect to the analogue lacking the entire N-terminus, which suggests that the N- and C-domains of Id3 contribute to the folding in a different manner, with the N-domain favoring β-strand elements while the C-domain stabilizing helix elements.

The double mutant **IV.13** in which the loop residues Pro-59 and Gln-63 were exchanged showed a conformation similar to the corresponding native sequence (Figure IV.11, left). This suggests that the loop region is flexible enough to tolerate the shift of the proline residue from the N-end to the C-end. Moreover, the beginning of helix-2 must be located after residue 63, which confirms the results of the sequence alignment studies.

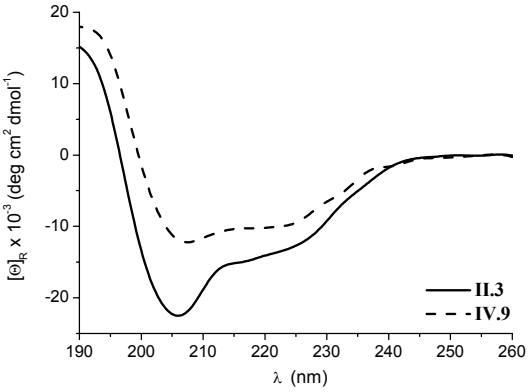


Figure IV.10. Comparison of the CD spectra of the Id3 polypeptides **II.3** (residues 41-119, 30 μM) and **IV.9** (residues 1-81, 23 μM) lacking the N-terminus and C-terminus, respectively.

Surprisingly, the mutation of Cys-47 in helix-1, which was found to be deleterious for the N-terminal peptide **IV.12**, was well tolerated in the C-terminal peptide **IV.14**: indeed, the CD spectra of the mutant **IV.14** and of the unmodified sequence **II.3** were almost superimposable (Figure IV.11, right). This different behavior might be explained by the fact that the local environment of the residue at position 47 is not the same in the Id3 sequences 1-81 and 41-119.

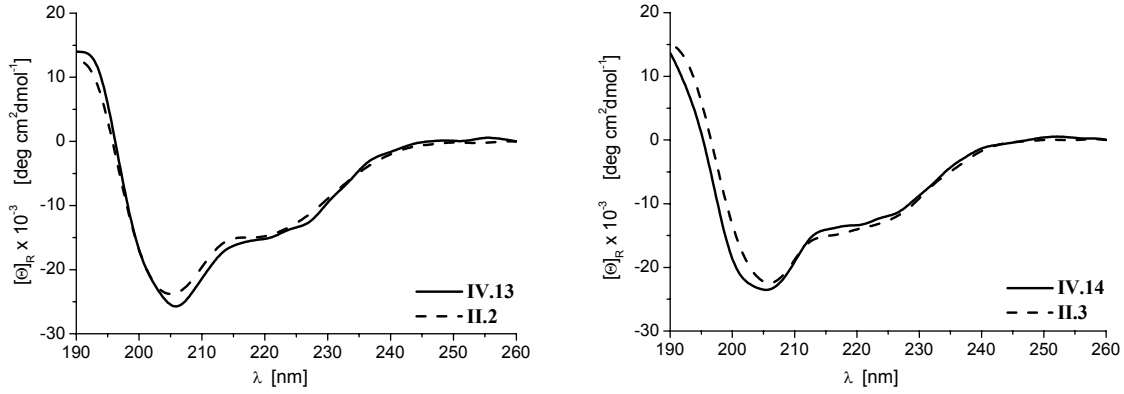


Figure IV.11. CD spectrum of the C-terminal double mutant **IV.13** in comparison with the native sequence **II.2** in phosphate buffer (0.1 M, pH 7.25) (*left*). CD spectrum of the C-terminal Cys-47 mutant **IV.14** in comparison with the native sequence **II.3** in phosphate buffer (0.1 M, pH 7.25) (*right*). The peptide concentration was 30 μM.

Table IV.2. Conformational properties of the N- and C-terminally truncated Id3 sequences and their analogues in phosphate buffer (0.1 M, pH 7.25) and in water.

Product	No. of residues	Concen-tration (μM)	% Helix (no. of residues)[a]	% Secondary structure elements by CONTIN[b] (no. of residues)			
				helix	β-strand	turns	unordered
IV.1	40	30	15 (6)	14 (6)	18 (7)	27 (11)	40 (16)
IV.2	38	100	5 (2)	8 (3)	28 (11)	23 (9)	41 (16)
IV.3	45	30	31 (14)	38 (17)	15 (7)	18 (8)	29 (13)
IV.4	45	30	13 (6)	16 (7)	24 (11)	24 (11)	36 (16)
IV.5	61	30	43 (26)	49 (30)	5 (3)	18 (11)	29 (18)
IV.6	16	100	13 (2)	10 (2)	5 (1)	32 (5)	53 (8)
IV.7	85	30	15 (13)	18 (15)	27 (23)	21 (18)	34 (29)
IV.8	61	30	21 (13)	29 (18)	21 (13)	19 (11)	31 (19)
IV.9	81	23	27 (22)	34 (28)	15 (12)	22 (18)	29 (23)
IV.9[c]	81	30	42 (34)	45 (36)	5 (4)	22 (18)	28 (23)
IV.10	81	35	11 (9)	21 (17)	28 (23)	18 (15)	33 (27)
IV.11	81	23	28 (23)	34 (28)	14 (11)	22 (18)	30 (24)
IV.11[c]	81	30	41 (33)	44 (36)	6 (5)	21 (17)	29 (23)
IV.12	81	32	7 (6)	8 (6)	35 (28)	22 (18)	35 (28)
IV.12[c]	81	30	30 (24)	24 (19)	20 (16)	19 (15)	37 (30)
IV.13	63	30	43 (27)	49 (31)	4 (3)	21 (13)	26 (16)
IV.14	79	30	35 (28)	41 (32)	4 (3)	23 (18)	32 (25)
IV.15	38	43	13 (5)	8 (3)	10 (4)	32 (12)	50 (19)
III.1	41	30	37 (15)	59 (24)	12 (5)	18 (7)	11 (5)
III.8	17	100	29 (5)	9 (2)	7 (1)	30 (5)	54 (9)

[a] The helix percentage was estimated by the ellipticity value at 222 nm using the procedure of Chen and Yang [22].

[b] The CD spectra were analyzed by the CONTIN algorithm.

[c] The peptide was measured in water.

IV.4 Discussion and Conclusion

The conformational Id3 fragment analysis has revealed some interesting properties of the N- and C-terminal domains: (*i*) the presence of the N-terminal region seems to reduce the α/β ratio (peptides **IV.2**, **IV.7** and **IV.8**), as does the removal of the C-terminal tail (peptide **IV.9**). (*ii*) In contrast, the C-terminus or part of it stabilize the helix character (e.g., peptide **IV.5**), even in the case of helix-1 destabilization by Cys-47 mutation (peptide **IV.14**). (*iii*) Phosphorylation of the N-terminus at position 5 enhanced the β-strand propensity of the N-terminal fragment 1-81 (peptide **IV.10**). (*iv*) The C-terminal proline-rich motif is capable to adopt a PPII helix stabilized by low temperature or urea addition.

Altogether, these results suggest that the N- and C-domains of the Id3 protein can be critical for the folding of the central HLH motif and for its dimerization behavior, the intrinsic flexibility of the N- and C-tails is probably a structural prerequisite for the modulation of the protein fold during the dimerization processes with different bHLH transcription factors.

IV.5 Literature

[1] Kiewitz, S. D., and Cabrele, C. (2005) Synthesis and conformational properties of protein fragments based on the Id family of DNA-binding and cell-differentiation inhibitors. *Biopolymers 80*, 762-74.

[2] Crisma, M., Formaggio, F., Moretto, A., and Toniolo, C. (2006) Peptide helices based on alpha-amino acids. *Biopolymers 84*, 3-12.

[3] Toniolo, C., Crisma, M., Formaggio, F., Peggion, C., Broxterman, Q. B., and Kaptein, B. (2004) Molecular spacers for physicochemical investigations based on novel helical and extended peptide structures. *Biopolymers 76*, 162-76.

[4] Toniolo, C., Formaggio, F., Tognon, S., Broxterman, Q. B., Kaptein, B., Huang, R., Setnicka, V., Keiderling, T. A., McColl, I. H., Hecht, L., and Barron, L. D. (2004) The complete chirospectroscopic signature of the peptide 3(10)-helix in aqueous solution. *Biopolymers 75*, 32-45.

[5] Karle, I. L. (2001) Controls exerted by the Aib residue: helix formation and helix reversal. *Biopolymers 60*, 351-65.

[6] Biron, Z., Khare, S., Samson, A. O., Hayek, Y., Naider, F., and Anglister, J. (2002) A monomeric 3(10)-helix is formed in water by a 13-residue peptide representing the neutralizing determinant of HIV-1 on gp41. *Biochemistry 41*, 12687-96.

[7] Pal, L., and Basu, G. (1999) Novel protein structural motifs containing two-turn and longer 3(10)-helices. *Protein Eng 12*, 811-4.

[8] Deed, R. W., Hara, E., Atherton, G. T., Peters, G., and Norton, J. D. (1997) Regulation of Id3 cell cycle function by Cdk-2-dependent phosphorylation. *Mol Cell Biol 17*, 6815-21.

[9] Forrest, S. T., Taylor, A. M., Sarembock, I. J., Perlegas, D., and McNamara, C. A. (2004) Phosphorylation regulates Id3 function in vascular smooth muscle cells. *Circ Res 95*, 557-9.

[10] Yokota, Y., and Mori, S. (2002) Role of Id family proteins in growth control. *J Cell Physiol 190*, 21-8.

[11] Norton, J. D., and Atherton, G. T. (1998) Coupling of cell growth control and apoptosis functions of Id proteins. *Mol Cell Biol 18*, 2371-81.

[12] Pujals, S., Fernandez-Carneado, J., Kogan, M. J., Martinez, J., Cavelier, F., and Giralt, E. (2006) Replacement of a proline with silaproline causes a 20-fold increase in the cellular uptake of a Pro-rich peptide. *J Am Chem Soc 128*, 8479-83.

[13] Fillon, Y. A., Anderson, J. P., and Chmielewski, J. (2005) Cell penetrating agents based on a polyproline helix scaffold. *J Am Chem Soc 127*, 11798-803.

[14] Kelly, M. A., Chellgren, B. W., Rucker, A. L., Troutman, J. M., Fried, M. G., Miller, A. F., and Creamer, T. P. (2001) Host-guest study of left-handed polyproline II helix formation. *Biochemistry 40*, 14376-83.

[15] Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (1999) Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci 8*, 370-80.

[16] Schievano, E., Pagano, K., Mammi, S., and Peggion, E. (2005) Conformational studies of Aib-rich peptides containing lactam-bridged side chains: evidence of 3(10)-helix formation. *Biopolymers 80*, 294-302.

[17] Prabhu, S., Ignatova, A., Park, S. T., and Sun, X. H. (1997) Regulation of the expression of cyclin-dependent kinase inhibitor p21 by E2A and Id proteins. *Mol Cell Biol 17*, 5888-96.

[18] Whittington, S. J., Chellgren, B. W., Hermann, V. M., and Creamer, T. P. (2005) Urea promotes polyproline II helix formation: implications for protein denatured states. *Biochemistry 44*, 6269-75.

[19] Sreerama, N., and Woody, R. W. (2004) Computation and analysis of protein circular dichroism spectra. *Methods Enzymol 383*, 318-51.

[20] Sreerama, N., and Woody, R. W. (1999) Molecular dynamics simulations of polypeptide conformations in water: A comparison of alpha, beta, and poly(pro)II conformations. *Proteins 36*, 400-6.

[21] Danielsson, J., Jarvet, J., Damberg, P., and Graslund, A. (2005) The Alzheimer beta-peptide shows temperature-dependent transitions between left-handed 3-helix, beta-strand and random coil secondary structures. *Febs J 272*, 3938-49.

[22] Chen, Y. H., Yang, J. T., and Chau, K. H. (1974) Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry 13*, 3350-9.

V. Fluorescence Resonance Energy Transfer (FRET) Studies of Labeled Id3 Fragments

V.1 Introduction

The conformational studies described in previous chapters have shown that the Id3 HLH peptide has high intrinsic helix propensity, whereas the N- and C-flanking regions do not adopt any preferred conformation as isolated domains. However, the N- [1, 2] and C-terminal domains [3, 4] of the Id3 protein are important for the control of the protein dimerization events mediated by the HLH domain, thus regulating the Id3 inhibitory activity. Accordingly, the conformation of larger Id3 fragments suggests that the regions preceding and following the HLH sequence are not completely flexible or extended, but are rather part of a compact fold. For example, the Id3 fragment 41-119 reproducing the HLH domain connected to the C-terminus displayed superior helical character with respect to the isolated HLH domain (~ 35 vs. ~ 24 helical residues, as estimated by the analysis of the CD spectra with the CONTIN algorithm), but comparable to that of the synthetic full-length protein (~ 40 helical residues) [5]. Thus, the C-terminal flanking region seems to exert a stabilizing effect on the Id3 HLH helicity. In contrast, the N-terminus seems to disfavor the helicity: an Id3 analogue lacking the first 20 N-terminal residues was found to form a significant amount of β-strands at the expense of α-helix, which was assumed to be caused by a probable folding back of the shortened N-terminus onto the HLH domain [5]. Large Id3 protein fragments fluorescently labeled by tryptophan/dansyl (Trp/Dns) donor/acceptor pair [6] at different positions along the sequence were prepared to further investigate how the Id3 HLH motif and its flanking regions are spatially arranged in solution. The efficiency of the FRET which is according to the Förster's theory [7] related to the distance between the two fluorophores was examined. The effect of the FRET pair introduction on the conformation of the native sequence was also investigated by CD spectroscopy. Altogether, the fluorescence and CD data confirm that the regions in proximity of the N- and C-termini of the HLH domain are not completely flexible and extended, but rather point toward the HLH core sequence, thus generating a compact fold. Interestingly, deletion of the C-terminus was found to increase the β-strand structure amount while reducing the helix content, as it was previously observed in the absence of the N-terminal residues 1-20 [5]. Even though the regions preceding and following the Id3 HLH domain have no secondary structure propensity as isolated fragments, they seem to influence the folding, in particular the α/β ratio, of the Id3 protein.

V.2 Design and Solid-Phase Synthesis of the Fluorescently Labeled Id3 Protein Fragments

A FRET active probe containing a donor/acceptor pair is characterized by the decrease in emission intensity of the donor in the presence of the acceptor, and can be quantified on the basis of the emission intensities of the donor in the presence and in the absence of the acceptor. The Id3 peptides containing the donor/acceptor pair Trp/Dns used in this study are summarized in Figure V.1. For each FRET peptide, two more analogues were synthesized, the first containing the donor (Trp) but not the acceptor (Dns) to measure the fluorescence of the donor in the absence of the acceptor, and the other reproducing the native Id3 sequence to examine the effect of the donor/acceptor pair on the conformation (Table V.1). The HLH domain (residues 41-81) adopts a stable helical conformation which should include about 24 residues, as estimated from the CD spectrum. We have however shown that the N-terminal helix-1 and the C-terminal helix-2 have different intrinsic helix propensity, the former being unordered and the latter helical in solution (Chapter III). Thus, it is probable that helix-1 becomes stable only upon interhelical contacts with helix-2, which leads to the HLH fold. To investigate such helix fold, the HLH peptide was labeled with a FRET pair (peptide **V.1**). We chose the Trp/Dns pair because of the relatively small size of the two fluorophores and the facile selective introduction during the solid-phase peptide synthesis. Indeed, if not naturally occurring, Trp can be added as a terminal residue or inserted in place of another naturally occurring hydrophobic amino acid (e.g. Tyr, Phe, Leu, Ile, Val), whereas the Dns group can be conjugated to the free N-terminus as well as to the side chain of a naturally occurring or properly introduced lysine residue. In our case, it was synthetically convenient to introduce Trp at the C-terminal position 82, which is naturally occupied by valine, whereas the Dns group was attached to the α-amino group of the N-terminal Gly-40, which was introduced as a short spacer between the fluorophore and the N-end of helix-1. Thus, while the native HLH sequence was not modified, the N- and C-ends became sterically more demanding and more hydrophobic than in the corresponding unlabeled peptide.

To further investigate the HLH folding in the presence of the regions proximal to its N- and C-termini, we designed peptides **V.2** and **V.3**, which contain the HLH motif, a 20-residue long N-tail or a 24-residue long N-tail together with a 20-residue long C-tail, respectively. In both peptides, Trp replaced the natural Tyr-76. The Dns group in peptide **V.2** was introduced as Lys(N^{ε}-Dns) in place of the natural Ile-24. Lys-24 was first introduced as Lys(N^{ε}-Mtt), which allows the selective on-resin deprotection of the side chain by treatment with 1% TFA in dichloromethane (DCM) in the presence of 1% TIS. Subsequent treatment of

the peptidyl-resin with Dns-Cl yielded the selectively labeled product. In case of peptide **V.3**, the Dns group was directly attached to the N-terminal residue Leu-17. The observation of FRET activity of these two analogues would suggest a close spatial arrangement between helix-2 and the N-terminal flanking region.

Instead, peptides **V.4-6** were designed to examine the HLH folding in the presence of the C-terminal domain. Similar to in peptide **V.1**, the Dns group was introduced at the N-terminal Gly-40, whereas the position of Trp varied: in peptide **V.4** it was introduced as an additional C-terminal residue, whereas in peptides **V.5** and **V.6** it replaced Leu-84 and Tyr-76, respectively.

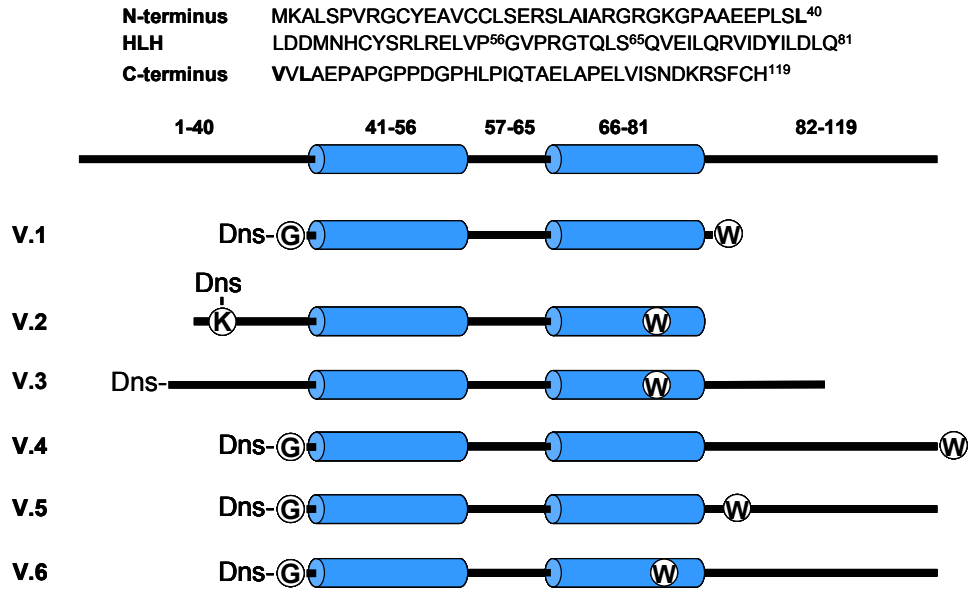


Figure V.1. Sequence and schematic representation of the Id3 protein domains (N-terminus, HLH motif and C-terminus) and of the fluorescently labeled fragments used in this study (the mutated positions are indicated in bold).

Peptide chain assembly was performed under standard Fmoc solid-phase conditions, including an initial loading of 0.6 mmol/g for Fmoc-His(trityl)-Wang resin (peptides **V.5** and **V.6**) as well as for Rink amide MBHA resin (peptides **V.1-4**), double coupling of the N^{α}-Fmoc-protected amino acids and activation *in situ* by HBTU/HOBt/DIPEA. The introduction of the Dns group was accomplished using its sulfonyl chloride derivative in the presence of DIPEA. After the synthesis was completed, cleavage from the solid support and removal of the side chain protecting groups were accomplished simultaneously by treatment with TFA in the presence of 10% scavengers. Crude products were purified by preparative RP-HPLC and characterized by analytical RP-HPLC and MALDI-TOF mass spectrometry (Table V.1).

Table V.1. Analytical data of the fluorescently labeled Id3 polypeptides **V.1**-**6**.

Fluorescently labeled Id3 polypeptides				
Product	**Sequence**	**$MW_{calcd.}$ (Da)**	**MW_{found} (Da)**	**t_R (min)**[a]
V.1	Dns-[G-40, W-82]-(40-82)-NH_2	5273.2	5270.4	24.1
V.1a	Ac-[G-40, W-82]-(40-82)-NH_2	5081.9	5085.3	22.6
V.1b	Dns-[G-40]-(40-81)-NH_2	5086.9	5089.3	23.4
V.2	Ac-[K(Dns)-24, W-76]-(21-81)-NH_2	7062.2	7098.5[b]	23.6
V.2a	Ac-[K-24, W-76]-(21-81)-NH_2	6828.9	6831.3	21.8
V.3	Dns-[W-76]-(17-101)-NH_2	9501.1	9502.6	22.2
V.3a	Ac-[W-76]-(17-101)-NH_2	9309.8	9307.1	21.5
V.4	Dns-[G-40, W-120]-(40-119)-NH_2	9271.8	9273.5	23.1
V.4a	Ac-[G-40, W-120]-(40-119)-NH_2	9080.5	9080.4	23.0
V.5	Dns-[G-40, W-84]-(40-119)-OH	9159.6	9159.5	22.9
V.5a	Ac-[G-40, W-84]-(40-119)-OH	8968.3	8966.1	21.7
V.6	Dns-[G-40, W-76]-(40-119)-OH	9109.6	9112.5	24.5
V.6a	Ac-[G-40, W-76]-(40-119)-OH	8918.3	8919.3	23.7

[a] Gradient: 10% ACN for 3 min, 10-70% ACN in 40 min.
[b] $(M+K)^+$.

V.3 CD Spectroscopy

Secondary structure investigation of the synthesized Id3 peptides was performed by CD spectroscopy in phosphate buffer (0.1 M, pH 7.25). The impact of the fluorescence labeling on the conformation of the native sequence was examined by comparing the CD spectra of the mono-labeled (only Trp), double-labeled (both Dns and Trp) and unlabeled peptides. Athough flexibility of the N- and C-termini of the HLH peptide **III.1** was assumed, their labeling to yield peptides **V.1**, **V.1a** and **V.1b** unexpectedly affected their properties. The double-labeled peptide **V.1** was poorly soluble in phosphate buffer and required the addition of at least 15% TFE to be dissolved. Despite the presence of TFE, peptide **V.1** gave a α-helix-like CD spectrum with very weak intensity in comparison to that of the native sequence **III.1**. Even the mono-labeled analogues **V.1a** and **V.1b** showed CD spectra with reduced intensity with respect to the unlabeled peptide: especially the N-terminal labeling was found to significantly affect the helix stability, leading to a conformational transition from helix to β-strands and turns, whereas the presence of the aromatic residue at the C-end caused a more moderate helix destabilization (Figure V.2). The fact that the HLH peptide is sensitive to the modification of

its N- and C-ends supports the presence of a compact fold which includes also the peripheral residues. Indeed, it should be pointed out that the Id HLH motif exists in solution mostly as an oligomer (dimer), similarly to all bHLH domains [8, 9], in which the helix bundle requires an ideal juxtaposition of the helices. Taking this into account, the behavior of the N- and/or C-terminally labeled Id3 HLH domain is likely to reflect an unfavorable helix packing as a consequence of the increased steric hindrance at the periphery, leading to a misfolded state.

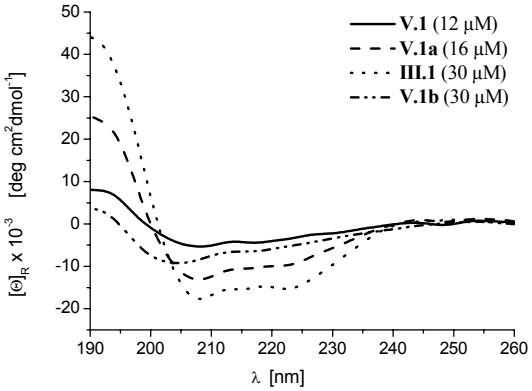


Figure V.2. CD spectra of the unlabeled (**III.1**) and fluorescence-labeled (**V.1**, **V.1a** and **V.1b**) Id3 HLH motif in phosphate buffer/TFE 85:15 (v/v).

The conformation of the fluorescence-labeled analogues of peptide **IV.8** (residues 21-81) and peptide **IV.7** (residues 17-101) was examined. Peptides **V.2** and **V.2a** gave CD curves similar to that of the unlabeled fragment 21-81, but were more intense and a slight blue-shift of the minimum close to 205 nm was observed, which was accompanied by a decrease in the intensity of the maximum (Figure V.3, left). This might be due to a negative effect of the substitution of Ile-24 with lysine in **V.2a** and lysine(Dns) in **V.2**, and/or to the substitution of Tyr-76 with Trp. However, despite the small differences in the dichroic data, the composition of secondary structure elements estimated with the CONTIN algorithm did not dramatically change within the peptides **IV.8**, **V.2** and **V.2a** (Table V.2).

In contrast, the mono- and double-labeled analogues of peptide **IV.7** (residues 17-101) were both found to be much more helical than the unlabeled form (Figure V.3, right). The helix fraction increased at the expense of the β-strand fraction, which must be attributed exclusively to the substitution of Tyr-76 with Trp in helix-2, as the CD curves of the mono- and double-labeled peptides were comparable in shape and intensity. It is interesting to note that Trp-76 does not have such a significant influence on the conformation of the peptide **V.2a**, which suggests that the local environment around position 76 is different in peptides **IV.7** and **IV.8**.

However, more striking is that the presence of the hydrophobic Trp residue not only increased the helicity of the Id3 fragment 17-101 (**IV.7**), but also allowed recovery of the helicity lost upon N-elongation of the sequence 41-101 (**IV.5**) to residue 17. Moreover, peptide **V.3a** tolerated the N-terminal dansylation, which just slightly enhanced the β-strand content (peptide **V.3**). Therefore, these last two peptides can be reasonably considered models of the Id3 HLH fold, whereas peptides **V.2** and **V.2a** are rather models for a misfolded state.

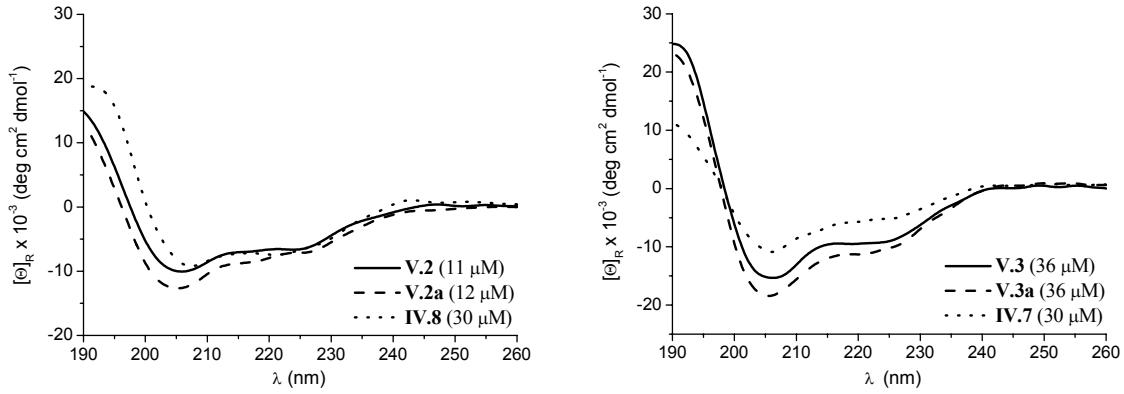


Figure V.3. CD spectra of the unlabeled and fluorescence-labeled forms of the Id3 fragments 21-81 (**IV.8**, *left*) and 17-101 (**IV.7**, *right*) in phosphate buffer (0.1 M, pH 7.25).

The conformation of the Id3 protein lacking the complete N-terminal domain (peptide **II.3**) has been described in Chapter II [5]. Here, we present fluorescence-labeled analogues of peptide **II.3**, which have been prepared to investigate the spatial arrangement of the long C-terminus with respect to the HLH motif. All peptides bear a Dns moiety at the N-end, whereas the position of the Trp residue has been moved along the sequence: peptides **V.6** and **V.6a** contain the donor in place of Tyr-76, peptides **V.5** and **V.5a** in place of Leu-84, and peptides **V.4** and **V.4a** at the C-end as an additional residue. Surprisingly, the presence of Trp at the C-end induced a significant conformational transition at the expense of the helix structure, as indicated by the appearance of a broad negative CD band centered near 203 nm which characterized both the mono-labeled and double-labeled analogues **V.4a** and **V.4** (Figure V.4, top-left). This behavior suggests that the C-terminal tail is not completely flexible or extended, but probably folds into a compact structure. It should be mentioned that with the introduction of the C-terminal Trp-amide, not only the ionizable carboxy-terminus was lost, but a bulky hydrophobic side chain was also present in a region otherwise rich in polar, acidic and basic side chains (S^{111}NDKRSFCH^{119}). Consequently, decreasing the polar character of the C-terminal tail might have compromised its stabilizing role. The C-terminal

residues 102-119 play a role in the folding as previously shown by comparing peptide **IV.3** (residues 57-101) with peptide **II.2** (residues 57-119), and the latter was found to be much more helical than the former.

When Trp was introduced in place of Leu-84, the double-labeled peptide **V.5** exhibited a CD spectrum similar to that of the unlabeled peptide **II.3**, whereas the curve of the mono-labeled peptide **V.5a** was more intensive (Figure V.4, top-right). These observations are interesting, because they suggest that a fold with superior helicity as the one induced by the bulky hydrophobic Trp side chain at position 84 is not compatible with the presence of an as much bulky hydrophobic Dns group at the N-end. One explanation might be that Trp-84 allowed the helix-2 to be C-terminally prolonged, which led both helix-1 and helix-2 to rearrange in a way that was incompatible with the presence of bulky N-terminal groups.

Unlike the Trp-84 containing peptides **V.5** and **V.5a**, the corresponding Trp-76 containing analogues **V.6** and **V.6a** were comparable with the unlabeled sequence **II.3** (Figure V.4, bottom), thus providing a valuable fluorescent model for the Id3 HLH fold in the presence of the C-terminal domain.

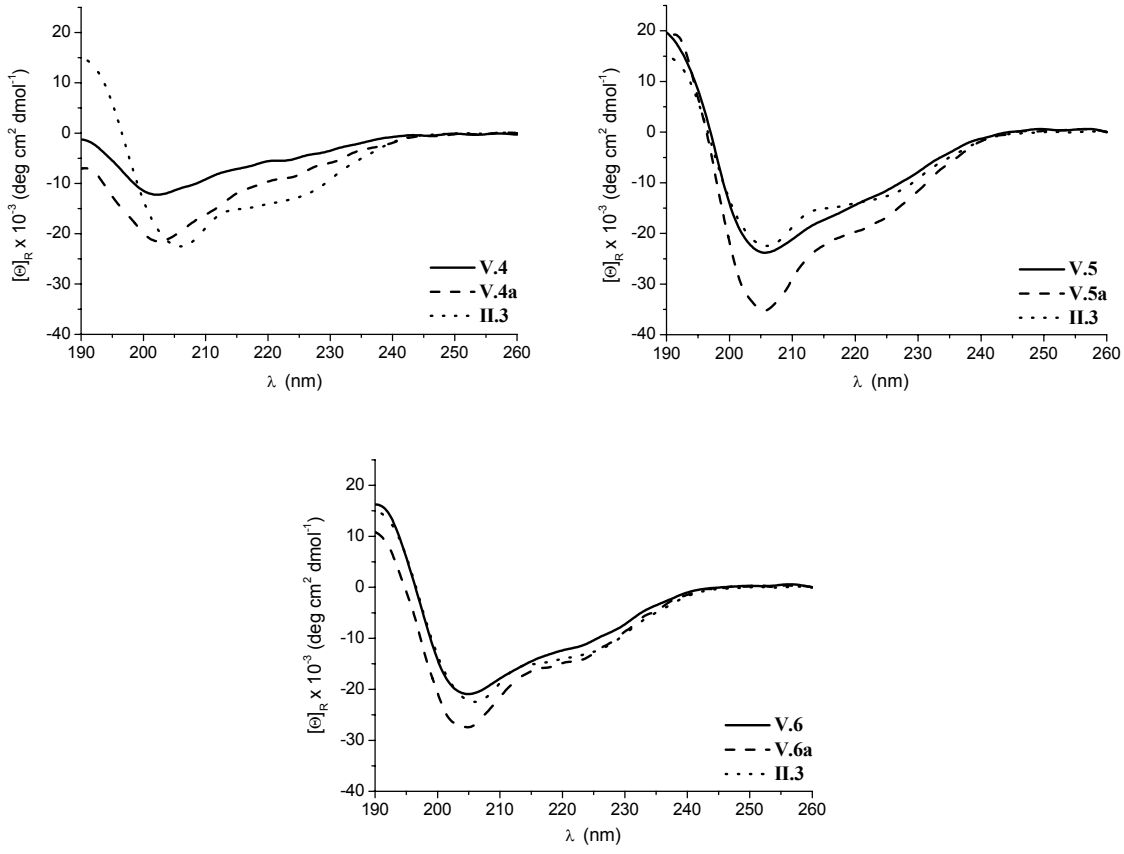


Figure V.4. CD spectra of the N-terminally truncated Id3 protein **II.3** and of its fluorescence-labeled analogues at ~ 30 μM concentration in phosphate buffer (0.1 M, pH 7.25).

Table V.2. Conformational properties of the fluorescence-labeled Id3 polypeptides **V.1-6** and comparison with the corresponding unlabeled polypeptides **III.1**, **IV.7-8** and **II.3**.

Product	No. of residues	Concentration (μM)	% Helix (no. of residues)[a]	% Secondary structure elements by CONTIN[b] (no. of residues)			
				helix	β-strand	turns	unordered
V.1	43	12	9 (4)	11 (5)	35 (15)	21 (9)	33 (14)
V.1a	43	16	28 (12)	35 (15)	18 (8)	17 (7)	30 (13)
III.1	41	30	44 (18)	60 (25)	6 (2)	11 (5)	23 (9)
V.1b	42	30	14 (16)	16 (7)	28 (12)	25 (10)	30 (13)
V.2	61	11	18 (11)	21 (13)	25 (15)	22 (13)	32 (20)
V.2a	61	12	23 (14)	26 (16)	20 (12)	22 (13)	32 (20)
IV.8	61	30	21 (13)	29 (18)	21 (13)	19 (11)	31 (19)
V.3	85	36	22 (19)	36 (31)	16 (13)	15 (13)	33 (28)
V.3a	85	36	31 (26)	39 (33)	10 (9)	19 (16)	32 (27)
IV.7	85	30	15 (13)	18 (15)	27 (23)	21 (18)	34 (29)
V.4	81	26	15 (12)	14 (11)	24 (19)	24 (19)	38 (31)
V.4a	81	31	28 (23)	12 (10)	4 (3)	30 (24)	54 (44)
II.3	79	30	38 (30)	44 (35)	5 (4)	21 (17)	30 (24)
V.5	80	30	35 (28)	38 (31)	9 (7)	20 (16)	33 (26)
V.5a	80	30	50 (40)	59 (47)	3 (2)	16 (13)	22 (18)
V.6	80	30	31 (25)	35 (28)	11 (9)	22 (17)	32 (26)
V.6a	80	30	41 (33)	41 (33)	5 (4)	23 (18)	31 (25)

[a] The helix percentage was estimated by the ellipticity value at 222 nm using the procedure of Chen and Yang [10].
[b] The CD spectra were analyzed by the CONTIN algorithm.

V.4 Fluorescence Spectroscopy

The labeled peptides were studied by fluorescence spectroscopy in phosphate buffer (0.1 M, pH 7.25) at 5 μM concentration. The emission spectra were collected upon excitation of the Trp side chain at 295 nm at 25 °C. The emission peak of Trp was centered at 350 nm, whereas that of the Dns moiety was observed around 500 nm when FRET occurred.

The fluorescence intensities of Trp in the mono-labeled analogue (in the absence of the Dns group), F_D, and in the corresponding double-labeled analogue (in the presence of the Dns group), F_{DA}, were used to calculate the FRET efficiency E, applying the following formula:

$$E = \left(1 - \frac{F_{DA}}{F_D}\right) \times 100\% \qquad \text{Eq. V.1}$$

Obviously, the E values are reliable only if the difference in the Trp emission observed between the mono- and double-labeled analogues exclusively depends on the proximity of the acceptor and not on other factors, such as changes in the local environment of the Trp side chain. To gain further insight in the conformation of the Id3 fragments, it is necessary that the introduction of the fluorescent labels is not accompanied by dramatic structural changes. Based on the CD data discussed above, this did not unfortunately hold true for all fluorescence-labeled peptides synthesized: for example, the chemical and structural properties of the mono- and, particularly, of the double-labeled analogues of the HLH fold (**V.1**, **V.1a** and **V.1b**) were completely different from those of the unlabeled form (**III.1**), which makes them useless as Id3 HLH models. Moreover, the mono- and double-labeled peptides were not structurally similar, which prevents the estimation of the FRET efficiency observed in the double-labeled peptide **V.1** (Figure V.5, left). Indeed, the reduced emission of Trp in the presence of Dns (indicated by the red arrow) might reflect either FRET or a different local environment of Trp, as the composition of secondary structure elements of the mono- and double-labeled sequences vary significantly. However, the double-labeled peptide is obviously FRET active, as the emission of Dns could be clearly detected (indicated by the green arrow). Interestingly, an equimolar mixture of the two peptides labeled only with the donor (**V.1a**) or acceptor (**V.1b**) did not show any intermolecular FRET (Figure V.5, right); this suggests not only that the FRET observed for peptide **V.1** occurred intramolecularly, but also that the mono-labeled species were not able to dimerize, thus supporting the idea that the presence of sterically demanding N- and C-ends is incompatible with the formation of a stable HLH dimer.

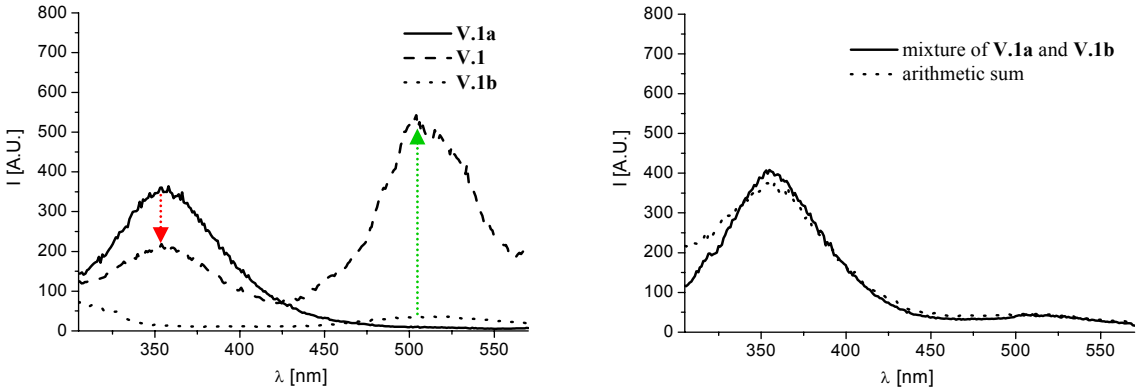


Figure V.5. Fluorescence spectra of the peptides **V.1a**, **V.1b** and **V.1** (each 5 μM, *left*) and of the equimolar mixture of **V.1a** and **V.1b** (each 5 μM) compared to the arithmetic sum of the two single spectra (*right*). Spectra were recorded in phosphate buffer/TFE 85:15 with an excitation wavelength of 295 nm at 25 °C.

As discussed above, peptide **IV.8** can be considered a misfolded state of the Id3 HLH motif, in which the number of α-helical and β-strand residues was comparable. Analysis of the corresponding fluorescent analogues **V.2a** and **V.2** revealed no FRET between Trp-76 and Lys(Dns)-24 (Figure V.6, left). However, the presence of a non-negligible Dns emission might reflect an energy transfer involving Tyr-48. These data suggest that the N-terminal tail and helix-2 are far away from each other in peptide **V.2**, as a consequence of the loss of the HLH fold.

In contrast to the unlabeled Id3 fragment 17-101 (peptide **IV.7**) which displayed a misfolded state rich in β-strands, the corresponding mono- and double-labeled peptides **V.3a** and **V.3** were characterized by a number of helical residues similar to that found for the Id3 fragments 41-101 (**IV.5**), 41-119 (**II.3**) and even the full-length Id3 protein, this indicates that the HLH fold is present again. Therefore, they can be used as fluorescent tools for the study of a "miniaturized" Id3 protein shortened at its N- and C-tails. Emission of Trp in the double-labeled peptide **V.3** turned out to be much lower than in the mono-labeled peptide **V.3a**, and a pronounced emission band was detected in the Dns region (Figure V.6, right). This indicates that peptide **V.3** is FRET active, with an efficiency of 49% which would correspond to a donor-acceptor distance R_{DA} of ~ 25 Å, as calculated by the following equation:

$$R_{DA} = R_0 \times \sqrt[6]{\frac{1}{E} - 1} \qquad \text{Eq. V.2}$$

where R_0 is the Förster radius that is equal to 24.5 Å for the FRET pair Dns-Trp in phosphate buffer (0.1 M, pH 7.25) at 25 °C [11].

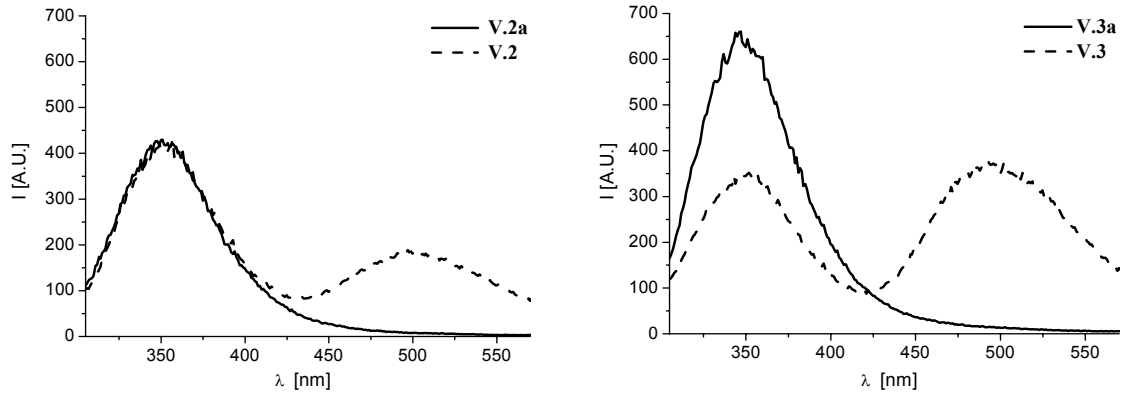


Figure V.6. Fluorescence spectra of the peptides **V.2a** and **V.2** (*left*) and **V.3a** and **V.3** (*right*) (each 5 μM). Spectra were recorded upon excitation at 295 nm in 0.1 M phosphate buffer (pH 7.25) at 25 °C.

Among the fluorescence-labeled analogues of the Id3 fragment 41-119 (**II.3**), those having Trp at position 76 in the presence and in the absence of the Dns group (peptides **V.6** and **V.6a**) were the only ones which maintained a conformation similar to that of the unlabeled peptide. In contrast, among analogues containing Trp-84, only the double-labeled one (peptide **V.5**) was comparable to the unlabeled form, whereas the mono-labeled one (peptide **V.5a**) showed increased helicity. This means that peptide **V.5** can be used as a fluorescent model for the unmodified sequence, but even if FRET occurs, it is not possible to estimate its efficiency, as the Trp environment is probably different in the mono- and double-labeled species. The two peptides **V.4** and **V.4a** bearing Trp-amide at the C-end were useless for further studies on the Id3 fragment 41-119, as they were mostly disordered.

Both double-labeled peptides **V.5** and **V.6** showed a pronounced emission maximum in the Dns region, accompanied by a reduction of the emission maximum in the Trp region (Figure V.7). Using equations V.1 and V.2, a FRET efficiency of 77% and a distance of ~ 20 Å between the N-terminal Dns group and Trp-76 were calculated. No estimation of the distance between the N-terminal Dns group and Trp-84 could be done, due to the reason discussed above.

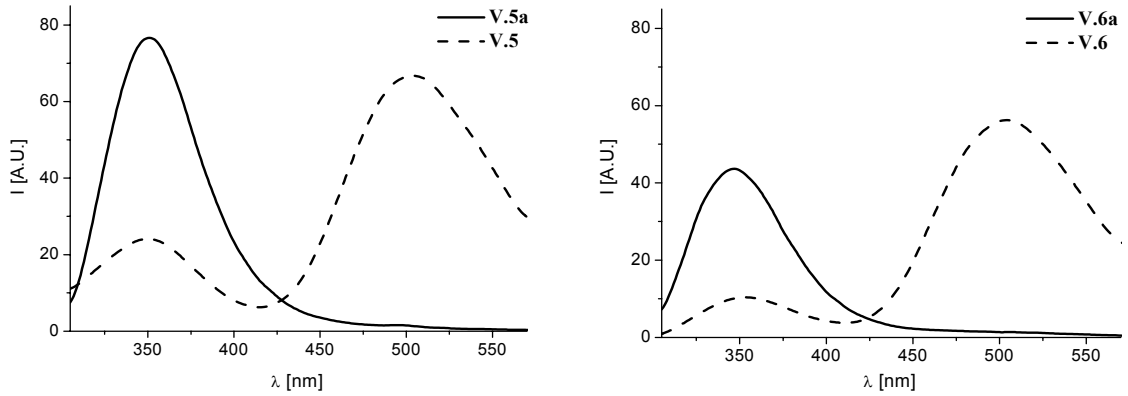


Figure V.7. Fluorescence spectra of the peptides **V.5** and **V.5a** (left, each 5 μM), and **V.6** and **V.6a** (right, each 5 μM). Spectra were recorded upon excitation at 295 nm in 0.1 M phosphate buffer (pH 7.25) at 25 °C.

V.5 Discussion and Conclusion

The Id3 HLH motif (residues 41-81) represents a highly compact structure which can apparently be disturbed by the insertion of a Trp residue at the C-end (corresponding to position 82, in which a valine residue is naturally present). The impact of a sterically demanding aromatic moiety located at the very end of the helix-2 was actually unexpected, as the tails of peptide helices display in general high flexibility, which should allow the accommodation of large groups. In Chapter III we have discussed the intrinsic helix propensity of helix-2. Helical wheel projection of the Id3 fragment 66-81 (corresponding to helix-2) predicts an amphipathic character with a wide and continuous hydrophobic surface (the only polar residues are located at the margins of the surface, being Gln-66 and Gln-81), and by the hydrophilic face including Glu-68, Gln-71, Arg-72, Asp-75 and Asp-79 (Figure V.8). The length and angle of the hydrophobic moment μ of the helix are 5.33 and 10.4. The length of μ can be related to the helix stability: the higher the length of μ, the more stable is the helix. Orientation of μ can reveal information about the helix packing. Elongation of the helix-2 by the residue 82 has not much influence on μ in the case of the natural Val-82, but it has a dramatic effect in the case of Trp-82, which leads to a decrease in the length of μ and change of the angle. Thus, peptides **V.1** and **V.1a** likely contain not only a less stable helix-2 due to its lowered amphipathic character, but also a different packing between helix-1 and helix-2, which is apparently less favorable than that in peptide **III.1**.

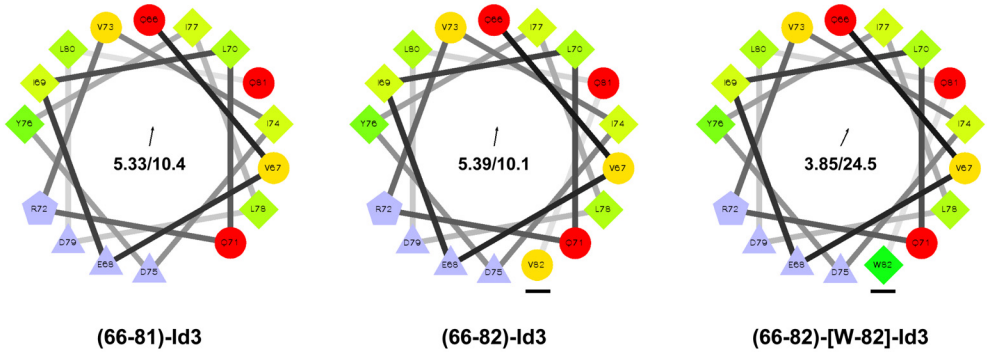


Figure V.8. Helical wheel projections of the Id3 fragments reproducing helix-2, residues 66-81 (*left*), and its elongation by residue 82 (indicated by underlined residue) as the natural Val (*middle*) or Trp (*right*). The projections have been created with the program available online [12] (hydrophilic residues as red circles, hydrophobic residues as diamonds, potentially negatively charged as triangles, and potentially positively charged as pentagons. Hydrophobicity is color coded as well: the most hydrophobic residue is green, and the amount of green is decreasing proportionally to the hydrophobicity, with zero hydrophobicity coded as yellow. Hydrophilic residues are coded red with pure red being the most hydrophilic (uncharged) residue, and the amount of red decreasing proportionally to the hydrophilicity. The potentially charged residues are light blue). The arrow inside the wheel represents the hydrophobic moment of the helix, μ (the numbers refer to the length/angle of μ).

Even more surprising than the impact of Trp-82 on the conformation of the Id3 HLH fold was that of the N-terminal Dns group (*cf.* peptide **V.1b**). In Chapter III we have already described poor intrinsic helix propensity of helix-1, which is actually disordered in solution. Labeling of its N-end through a glycine residue as a short spacer should have therefore been tolerated. Instead, the effect of the N-terminal Dns group was significantly more deleterious than that of the C-terminal Trp residue, leading to an almost complete loss of helicity (*cf.* peptides **V.1** and **V.1b** in Table V.2). The high sensitivity of the N- and C-ends of the Id3 HLH peptide to primary structure mutations suggests that they are not flexible tails of a compact core, but rather stabilize the folding.

Interestingly, the labeling of the N-end of the Id3 HLH motif was tolerated in the case of the larger fragment reproducing the HLH region and the complete C-terminus (peptides **V.5** and **V.6**). This different behavior between the short HLH peptide and the C-terminally elongated analogues might be explained by the fact that the latter folded into a less compact structure, in which the N-terminal Dns groups can be easily located. These analogues displayed FRET activity between the Trp residue at position 76 or 84, and the N-terminal Dns group.

The distance calculated by the measured FRET efficiency and the Förster radius was ~ 20 Å between the N-end and Trp-76. Considering the crystal structure of the DNA-bound MyoD dimer, this distance is slightly shorter than those measured intra- and intermolecularly between the N-end of the MyoD helix-1 (Arg-121, corresponding to Id3 Leu-40) and Tyr-156 of the helix-2 (corresponding to Id3 Tyr-76), which are ~ 26 Å and ~ 22 Å, respectively (Figure V.9). The Id3 HLH dimer probably adopts more compact structure than the DNA-binding MyoD dimer.

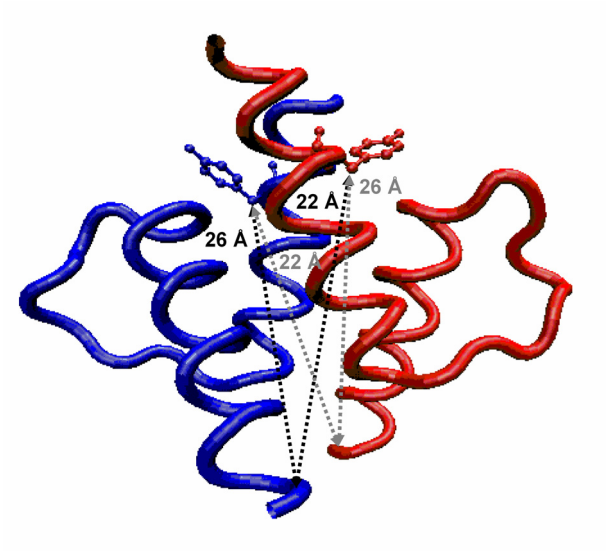


Figure V.9. View of the four-helix bundle of the MyoD bHLH motif bound to the DNA (PDB ID: 1MDY) [13]. Each subunit corresponds to the MyoD residues 121-161. The side chain of Tyr-156, which corresponds to Tyr-76 of the Id3 HLH motif, is shown. The distances have been measured by using the α-carbon atom of Arg-121 (Id3 Leu-40) and the β-carbon of Tyr-156 (Id3 Tyr-76).

The modification of C-terminus by amidation with Trp-amide caused a dramatic effect on spatial arrangement, which indicates that the C-end of the Id3 protein is important. The design and synthesis of other fluorescent analogues of the Id3 fragment 41-119 are necessary to analyze the structural role of this flanking region.

FRET properties of the Id3 fragment 17-101, which reproduces the HLH motif preceded and followed by two flanking regions (peptide **V.3**), suggest that the flanking region N-terminal to the HLH motif is not extended but rather folded toward the HLH core, as the distance between positions 17 and 76 in peptide **V.3** was only ~ 5 Å longer than the distance between positions 40 and 76 in peptide **V.6**.

Summarizing the results presented above, it can be concluded that the Id3 HLH fold adopts a compact structure in which the helix packing should be similar to that found for the MyoD HLH fold. Moreover, the region N-terminal to the HLH domain is likely to build an inverse loop that brings position 17 in proximity of the beginning of helix-1 (Figure V.10).

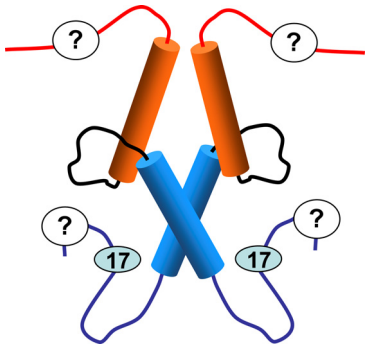


Figure V.10. Proposed model of the Id3 dimer based on the FRET analogues.

V.6 Literature

[1] Deed, R. W., Hara, E., Atherton, G. T., Peters, G., and Norton, J. D. (1997) Regulation of Id3 cell cycle function by Cdk-2-dependent phosphorylation. *Mol Cell Biol 17*, 6815-21.

[2] Forrest, S. T., Taylor, A. M., Sarembock, I. J., Perlegas, D., and McNamara, C. A. (2004) Phosphorylation regulates Id3 function in vascular smooth muscle cells. *Circ Res 95*, 557-9.

[3] Chen, B., Han, B. H., Sun, X. H., and Lim, R. W. (1997) Inhibition of muscle-specific gene expression by Id3: requirement of the C-terminal region of the protein for stable expression and function. *Nucleic Acids Res 25*, 423-30.

[4] Hata, K., Yoshimoto, T., and Mizuguchi, J. (2004) CD40 ligand rescues inhibitor of differentiation 3-mediated G1 arrest induced by anti-IgM in WEHI-231 B lymphoma cells. *J Immunol 173*, 2453-61.

[5] Svobodova, J., and Cabrele, C. (2006) Stepwise solid-phase synthesis and spontaneous homodimerization of the helix-loop-helix protein Id3. *Chembiochem 7*, 1164-1168.

[6] Bettio, A., Dinger, M. C., and Beck-Sickinger, A. G. (2002) The neuropeptide Y monomer in solution is not folded in the pancreatic-polypeptide fold. *Protein Sci 11*, 1834-44.

[7] Stryer, L. (1978) Fluorescence energy transfer as a spectroscopic ruler. *Annu Rev Biochem 47*, 819-46.

[8] Kiewitz, S. D., and Cabrele, C. (2005) Synthesis and conformational properties of protein fragments based on the Id family of DNA-binding and cell-differentiation inhibitors. *Biopolymers 80*, 762-74.

[9] Fairman, R., Beran-Steed, R. K., Anthony-Cahill, S. J., Lear, J. D., Stafford, W. F., 3rd, DeGrado, W. F., Benfield, P. A., and Brenner, S. L. (1993) Multiple oligomeric states regulate the DNA binding of helix-loop-helix peptides. *Proc Natl Acad Sci USA 90*, 10429-33.

[10] Chen, Y. H., Yang, J. T., and Chau, K. H. (1974) Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry 13*, 3350-9.

[11] Schiller, P. W. (1972) Study of adrenocorticotropic hormone conformation by evaluation of intramolecular resonance energy transfer in N -dansyllysine 21 -ACTH-(1-24)-tetrakosipeptide. *Proc Natl Acad Sci USA 69*, 975-9.

[12] http://rzlab.ucr.edu/scripts/wheel.18th January 2007.

[13] Ma, P. C., Rould, M. A., Weintraub, H., and Pabo, C. O. (1994) Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell 77*, 451-9.

VI. Synthetic Studies on Id3 Peptide Thioesters as Tools for Native Chemical Ligation

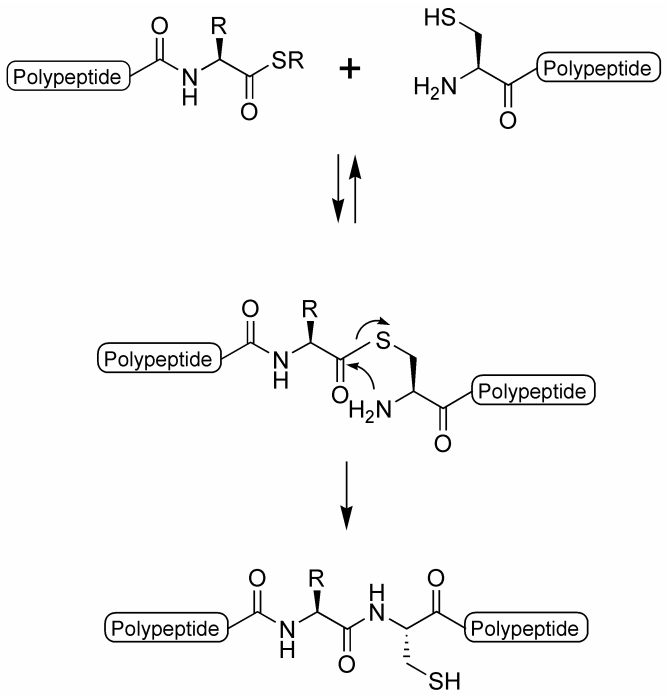
VI.1 Introduction

The solid-phase methodology is widely applied for the chemical synthesis of peptides and protein fragments. In general, the number of amino acid residues that can be assembled in a stepwise manner is around 50, but there are also examples of larger polypeptides, including the 119-residue long Id3 protein shown in Chapter II. The success of a solid-phase synthesis is strongly related to the amino acid sequence which can be classified as "good" or "difficult", depending on the degree of homogeneity of the crude product. The most critical problem in stepwise SPPS is the incompleteness of coupling and deprotection reactions, which are very often caused by the formation of aggregates of the growing peptide chains. Accordingly, sequences predicted to form β-sheets are usually difficult to prepare. The amino acid composition also plays a role: the presence of a considerable number of β-branched amino acids like Val, Ile, Leu, Thr, as well as of Arg and Lys can compromise the synthesis.

When the desired peptide chain length can not be reached by stepwise SPPS, an alternative is the use of the peptide fragment condensation in solution. However, one disadvantage of this approach is the fact that the two fragments must be fully-protected to allow a regioselective reaction between the C-end of one fragment and the N-end of the second fragment. As the purification of fully-protected peptides by preparative RP-HPLC is not easily practicable, the solution condensation is mostly limited to those cases, where the crude fully-protected species are obtained with good homogeneity. Another problem in the condensation of peptide fragments is the high risk of epimerization of the α-carbon of the activated C-end. An alternative to the latter method is represented by the so-called native chemical ligation (NCL) of unprotected peptides [1-3]. This procedure does not require coupling reagents or protection schemes, but is achieved through a reversible chemoselective capture step followed by an irreversible intramolecular acyl transfer. NCL involves the reaction of two unprotected peptides, the first in form of a C-terminal thioester and the other one having an N-terminal Cys, which eventually leads to a naturally occurring amide bond at the ligation site [4]. The peptide segments react by means of a reversible thiol/thioester exchange to give the initial thioester-linked product. The latter spontaneously undergoes an irreversible nucleophilic rearrangement by a highly favored intramolecular mechanism which results into the final amide-linked product (Scheme VI.1). The reaction is performed in aqueous solution at neutral or lightly basic

pH, and generally under denaturing conditions. Folding-assisted NCL has been used in special cases to accelerate the ligation step [5]. Although the ligation step is usually accomplished in solution, some examples are reported in the literature, in which one of the two segments is bound to a water-compatible resin (e.g. cellulose), while the other is applied in aqueous solution [6].

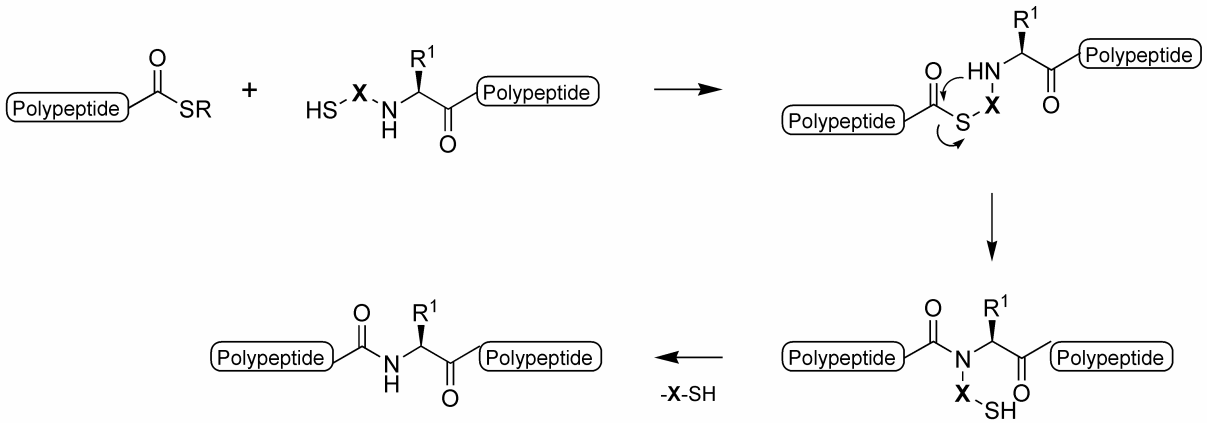
Scheme VI.1. NCL between a C-terminal polypeptide thioester and an N-terminal Cys-containing polypeptide.



Expressed protein ligation (EPL), also called intein-mediated protein ligation, is a variant of NCL [7-9]. The difference arises from the preparation of the C-terminal thioester: in the NCL, it is obtained by chemical methods, whereas in the EPL it is first expressed as a protein fused to a mutated intein which is then subjected to an intramolecular N→S acyl shift and an intermolecular transthioesterification.

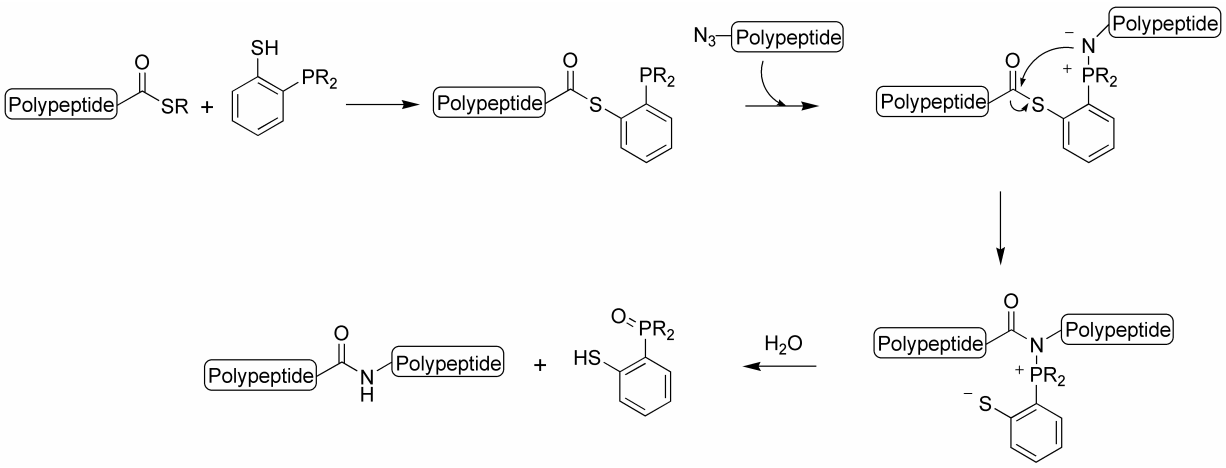
One disadvantage of the NCL and EPL methods is the necessity of a Cys residue at the ligation site [10]. The occurrence of this amino acid in globular proteins is very low and the insertion of additional Cys residues can alter the protein structure and function by the formation of unnatural disulfide bridges. Thus, a lot of efforts have been made to develop strategies which do not require an N-terminal Cys residue at the ligation point, including the use of removable N-terminal thiol auxiliaries, such as the mercaptoethoxy- [11] and the mono- or dimethoxy-substituted 2-mercaptobenzyl group [2, 12] (Scheme VI.2.).

Scheme VI.2. Thiol auxiliary ligation (HS-**X**: mercaptoethoxy-, mono- or dimethoxy-substituted benzyl thiol auxiliary).



Another elegant ligation method has been inspired by the Staudinger reaction, in which a phosphine is used to reduce an azide to an amine [13-15]. In this method a peptide bearing a C-terminal phosphinothioester is coupled to another peptide with an N-terminal α-azido group to form an intermediate iminophosphorane that possesses a nucleophilic nitrogen capable of intramolecular reaction with an acyl donor and give the desired amide bond (Scheme VI.3). This method may also be combined with the classic NCL for tandem ligation reactions [16, 17].

Scheme VI.3. The Staudinger ligation.

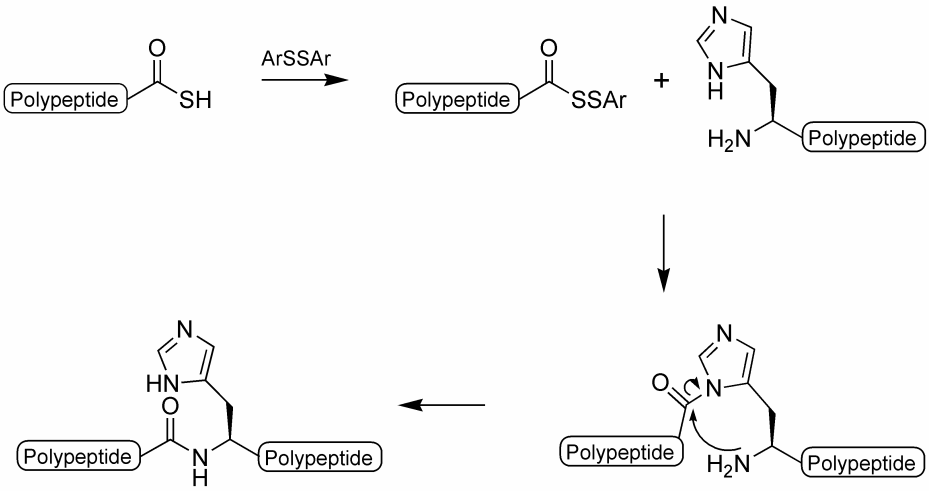


The principle of classical Cys ligation has been successfully extended to obtain an Ala [18] or a Met residue at the ligation site [2, 19]. The Ala residue, which is highly frequent in peptides and proteins, is chosen for the ligation site and substituted with a Cys residue. After classical NCL, the Cys side chain is converted to a methyl group by desulfurization using palladium or Raney-nickel and hydrogen. In the Met ligation, homo-Cys instead of Cys

is used for the transthioesterification step. After the S→N-acyl migration through a favorable six-membered ring transition state to form the homo-cysteinyl product, an additional S-methylation step follows.

The His ligation has been also proposed, as the imidazole group of the His side chain can be an alternative nucleophile to the thiol group [2, 20]. In this strategy, the acyl segment which bears the thioacid is activated by a thiophilic promoter (e.g. Ellman's reagent) and then captured by the imidazole of an N-terminal His, subsequently leading to the $N^{im}\rightarrow N^{\alpha}$-acyl transfer to form a peptide bond (Scheme VI.4).

Scheme VI.4. The His ligation.



The folding-assisted protein ligation using C-terminal peptide thioesters is another example of Cys-free ligation [5]. In this case, the protein conformation determines the regioselective coupling reaction, provided that the protein is able to correctly fold in a noncovalent manner.

VI.2 Chemical Synthesis of C-Terminal Peptide Thioesters

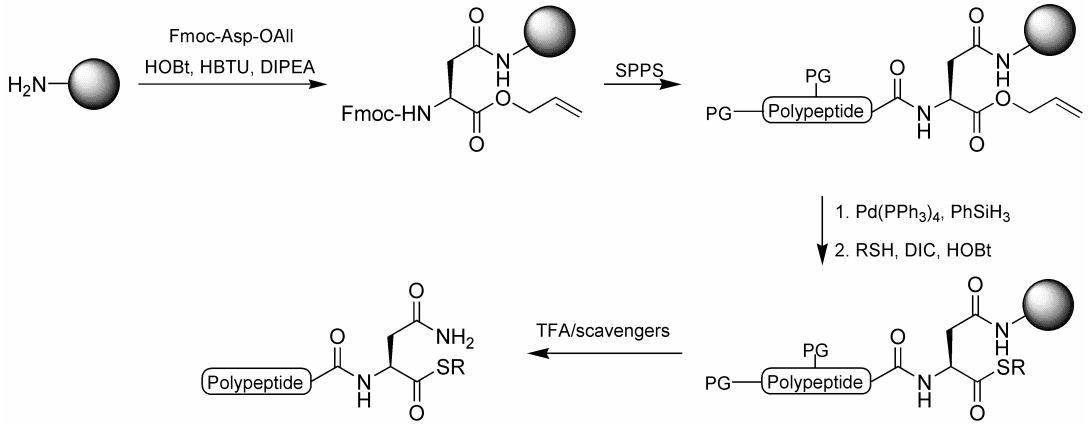
C-terminal peptide thioesters are essential intermediates for chemical ligation [21]. They are typically prepared either by SPPS [22] or biosynthetically, especially in the case of large thioesters, by the intein technology [23]. To avoid the uncomfortable Boc chemistry which requires the use of strong acids like HF to remove the peptide from the solid support, several research groups have recently explored the possibility of applying the Fmoc chemistry for the preparation of thioesters. One problem that must be solved is the lability of the thioester bond toward bases; this means that the repetitive treatment with piperidine to cleave the Fmoc protecting group [24] is not compatible with the presence of a C-terminal thioester

bond. Nonetheless, the Fmoc/*t*-Bu strategy would offer some advantages with respect to the Boc/benzyl one, including the possibility to achieve milder overall reaction and cleavage conditions, the compatibility with a variety of solvents and solid supports, and the real-time UV monitoring of the reactions. In contrast to the Boc methodology, the Fmoc methodology also provides the synthesis of phospho- and glycopeptides in good yields [25]. Fmoc-chemistry-based protocols for the preparation of peptide thioesters are now available, which include the use of new cleavage cocktails and different types of resin-linkers [21]. For example, the non-nucleophilic but strongly basic diazabicyclo[5,4,0]-undec-7-ene (DBU) in conjunction with *N*-hydroxybenzotriazole (HOBt) has been shown to be able to remove efficiently the Fmoc group without causing aminolysis of the thioester bond [26, 27].

Solid-phase synthesis using a backbone amide linker (BAL) strategy relies on the anchoring of the first amino acid protected at the α-carboxylic group as allyl ester to the linker by reductive amination [28, 29]. After peptide chain assembly, the orthogonal allylic protection is removed and the free carboxylic acid is coupled to an amino acid thioester. The final cleavage/deprotection step releases the peptide thioester in solution. However, this procedure has the disadvantage that a diketopiperazine (DKP) can be formed at the dipeptide stage and that the C-terminal residue can epimerize upon coupling with the thioester. To prevent the formation of DKP, an amino trithioorthoester derived from Gly can be first anchored to the linker by reductive amination [30]. As the trithioorthoester is not susceptible to nucleophilic attack, the formation of DKP is avoided. After standard Fmoc-based assembly of the peptide, treatment of the peptidyl-resin with TFA leads directly to the corresponding peptide thioester. One limitation of this procedure is the fact that only C-terminal Gly thioesters can be synthesized.

Similarly to the BAL strategy, the side chain anchoring of Fmoc-Glu-allyl or Fmoc-Asp-allyl to hydroxyl or amine linkers has also been used for the solid-phase synthesis of thioesters as well as of circular peptides [31-34]. Deprotection of the allyl group is achieved by treatment of the peptidyl-resin with tetrakis(triphenylphosphine)palladium in the presence of *N*-methylmorpholine or phenylsilane (Scheme VI.5) [34].

Scheme VI.5. Side chain anchoring strategy combined to thioesterification on the solid support (PG: protecting group).

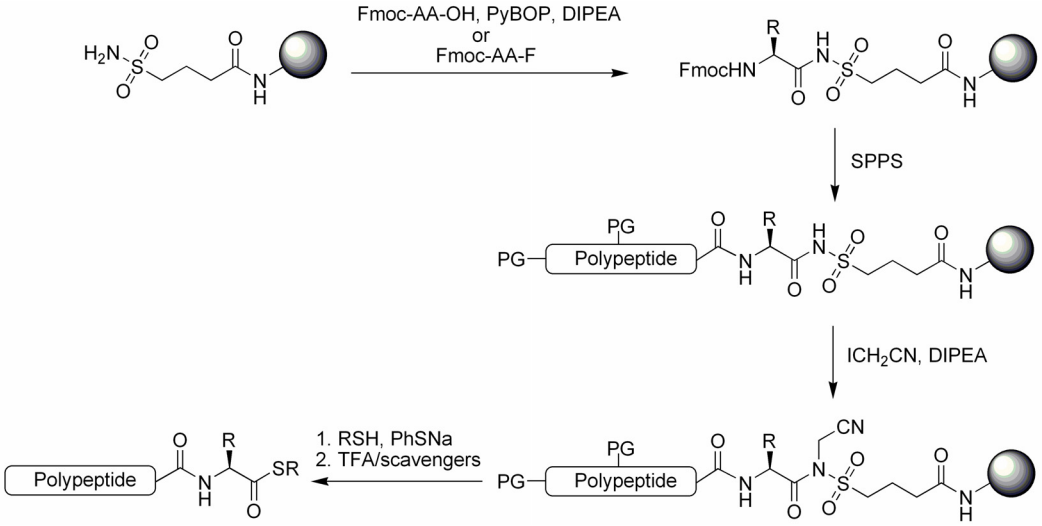


C-terminal peptide esters anchored to Wang or PAM resins have been converted into thioesters by treatment with an excess of Me_2AlCl/RSH before TFA cleavage [35]. In this way, the use of special linkers is not required; however, epimerization of the C-terminal amino acid currently limits this application to the preparation of peptide thioesters having a C-terminal Gly [36].

Sophisticated linkers have been also developed for the synthesis of peptide thioesters using Fmoc chemistry. One example is the aryl hydrazine safety-catch linker which, at the end of the peptide chain assembly, can be activated by mild oxidation with *N*-bromosuccinimide in the presence of pyridine in DCM [37]. The reactive acyl diazene is then efficiently cleaved with an amino acid S-alkyl thioester to yield the corresponding fully-protected peptide thioester. Subsequent side chain deprotection with TFA provides the desired peptide thioester.

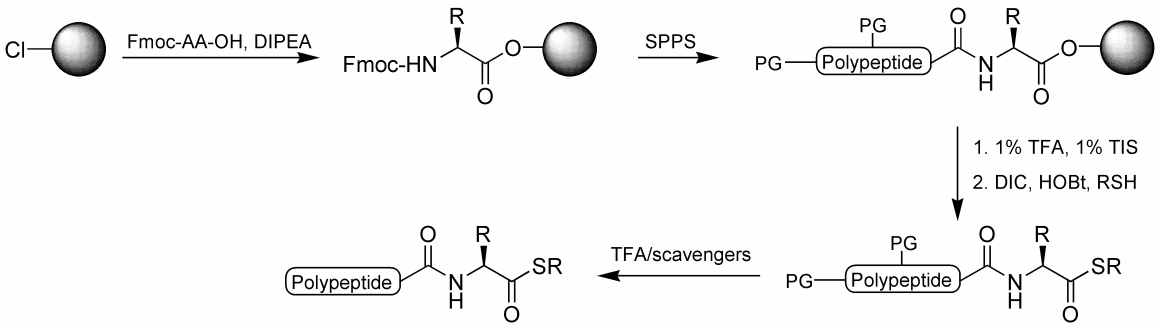
In the case of the Kenner's sulfonamide safety-catch linker [38], the C-terminus of the growing peptide chain is anchored to the resin via an *N*-acyl-sulfonamide which is stable both in basic and strongly nucleophilic reaction conditions. The first Fmoc-amino acid is attached to the sulfonamide resin using PyBOP/DIPEA at low temperatures to avoid racemization [39], or using the acyl fluoride that has been shown to give high loading in short times [40]. The *N*-acyl-sulfonamide is preactivated after peptide synthesis by N-alkylation with trimethylsilyldiazomethane or iodoacetonitrile [25, 41], followed by thiolysis (Scheme VI.6).

Scheme VI.6. The sulfamylbutyryl (SAB) safety-catch resin and its application to the synthesis of C-terminal peptide thioesters (PG: protecting group).



C-terminal peptide thioesters can be prepared also in solution using protected peptides [42]. The peptide is first assembled on the highly acid-labile 2-chlorotrityl chloride resin. After cleavage with diluted acid, the free C-terminal carboxylic group of the otherwise fully-protected peptide undergoes thioesterification with the appropriate thiol using DIC and HOBt as activating reagents [43]. Final treatment of the product with concentrated TFA provides the fully-unprotected peptide thioester (Scheme VI.7).

Scheme VI.7. Thioesterification of protected peptides in solution. (PG: protecting group, TIS: triisopropyl-silane).



Before the stepwise solid-phase synthesis of the Id3 protein was achieved, we also considered the possibility of using the NCL approach, if the assembly of large fragments or of the entire Id3 sequence had been prevented by incomplete acylation/deprotection reactions during the chain growth. Here, we present the results of a synthetic study of peptide thioesters related to Id3 fragments.

VI.3 Synthesis of Id3 Polypeptide Thioesters

The Id3 protein contains five natural Cys residues at positions 10, 15, 16, 47 and 118; however, the only Cys residue that may be suitable for a NCL procedure is Cys-47, as positions 10, 15, 16 and 118 are too close to the termini, thus providing no benefit. Alternatively, beside the Cys-mediated ligation, also the thiol auxiliary- or His-mediated ligations may be considered. To evaluate the synthetic accessibility of peptide thioesters for their use in NCL experiments, we decided to prepare a number of Id3 fragments to be converted into thioesters either in solution (method *a*) or on the solid support (methods *b* and *c*). The Id3 peptide thioesters which have been prepared in this study are reported in Table VI.1. We took short (up to 16 residues), medium (up to 29 residues) or large portions (up to 46 residues) of three Id3 subdomains (the HLH motif and the N- and C-termini), in order to test sequences which exhibit various difficulties. We also analyzed the influence of the thiol on the thioesterification reaction and on the stability of the final thioester: 3-mercapto-propionic acid ethyl ester (*HS*PAE), benzylmercaptane (*HS*Bn) and *p*-acetamidothiophenol (*HS*PAA) were used.

VI.3.1 Thioester Formation in Solution (Method *a*)

Three different Id3 sequences were used to accomplish the synthesis of the corresponding thioesters in solution: 1-46, 41-56 and 48-56. The sequences were first assembled on a 2-chlorotrityl resin, from which they were cleaved as fully-protected C-terminal peptide acids.

Table VI.1. Analytical data of the synthetic Id3 polypeptide thioesters.

Human Id3[a]					
M[1]KALSPVRGCYEAVCCLSERSLAIARGRGKGPAAEEPLSL[40] LDDM**NH**CYSRLRELV**P**[56]**G**VPR**G**T**Q**LS[65]QVEILQRVIDYIL**D**LQ[81] VVLAEPAPGPPDGPHLPIQTAELAPELVIS**N**DKRSFCH[119]					
Synthetic Id3 peptide thioesters					
Method	**Peptide**	**Sequence**	**$MW_{calcd.}$ (Da)**	**MW_{found} (Da)**	**t_R (min)**[b]
a	**VI.1**	H-(1-46)-*S*PAE	5032.9	5034.3	17.1
	VI.2	H-(1-46)-*S*Bn	5022.9	5022.6	16.6
	VI.3	H-(1-46)-*S*PAA	5065.9	4895.0[c]	ca. 15
	VI.4	Ac-(41-56)-*S*PAE	2119.5	2119.0	17.9
	VI.5	Ac-(48-56)-*S*PAE	1290.6	1290.5	16.4
	VI.6	Ac-(48-56)-*S*Bn	1280.6	1279.7	17.5
b	**VI.7**	Ac-(17-45)-*S*PAE	3225.7	3225.5	15.2
	VI.8	Ac-(17-45)-*S*Bn	3215.7	3216.0	16.9
	VI.9	Ac-(17-45)-*S*PAA	3258.8	3256.9	14.8
	VI.10	H-(82-112)-*S*PAE	3258.8	3256.9	17.7
	VI.11	H-(82-112)-*S*Bn	3248.8	3251.6	19.2
	VI.12	Ac-(41-79)-*S*Bn	4704.5	4702.6	23.4
	VI.13	Ac-[γ-Glu(α-C(O)NH_2)-63]-(41-63)-*S*Bn	2805.3	2807.7	16.5
c	**VI.14**	Ac-[Cys(S*t*Bu)-47]-(41-61)-*S*Bn	2574.2	2574.8	nd[d]
	VI.15	Ac-[Cys(S*t*Bu)-47]-(38-57)-*S*Bn	2480.0	2479.9	nd[d]

[a] Underlined sequences: helix-1 (41-56) and helix-2 (66-81); double underlined sequence: loop (57-65). The positions chosen as the C-terminal thioesters are indicated in bold.

[b] Gradient: 10% ACN for 3 min, 10-70% ACN in 40 min.

[c] The peptide corresponds to a mixture of cyclic isomers obtained by aminolysis/thiolysis of the thioester.

[d] nd: not determined.

To control the quality of the peptide chains after SPPS, a small portion of each peptide was cleaved from the resin and simultaneously fully-deprotected. Whereas the short sequences 41-56 and 48-56 exhibited good homogeneity, the larger sequence 1-46 was contaminated by a conspicuous amount of impurities. As the peptide contains air-sensitive amino acids (three Cys and two Met residues), the crude product was treated with DTT to reduce potential disulfide bonds; however, the HPLC profile did not change, thus excluding the presence of byproducts due to Cys oxidation. By contrast, when the crude product was treated with TMSBr/EDT/TFA to reduce potentially oxidized Met residues, one of the two major HPLC peaks was significantly reduced, thus confirming the presence of a large amount of sequences containing Met(O) (Figure VI.1).

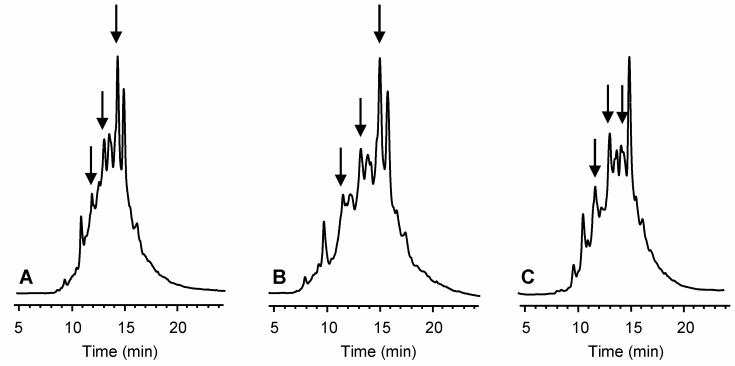


Figure VI.1. HPLC profile of the crude Id3 peptide acid H-(1-46)-OH as obtained after TFA cleavage (A) and upon additional treatment with DTT (B) or with TMSBr/EDT/TFA for partial Met(O) reduction (C). The arrows indicate the peaks corresponding to the sequences containing oxidized Met.

It should be pointed out that the poor homogeneity of the fully-deprotected sequence 1-46 might not reflect the quality of the peptide chain bound to the resin, as a large number of by-products can be formed during the cleavage step with concentrated TFA. Thus, we decided to convert the fully-protected sequence 1-46 into the corresponding thioesters (peptides **VI.1**-**3**). The peptide acid was activated by DIC/HOBt in DCM and then treated with the thiol. The formation of the thioester was monitored by analytical RP-HPLC and MALDI-TOF mass spectrometry. It was observed that the rate of the thioesterification increased with the nucleophilic character of the thiol: *HS*PAE < *HS*Bn < *HS*PAA. When the reaction was completed (after 2 days with *HS*PAA, 4 days with *HS*Bn and 10 days with *HS*PAE), the side chain protecting groups were cleaved using TFA with 10% (v/v) scavengers, and the obtained crude products were analyzed by RP-HPLC and MALDI-TOF mass spectrometry. In all three cases, the desired peptide thioester could be detected, and the crude products were thus subjected to

preparative RP-HPLC. The peptide thioesters **VI.1** and **VI.2** could be obtained with good purity, whereas the fraction isolated for the thioester **VI.3** showed a complex analytical HPLC profile (Figure VI.2). Moreover, mass spectrometry analysis revealed that no peptide thioester was present, but rather a mixture of isomers with a molecular mass smaller than the one expected for the thioester and also for the initial acid. The mass difference was attributed to the loss of the *S*PAA moiety upon intramolecular nucleophilic attack by an amino group (N-terminus or Lys side chain) or a thiol group from one of the three internal Cys residues. Such cyclization must have taken place during the preparative HPLC steps (probably during evaporation of acetonitrile from the isolated fractions before freezing and lyophilization).

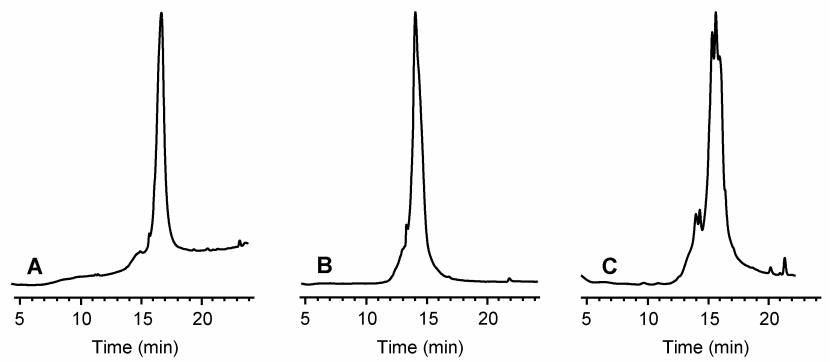


Figure VI.2. HPLC profiles of the Id3 peptide thioesters H-(1-46)-*S*PAE (A) and H-(1-46)-*S*Bn (B) after preparative RP-HPLC, and of the cyclic isomers of the Id3 sequence 1-46 obtained by aminolysis/thiolysis of the corresponding thioester **VI.3** during purification by preparative RP-HPLC (C).

These results have shown that *HS*PAA has the advantage of forming moderately quickly the corresponding thioester, which is however very reactive and highly sensitive to the presence of nucleophiles. Therefore, its use should be combined with masked side chains for Cys and Lys as well as with a protected N-terminus.

The preparation of the thioesters **VI.4-6** in solution went smoothly, as shown by the HPLC profiles of the crude products (the HPLC profile of the crude peptide **VI.5** is reported in Figure VI.3A as an example), which were then purified by preparative RP-HPLC (Figure VI.3B-C).

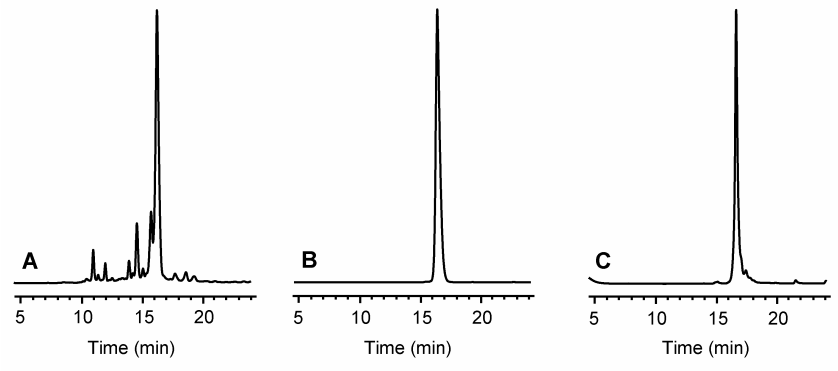


Figure VI.3. HPLC profiles of Ac-(48-56)-*S*PAE before purification (A), and of Ac-(48-56)-*S*Bn (B) and Ac-(41-56)-*S*PAE (C) after purification.

VI.3.2 Thioester Formation on the Solid Support (Methods *b* and *c*)

The thioester formation on the solid support brings the advantage of avoiding the handling of fully-protected peptide intermediates in solution. We used two different approaches, the first based on the side chain anchoring of the C-terminal amino acid to the resin (method *b*), and the other based on the application of a safety-catch resin linker (method *c*).

With the method *b*, we synthesized the Id3 peptide thioesters **VI.7-13**. The amino acid sequences 17-45 and 82-112 were first assembled on Rink amide resin preloaded with Fmoc-Asp-OAll, to obtain the native Asn-45 and Asn-112, respectively, after the resin cleavage. The sequence 41-79 was assembled on Wang resin preloaded with Fmoc-Asp-OAll, to obtain the native Asp-79 after the resin cleavage. Finally, the sequence 41-63 was assembled on Rink amide resin preloaded with Fmoc-Glu(OAll)-OH, to obtain an unnatural γ-amino acid residue at position 63, in place of Gln, after the resin cleavage (such backbone modification might be useful to design an Id3 HLH analogue with an altered loop structure). After the peptide chain assembly was completed, the allyl group was removed by $Pd(PPh_3)_4$ and $PhSiH_3$ or *N*-methylmorpholine in DCM. The free acid was then activated with DIC/HOBt and treated with the desired thiol. The reaction ran for 18 hours, and the peptide thioester was then cleaved from the resin, purified by preparative RP-HPLC and characterized by analytical RP-HPLC and MALDI-TOF mass spectrometry. The RP-HPLC and mass spectrometry data indicated that the crude products obtained from the syntheses of **VI.7-9** were contaminated by a large fraction of Met(O)-44-containing chains, and an additional reductive treatment was

therefore carried out before purification (Figure VI.4). It is worthwhile to note that the PAA thioester of the N-acetylated Id3 sequence 17-45 (peptide **VI.9**) was found to be stable enough during purification, in contrast to the PAA thioester of the Id3 sequence 1-46 (**VI.3**). This confirms the sensitivity of the PAA thioester towards free Cys side chains and the N-terminus, whereas the unmasked Lys side chain does not seem to cause aminolysis of the PAA thioester.

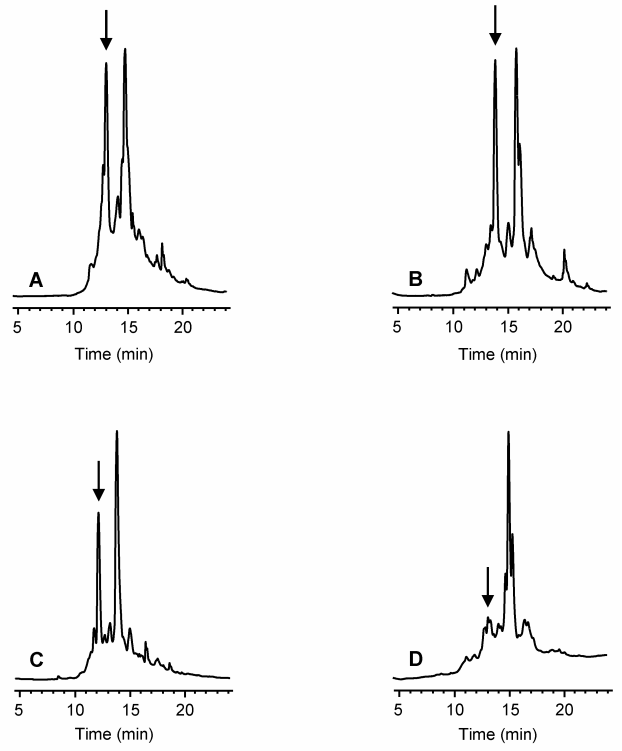


Figure VI.4. HPLC profiles of the crude Id3 peptide thioesters Ac-(17-45)-*S*R, whit R = PAE (A), Bn (B) and PAA (C). (D) HPLC profile of Ac-(17-45)-*S*PAE after reduction of Met(O)-44. The arrows indicate the peaks corresponding to the peptide fraction containing the oxidized Met.

The other peptide thioesters **VI.10-13** were obtained with high homogeneity even as crude products (Figure VI.5). Interestingly, Met-44 was not oxidized during the syntheses of peptides **VI.12-13**, contrarily to the observation during the syntheses of peptides **VI.7-9**. This may be explained by the fact that Met-44 remained exposed to air for much longer time during the latter syntheses.

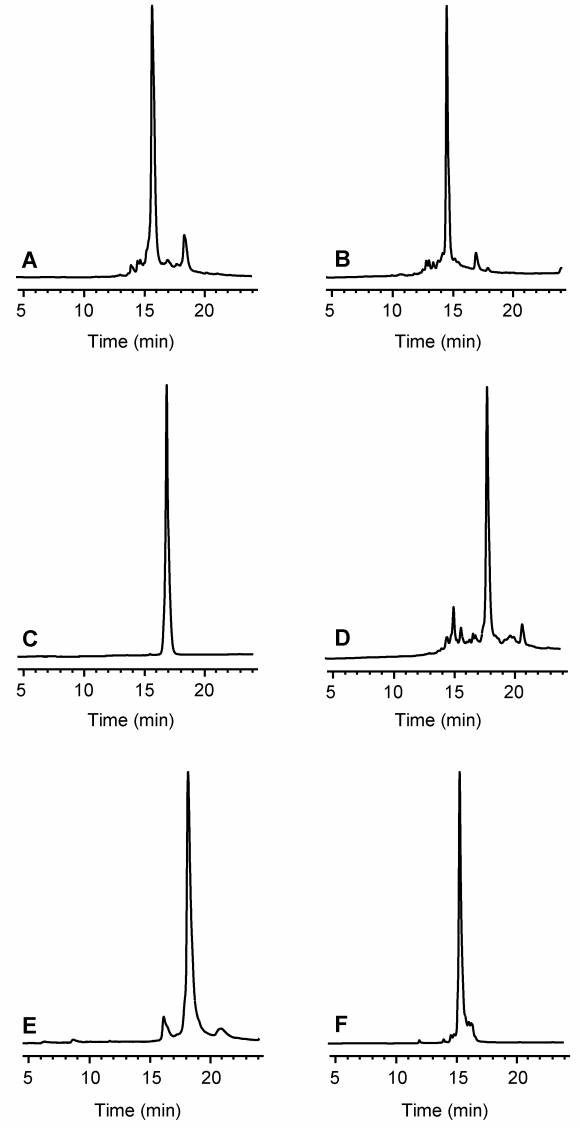


Figure VI.5. HPLC profiles of the following Id3 peptides: (A) H-(82-112)-OAll (crude), (B) H-(82-112)-OH (crude), (C) H-(82-112)-*S*PAE (purified), (D) H-(82-112)-*S*Bn (crude), (E) Ac-(41-79)-*S*Bn (purified) and (F) Ac-(41-63)-*S*Bn (purified).

By contrast, the two syntheses carried out with the method *c* using the safety-catch resin linker SAB did not run smoothly. The desired peptide thioesters **VI.14-15** were obtained as minor products accompanied by abundant and complex byproducts which could not be identified by mass spectrometry analysis (Figure VI.6).

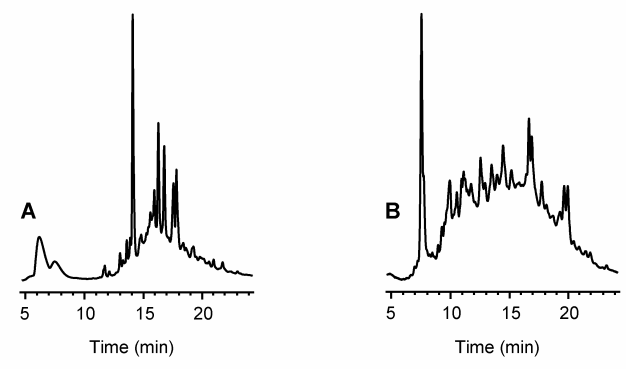


Figure VI.6. HPLC profiles of the crude products obtained from the syntheses of peptides **VI.14** (A) and **VI.15** (B).

VI.4 Conclusion

Thioester formation and subsequent chemical ligation is an alternative method to the stepwise SPPS to prepare large Id3 protein fragments or even the full-length protein. Three different kinds of peptide thioesters were prepared (R-*S*PAE, R-*S*Bn and R-*S*PAA) by three different thioesterification routes. The most effective one was the thioesterification on the solid support with anchored side chain of the C-terminal residue: indeed, the reaction was fast and clean and the work-up was easy.

It should be kept in mind that sequences containing unmasked thiol and amino groups could be critical because of intramolecular transthioesterification or amidation; it would be therefore worthwhile to mask these groups with orthogonal protecting groups.

The synthesized Id3 peptide thioesters may be applied to different types of ligation to form large Id3 fragments: for example, Cys-mediated ligation can be used with the thioesters of the sequence 1-46. His-mediated ligation can be used with the thioesters of the sequence 17-45, whereas thiol-auxiliary-mediated ligation can be used with the remaining thioesters. The synthetic study of Id3 peptide thioester was carried out in parallel with the stepwise SPPS of the Id3 protein. The latter was successful, as described in Chapter II, and the NCL experiments were therefore not carried out. However, the peptide thioesters presented here provide valuable intermediates for the future syntheses of Id3 protein related analogues.

VI.5 Literature

[1] Nilsson, B. L., Soellner, M. B., and Raines, R. T. (2005) Chemical synthesis of proteins. *Annu Rev Biophys Biomol Struct 34*, 91-118.

[2] Tam, J. P., Xu, J., and Eom, K. D. (2001) Methods and strategies of peptide ligation. *Biopolymers 60*, 194-205.

[3] Dawson, P. E., and Kent, S. B. (2000) Synthesis of native proteins by chemical ligation. *Annu Rev Biochem 69*, 923-60.

[4] Hackeng, T. M., Griffin, J. H., and Dawson, P. E. (1999) Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc Natl Acad Sci USA 96*, 10068-73.

[5] Beligere, G. S., and Dawson, P. E. (1999) Conformationally assisted protein ligation using C-terminal thioester peptides. *J Am Chem Soc 121*, 6332-3.

[6] Canne, L. E., Botti, P., Simon, R. J., Chen, Y., Dennis, E. A., and Kent, S. B. H. (1999) Chemical protein synthesis by solid phase ligation of unprotected peptide segments. *J Am Chem Soc 121*, 8720-7.

[7] Schwarzer, D., and Cole, P. A. (2005) Protein semisynthesis and expressed protein ligation: chasing a protein's tail. *Curr Opin Chem Biol 9*, 561-9.

[8] David, R., Richter, M. P., and Beck-Sickinger, A. G. (2004) Expressed protein ligation. Method and applications. *Eur J Biochem 271*, 663-77.

[9] Muir, T. W. (2003) Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem 72*, 249-89.

[10] Quaderer, R., and Hilvert, D. (2002) Selenocysteine-mediated backbone cyclization of unprotected peptides followed by alkylation, oxidative elimination or reduction of the selenol. *Chem Commun (Camb)*, 2620-1.

[11] Canne, L. E. (1996) Extending the applicability of native chemical ligation. *J Am Chem Soc 118*, 5891-6.

[12] Offer, J., and Dawson, P. E. (2000) N-alpha-2-mercaptobenzylamine-assisted chemical ligation. *Org Lett 2*, 23-6.

[13] Kim, H., Cho, J. K., Aimoto, S., and Lee, Y. S. (2006) Solid-phase Staudinger ligation from a novel core-shell-type resin: a tool for facile condensation of small peptide fragments. *Org Lett 8*, 1149-51.

[14] Nilsson, B. L., Hondal, R. J., Soellner, M. B., and Raines, R. T. (2003) Protein assembly by orthogonal chemical ligation methods. *J Am Chem Soc 125*, 5268-9.

[15] Nilsson, B. L., Kiessling, L. L., and Raines, R. T. (2000) Staudinger ligation: a peptide from a thioester and azide. *Org Lett 2*, 1939-41.

[16] Eom, K. D., Miao, Z., Yang, J.-L., and Tam, J. P. (2003) Tandem ligation of multipartite peptides with cell-permeable activity. *J Am Chem Soc 125*, 73-82.

[17] Tam, J. P., Yu, Q., and Lu, Y. A. (2001) Tandem peptide ligation for synthetic and natural biologicals. *Biologicals 29*, 189-96.

[18] Yan, L. Z., and Dawson, P. E. (2001) Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J Am Chem Soc 123*, 526-33.

[19] Tam, J. P., and Yu, Q. (1998) Methionine ligation strategy in the biomimetic synthesis of parathyroid hormones. *Biopolymers 46*, 319-27.

[20] Zhang, L., and Tam, J. P. (1997) Orthogonal coupling of unprotected peptide segments through histidyl amino terminus. *Tetrahedron Lett 38*, 3-6.

[21] Camarero, J. A., and Mitchell, A. R. (2005) Synthesis of proteins by native chemical ligation using Fmoc-based chemistry. *Protein Pept Lett 12*, 723-8.

[22] Albericio, F. (2004) Developments in peptide and amide synthesis. *Curr Opin Chem Biol 8*, 211-21.

[23] Noren, C. J., Wang, J., and Perler, F. B. (2000) Dissecting the chemistry of protein splicing and its applications. *Angew Chem Int Ed 39*, 450-66.

[24] Sheppard, R. (2003) The fluorenylmethoxycarbonyl group in solid phase synthesis. *J Pept Sci 9*, 545-52.

[25] Shin, Y., Winans, K. A., Backes, B. J., Kent, S. B. H., Ellman, J. A., and Bertozzi, C. R. (1999) Fmoc-based synthesis of peptide-$^{\alpha}$thioesters: application to the total chemical synthesis of a glycoprotein by native chemical ligation. *J Am Chem Soc 121*, 11684-9.

[26] Clippingdale, A. B., Barrow, C. J., and Wade, J. D. (2000) Peptide thioester preparation by Fmoc solid phase peptide synthesis for use in native chemical ligation. *J Pept Sci 6*, 225-34.

[27] Bu, X., Xie, G., Law, C. W., and Guo, Z. (2002) An improved deblocking agent for direct Fmoc solid-phase synthesis of peptide thioesters. *Tetrahedron Lett 43*, 2419-22.

[28] Alsina, J., Yokum, T. S., Albericio, F., and Barany, G. (1999) Backbone amide linker (BAL) strategy for N(alpha)-9-fluorenylmethoxycarbonyl (Fmoc) solid-phase synthesis of unprotected peptide p-nitroanilides and thioesters. *J Org Chem 64*, 8761-9.

[29] Gross, C. M., Lelievre, D., Woodward, C. K., and Barany, G. (2005) Preparation of protected peptidyl thioester intermediates for native chemical ligation by N-alpha-9-fluorenylmethoxycarbonyl (Fmoc) chemistry: considerations of side-chain and backbone anchoring strategies, and compatible protection for N-terminal cysteine. *J Pept Res 65*, 395-410.

[30] Brask, J., Albericio, F., and Jensen, K. J. (2003) Fmoc solid-phase synthesis of peptide thioesters by masking as trithioortho esters. *Org Lett 5*, 2951-3.

[31] Wang, P., and Miranda, L. P. (2005) Fmoc-protein synthesis: Preparation of peptide thioesters using a side-chain anchoring strategy. *Int J Pept Res Ther 11*, 117-23.

[32] Dolphin, G. T. (2006) A designed well-folded monomeric four-helix bundle protein prepared by Fmoc solid-phase peptide synthesis and native chemical ligation. *Chemistry 12*, 1436-47.

[33] Tulla-Puche, J., and Barany, G. (2004) On-resin native chemical ligation for cyclic peptide synthesis. *J Org Chem 69*, 4101-7.

[34] Grieco, P., Gitu, P. M., and Hruby, V. J. (2001) Preparation of 'side-chain-to-side-chain' cyclic peptides by allyl and alloc strategy: potential for library synthesis. *J Pept Res 57*, 250-6.

[35] Swinnen, D., and Hilvert, D. (2000) Facile, Fmoc-compatible solid-phase synthesis of peptide C-terminal thioesters. *Org Lett 2*, 2439-42.

[36] Sewing, A., and Hilvert, D. (2001) Fmoc-compatible solid-phase peptide synthesis of long C-terminal peptide thioesters. *Angew Chem Int Ed 40*, 3395-6.

[37] Camarero, J. A., Hackel, B. J., de Yoreo, J. J., and Mitchell, A. R. (2004) Fmoc-based synthesis of peptide alpha-thioesters using an aryl hydrazine support. *J Org Chem 69*, 4145-51.

[38] Backes, B. J., Virgilio, A. A., and Ellman, J. A. (1996) Activation method to prepare a highly reactive acylsulfonamide 'safety-catch' linker for solid-phase synthesis. *J Am Chem Soc 118*, 3055-6.

[39] Backes, B. J., and Ellman, J. A. (1999) An alkanesulfonamide "safety-catch" linker for solid-phase synthesis. *J Org Chem 64*, 2322-30.

[40] Ingenito, R., Dreznjak, D., Guffler, S., and Wenschuh, H. (2002) Efficient loading of sulfonamide safety-catch linkers by Fmoc amino acid fluorides. *Org Lett 4*, 1187-8.

[41] Ingenito, R., Bianchi, E., Fattori, D., and Pessi, A. (1999) Solid phase synthesis of peptide C-terminal thioesters by Fmoc/*t*-Bu chemistry. *J Am Chem Soc 121*, 11369-74.

[42] Futaki, S., Sogawa, K., Maruyama, J., Asahara, T., and Niwa, M. (1997) Preparation of peptide thioesters using Fmoc-solid-phase peptide synthesis and its application to the construction of a template-assembled synthetic protein (TASP). *Tetrahedron Lett 38*, 6237-40.

[43] von Eggelkraut-Gottanka, R., A., K., Beck-Sickinger, A. G., and Beyermann, M. (2003) Peptide $^{\alpha}$thioester formation using standard Fmoc-chemistry. *Tetrahedron Lett 44*, 3551-4.

VII. Design and Synthesis of Fluorogenic MMP-2 Substrates

VII.1 Introduction

A number of biological studies have clearly shown that the Id proteins are required for tumor angiogenesis and vascularization. For example, observations made by Benezra and co-workers suggest that the expression of $\alpha_v\beta_3$ integrin during tumor-induced angiogenesis and the recruitment and/or expression of the matrix metalloproteinase MMP-2 required for tumor growth and metastasis may be directly or indirectly controlled by the Id proteins [1]. The same research group demonstrated that loss of Id proteins in tumor endothelial cells affects the expression of several proangiogenic genes, including $\alpha_6\beta_4$ integrins, MMPs and fibroblast growth factor receptor-1 [2]. The Id proteins have been also shown to activate the expression of a 120-kDa gelatinase in breast epithelia which enhances the invasiveness of these cells [3]. These and other data support the idea that a relationship may exist between the Id proteins and the MMPs. Therefore, it would be interesting to see whether the Id protein activity correlates with the enzymatic activity of the MMPs. Here, we present the design and synthesis of fluorogenic probes for the detection of the MMP activity. Such substrates might be useful tools for MMP assays *in vitro* as well as for *in vivo* imaging. Beside their fluorescence properties, which should reflect the MMP activity, these substrates should also be able to enter the cells in order to correlate the fluorescence signal with the MMP localization. Before presenting our results, biology of the MMPs and cell-membrane penetration approaches will be briefly described

VII.1.1 The MMP Family

The MMP family consists of structurally and functionally related zinc-dependent endopeptidases which are classified into subfamilies based on their substrate preferences: gelatinases, collagenases, stromelysins, matrilysins, membrane-type MMPs and other MMPs. The gelatinase subfamily consists of pro-MMP-2 (gelatinase A, 72 kDa) and pro-MMP-9 (gelatinase B, 92 kDa), and of their active forms which are responsible for physiological degradation of basement membrane proteins such as gelatin, collagen, elastin and fibronectin [4]. The MMPs are secreted as inactive proenzymes (pro-MMPs) and require extracellular activation by protease-catalyzed cleavage of the N-terminal pro-peptide [5]. MMP activity is regulated at various levels, including gene expression, spatial localization, activation from the latent form and inhibition by tissue inhibitors of MMPs (TIMPs). Whereas MMP-2 appears to be

constitutively expressed by many cell types in culture, MMP-9 expression is induced by cytokines, growth factors and cell/stroma interactions [6]. The balance between the levels of activated MMPs and the free form of TIMPs determines the net MMP activity. The MMPs are expressed in low amounts by normal cells. They selectively degrade components of the extracellular matrix and such controlled degradation is essential in many physiological situations, including developmental tissue remodeling, angiogenesis, tissue repair (wound healing), and normal turnover of extracellular matrix (female reproductive cycle, bone remodeling, nerve growth, hair follicle cycling, immune response and apoptosis).

In pathological conditions, MMPs are overexpressed [7, 8] and involved in inflammation, degradation of bone, autoimmune disease, gastric ulcer [4, 6], vascular diseases [9] and in the invasive migration of cancer cells across the basement membrane, like in tumor metastasis [10]. Malignant tumors use MMPs for tumor growth and spreading. Therefore, MMPs present an interesting target not only for anti-tumor therapy but also for tumor diagnostics. Synthetic MMP inhibitors and substrates should lead to the development of a new generation of anticancer agents [11, 12].

VII.1.2 Cell-Penetrating Systems

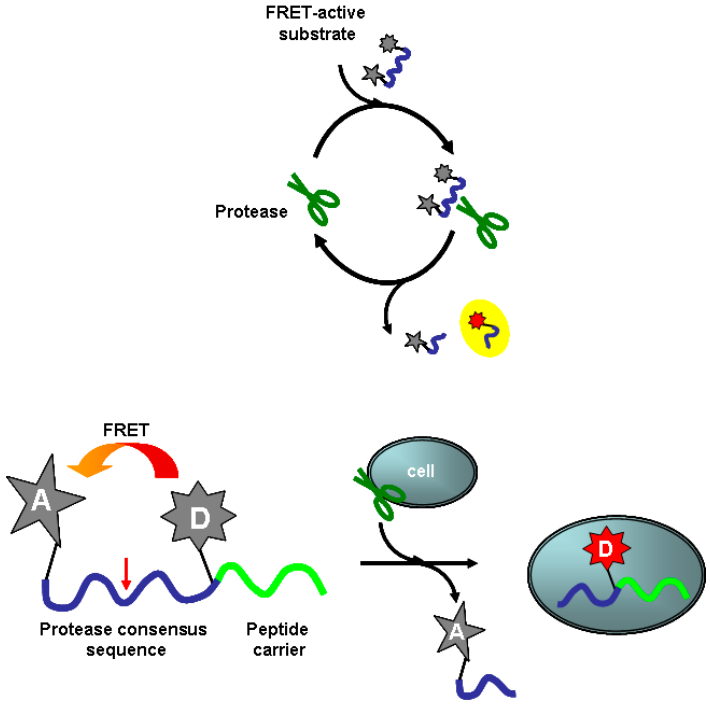
The conjugation of a membrane-permeable peptide vector (ca. 10 amino acid residues) with cargo molecules enables the transport of various bioactive molecules (oligopeptides, oligonucleotides, liposomes, phages, etc.) [13] which would be otherwise unable to translocate into the cells through the membranes and exert their functions [14]. One of the most typical peptide vectors is a short Arg-rich peptide segment derived from the human immunodeficiency virus (HIV)-1 Tat protein [15]. Various Arg-rich oligopeptides also possess very similar characteristics as delivery vectors [16]. No significant damage of the plasma membranes and little toxicity have been observed during the treatment of the cells with Arg-rich peptides [14]. Electrostatic interactions with the negatively charged glycosaminoglycans supposedly concentrate these peptides on the cell surface, and this would eventually accelerate their cellular uptake [16]. The formation of bivalent hydrogen bonds between the Arg side chains and the phosphate groups is also proposed: the guanidinium moiety can form ideal hydrogen bonding structures with phospholipids or other bivalent anions present in the lipid bilayers [17]. Such hydrogen bonding formation and the highly basic nature of the guanidinium function ($pK_a \sim 12.5$) are important factors controlling the translocation of these peptides. Whereas the guanidinium head group of Arg is critical for the transport, the chirality

of the amino acid [18], the length of the side chain and the spacing and composition of the backbone can all be varied [19].

VII.2 Design and Synthesis of MMP-2 Substrates

Targeting of proteases for diagnostics offers some advantages with respect to the targeting of other classes of biomolecules like receptors and antigens, the most important one being the possibility of the signal amplification due to the fact that one single protease molecule can bind and hydrolyze a large number of substrate molecules, thus eliciting the sensitivity of the detection method. For example, in the case of a substrate bearing a FRET-active donor/acceptor (D/A) system which can be deactivated upon protease hydrolysis, the fluorescence signal due to the piece of substrate bearing the donor can be amplified by the iterative catalytic cycle, as shown in Scheme VII.1.

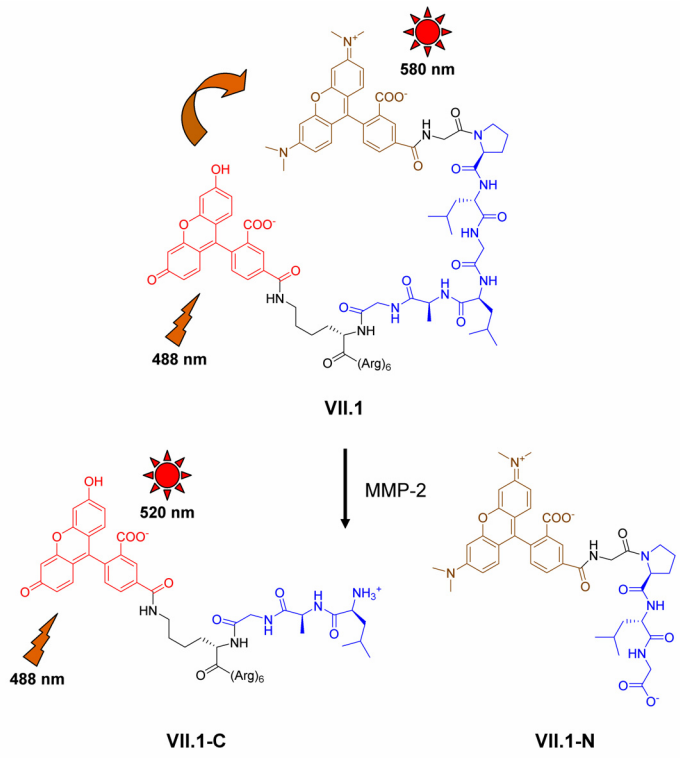
Scheme VII.1. Catalytic cycle of the hydrolysis of a FRET-active substrate (*top*), and schematic view of its proteolysis and cellular uptake (*bottom*).



We were interested in the targeting of MMPs, in particular of MMP-2, which are overexpressed in tumor cells together with the Id proteins. We designed a probe (**VII.1**) containing the D/A FRET pair 5-carboxyfluorescein (FAM)/5-carboxytetramethylrhodamine (TAMRA), the consensus sequence for MMP-2 (and, to a minor extent, for MMP-9) Pro-Leu-Gly*Leu-Ala-Gly (the asterisk represents the cleavage site), and a C-terminal tail with six Arg residues as

a peptide carrier to improve the penetration into the cells (Scheme VII.1, bottom). The acceptor was attached through an N-terminal Gly residue preceding the MMP-2 consensus sequence, whereas the donor was introduced through the side chain of a Lys residue following the consensus sequence. MMP-2 (and, to a minor extent, MMP-9) should be able to recognize the consensus sequence and cut the substrates in two pieces: the N-terminal part containing TAMRA (**VII.1-N**) and the C-terminal part containing FAM (**VII.1-C**) (Scheme VII.2). Before proteolysis, the presence of the two chromophores in the small peptide should provide a FRET-active system which can be "switched off" upon MMP-catalyzed hydrolysis.

Scheme VII.2. Chemical structure of the MMP-2 substrate **VII.1** and MMP-2-catalyzed hydrolysis.



The C-terminally amidated peptide chain of **VII.1** was assembled on Rink amide resin using the Fmoc chemistry. The side chain of the Lys residue was protected by Mtt group. Before removal of the Fmoc group from the N-terminal Gly, the Mtt group was selectively cleaved by treating the peptidyl-resin with 1% TFA in DCM in the presence of TIS. The free Lys side chain was then coupled to FAM. Fmoc cleavage and coupling of TAMRA followed (route *a*). Alternatively, the coupling of TAMRA to the unprotected N-terminal Gly was performed at first, followed by Mtt cleavage and coupling of FAM (route *b*). However, the route *a* provided a crude product with superior homogeneity, as shown by RP-HPLC profile (Figure VII.1).

The double-labeled peptide was cleaved from the resin and the Arg side chain protecting groups (Pbf) were simultaneously removed using a TFA cleavage mixture with 10% (v/v) scavengers (water/thioanisole/EDT) for three hours. Thioanisole was found to be more efficient than TIS in removing all six Pbf groups. The crude product was analyzed by RP-HPLC and MALDI-TOF mass spectrometry. The desired peptide was found as the major product (MS_{calcd}: 2418.8 Da, MS_{found}: 2419.0 Da). Minor side products were the peptide containing two FAM groups (MS_{calcd}: 2364.6 Da, MS_{found}: 2364.7 Da) or two TAMRA groups (MS_{calcd}: 2473.0 Da, MS_{found}: 2473.6 Da), and the TAMRA-mono-labeled one (MS_{calcd}: 2060.5 Da, MS_{found}: 2060.6 Da).

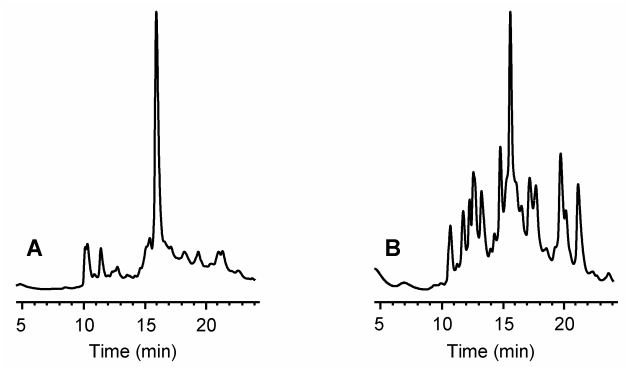


Figure VII.1. RP-HPLC profiles of the crude peptide **VII.1**, obtained by route *a* (A) or route *b* (B) (see text for details).

VII.3 Spectroscopic Properties of Compound VII.1

Compound **VII.1** was dissolved in Tris-HCl buffer (50 mM, pH 7.25) containing 5 mM $CaCl_2$ and 0.01 mM $ZnCl_2$. The UV spectrum of the solution showed two absorption maxima at 500 nm for FAM group and 560 nm for TAMRA group (Figure VII.2A). The fluorescence spectrum obtained with λ_{exc} = 488 nm was characterized by a large peak centered at 520 nm. A second peak was expected around 580 nm for the TAMRA emission as a consequence of FRET; however, no well-defined emission band was present in this part of the spectrum (Figure VII.2B, solid line). This could be due to *(a)* low intensity of TAMRA emission band with respect to that of FAM, *(b)* the absence of FRET, or *(c)* significant quenching of the TAMRA emission by FAM group. The first hypothesis is supported by the fact that the emission intensities of the intact probe (before MMP hydrolysis) around 580 nm after excitation at 550 nm and 488 nm were similar (Figure VII.2C-D, solid line); thus, detection of the TAMRA peak is not possible, as the emission band of TAMRA could be hidden under the FAM emis-

sion band. The second hypothesis can be ruled out: the small dimensions of the probe bring the two dyes close enough to allow FRET, unless they are unfavorably oriented. The donor and acceptor transition dipoles must be approximately parallel for an efficient FRET. In addition, the emission of FAM increased after proteolysis (Figure VII.2B), thus indicating that either FRET or the static fluorescence quenching had occurred. The latter would assume the reciprocal quenching of the donor and acceptor due to the fact that the two dyes form an intramolecular dimer. This has been reported for other peptide-based molecules bearing FAM and TAMRA and exhibiting low fluorescence intensity of both dyes [20-23]. Such dimerization leads to a ground-state complex and intramolecular dipole-dipole interaction within the dyes, which can be interpreted in terms of the exciton theory [24]. The formation of an intramolecular dimer between FAM and TAMRA is plausible in the substrate **VII.1**, as the consensus sequence for MMP-2 adopts a hairpin-like structure in solution [25] bringing the fluorescent dyes into close proximity. In addition, after MMP-catalyzed hydrolysis of **VII.1**, not only the intensity of FAM emission increased, but also that of TAMRA, as shown in Figure VII.2C-D. These results support the idea that a static (ground-state complex) rather than a dynamic (FRET) fluorescence quenching occurs in the intact probe **VII.1**.

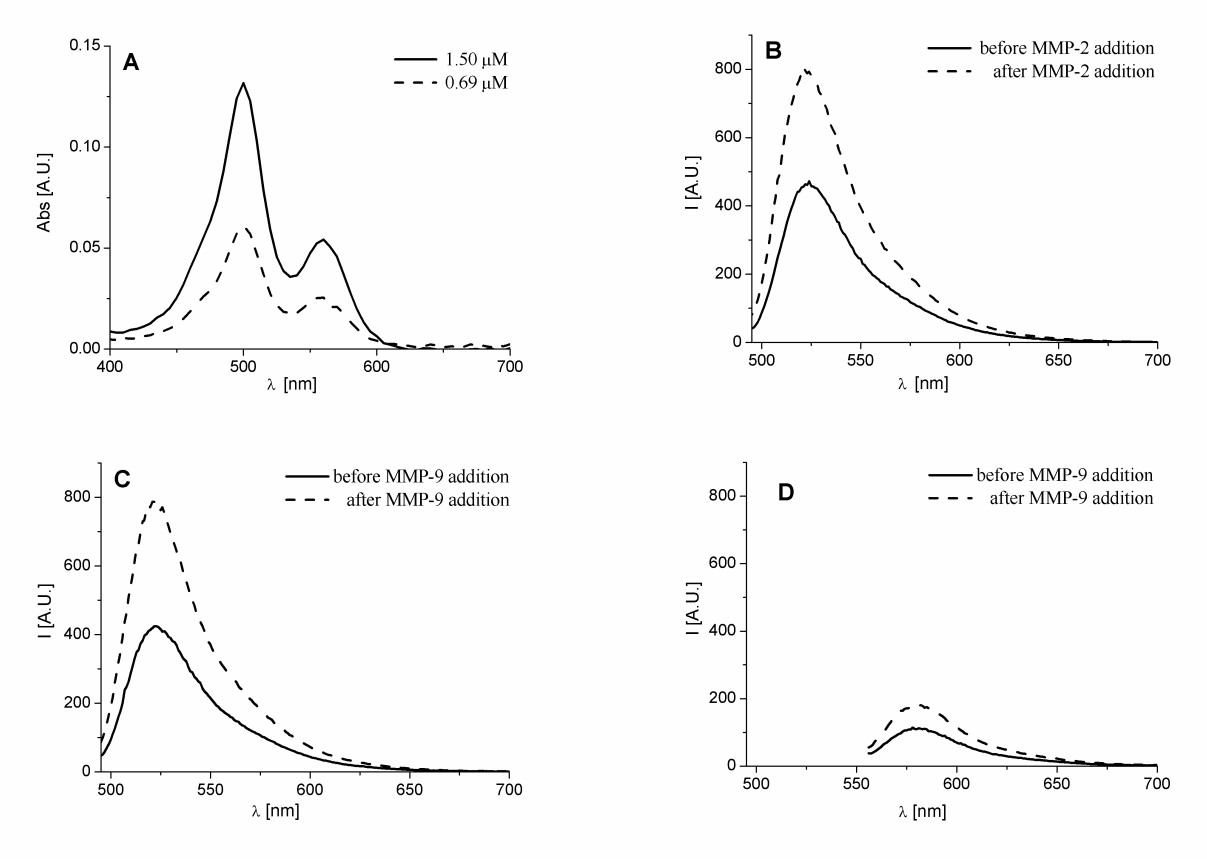


Figure VII.2. Spectroscopic properties of the substrate **VII.1**. (A) UV spectrum at two different concentrations; (B) emission spectrum at the peptide concentration of 0.69 µM (λ_{exc} = 488 nm) before and after MMP-2-mediated hydrolysis; (C) emission spectrum at the peptide concentration of 0.67 µM (λ_{exc} = 488 nm) before and after MMP-9-mediated hydrolysis; (D) emission spectrum at the peptide concentration of 0.67 µM (λ_{exc} = 550 nm) before and after MMP-9-mediated hydrolysis.

VII.4 Enzymatic Activity of MMP-2/MMP-9 in the Presence of Peptide VII.1

To investigate the proteolytic function of MMP-2 on compound **VII.1**, the latter was incubated with the active protease in the ratio 230:1 (**VII.1**:protease) and the emission spectra (λ_{exc} = 488 nm) were recorded over three hours (Figure VII.3). The fluorescence intensity at 520 nm increased until reaching an asymptotic value, indicating that the reaction ran to completion after about one hour. The hydrolysis products were characterized by MALDI-TOF mass spectrometry which confirmed the presence of the expected fragments **VII.1-N** and **VII.1-C**.

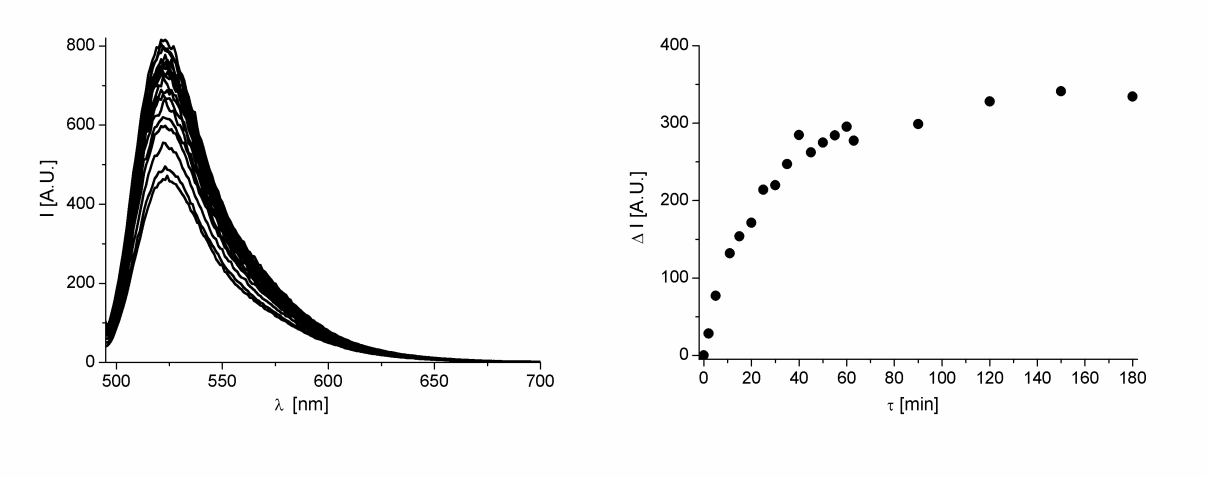


Figure VII.3. Enzymatic activity of MMP-2. Emission spectra of peptide **VII.1** at the concentration of 0.69 μM (λ_{exc} = 488 nm) recorded over three hours (*left*). Fluorescence intensity at 520 nm as a function of time: λ_{exc} = 488 nm; $\Delta I = I_{520}(\tau) - I_{520}(\tau_0)$, (*right*).

The same experiment was repeated using MMP-9 instead of MMP-2. However, it was clear that this enzymatic reaction was much slower than the one with MMP-2: only after three days, the emission curve of FAM at 520 nm reached an asymptote (Figure VII.4). Therefore, probe **VII.1** is a good MMP-2 substrate, but only a very poor one for MMP-9, as expected.

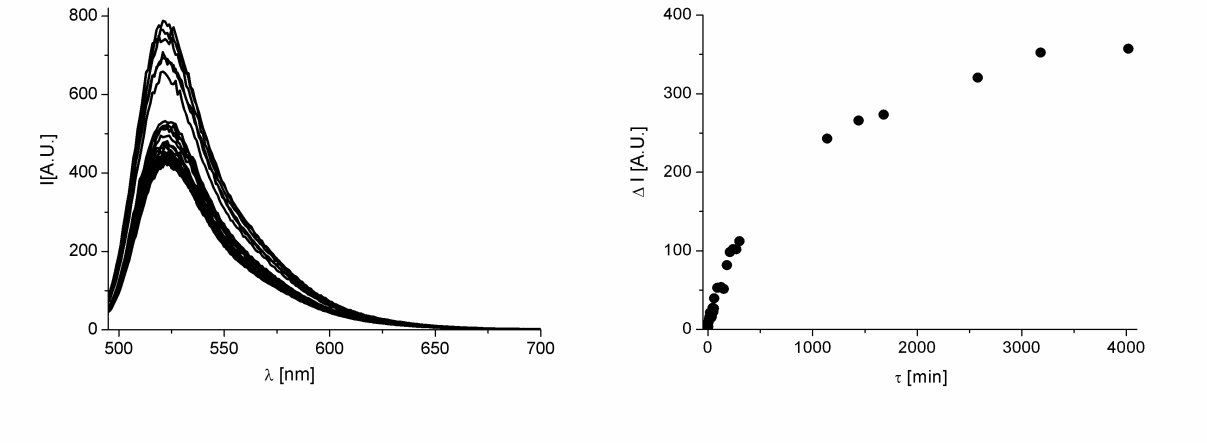
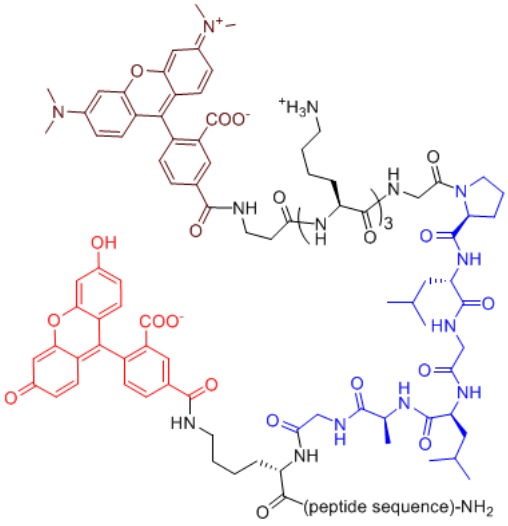


Figure VII.4. Enzymatic activity of MMP-9. Emission spectra of peptide **VII.1** at the concentration of 0.67 μM (λ_{exc} = 488 nm) recorded over three days (*left*). Fluorescence intensity at 520 nm as a function of time: λ_{exc} = 488 nm; $\Delta I = I_{520}(\tau) - I_{520}(\tau_0)$, (*right*).

VII.5 Design and Synthesis of MMP-2 Substrates Containing Fibrillogenic Domains

The probe **VII.1** described above has been designed to undergo an enzymatic process which leads to a change of the fluorescence properties of the conjugated chromophore (in this case to an increase in the FAM emission). The possibility to use the enzymatic reaction not only to affect the spectroscopic properties of a probe, but also to trigger simultaneously other events, such as conformational transitions, aggregation or even precipitation, was also studied. In the latter case, the enzymatic activity could be detected by the naked eye. An ideal probe should aggregate, but it should be possible to reduce or suppress this tendency. Thus, we decided to use peptides known for their tendency to form β-sheet aggregates, which can be however partially neutralized by the presence of positively charged regions: for example, aggressive precipitation of the Id2 protein fragment 103-124 can be prevented by conjugation with an N-terminal Lys-tag [26]. For this reason, we designed the peptide **VII.2**, in which the Id2 fragment 103-124 is linked to an N-terminal Lys-tag through the MMP-2 consensus sequence (Figure VII.5). The presence of the Lys-tag should avoid aggressive precipitation of the peptide; however, loss of the Lys-tag after MMP-2-catalyzed hydrolysis should cause peptide aggregation with consequent precipitation. We also designed another peptide based on the same principle (peptide **VII.3**). This time, instead of the Id2 fragment, we chose a short sequence from the islet amyloid polypeptide (NFGAIL), which was identified as the minimum length sequence sufficient to self-associate into β-sheet-containing amyloid fibrils [27, 28]. Further, we labeled these peptides with the FRET pair FAM/TAMRA as shown in Figure VII.5: if FRET occurs, the emission of FAM should be quenched before, but increased after hydrolysis. The FAM group is connected to the C-terminal fragment, which contains the fibrillogenic sequence influencing its solubility properties and causing the formation of high-order oligomers or even insoluble aggregates. This might lead to self-quenching of FAM groups located in close proximity due to aggregation. Therefore, a decrease of the FAM emission during enzymatic hydrolysis should be related to the association process of the fibrillogenic fragments.



VII.2 peptide sequence: LTTLNTDISILSLQASEFPSEL (103-124)-Id2

VII.3 peptide sequence: NFGAIL

Figure VII.5. Chemical structure of the MMP-2 substrates **VII.2-3**.

Peptides **VII.2-3** were prepared as described before for peptide **VII.1**, using the route *a* for the fluorescent labeling, and were then purified by preparative RP-HPLC (for details see the Experimental Part) and analyzed by MALDI-TOF mass spectrometry.

VII.6 Spectroscopic Properties of Compounds VII.2-3

Compounds **VII.2-3** were dissolved in Tris-HCl buffer (50 mM, pH 7.25) containing 5 mM $CaCl_2$ and 0.01 mM $ZnCl_2$ (10% DMSO was added to completely dissolve **VII.2**). The UV spectra of the two solutions showed two absorption maxima at 500 nm for FAM and 560 nm for TAMRA. The fluorescence spectra obtained with λ_{exc} = 488 nm were characterized by a large peak centered at 522 nm (Figure VII.6). As already discussed for peptide **VII.1**, the absence of a well-defined peak around 580 nm might be due to (*a*) low intensity of TAMRA emission band with respect to that of FAM, *(b)* lack of FRET, or *(c)* significant quenching of the TAMRA emission. The first hypothesis is supported as mentioned for substrate **VII.1**. The second hypothesis can be ruled out because the distance between the two dyes should certainly allow FRET in both substrates **VII.2** and **VII.3**. Intensities of the two dyes in the intact probe **VII.2** did not significantly vary from those found after hydrolysis, which excludes both FRET activity and/or reciprocal quenching of FAM and TAMRA groups (Figure VII.6A-B). This suggests that the two chromophores are not close enough to form an intramolecular dimer as in peptide **VII.1**, and, moreover, adopt unfavorable orientation for dynamic energy transfer. This also implies a reduced flexibility of the dye, which probably

reflects an oligomeric state of the probe **VII.2**, due to the presence of the fibrillogenic C-terminal domain. This can also explain the poor solubility of this peptide in Tris buffer. By contrast, the emissions of both dyes increased after hydrolysis of the probe **VII.3**, suggesting FRET activity and/or intramolecular quenching (Figure VII.6C-D).

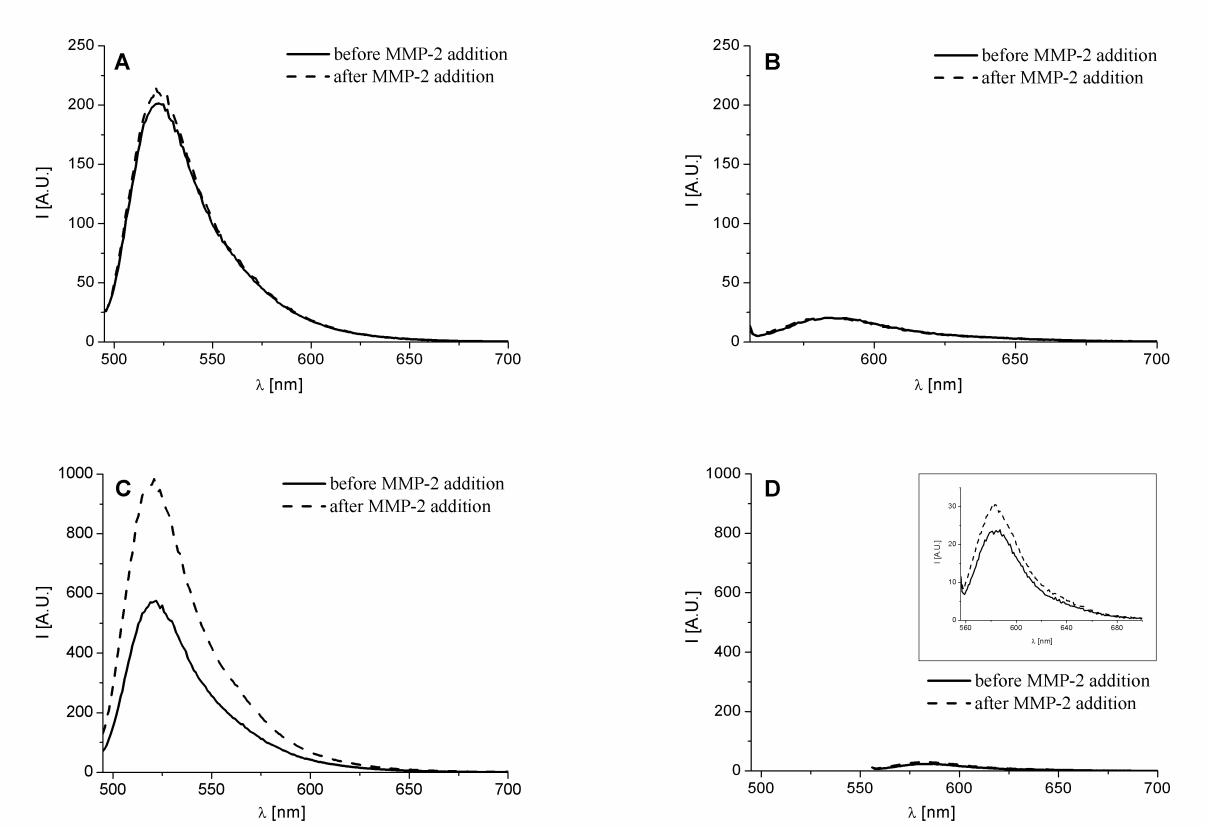


Figure VII.6. Emission spectrum of **VII.2** at the concentration of 0.7 µM before and after MMP-2-mediated hydrolysis: (A) λ_{exc} = 488 nm; (B) λ_{exc} = 550 nm. Emission spectrum of **VII.3** at the concentration of 0.6 µM before and after MMP-2-mediated hydrolysis: (C) λ_{exc} = 488 nm; (D) λ_{exc} = 550 nm.

VII.7 Enzymatic Activity of MMP-2 in the Presence of Peptides VII.2-3

To investigate the proteolytic function of MMP-2 on compounds **VII.2-3**, the latter were incubated with the active protease in the ratio 230:1 and 200:1, respectively, and the emission spectra (λ_{exc} = 488 nm) were recorded over three days (Figure VII.7). Both peptides exhibited a slight decrease in the intensity of FAM within the first hour (10% for **VII.2** and 5% for **VII.3**), followed by a recovery of the initial intensity after ~ 30 hours for **VII.2** and ~ 2 hours for **VII.3**. Whereas no significant changes were observed for the peptide **VII.2** in the following hours (Figure VII.7B), the emission of the peptide **VII.3** reached a ~ 70% increment after ~ 73 hours (Figure VII.7D).

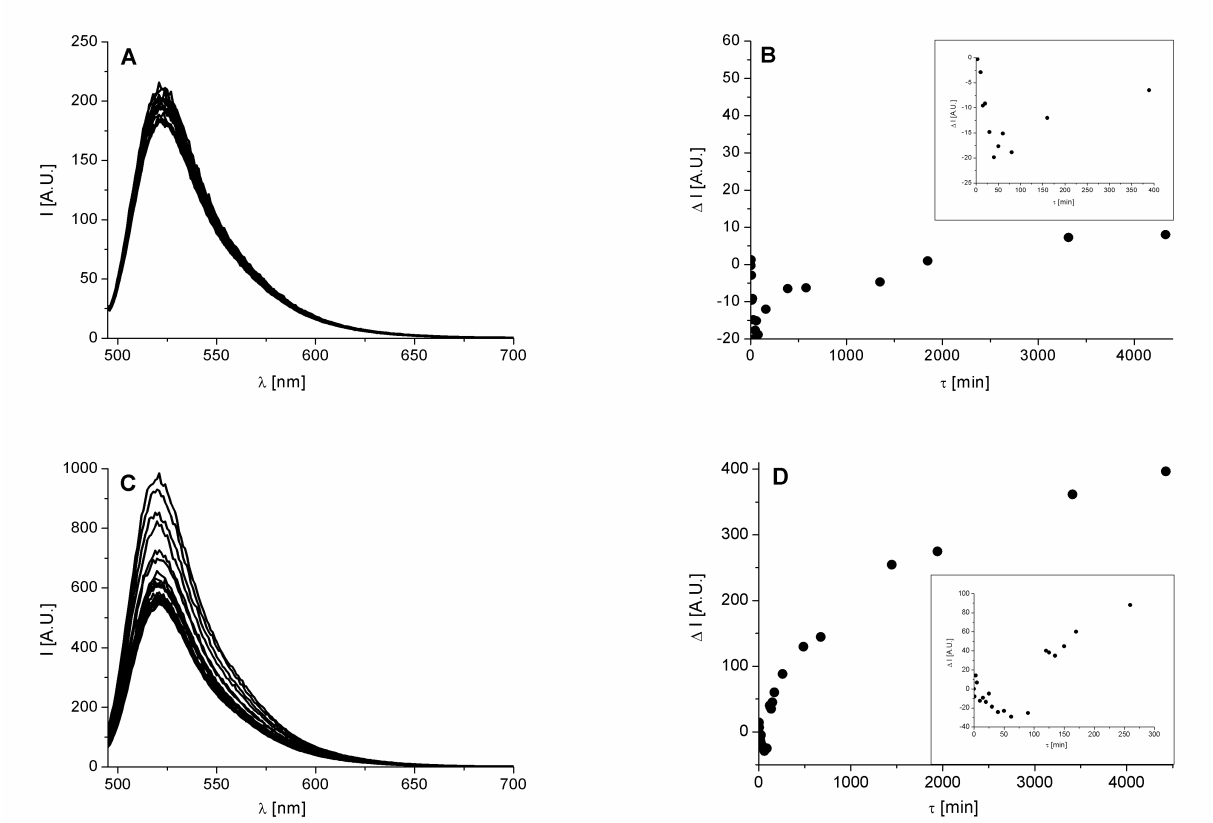


Figure VII.7. Enzymatic activity of MMP-2 on peptides **VII.2-3**. (A) Emission spectra of peptide **VII.2** at the concentration of 0.7 μM (λ_{exc} = 488 nm) recorded over 73 hours. (B) Fluorescence intensity of **VII.2** at 520 nm as a function of time (λ_{exc} = 488 nm; $\Delta I = I_{520}(\tau) - I_{520}(\tau_0)$). (C) Emission spectra of peptide **VII.3** at the concentration of 0.6 μM (λ_{exc} = 488 nm) recorded over 73 hours. (D) Fluorescence intensity of **VII.3** at 520 nm as a function of time (λ_{exc} = 488 nm; $\Delta I = I_{520}(\tau) - I_{520}(\tau_0)$).

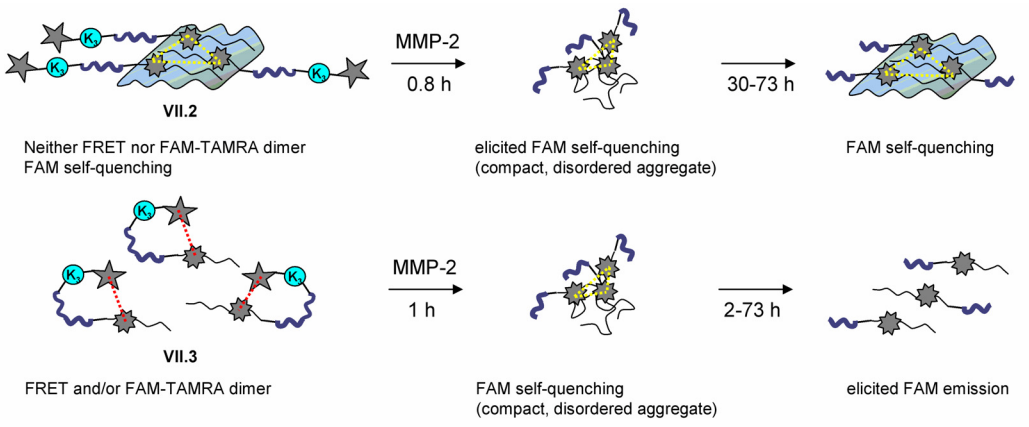
VII.8 Discussion and Conclusion

We have presented the synthesis of fluorogenic MMP-2 substrates containing the fluorescent labels FAM and TAMRA. The enzymatic activity was monitoring by the fluorescence of FAM group in Tris-HCl buffer solution containing the substrate **VII.1**, **VII.2** or **VII.3** and MMP-2 (or, for comparison, MMP-9) in the ratio 230:1 (peptides **VII.1** and **VII.2**) or 200:1 (peptide **VII.3**).

It is known that, when FAM and TAMRA are used together to label a small peptide, not only a dynamic (FRET), but also a static (ground-state complex) quenching between FAM and TAMRA may occur: in the first case, a decrease in the emission intensity of FAM should be accompanied by TAMRA emission, whereas in the second case both fluorophores will be quenched. Spectroscopic behavior of peptides **VII.1** and **VII.3** indicates the formation of a ground-state complex between the two dyes. On the other hand, peptide **VII.2** exhibited

neither a dynamic nor a static quenching of the dyes, probably due to their unfavorable orientation. Peptide **VII.1** was found to be a good fluorogenic substrate for MMP-2, and could be completely hydrolyzed within 1 hour giving rise to ~ 65% increase in the emission of FAM. In contrast, the two peptides containing a C-terminal fibrillogenic domain showed a slight decrease in the FAM emission within the first hour, which was after 73 hours accompanied by a complete recovery in the case of **VII.2** and by a ~ 70% increase of the FAM emission in the case of **VII.3**. An explanation of the behavior of the two peptides is given in Scheme VII.3: peptide **VII.2** supposedly aggregates under the used conditions, despite the presence of 10% DMSO which is necessary to completely dissolve the peptide. Thus, it is likely that the two dyes (especially FAM that is closer to the fibrillogenic sequence than TAMRA) suffer from reduced spatial freedom with consequent high probability to adopt an orientation unsuitable for energy transfer. However, FAM could undergo self-quenching due to aggregation. After MMP-2 hydrolysis, the N-terminal domain of **VII.2** is removed together with TAMRA group and three lysine residues. Therefore, the N-terminally truncated analogue is expected to loose some solubility that can cause a conformational change and give rise to more compact and larger aggregates, in which the FAM self-quenching could be elicited. These, however, are formed only transiently and convert into aggregates similar to those of the intact probe.

Scheme VII.3. Proposed model of MMP-2 hydrolysis of the fibrillogenic substrates **VII.2-3**.



By contrast, peptide **VII.3** is supposed to be a monomer in which FAM and TAMRA form an intramolecular dimer, similarly to peptide **VII.1**. After MMP-2 hydrolysis, the N-terminally truncated analogue probably forms transient oligomers with consequent FAM self-quenching, similarly to **VII.2**. Dissociation follows, leading to monomers characterized by elicited FAM emission with respect to the initial intact probe, as neither static nor dynamic quenching can take place any longer.

VII.9 Literature

[1] Lyden, D., Young, A. Z., Zagzag, D., Yan, W., Gerald, W., O'Reilly, R., Bader, B. L., Hynes, R. O., Zhuang, Y., Manova, K., and Benezra, R. (1999) Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature 401*, 670-7.

[2] Ruzinova, M. B., Schoer, R. A., Gerald, W., Egan, J. E., Pandolfi, P. P., Rafii, S., Manova, K., Mittal, V., and Benezra, R. (2003) Effect of angiogenesis inhibition by Id loss and the contribution of bone-marrow-derived endothelial cells in spontaneous murine tumors. *Cancer Cell 4*, 277-89.

[3] Desprez, P. Y., Lin, C. Q., Thomasset, N., Sympson, C. J., Bissell, M. J., and Campisi, J. (1998) A novel pathway for mammary epithelial cell invasion induced by the helix-loop-helix protein Id-1. *Mol Cell Biol 18*, 4577-88.

[4] Ganguly, K., Maity, P., Reiter, R. J., and Swarnakar, S. (2005) Effect of melatonin on secreted and induced matrix metalloproteinase-9 and -2 activity during prevention of indomethacin-induced gastric ulcer. *J Pineal Res 39*, 307-15.

[5] Tchougounova, E., Lundequist, A., Fajardo, I., Winberg, J. O., Abrink, M., and Pejler, G. (2005) A key role for mast cell chymase in the activation of pro-matrix metalloprotease-9 and pro-matrix metalloprotease-2. *J Biol Chem 280*, 9291-6.

[6] Swarnakar, S., Ganguly, K., Kundu, P., Banerjee, A., Maity, P., and Sharma, A. V. (2005) Curcumin regulates expression and activity of matrix metalloproteinases 9 and 2 during prevention and healing of indomethacin-induced gastric ulcer. *J Biol Chem 280*, 9409-15.

[7] Talvensaari-Mattila, A., Santala, M., Soini, Y., and Turpeenniemi-Hujanen, T. (2005) Prognostic value of matrix metalloproteinase-2 (MMP-2) expression in endometrial endometrioid adenocarcinoma. *Anticancer Res 25*, 4101-5.

[8] Talvensaari-Mattila, A., and Turpeenniemi-Hujanen, T. (2006) Matrix metalloproteinase 9 in the uterine cervix during tumor progression. *Int J Gynaecol Obstet 92*, 83-4.

[9] Kim, C. H., and Moon, S. K. (2005) Epigallocatechin-3-gallate causes the p21/WAF1-mediated G(1)-phase arrest of cell cycle and inhibits matrix metalloproteinase-9 expression in TNF-alpha-induced vascular smooth muscle cells. *Arch Biochem Biophys 435*, 264-72.

[10] Coussens, L. M., Raymond, W. W., Bergers, G., Laig-Webster, M., Behrendtsen, O., Werb, Z., Caughey, G. H., and Hanahan, D. (1999) Inflammatory mast cells up-regulate angiogenesis during squamous epithelial carcinogenesis. *Genes Dev 13*, 1382-97.

[11] Rossello, A., Nuti, E., Catalani, M. P., Carelli, P., Orlandini, E., Rapposelli, S., Tuccinardi, T., Atkinson, S. J., Murphy, G., and Balsamo, A. (2005) A new development of matrix metalloproteinase inhibitors: twin hydroxamic acids as potent inhibitors of MMPs. *Bioorg Med Chem Lett 15*, 2311-4.

[12] Rossello, A., Nuti, E., Carelli, P., Orlandini, E., Macchia, M., Nencetti, S., Zandomeneghi, M., Balzano, F., Uccello Barretta, G., Albini, A., Benelli, R., Cercignani, G., Murphy, G., and Balsamo, A. (2005) N-i-Propoxy-N-biphenylsulfonylaminobutylhydroxamic acids as potent and selective inhibitors of MMP-2 and MT1-MMP. *Bioorg Med Chem Lett 15*, 1321-6.

[13] Prive, G. G., and Melnick, A. (2006) Specific peptides for the therapeutic targeting of oncogenes. *Curr Opin Genet Dev 16*, 71-7.

[14] Futaki, S. (2006) Oligoarginine vectors for intracellular delivery: design and cellular-uptake mechanisms. *Biopolymers 84*, 241-9.

[15] Futaki, S., Nakase, I., Suzuki, T., Nameki, D., Kodama, E., Matsuoka, M., and Sugiura, Y. (2005) RNase S complex bearing arginine-rich peptide and anti-HIV activity. *J Mol Recognit 18*, 169-74.

[16] Futaki, S. (2005) Membrane-permeable arginine-rich peptides and the translocation mechanisms. *Adv Drug Deliv Rev 57*, 547-58.

[17] Sakai, N., Takeuchi, T., Futaki, S., and Matile, S. (2005) Direct observation of anion-mediated translocation of fluorescent oligoarginine carriers into and across bulk liquid and anionic bilayer membranes. *Chembiochem 6*, 114-22.

[18] Gammon, S. T., Villalobos, V. M., Prior, J. L., Sharma, V., and Piwnica-Worms, D. (2003) Quantitative analysis of permeation peptide complexes labeled with Technetium-99m: chiral and sequence-specific effects on net cell uptake. *Bioconjug Chem 14*, 368-76.

[19] Rothbard, J. B., Kreider, E., VanDeusen, C. L., Wright, L., Wylie, B. L., and Wender, P. A. (2002) Arginine-rich molecular transporters for drug delivery: role of backbone spacing in cellular uptake. *J Med Chem 45*, 3612-8.

[20] Law, B., Weissleder, R., and Tung, C. H. (2005) Mechanism-based fluorescent reporter for protein kinase A detection. *Chembiochem 6*, 1361-7.

[21] Packard, B. Z., Toptygin, D. D., Komoriya, A., and Brand, L. (1996) Profluorescent protease substrates: intramolecular dimers described by the exciton model. *Proc Natl Acad Sci USA 93*, 11640-5.

[22] Hoogerhout, P., Stittelaar, K. J., Brugghe, H. F., Timmermans, J. A., ten Hove, G. J., Jiskoot, W., Hoekman, J. H., and Roholl, P. J. (1999) Solid-phase synthesis and application of double-fluorescent-labeled lipopeptides, containing a CTL-epitope from the measles fusion protein. *J Pept Res 54*, 436-43.

[23] Wei, A. P., Blumenthal, D. K., and Herron, J. N. (1994) Antibody-mediated fluorescence enhancement based on shifting the intramolecular dimer<-->monomer equilibrium of fluorescent dyes. *Anal Chem 66*, 1500-6.

[24] Kasha, M. (1963) Energy transfer mechanisms and the molecular exciton model for molecular aggregates. *Radiat Res 20*, 55-70.

[25] Jiang, T., Olson, E. S., Nguyen, Q. T., Roy, M., Jennings, P. A., and Tsien, R. Y. (2004) Tumor imaging by means of proteolytic activation of cell-penetrating peptides. *Proc Natl Acad Sci USA 101*, 17867-72.

[26] Colombo, N., Schroeder, J., and Cabrele, C. (2006) A short Id2 protein fragment containing the nuclear export signal forms amyloid-like fibrils. *Biochem Biophys Res Commun 346*, 182-7.

[27] Tatarek-Nossol, M., Yan, L. M., Schmauder, A., Tenidis, K., Westermark, G., and Kapurniotu, A. (2005) Inhibition of hIAPP amyloid-fibril formation and apoptotic cell death by a designed hIAPP amyloid- core-containing hexapeptide. *Chem Biol 12*, 797-809.

[28] Kapurniotu, A., Schmauder, A., and Tenidis, K. (2002) Structure-based design and study of non-amyloidogenic, double N-methylated IAPP amyloid core sequences as inhibitors of IAPP amyloid formation and cytotoxicity. *J Mol Biol 315*, 339-50.

VIII. Summary

The Id3 protein is a member of the Id family of negative regulators of the bHLH transcription factors, especially of the ubiquitous E-proteins and the tissue-specific myogenin-regulating factor MyoD. By forming heterodimers with the bHLH proteins, Id3 prevents their binding to DNA and, consequently, the transcriptional activation. This eventually leads to increased cell proliferation and reduced cell differentiation. E47 and E12 preferentially associate with Id3 rather than with any other Id protein both *in vitro* and *in vivo*. In general, the Id proteins are overexpressed in proliferating, undifferentiated normal and cancer cells, while their expression level is low or absent in non-proliferating, terminally differentiated cells. The expression pattern of Id3 overlaps with that of Id1 during embryogenesis—both proteins are required for neurogenesis, angiogenesis and vascularization of tumor.

With 119 residues (13 kDa), Id3 is the smallest member of the Id protein family. The 41-residue long HLH motif (spanning the Id3 residues 41-81) is responsible for the interaction with the bHLH proteins. However, the N- and C-termini are also important for the Id3 function: for example, the N-terminus contains a consensus site for CDK2 which phosphorylates a serine residue at position 5 and thus modulates the Id3 affinity toward the E-proteins. Moreover, the N-terminal amino group is important for ubiquitination, whereas the C-terminal tail is critical for the full inhibitory activity of Id3 on the MCK reporter gene.

The Id3 protein has been shown to be a valuable target for solid-phase synthesis using Fmoc strategy; therefore, preparation of further chemically modified analogues of Id3 which enable investigation of its biophysical and conformational features is encouraged. An example thereof is the Id3 protein analogue which contains two amino acid substitutions in the loop region. Altering the position of a potential proline-induced kink of the backbone within the loop does not cause significant structural changes, and shows that this region of the Id3 protein is flexible enough to tolerate amino acid mutations. The Id3 protein and its analogue—obtained by chemical methods—spontaneously undergo self-association and form disulfide-bonded dimers. The value of $\Delta G^0_{H_2O}$ for the unfolding of the Id3 HLH domain was found to be smaller than those reported in the literature for the bHLH domains of the Id3 dimerization partners E47 and MyoD, which is in agreement with the observation that the Id3 forms less stable homodimers then the corresponding binding partners.

The helical character of the HLH motif of Id3 was found to be sensitive to amino acid replacement within the two helices. In particular, mutation of the two tyrosine residues—one

in helix-1 and the other in helix-2—which are conserved within the Id family, but not in the HLH family, has shown that the hydroxyl group of the tyrosine residue in helix-1 is not important for the conformation, whereas the tyrosine residue in helix-2 must be involved in hydrogen bonds stabilizing the Id3 HLH motif within the dimer. Also the cysteine residue in helix-1, which is conserved within the Id family, seems to be important for the helix stability, as removal of its side chain led to dramatic helix destabilization.

Whereas the Id3 HLH motif is highly helical, the regions preceding and following it do not adopt any preferred conformation as isolated fragments. However, they influence the α/β ratio of the HLH fold in an opposite manner, as the N-terminus favors the formation of β-sheet strands, while the C-terminus stabilizes the helix component. Phosphorylation of the N-terminus at position 5 (CDK2 phosphorylation site) is also likely to induce a conformational change, as shown by the enhanced β-strand propensity of the N-terminal fragment 1-81 after modification. Altogether, these results suggest that the N- and C-domains of the Id3 protein are critical for the folding of the central HLH motif and for its dimerization behavior. The intrinsic flexibility of the N- and C-tails is probably a structural prerequisite for the modulation of the protein fold during the dimerization processes with different bHLH transcription factors.

To further investigate how the Id3 HLH motif and its flanking regions are spatially arranged in solution, large Id3 protein fragments (namely 41-81, 21-81, 17-101, and 41-119) were fluorescently labeled by the introduction of the donor/acceptor pair tryptophan/dansyl at different positions along the sequence, and the efficiency of FRET which is related to the distance between the two fluorophores was examined. It should be pointed out that the Id HLH motif exists mostly as a dimer in the solution, similarly to all bHLH domains, in which the helix bundle is stabilized by an ideal juxtaposition of the helices. The negative impact of the fluorescent labeling of the N- and C-ends of the Id3 HLH peptide on its fold is likely to reflect an unfavorable helix packing as resulting from the increased steric hindrance at the periphery, which leads to a misfolded state. Interestingly, the labeling of the N-end of the Id3 HLH motif was tolerated in the case of the larger fragment 41-119, which spans both the HLH region and the entire C-terminus. This difference between the short HLH peptide and the C-terminally elongated analogue might be explained by the fact that the latter folded into a less compact structure, in which the N-terminal Dns groups could be easily accomodated. Contrarily to the fragment 41-119, the fragment 21-81 can be considered a misfolded state of the Id3 HLH motif, in which the number of α-helical and β-sheet residues was comparable. Analysis of the corresponding fluorescent analogues revealed no FRET between Trp-76 and

Lys(Dns)-24. This suggests that the N-terminal tail and helix-2 were far away from each other as a consequence of the loss of the HLH fold. In contrast to the unlabeled Id3 fragment 17-101 which displayed a misfolded state rich in β-strands, the corresponding mono- and double-labeled peptides (Trp-76 and Dns attached to the N-terminal residue Leu-17) were characterized by a number of helical residues similar to that found for the Id3 fragments 41-101, 41-119 and even the full-length Id3 protein, suggesting that the HLH fold was present. Therefore, they can be used as fluorescent tools for the study of a "miniaturized" Id3 protein shortened at its N- and C-tails. The FRET properties of the Id3 fragment 17-101 suggest that the flanking region N-terminal to the HLH motif is not extended but rather folded toward the HLH core.
In the case of the double-labeled Id3 fragment 41-119 having Trp at position 76 and the Dns group at the N-end, the distance between the donor and the acceptor was estimated to 20 Å. Considering the crystal structure of the DNA-bound MyoD dimer, this distance is slightly smaller than those measured (intra- and intermolecularly) between the N-end of the MyoD helix-1 and Tyr-156 of the helix-2 (corresponding to Id3 Tyr-76), which are ~ 26 Å and ~ 22 Å, respectively. The Id3 HLH dimer probably adopts a more compact structure than the DNA-binding MyoD dimer.

All Id3 polypeptides used for the conformational and fluorescence studies were prepared by stepwise SPPS. An alternative method would consist in the selective ligation of C-terminal peptide thioesters with N-terminal thiol-containing peptides. To evaluate the synthetic accessibility of peptide thioesters, we prepared a number of Id3 fragments to be converted into thioesters either in solution or on the solid support. Short (up to 16 residues), medium (up to 29 residues) or large portions (up to 46 residues) of the three Id3 subdomains (the HLH motif and the N- and C-termini) were prepared in order to test sequences which exhibited various difficulties. Three different thioesterification routes were applied, and the one performed on solid support with side chain anchored peptides was the most effective. Indeed, the reaction was fast and clean and the work-up was easy. The herein described peptide thioesters provide valuable intermediates for the future syntheses of Id3 protein related analogues.

The Id proteins are required for tumor angiogenesis and vascularization. The expression of $\alpha_v\beta_3$ integrin during tumor-induced angiogenesis, and the recruitment and/or expression of the matrix metalloproteinase MMP-2 required for tumor growth and metastasis may be directly or indirectly controlled by the Id proteins. We were interested in the development of fluorescent tools for the detection of the MMPs activity (in particular of MMP-2) which are

overexpressed in tumor cells together with the Id proteins. Herein we presented the synthesis of a fluorogenic MMP-2 substrate displaying the FRET pair FAM/TAMRA and a C-terminal oligo-arginine to favor cell membrane penetration. Additionally, two probes were designed, in which a fibrillogenic sequence (from the Id2 protein or the islet amyloid polypeptide) was linked to an N-terminal Lys-tag through the MMP-2 consensus sequence in order to use the enzymatic process not only to modulate the spectroscopic properties of the probe, but also to trigger other events, such as conformational transitions and aggregation. The presence of the Lys-tag should avoid aggressive precipitation of the aggregation-prone domain, while loss of the Lys-tag after MMP-2–catalyzed hydrolysis should accelerate the aggregation with consequent precipitation of the probe.

When FAM and TAMRA are used together to label a small peptide, not only a dynamic (FRET) but also a static (ground-state complex) quenching due to formation of an intramolecular FAM/TAMRA dimer can occur. The spectroscopic properties of the two substrates containing the oligo-arginine and the amyloid sequence, respectively, suggested the formation of a ground-state complex between the two dyes. In contrast, the substrate containing the Id2 fragment exhibited neither dynamic nor static quenching of the dyes, presumably due to their unfavorable orientation. The oligo-arginine-containing probe was found to be a good fluorogenic substrate for MMP-2: its complete hydrolysis within 1 hour was accompanied by ~ 65% increase in the FAM emission. In contrast, the two peptides containing a C-terminal fibrillogenic domain exhibited a slight decrease in the FAM emission within the first hour, which is likely to reflect the triggering of the aggregation process with consequent self-quenching of FAM.

In conclusion, we could show that (*i*) the SPPS approach using Fmoc chemistry can be successfully applied for the synthesis of the full-length Id3 protein as well as of truncated and modified analogues. (*ii*) The conformational properties of all synthesized Id3 polypeptides were studied by circular dichroism spectroscopy and revealed a complex interplay between the Id3 subdomains; moreover, (*iii*) the FRET properties of some fluorescence-labeled Id3 segments suggest that the N-terminus folds back to the compact HLH motif. These results contribute to the conformational characterization of this biologically highly relevant biomolecule and build a solid ground for future investigations, which may include the study of the structural arrangement of the Id3 C-terminus and of the role of individual Cys residues in the Id3 folding and dimerization processes.

IX. Experimental Part

IX.1 Materials

N^{α}-Fmoc-protected amino acids were purchased from MultiSynTech (Witten, Germany) and Novabiochem (Merck Biosciences GmbH, Schwalbach/Ts., Germany). Rink amide 4-methylbenzhydrylamine (MBHA) resin crosslinked with 1% DVB (loading 0.6 mmol/g, 100-200 mesh) and HBTU were from MultiSynTech. Wang resin (loading 1.1 mmol/g, 200-400 mesh), sulfamylbutyryl AM resin crosslinked with 1% DVB (loading 1.1 mmol/g, 200-400 mesh) were from Novabiochem. 2-Chlorotrityl chloride PS resin crosslinked with 1% DVB (loading 1.6 mmol/g, 200-400 mesh) was purchased from Senn Chemicals (Dielsdorf, Switzerland). HOBt, DIPEA, TFA, trimethylsilyldiazomethane, sodium thiophenolate, GnHCl, *N*-methylmorpholine, dansyl chloride, DCM, tetrakis(triphenylphosphine)palladium, phenylsilane, iodoacetonitrile, DMSO, SDS, DTT, α-cyano-4-hydroxycinnamic acid, and sinapinic acid were from Fluka (Taufkirchen, Germany).

The following peptide-synthesis-grade reagents were purchased from Biosolve (Valkenswaard, the Netherlands): piperidine, NMP, DMF and diethyl ether. HPLC-grade acetonitrile (ACN) and TFA for UV spectroscopy were from Biosolve. DIC, *HS*PAA and TFE were from Acros (Geel, Belgium). Sodium hydrogenphosphate and dihydrogenphosphate, acetic anhydride, acetic acid, EDT, TMSBr, *HS*Bn, thioanisole, zinc chloride, THF and hexane were obtained from Merck (Darmstadt, Germany). Methanol was purchased from J. T. Baker (Deventer, the Netherlands). *HS*PEA and TIS were from Aldrich (Steinheim, Germany). Ammonium carbonate was from AppliChem (Darmstadt, Germany). Tris was from USB Corporation (Cleveland, USA). Calcium chloride was obtained from Carl Roth (Karlsruhe, Germany). Boc anhydride was from Senn Chemicals. MMP-2 and MMP-9 were purchased from Merck Biosciences (Darmstadt, Germany). 5-FAM and 5-TAMRA were obtained from Novabiochem.

IX.2 Methods

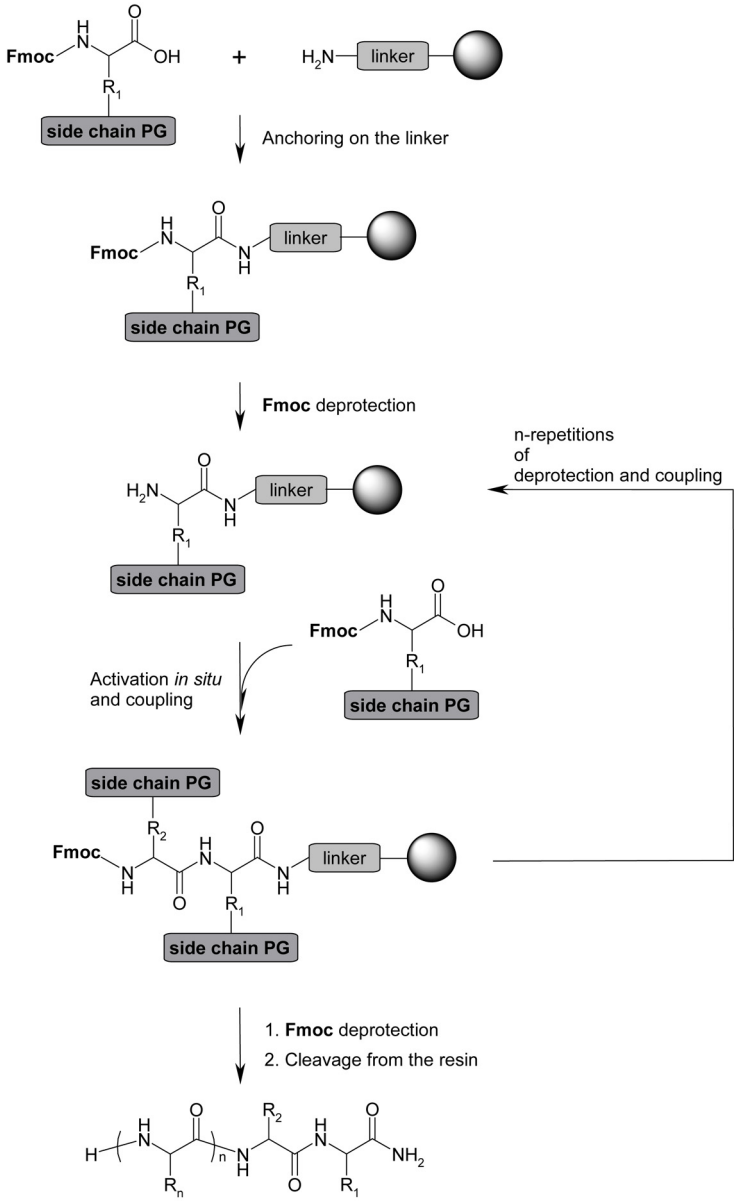
IX.2.1 Solid-Phase Peptide Synthesis (SPPS)

The SPPS methodology was developed by R. Bruce Merrifield [1] in the early 1960's, and has proven to be a powerful method for the preparation of peptides and small proteins. The concept of SPPS is based on the build-up of the peptide backbone from the C-terminal to

the N-terminal end, with the C-terminus anchored to a solid support through a linker. The chemistry used by Merrifield was based on the Boc strategy. To avoid the repetitive acidolysis required in Boc SPPS, new methods were developed. The most successful and extensively explored chemistry for SPPS utilizes the base-labile Fmoc group. In the Fmoc SPPS the initial N^{α}-Fmoc and, if required, side chain protected amino acid is loaded onto the insoluble support via an acid-labile linker [2, 3] (Scheme IX.1). The side chain protecting groups are also acid-labile, like the trityl, Boc and *t*-Bu groups. The support-bound Fmoc amino acid is deprotected by a base (usually piperidine). The second N^{α}-Fmoc amino acid is activated *in situ*. Several different activators may be utilized [4, 5]. The most widely used coupling method in Fmoc SPPS is the active ester method employing HOBt and its corresponding uronium salt analogue HBTU. The deprotection/coupling cycles are repeated till the sequence is complete. Finally, the crude product is cleaved from the solid support and simultaneously side chain deprotected by concentrated acid solution (TFA) to obtain the peptide as a free acid or amide, depending on the chemical nature of the linker.

In this work, for the preparation of peptide amides, the Rink amide resin has been used. As the MBHA linker of the Rink amide resin is easily acylated, no preloading is necessary and the resin can be directly used for the automated synthesis. Instead, the two resins applied in the synthesis of peptide acids, the 4-hydroxybenzyl (Wang) and 2-chlorotrityl chloride resins, and the resin used for peptide thioester formation, the sulfamylbutyryl resin, must be manually preloaded and capped before starting the automated peptide chain assembly. The Wang linker is labile in concentrated TFA (> 80%), whereas the trityl linker is labile already in the presence of 1% TFA. Therefore, the 2-chlorotrityl chloride resin can be used to prepare fully-protected peptide acids which may be further modified at the C-terminus in solution [6]. In contrast, the peptidyl-sulfonamide is completely stable in the presence of basic or strongly nucleophilic reagents. Alkylation of the sulfonamide by trimethylsilyldiazomethane or iodoacetonitrile is therefore required, followed by nucleophilic cleavage.

Scheme IX.1. General approach of the Fmoc SPPS.



Fmoc SPPS is primarily designed for acidolysis of the peptide-linker. General side chain protection by the *t*-Bu group (for Asp, Glu, Ser, Thr, and Tyr), Boc group (for Lys and Trp), and trityl group (for Arg, Asn, Gln, Cys, and His) allows simultaneous deprotection and peptide-linker cleavage by TFA. The cleavage of the *t*-Bu and Boc groups in TFA results in *t*-Bu cation and *t*-Bu-trifluoroacetate formation, which can lead to *tert*-butylation of Trp, Tyr and Met residues. These undesired modifications can be minimized by utilizing effective combinations of scavengers (including thiol compounds, water, phenol, TIS) [7].

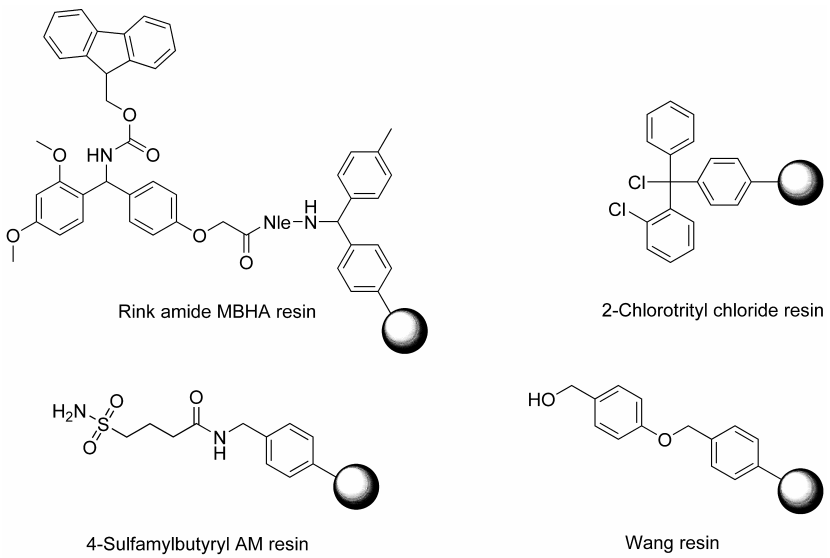


Figure IX.1. Chemical structures of the four different resin linkers used in this work for SPPS (the Rink amide MBHA resin was purchased as Fmoc-protected).

The Fmoc protecting group is highly labile to bases, especially to secondary amines (e.g. piperidine), which can remove the acidic proton of the fluorene ring system and cause the formation of a reactive dibenzofulvene. This reacts with the used secondary amine to form a stable adduct (e.g. fluorenylmethyl-piperidine).

IX.2.2 Circular Dichroism Spectroscopy

Circular dichroism [8] (CD) is the most widely used form of chiroptical spectroscopy which relies on different absorption of left- and right-circularly polarized light by chiral chromophores which either possess intrinsic chirality or are placed in a chiral environment. In practice, the plane-polarized radiation is split into two (left and right) circularly polarized components. If one of the components is absorbed by the sample to a greater extent than the other, the resultant radiation will be elliptically polarized. Dichroism is expressed as the difference in absorbance of the two components as a function of wavelength:

$$\Delta A = A_L - A_R = \varepsilon_L \times c \times l - \varepsilon_R \times c \times l = \Delta\varepsilon \times c \times l \qquad \text{Eq. IX.1}$$

where A_L and A_R are the absorbances for the left- and right-circularly polarized light, respectively, ε_L and ε_R are the molar extinction coefficients for the left- and right-circularly polarized light, respectively, $\Delta\varepsilon$ is the differential molar extinction coefficient ($M^{-1}cm^{-1}$), c is concentration (M) and l is the path length (cm).

The CD signal is also expressed as the ellipticity (Θ) in degrees. The numerical relationship between ΔA and Θ is:

$$\Theta = 32.98 \times \Delta A \qquad \text{Eq. IX.2}$$

For peptides and proteins, it is helpful to convert the ellipticity Θ to the mean-residue molar ellipticity $[\Theta]_R$ (deg cm^2 $dmol^{-1}$):

$$[\Theta]_R = \frac{100 \times \Theta}{l \times c \times n} \qquad \text{Eq. IX.3}$$

where n is the number of amino acid residues. A CD spectrum is obtained when the dichroism is recorded as a function of wavelength.

The CD spectroscopy is a valuable technique to study protein structure in solution [9]. In the far UV region (190-240 nm), the peptide bond is the main absorbing group, which can be used to estimate the content of secondary structural motifs including α-helix, β-sheet, unordered structure, PPII helix, and turns (Figure IX.2). The α-helix displays intensive CD amide bands attributed to the n→π* (negative ellipticity at 222 nm) and π→π* transitions, the latter is split into two bands of opposite sign (negative at 208 nm and positive at 193 nm) because of exciton coupling. The β-sheet exhibits a broad negative band near 218 nm (attributed to the n→π* transition) and a positive band near 195 nm (attributed to the π→π* transition). Unordered structure and the PPII helix, which is a component of unordered structures, have an intensive negative band near 200 nm (attributed to the π→π* transition) and the PPII helix has also an additional weak positive band near 225 nm (attributed to the n→π* transition).

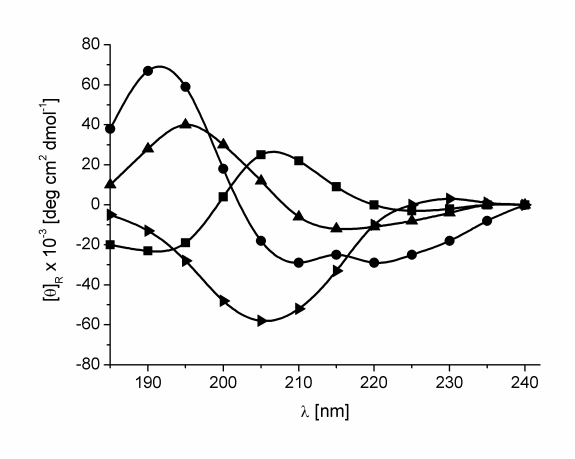


Figure IX.2. Characteristic CD spectra of polypeptides in the α-helical (●), β-sheet (▲), β-turn (■), and unordered conformation (►) [10].

The CD spectrum is a sum of the characteristic individual spectra arising from each type of secondary structure present in a protein sample. Typically, empirical analysis methods utilize a reference database [10-13]comprising spectra of proteins whose crystal structure (and therefore their secondary structures) are known and can provide an estimate of α-helix, β-sheet, β-turn and unordered structure content. In addition, the CD spectrum in the near UV (250-300 nm) reflects the environments of the aromatic amino acid side chains (Phe, Tyr, and Trp) and thus gives information about the tertiary structure.

Protein Folding

The CD spectroscopy can be used to monitor protein denaturation because it is a quantitative technique [8, 9]. Like other optical measurements, CD measurements obey the Beer's Law: the intensity of the CD signal of each component in the solution is directly proportional to its concentration, and the total spectrum represents the sum of all the contributing spectra. Thus, the change in the CD spectrum of a protein upon addition of a denaturant, ligand, heat, or pH change is directly proportional to the amount of protein changed by the perturbation. Following changes in the optical activity of a protein as a function of temperature or denaturant is a very convenient method to study the thermodynamics of folding. Indeed, the major contemporary application of CD is the determination of the thermodynamics of protein folding and the effects of mutations on protein stability.

IX.2.3 Fluorescence Steady-State Spectroscopy

Fluorescence energy transfer measurements [14] are used to estimate the average distances between sites on macromolecules. Measurements of the effects of conformational changes on these distances yield important structural information about the molecule.

Fluorescence resonance energy transfer (FRET) is the transfer of energy from an electronically-excited donor chromophore (D) to an acceptor chromophore (A) in the ground state through long-range dipole-dipole interactions without emission of a photon [15]. The acceptor absorbs energy at the emission wavelength of the donor, and if fluorescent, emission will be observed when the donor is excited (sensitized emission). The theory of energy transfer is based on the concept of fluorophore as an oscillating dipole, which can exchange energy with a second dipole having a similar resonance frequency [16]. In 1948, Förster proposed a theory for long-range molecular interactions by resonance energy transfer, which postulated that the rate of transfer depends on the inverse sixth power of the distance between the donor and the acceptor. Förster theory is based on very weak interactions between donor and acceptor, which should be in the distance range of 10 to 100 Å. The average D-A distance can be determined from the efficiency of energy transfer according to the following formula:

$$R_{DA} = R_0 \times \sqrt[6]{\frac{1}{E} - 1} \qquad \text{Eq. V.2}$$

where R_{DA} is the distance between the donor and the acceptor, and R_0 is the critical distance between the two fluorophores, called the Förster radius, at which the energy transfer efficiency (E) is 50%. E can be determined by steady-state measurements of the extent of donor quenching due to the acceptor

$$E = \left(1 - \frac{F_{DA}}{F_D}\right) \times 100\% \qquad \text{Eq. V.1}$$

where F_D and F_{DA} are relative average donor fluorescence intensities in the absence and presence of the acceptor, respectively. Typical R_0 values for different D-A pairs are reported in the literature [17]. Calculation of the Förster distance within a D-A pair depends upon the spectroscopic properties of D and A, including the transition dipole orientation factor κ^2 between D and A, the overlap integral between D emission and A absorption spectra $J(\lambda)$ (Figure IX.3), the refractive index of the medium n, and the quantum yield of the D (Q_D) in the

absence of A [18]. The greater the overlap of emission spectrum of D with the absorption spectrum of the A, the higher is the value of R_0.

$$R_0 = 9.78 \times 10^3 \times \sqrt[6]{\frac{\kappa^2 \times Q_D \times J(\lambda)}{n^4}} \quad \text{(in Å)} \qquad \text{Eq. IX.4}$$

Since the timescale of resonance energy transfer is in the order of nanoseconds, many processes that are time-averaged in other techniques can be resolved [17].

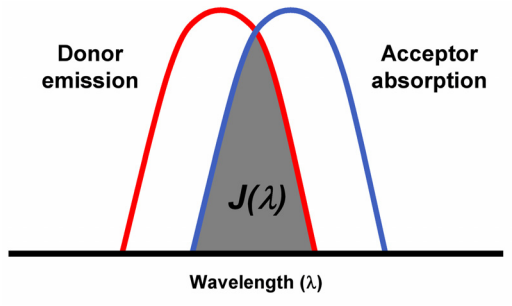


Figure IX.3. The spectral overlap between donor emission and acceptor absorption, defined as *J(λ)*.

IX.3 Standard Procedure for Polypeptide Synthesis

Peptide chain synthesis was performed at 0.018 mmol scale in 2 mL vessels using a *Syro-I* batchwise synthesizer (MultiSynTech, Witten, Germany), and applying Fmoc chemistry. The following side chain protecting groups were used: *t*-Bu for Asp, Glu, Ser, Thr, and Tyr; trityl for Cys, His, Asn, and Gln; Boc for Lys and Trp; 2,2,4,6,7-pentamethyl-dihydrobenzofurane-5-sulfonyl (Pbf) for Arg. The peptide amides were assembled on Rink amide MBHA resin (loading 0.6 mmol/g), whereas peptide acids were synthesized on Fmoc-His(trityl)-Wang resin (loading 0.6 mmol/g). Double couplings (40 min each) were performed by *in situ* activation of N^{α}-Fmoc-amino acids (5 eq) by HBTU/HOBt/DIPEA (5:5:10 eq) in DMF/NMP (70:30, v/v). Fmoc-Aib-OH, Fmoc-Lys(Mtt)-OH and Fmoc-*O*-benzylphosphoserine were also inserted into the sequence by these standard coupling conditions. The Fmoc group was cleaved by 40% piperidine (v/v) in a mixture of DMF/NMP (4:1, v/v) for 3 min, and then again with 20% piperidine for 10 min.

The N-terminally truncated analogues were acetylated at the N-terminus with acetic anhydride (10 eq) in the presence of DIPEA (10 eq) in DMF for 15 min. Finally, cleavage of the polypeptides from the resin with concomitant side chain deprotection was performed with TFA in the presence of 10% (v/v) scavengers (water/TIS/EDT 3:4:3, v/v). After 2.5 h, the

resin was filtered off and the crude products were precipitated with ice-cold diethyl ether and recovered by centrifugation at 3 °C for 8 min. Several ether washes/centrifugation cycles were carried out to efficiently remove the scavengers. When Met oxidation occurred, a reductive treatment by TFA/EDT/TMSBr solution (95:3:2, v/v) followed the cleavage of the Id3 polypeptides or whole proteins from the resin. After 1 hour, the reduced peptides were precipitated by ice-cold diethyl ether and recovered by centrifugation as mentioned above. The benzyl protecting group was cleaved from the side chain of the Ser residue simultaneously with the reduction of oxidized Met residues. The crude peptides were purified and analyzed by RP-HPLC (Figures IX.4-7) and mass spectrometry.

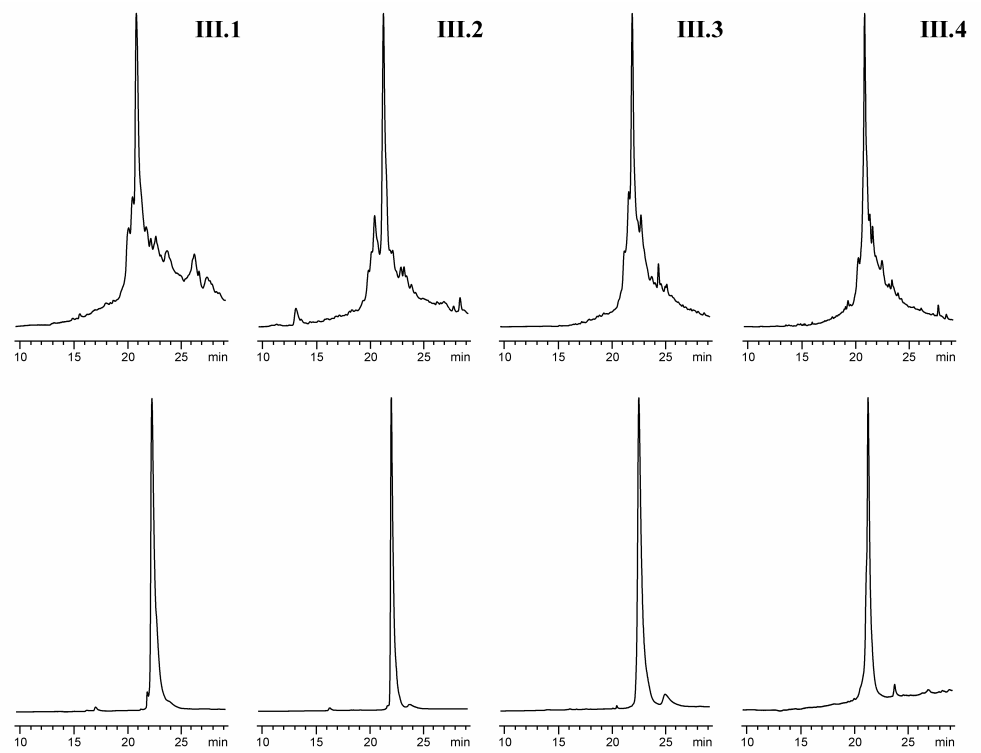


Figure IX.4. Analytical RP-HPLC profiles of crude (*top*) and purified (*bottom*) Id3 peptides **III.1-4**.

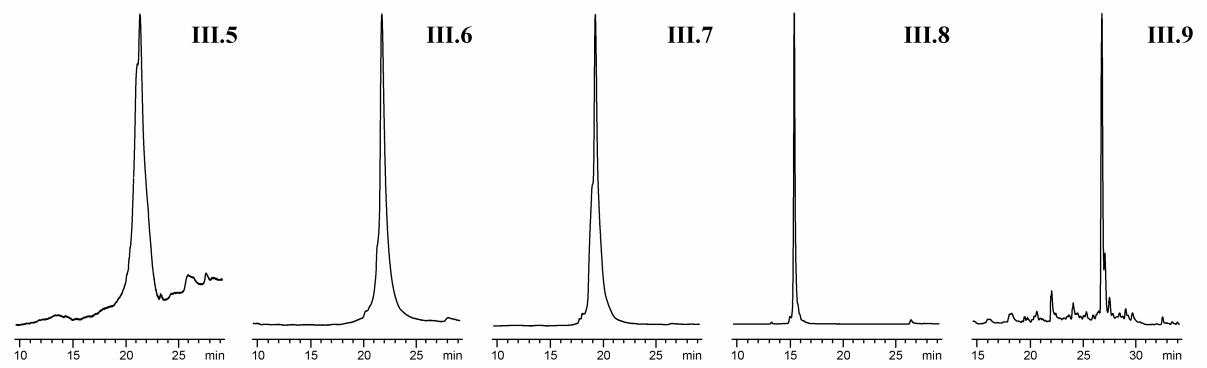


Figure IX.5. Analytical RP-HPLC profiles of purified peptides **III.5-7**, helix-1 (**III.8**) and crude helix-2 (**III.9**).

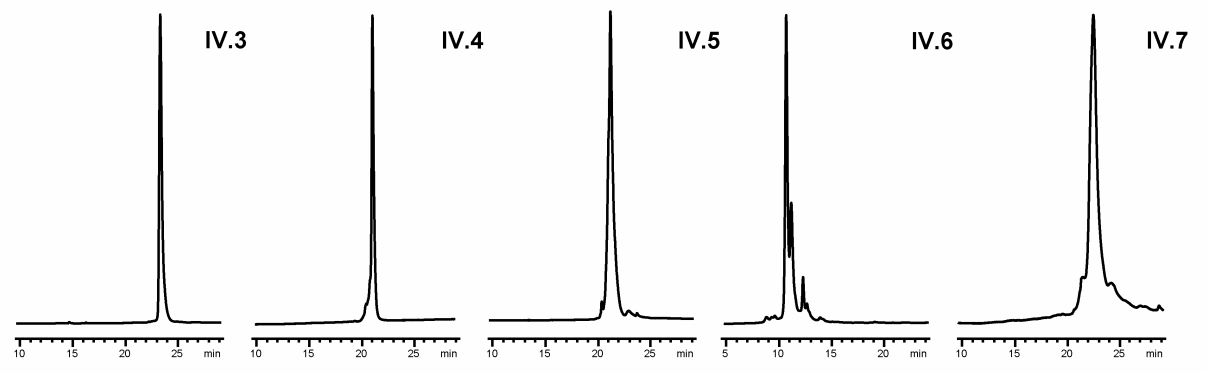


Figure IX.6. Analytical RP-HPLC profiles of purified Id3 peptides **IV.3-5** and **IV.7** and crude peptide **IV.6**.

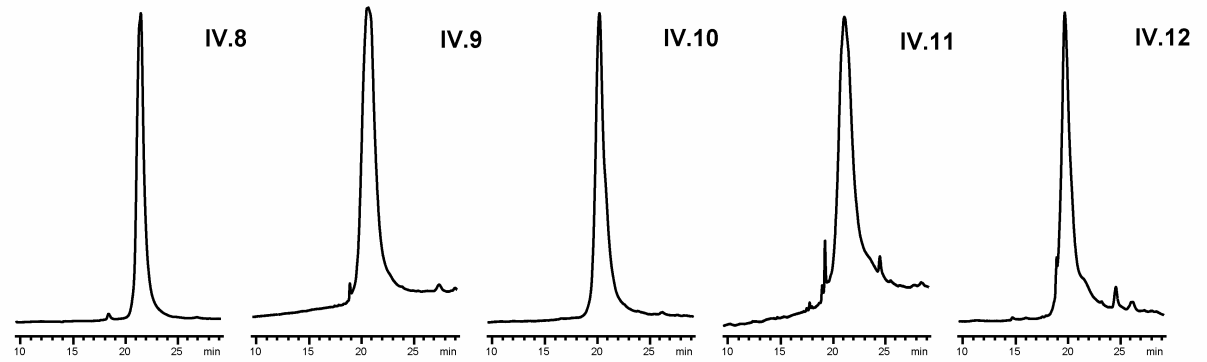


Figure IX.7. Analytical RP-HPLC profiles of purified Id3 peptides **IV.8-12**.

IX.3.1 Synthesis of the Full-Length Id3 Protein

Fmoc-His(trityl)-preloaded Wang resin (0.6 mmol/g) was prepared by treatment of pre-swelled Wang resin with Fmoc-His(trityl)-OH/HOBt/DIC (4 eq each) in DMF. After 30 min, DIPEA (4 eq) was added and the reaction was let run overnight. The resulting loading was determined by Fmoc cleavage with piperidine in DMF and UV measurement of the fluorenyl-methyl-piperidine adduct at 300 nm (ε = 7800 M^{-1} cm^{-1}) [19]. The remaining free hydroxyl groups of the Wang linker were acetylated with acetic anhydride (10 eq) in the presence of DIPEA (10 eq) in DMF.

Due to the increased swelling volume during the synthesis, the peptidyl-resins for **II.4**, **II.5** and **II.6** were split into two reactor vessels after about 60 acylation steps. The acylations after the resin splitting were accomplished by doubling the reagent excess. The HPLC profiles of the crude products of the native (**II.5**) and double mutated Id3 protein **II.6** were compared

(Figure IX.8A), both proteins were purified and characterized by MALDI-TOF mass spectrometry before (Figures IX.9-10) and after trypsin digestion. The ESI-MS method was used for the characterization of the four intermediate products **II.1-4** (Figures IX.11-14).

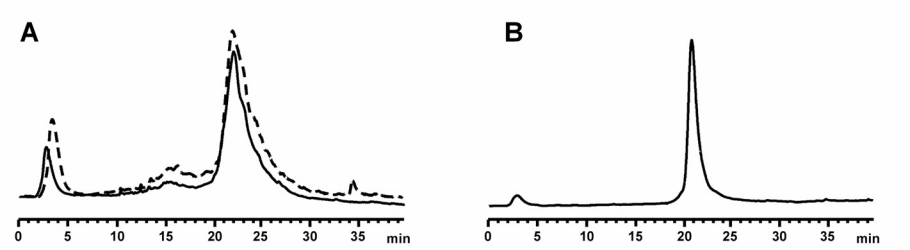


Figure IX.8. Analytical RP-HPLC profile of the Id3 double mutant **II.6**. (A) Comparison of the crude products of **II.6** (solid line) and **II.5** (dashed line). (B) Product **II.6** after preparative RP-HPLC purification.

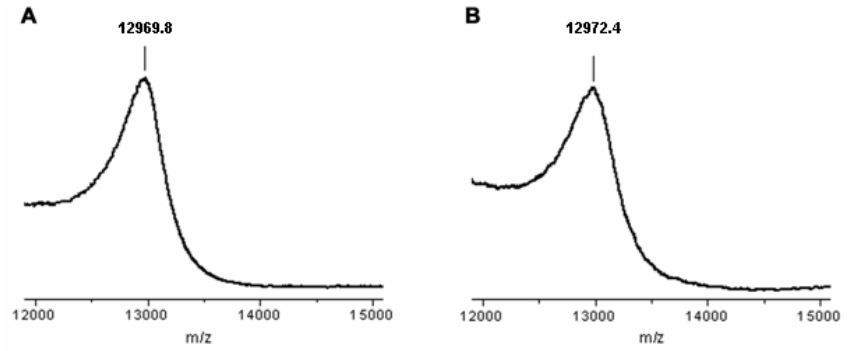


Figure IX.9. MALDI-TOF-MS spectra of two samples (A and B) of product **II.5** (average MW_{found}: 12971.1 ± 1.8 Da). The matrix α-cyano-4-hydroxycinnamic acid was used. The spectra were recorded on a mass spectrometer from GSG (Bruchsal, Germany).

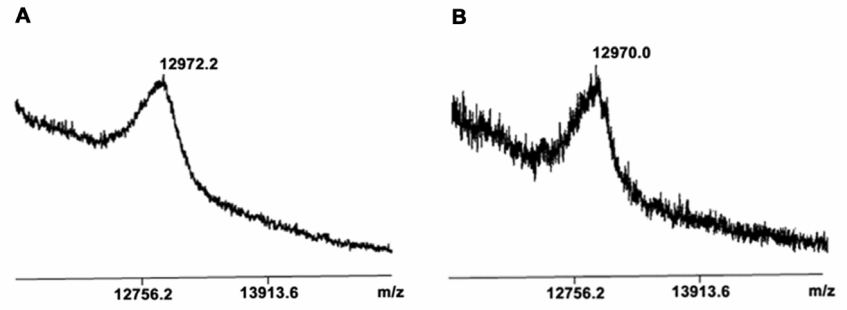


Figure IX.10. MALDI-TOF-MS spectra of two samples (A and B) of product **II.6** (average MW_{found}: 12971.1 ± 1.6 Da). The spectra were recorded on a mass spectrometer from Applied Biosystems.

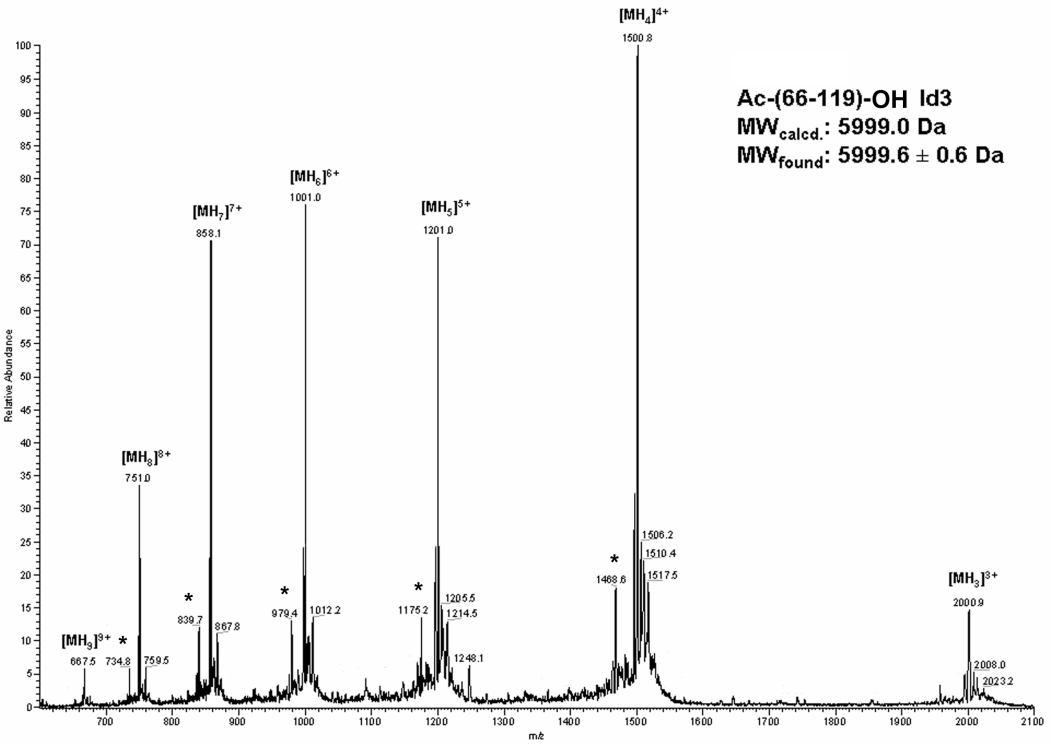


Figure IX.11. ESI-MS spectrum of peptide **II.1** (MW_{found} 5999.6 ± 0.6 Da was obtained from all the observed charge states). An impurity with molecular mass of 5870.6 ± 0.3 Da (129 Da, probably due to the absence of one glutamic acid residue) could be detected (*).

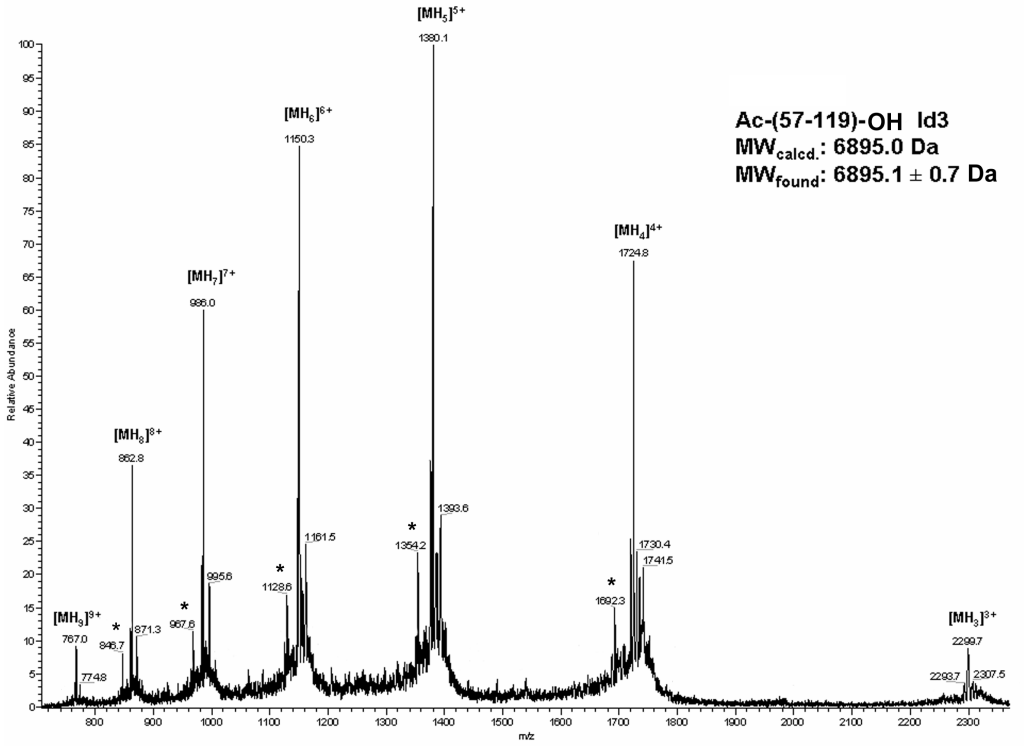


Figure IX.12. ESI-MS spectrum of peptide **II.2** (MW_{found} was obtained from all the observed charge states). An impurity with molecular mass of 6765.6 ± 0.5 Da (129 Da, probably due to the absence of one glutamic acid residue) could be detected (*).

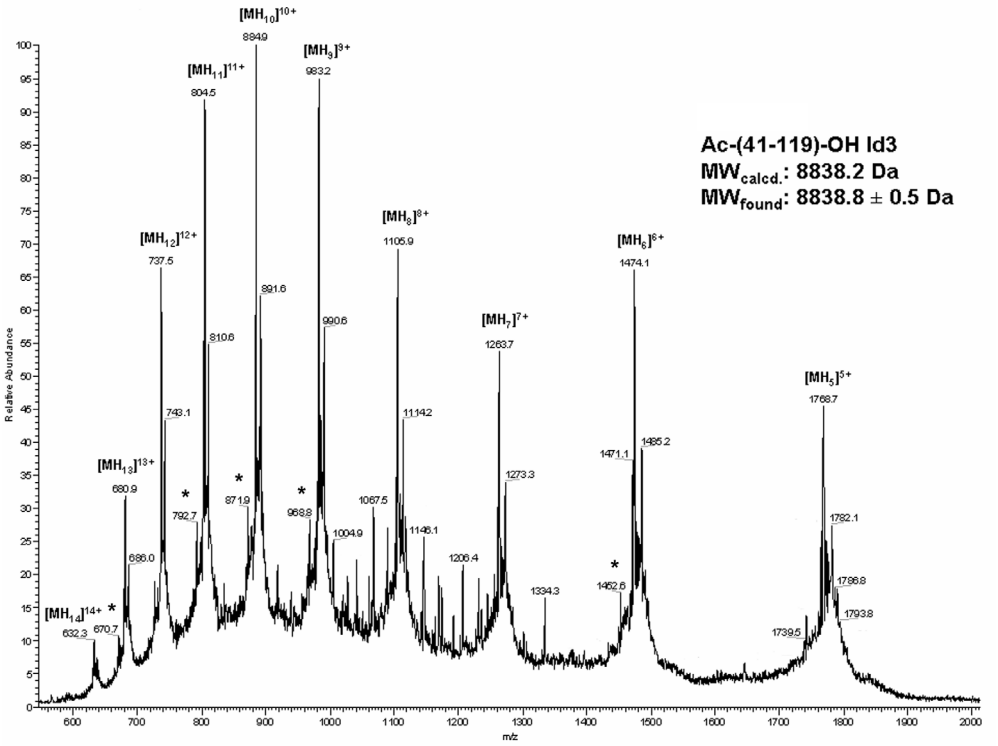


Figure IX.13. ESI-MS spectrum of peptide **II.3** (MW_{found} was obtained from all the observed charge states). An impurity with molecular mass of 8708.8 ± 1.6 Da (130 Da, probably due to the absence of one glutamic acid residue) could be detected (*).

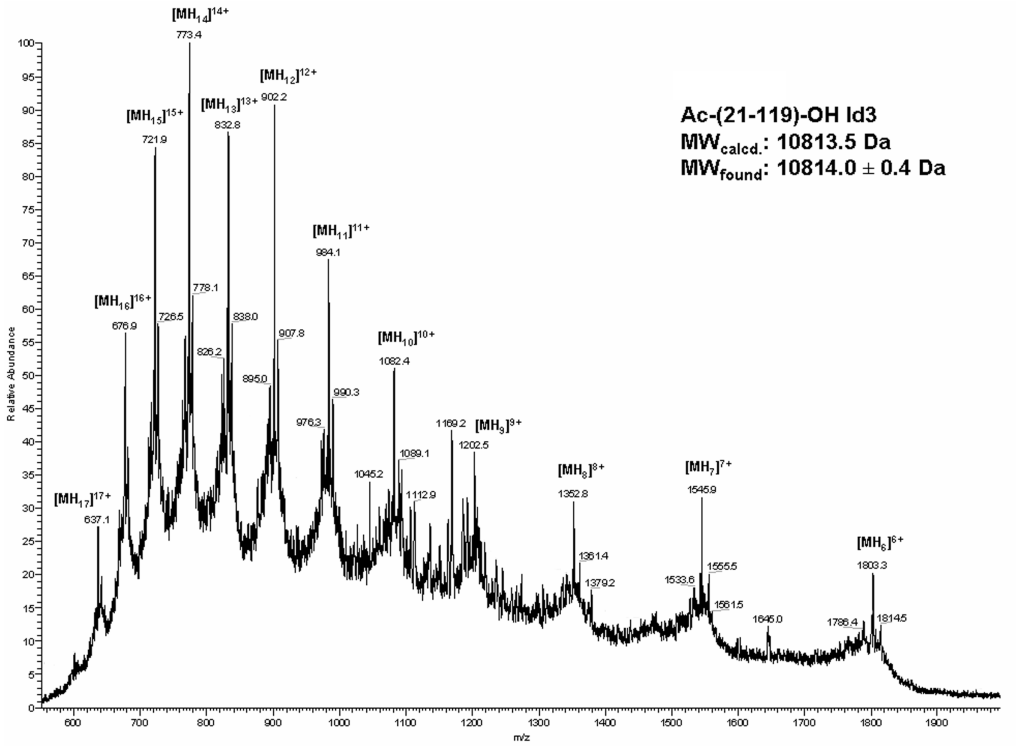


Figure IX.14. ESI-MS spectrum of peptide **II.4** (MW_{found} was obtained from all the observed charge states).

IX.3.2 Dansyl-Labeled Peptides

The dansyl group was connected either to the free N-terminus (**V.1**, **V.1b**, **V.3**, **V.4**, **V.5**, and **V.6**) or to the side chain of a Lys residue (**V.2**). In peptide **V.2**, the orthogonally protected lysine, (Fmoc-Lys(Mtt)-OH), was inserted instead of the natural Ile-24. When the synthesis of **V.2** was complete, the N-terminus was acetylated with acetic anhydride (10 eq) in the presence of DIPEA (10 eq) in DMF for 15 min, and the N^{ε}-4-methyltrityl (Mtt) protecting group was cleaved from the Lys-24 side chain by 1% TFA in dry DCM in the presence of 5% TIS in ten repetitive cycles (1 min each). After Mtt deprotection, the resin was washed with DIPEA to neutralize the Lys side chain group. The free amino group of the Lys side chain (peptide **V.2**) or of the N-terminus (**V.1**, **V.1b**, **V.3**, **V.4**, **V.5**, and **V.6**) was derivatized with dansyl chloride (5 eq) in the presence of DIPEA (5 eq) in DMF overnight. The reaction was checked in small scale by MALDI-TOF mass spectrometry. The labeled peptides were then cleaved from the resin and analyzed by RP-HPLC (Figure IX.15) and MALDI-TOF mass spectrometry.

IX.3.3 Labeling of MMP-2 Substrates

The Mtt-group was cleaved from the Lys side chain by 1% TFA in DCM in the presence of 5% TIS [20] in ten repetitive cycles (1 min each) before removal of the Fmoc group from the N-terminal Gly in peptide **VII.1**. The free amino group was then acylated with 5-FAM (5 eq) activated with DIC/HOBt (5 eq each) in DMF. The reaction ran for 30 min and its completion was checked qualitatively by the ninhydrin test, and quantitatively by performing a small scale TFA cleavage of the peptide from the resin and by HPLC and MALDI-TOF mass spectrometry analysis. Afterwards, the N-terminal Fmoc group was removed and the resulting free amino group of Gly was acylated with 5-TAMRA (3 eq) activated with DIC/HOBt (3 eq each) in DMF. The reaction was stopped after 40 min and the peptide was cleaved from the resin with concomitant side chain deprotection with the mixture of TFA and 10% (v/v) scavengers (water/thioanisole/EDT 2:4:4, v/v) for 3 hours. The crude peptide was precipitated from ice-cold diethyl ether and recovered by centrifugation at 3 °C for 8 min. Alternatively, the TAMRA group was first coupled to the unprotected N-terminus, followed

by Mtt cleavage from the Lys side chain and acylation by 5-FAM. However, the former strategy provided a crude product with superior homogeneity, as shown by the better HPLC profile. Labeled peptides **VII.2** and **VII.3** were thus prepared by the former strategy (Figures IX.16-17). The major products resulted to be the desired peptides with the expected molecular mass (MS_{calcd} for **II.2**: 4311.9 Da, MS_{found}: 4338.2 Da $(M+Na)^+$; MS_{calcd} for **II.3**: 2552.9 Da and MS_{found}: 2557.9 Da).

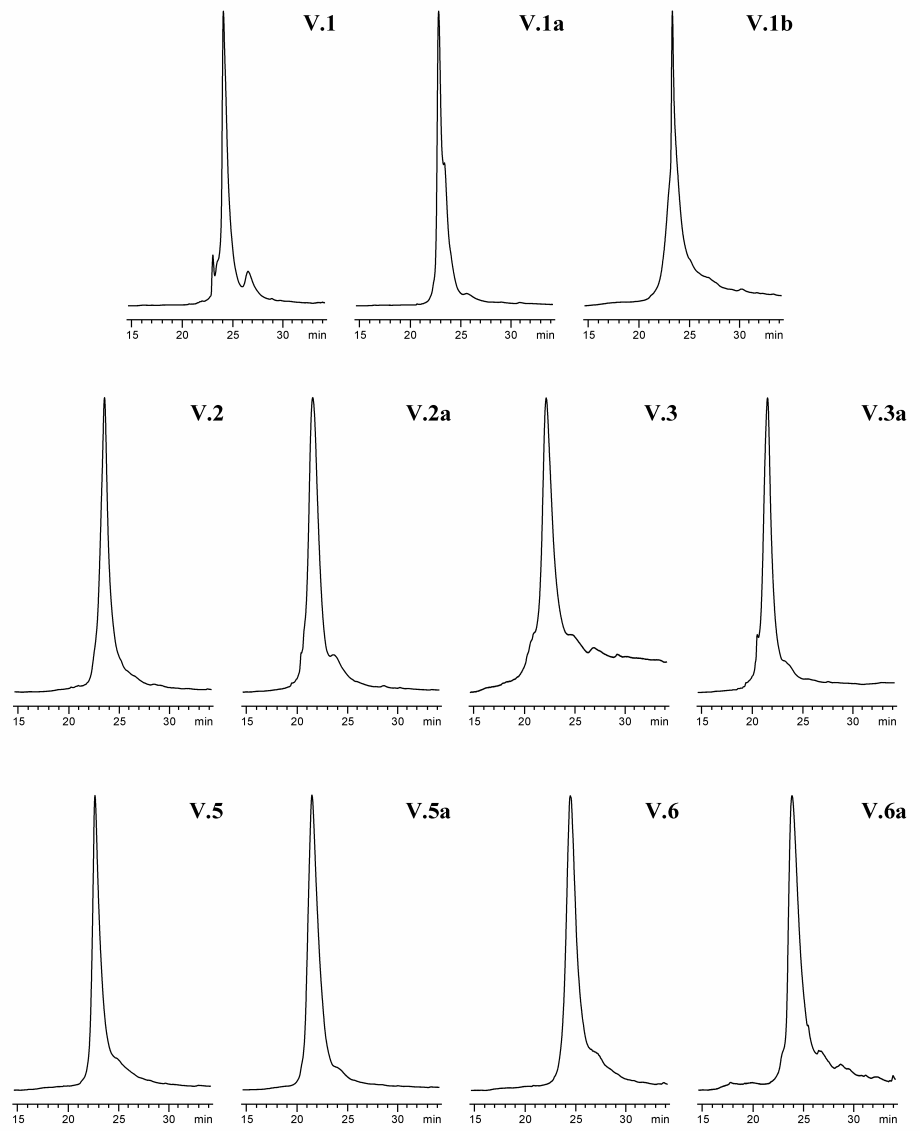


Figure IX.15. Analytical RP-HPLC profiles of dansyl-labeled purified peptides.

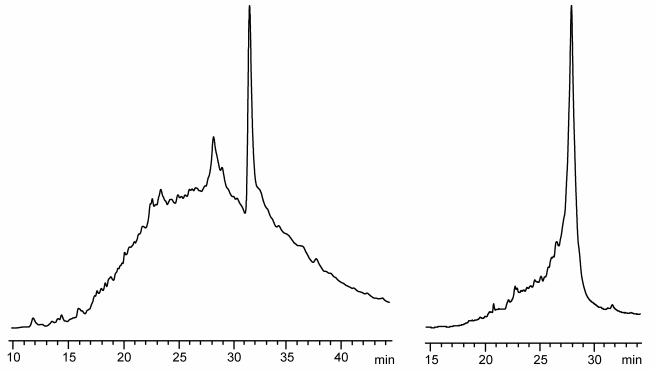


Figure IX.16. Peptide **VII.2** crude (*left*) and after preparative RP-HPLC (*right*).

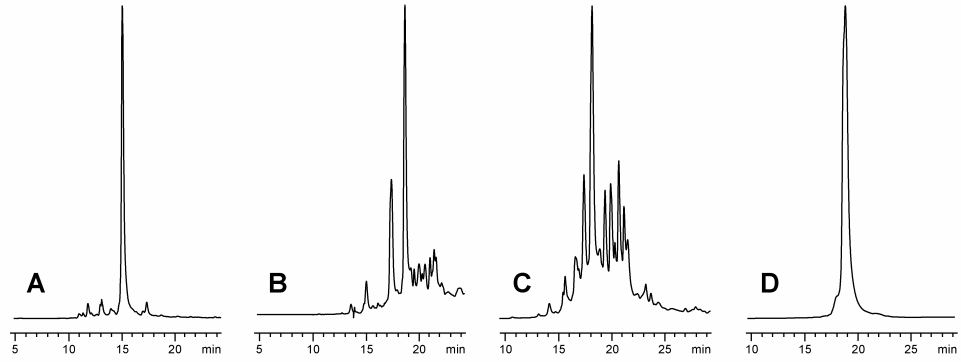


Figure IX.17. Peptide **VII.3**: unprotected and unlabeled (A), FAM mono-labeled crude (B), FAM/TAMRA double-labeled crude (C) and purified peptide **VII.3** (D).

IX.3.4 Synthesis of Id3 Peptide Thioesters

Thioester Formation in Solution

Fmoc-His(trityl)- and Fmoc-Pro-preloaded 2-chlorotrityl chloride resins (both 0.32 mmol/g) were prepared by treatment of the pre-swelled resin with Fmoc-His(Trt)-OH or Fmoc-Pro-OH (1 eq) and DIPEA (2 eq) in DMF for 2.5 hours. The resulting loading was determined by Fmoc cleavage with piperidine in DMF and UV measurement of the fluorenyl-methyl-piperidine adduct at 300 nm ($\varepsilon = 7800\ M^{-1}\ cm^{-1}$) [19]. The remaining chloride groups of the trityl linker were capped with the mixture DCM/MeOH/DIPEA 17:2:1 (v/v) in three repetitive cycles (10 min each).

When the synthesis was complete, the cleavage of the peptide from the resin was performed by 1% TFA with 1% TIS in dry DCM in approximately ten cycles (1 min each) to obtain the fully-protected peptide with the free C-terminal carboxylic acid. The presence of the peptide in the filtrates was checked by UV. The peptide-containing fractions were pooled

and the solvent was evaporated. The peptide was re-dissolved in a small amount of MeOH and precipitated by cold water. Thioesterification with HOBt/DIC/thiol (1.5:1.5:15 eq) in DCM followed. Three different thiols were used for the thioesterification: *HS*PAE, *HS*Bn and *HS*PAA. The reaction was monitored by analytical RP-HPLC and MALDI-TOF mass spectrometry of the fully-deprotected product which was obtained by treating small portions of the reaction mixture with TFA containing 10% (v/v) scavengers (water/TIS/EDT 3:4:3, v/v. When Cys and Met were absent in the sequence, EDT was omitted). After 2.5 h, the crude peptides were precipitated from ice-cold diethyl ether and recovered by centrifugation at 3 °C for 8 min. When the reaction was complete (after 2 days when *HS*PAA was used, after 4 days when *HS*Bn was used and after 10 days when *HS*PAE was used), DCM was evaporated, the peptides were re-dissolved in a small amount of MeOH and precipitated by cold water. The side chains were deprotected by TFA containing 10% (v/v) scavengers (water/TIS/EDT 3:4:3, v/v. When Cys and Met were absent in the sequence, EDT was omitted). After 2.5 h, the crude peptides were precipitated from ice-cold diethyl ether and recovered by centrifugation at 3 °C for 8 min. As the peptides **VI.1-3** contain air-sensitive amino acids (three Cys and two Met residues), they were further treated with TMSBr (TFA/EDT/TMSBr 95:3:2, v/v) and DTT (sulfhydryl reducing agent). The latter treatment proceeded with 15 eq of DTT in ammonium carbonate buffer (pH 8) at room temperature overnight.

Thioester Formation on Solid Support

Fmoc-Asp-OAll-preloaded Wang resin (0.55 mmol/g) was prepared by treatment of the pre-swelled Wang resin with Fmoc-Asp-OAll/HOBt/DIC (5 eq each) in DMF. After 30 min, DIPEA (5 eq) was added and the reaction was let run at room temperature overnight. The resulting loading was determined by Fmoc cleavage with piperidine in DMF and UV measurement of the fluorenylmethyl-piperidine adduct at 300 nm (ε = 7800 M^{-1} cm^{-1}) [19]. The remaining free hydroxyl groups of the Wang linker were acetylated with acetic anhydride (10 eq) in the presence of DIPEA (10 eq) in DMF. When the synthesis was complete the free N-terminal amino group was acetylated (peptides **VI.7-9** and **VI.12-13**) or Boc-protected (**VI.10-11**) by Boc anhydride (5 eq) in the presence of DIPEA (5 eq) in DCM. The reaction proceeded for 1 hour at room temperature.

The allyl group was deprotected by $Pd(PPh_3)_4$ (0.5 eq) with $PhSiH_3$ (20 eq) or *N*-methylmorpholine (20 eq) in DCM in five repetitive cycles (10 min each). The efficiency of the deprotection was analyzed by MALDI-TOF mass spectrometry using small scale TFA cleavage from the resin. The unprotected acid underwent thioesterification in a mixture of

HOBt/DIC/thiol (1.5:1.5:15 eq) in DCM on solid support. Three different thiols (*HS*PAE, *HS*Bn and *HS*PAA) were used for the thioesterification. After 18 hours, the polypeptides were cleaved from the resin with concomitant side chain deprotection by TFA containing 10% (v/v) scavengers (water/TIS/EDT 3:4:3, v/v). The Cys- and Met-free peptide thioesters were cleaved from the resin by an EDT-free TFA cleavage mixture containing 10% (v/v) scavengers (water/TIS 5:5, v/v). After 2.5 h, the crude peptides were precipitated from ice-cold diethyl ether and recovered by centrifugation at 3 °C for 8 min.

Thioester Formation Using Sulfamylbutyryl Resin

Fmoc-Gly-preloaded sulfamylbutyryl resin (0.5 mmol/g) was prepared by the treatment of pre-swelled sulfamylbutyryl resin with Fmoc-Gly-OH/PYBOP (4 eq each) and DIPEA (8 eq) in DMF overnight at room temperature. The resulting loading was determined by Fmoc cleavage with piperidine in DMF and UV measurement of the fluorenylmethyl-piperidine adduct at 300 nm ($\varepsilon = 7800\ M^{-1}\ cm^{-1}$) [19]. The remaining free sulfonamide groups of the linker were acetylated with acetic anhydride (10 eq) in the presence of DIPEA (10 eq) in DMF.

The Cys-47 residue was inserted into the sequence in the form of Fmoc-Cys(*t*-Buthio)-OH. The double coupling (2 h each) was performed using Fmoc-Cys(*t*-Buthio)-OH/HOBt/DIC (5 eq each) in DMF at room temperature. To check the quality of the growing chain, a small peptide portion was cleaved from the resin after the coupling of Tyr-48. Before cleavage, the N-terminus of Tyr-48 was acetylated by acetic anhydride (10 eq) in the presence of DIPEA (10 eq) in DMF, then the resin linker was alkylated by trimethylsilyldiazomethane (10 eq) in THF/hexane (1:1, v/v) for 2 h, and, finally, the peptide was cleaved from the resin by nuclephile (MeOH) in the presence of 10% (v/v) DIPEA at room temperature overnight. The solvent was evaporated by a stream of nitrogen and then the peptide methylester was deprotected by TFA containing 10% (v/v) scavengers (water/TIS 5:5, v/v). After 2.5 h, the crude peptide was precipitated from ice-cold diethyl ether and recovered by centrifugation at 3 °C for 8 min. The corresponding peptide methylesters were characterized by analytical RP-HPLC an MALDI-TOF mass spectrometry.

Once the synthesis was complete, the N-terminus of the peptidyl-resin was acetylated by acetic anhydride (10 eq) in the presence of DIPEA (10 eq) in DMF for 15 min and the resin was activated by ICH_2CN (35 eq) in the presence of DIPEA (15 eq) in 1 mL NMP. The reaction mixture was shaken in dark for 18 hours. The activated *N*-alkyl-*N*-acyl sulfonamide was cleaved by *HS*Bn (50 eq) in the presence of NaSPh in 700 µl of THF for 24 h, leading to the formation of the corresponding peptide benzylthioesters. The solvent was evaporated by

a stream of nitrogen, and then the peptide side chains were deprotected by TFA (1.5 mL) with 10% (v/v) scavengers (water/TIS/EDT 3:4:3, v/v). After 2.5 h, the crude peptides were precipitated from ice-cold diethyl ether and recovered by centrifugation at 3 °C for 8 min.

IX.4 Analytical and Preparative RP-HPLC

For purification, the crude products were dissolved in aqueous acetic acid (30%, v/v) or in the mixture with 50% (v/v) DMSO (peptide **VII.2**), and separated by preparative RP-HPLC using the Luna C18(2) column, 10 µm particle size, 250 × 21.2 mm (Phenomenex, Aschaffenburg, Germany). The binary gradient elution system was: (A) TFA (0.004%, v/v) in water, and (B) ACN. The following gradient was used: 15-65% B over 66.7 min at a flow rate of 21 mL/min. UV detection at 220 nm was used. Purity of the products after lyophilization was checked by analytical RP-HPLC using the Luna C18(2) column, 90 Å pore size, 3 µm particle size, 150 × 4.60 mm (Phenomenex). The samples were dissolved in 0.1% TFA (v/v) or in MeOH (peptide **VII.2**). The binary gradient elution system was (A) TFA (0.012%, v/v) in water and (B) TFA (0.01%, v/v) in ACN. The following gradient was used: 10% B for 3 min, then up to 70% B in 40 min at a flow rate of 1 mL/min. UV detection at 220 nm was used, except for the double-labeled, dark-red colored, peptide **VII.1**, which was analyzed by analytical RP-HPLC using the following gradient: 10-40% B over 30 min, with detection at 220 nm (peptide bond absorption), 480 nm (FAM absorption) and 550 nm (TAMRA absorption). The gradient 10-70% B over 30 min was used in RP-HPLC analysis for the labeled peptides **VII.2** and **VII.3**.

All purified peptides were obtained with purity higher than 90% (based on their analytical RP-HPLC profiles) and the overall yields were 8 to 16% (Chapter II).

IX.5 Mass Spectrometry

Molecular masses were recorded on a Future GSG spectrometer (Bruchsal, Germany) or on a mass spectrometer from Applied Biosystems for matrix-assisted laser desorption time-of-flight (MALDI-TOF) analysis, and on a ThermoQuest spectrometer (Finnigan) for electrospray ionization (ESI)**.** MALDI-TOF mass spectrometry samples were dissolved in 0.1% (v/v) TFA in water and mixed 1:1 (v/v) with a solution of α-cyano-4-hydroxycinnamic acid in MeOH/ACN (1:1, v/v).

Tryptic digestion of the full-length Id3 protein and its mutant was carried out overnight at 38 °C in ammonium carbonate buffer (50 mM, pH 8). After acidification by 0.1% (v/v)

TFA in water and mixing with the matrix solution, the resulting protein fragments were analyzed by MALDI-TOF mass spectrometry.

IX.6 SDS-PAGE

Id3 protein samples **II.5** and **II.6** (fresh, incubated overnight, air-oxidized, or crosslinked with glutaraldehyde) were diluted to a final concentration of 15 μM with Tris buffer (70 mM, pH 6.8) containing SDS (3%) with or without DTT (10 mM) (Figure IX.18). The solutions were heated to 95 °C for 3 min and loaded on a SDS-polyacrylamide gel (12.5%) with a stacking gel (4%). Staining was performed by Coomassie brilliant blue. The molecular weight marker was the peqGOLD Protein Marker 1.

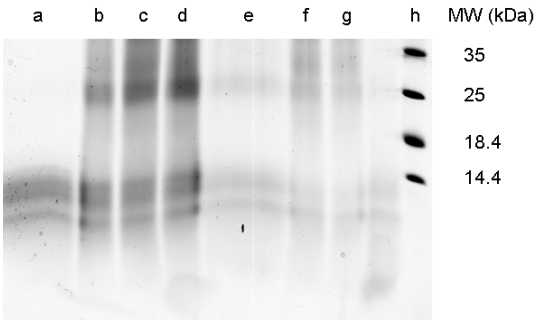


Figure IX.18. SDS-PAGE characterization of the homodimers formed by the Id3 protein **II.5**. The loaded samples correspond to the protein incubated overnight and then denatured in the presence (lane a) or in the absence (lane b) of DTT, and to the protein incubated overnight, air-oxidized and denatured in the absence of DTT (lanes c and d, with and without boiling, respectively). Lanes e-g correspond to the protein incubated overnight, crosslinked with glutaraldehyde, and then denatured in the presence (lane e) and in the absence (lanes f and g – the sample in g was not boiled) of DTT. The molecular weight marker was loaded in lane h.

IX.7 UV and CD Spectroscopy

All CD spectra were recorded on a Jasco J-710 CD spectropolarimeter (Groß-Umstadt, Germany) using a quartz cell of 0.02 cm path length (Hellma, Mühlheim, Germany). The peptides were dissolved in phosphate buffer (0.1 M; pH 7.25), in the mixture with TFE (15 or 30%), or in water, centrifuged, and the supernatant was analyzed by UV spectroscopy for the determination of the concentration on the basis of the absorption of the Tyr and/or Trp residues at 280 nm (ε = 1480 M^{-1} cm^{-1} for each Tyr [21], ε = 5540 M^{-1} cm^{-1} for each Trp [22]) on a Cary 100 Conc from Varian GmbH (Darmstadt, Germany). The UV cuvettes (1 mL and 0.5 mL with 1 cm path length) were purchased from Hellma. For the peptides **III.5** and **III.6** which did not contain any Tyr or Trp residue in the sequence, the concentration was

determined by weight considering a 70% peptide content. The CD samples were obtained by dilution of freshly prepared stock solutions of the peptides. For each spectrum, eight scans were accumulated using a step resolution of 0.2 nm, a bandwidth of 2 nm, a response time of 2 s, and a scan speed of 20 nm/min. The spectrum of the buffer was subtracted from that of the peptide to eliminate interferences from cell, solvent, and optical equipment. Noise reduction was obtained by a Fourier transform filter with the program Origin (OriginLab Corporation, Northampton, MA, USA). The ellipticity was expressed as mean-residue molar ellipticity. The helix percentage was estimated using the measured ellipticity at 222 nm and the predicted value for a 100% helix of *n* residues, which was calculated using the following equation [23, 24]:

$$\left[\Theta_{222}\right]_R = -37400 \times \left(1 - \frac{2.5}{n}\right) \qquad \text{Eq. IX.5}$$

In addition, composition of secondary structure elements was obtained by submitting the CD data to the Dichroweb online server [11, 25]. The best fit, characterized by the lowest RMSD value between the calculated and the experimental CD curves was given by the CONTIN algorithm [13, 26].

IX.7.1 Denaturation with Guanidinium Chloride

The experimental data obtained from the guanidinium chloride-induced unfolding, as monitored by CD at 222 nm, were fitted by a two-state model with Eq. IX.6, as described by Santoro and Bolen [27]:

$$\Theta_{obs} = \frac{\left(\Theta^0{}_N + m_N[D]\right) + \left(\Theta^0{}_U + m_U[D]\right) \times \exp\left\{-\frac{\Delta G^0{}_{H_2O} + m_G[D]}{RT}\right\}}{1 + \exp\left\{-\frac{\Delta G^0{}_{H_2O} + m_G[D]}{RT}\right\}} \qquad \text{Eq. IX.6}$$

where $\Theta^0{}_N$ and $\Theta^0{}_U$ are the values extrapolated to zero denaturant concentration [*D*] for the pre- and post-transitional baselines, respectively, m_N and m_U are the slopes of the baselines. $\Delta G^0{}_{H_2O}$ (free energy of unfolding in aqueous solution in the absence of denaturant) and m_G are the intercept and the slope, respectively, of the linear fit of the ΔG vs. [*D*]. Non-linear least-squares fitting of the experimental data was performed by MS Excel.

IX.7.2 Temperature-Dependent CD measurement

The CD spectra of peptide **IV.6** were collected from 5 to 45 °C with steps of 5 or 10 °C. For each spectrum, eight scans were accumulated using a step resolution of 0.2 nm, a band-width of 2 nm, a response time of 2 s, and a scan speed of 20 nm/min.

IX.8 Fluorescence Spectroscopy

IX.8.1 Fluorescence Spectroscopy of the Trp/Dns Id3 Peptides

The labeled polypeptides were dissolved in 0.1 M phosphate buffer (pH 7.25) or in the mixture thereof with 15% TFE, and the peptide concentration was determined from the UV absorbance of Trp and Tyr at 280 nm, as mentioned above. The 5 μM samples were obtained by dilution of the CD stock solutions prepared the same day. The emission spectra were recorded at 25 °C by excitation at 295 nm (λ_{exc} of Trp) on a Varian Cary Eclipse Fluorimeter. The efficiency of FRET and distances between Trp and Dns were calculated using Eq. V.1-2.

IX.8.2 Fluorescence Spectroscopy of the MMP-2 Substrates VII.1, VII.2 and VII.3

Compounds **VII.1** (dark-red colored), **VII.2** (dark-red colored) and **VII.3** (bright-orange colored) were dissolved in Tris-HCl buffer (50 mM, pH 7.25) containing 5 mM $CaCl_2$ and 0.01 mM $ZnCl_2$ (10% DMSO was added to completely dissolve **VII.2**) and the peptide concentration was determined from the UV absorbance of FAM at 495 nm (ε = 83000 M^{-1} cm^{-1} [28]) at 27 °C. The enzymatic activity was monitored by incubating the substrates **VII.1-3** and MMP-2 (or, for comparison, MMP-9) in the ratio 230:1 (peptides **VII.1** and **VII.2**) or 200:1 (peptide **VII.3**). The emission spectra were recorded at 27 °C by excitation at 488 nm (λ_{exc} of FAM) or 550 nm (λ_{exc} of TAMRA) in the absence and then in the presence of active MMP-2 (or, for comparison, MMP-9). The reaction was monitored for 3 hours (**VII.1** with MMP-2) or 3 days (**VII.1** with MMP-9, **VII.2** and **VII.3** with MMP-2) and the fluorescence emission spectra were collected every 5 min during the first hour, every 30 min during the next two hours, and then once a day (**VII.1** with MMP-9, **VII.2** and **VII.3** with MMP-2).

IX.9 Literature

[1] Merrifield, R. B. (1963) Solid phase peptide synthesis. I. The synthesis of a tetrapeptide. *J Am Chem Soc 85*, 2149-54.

[2] Sheppard, R. (2003) The fluorenylmethoxycarbonyl group in solid phase synthesis. *J Pept Sci 9*, 545-52.

[3] Fields, G. B., and Noble, R. L. (1990) Solid phase peptide synthesis utilizing 9-fluorenylmethoxycarbonyl amino acids. *Int J Pept Protein Res 35*, 161-214.

[4] Albericio, F. (2004) Developments in peptide and amide synthesis. *Curr Opin Chem Biol 8*, 211-21.

[5] Marder, O., and Albericio, F. (2003) Industrial application of coupling reagents in peptide synthesis. *Chimica Oggi 21*, 6-11.

[6] von Eggelkraut-Gottanka, R., A., K., Beck-Sickinger, A. G., and Beyermann, M. (2003) Peptide $^{\alpha}$thioester formation using standard Fmoc-chemistry. *Tetrahedron Lett 44*, 3551-4.

[7] Lundt, B. F., Johansen, N. L., Volund, A., and Markussen, J. (1978) Removal of t-butyl and t-butoxycarbonyl protecting groups with trifluoroacetic acid. Mechanisms, biproduct formation and evaluation of scavengers. *Int J Pept Protein Res 12*, 258-68.

[8] Greenfield, N. (1999) Applications of circular dichroism in protein and peptide analysis. *Trends Anal Chem 18*, 236-44.

[9] Kelly, S. M., and Price, N. C. (2000) The use of circular dichroism in the investigation of protein structure and function. *Curr Protein Pept Sci 1*, 349-84.

[10] Greenfield, N. J. (1996) Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal Biochem 235*, 1-10.

[11] Whitmore, L., and Wallace, B. A. (2004) DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res 32*, W668-73.

[12] Sreerama, N., Venyaminov, S. Y., and Woody, R. W. (1999) Estimation of the number of alpha-helical and beta-strand segments in proteins using circular dichroism spectroscopy. *Protein Sci 8*, 370-80.

[13] Provencher, S. W., and Glockner, J. (1981) Estimation of globular protein secondary structure from circular dichroism. *Biochemistry 20*, 33-7.

[14] Stryer, L. (1978) Fluorescence energy transfer as a spectroscopic ruler. *Annu Rev Biochem 47*, 819-46.

[15] Sapsford, K. E., Berti, L., and Medintz, I. L. (2006) Materials for fluorescence resonance energy transfer analysis: beyond traditional donor-acceptor combinations. *Angew Chem Int Ed 45*, 4562-89.

[16] http://www.olympusfluoview.com/applications/fretintro.html. 7th November 2006.

[17] Wu, P., and Brand, L. (1994) Resonance energy transfer: methods and applications. *Anal Biochem 218*, 1-13.

[18] dos Remedios, C. G., and Moens, P. D. (1995) Fluorescence resonance energy transfer spectroscopy is a reliable "ruler" for measuring structural changes in proteins. Dispelling the problem of the unknown orientation factor. *J Struct Biol 115*, 175-85.

[19] Meienhofer, J., Waki, M., Heimer, E. P., Lambros, T. J., Makofske, R. C., and Chang, C. D. (1979) Solid phase synthesis without repetitive acidolysis. Preparation of leucyl-alanyl-glycyl-valine using 9-fluorenylmethyloxycarbonylamino acids. *Int J Pept Protein Res 13*, 35-42.

[20] Hoogerhout, P., Stittelaar, K. J., Brugghe, H. F., Timmermans, J. A., ten Hove, G. J., Jiskoot, W., Hoekman, J. H., and Roholl, P. J. (1999) Solid-phase synthesis and application of double-fluorescent-labeled lipopeptides, containing a CTL-epitope from the measles fusion protein. *J Pept Res 54*, 436-43.

[21] Mach, H., Middaugh, C. R., and Lewis, R. V. (1992) Statistical determination of the average values of the extinction coefficients of tryptophan and tyrosine in native proteins. *Anal Biochem 200*, 74-80.

[22] Bettio, A., Dinger, M. C., and Beck-Sickinger, A. G. (2002) The neuropeptide Y monomer in solution is not folded in the pancreatic-polypeptide fold. *Protein Sci 11*, 1834-44.

[23] Chang, C. T., Wu, C. S., and Yang, J. T. (1978) Circular dichroic analysis of protein conformation: inclusion of the beta-turns. *Anal Biochem 91*, 13-31.

[24] Chen, Y. H., Yang, J. T., and Chau, K. H. (1974) Determination of the helix and beta form of proteins in aqueous solution by circular dichroism. *Biochemistry 13*, 3350-9.

[25] Lobley, A., Whitmore, L., and Wallace, B. A. (2002) DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics 18*, 211-2.

[26] van Stokkum, I. H., Spoelder, H. J., Bloemendal, M., van Grondelle, R., and Groen, F. C. (1990) Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal Biochem 191*, 110-8.

[27] Santoro, M. M., and Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants. *Biochemistry 27*, 8063-8.

[28] www.greiner-bio-one.co.jp/products/DNA/thermo_labeling.pdf. 5th November 2006.

Publications

- **Svobodová J**., Cabrele C. (2006) Stepwise solid-phase synthesis and spontaneous homo-dimerization of the helix-loop-helix protein Id3. *ChemBioChem 7*, 1164-1168.
- **Svobodová J**., Cabrele C. (2006) A synthetic and spectroscopic approach to study the conformational properties of the inhibitor of DNA binding Id3. *Peptide Sci*, 278.
- Cabrele C., Kiewitz S., **Havrdová J**., Colombo N. (2005) Studies on the chemical synthesis of the Id proteins and mutated analogs. *230th ACS National Meeting, Abstracts of Papers*.

Poster presentations

- **Havrdová J**., Cabrele C. (2005) Studies towards the synthesis and conformation of the Id3 protein. 7th German Peptide Symposium, Braunschweig (Germany)
- Colombo N., **Havrdová J**. and Cabrele C. (2004) Id Proteins. 2nd Summer School of Medicinal Chemistry, Regensburg (Germany)

Curriculum Vitae

Place of birth:	Turnov (Czech Republic), 27th May 1980
Education and studies:	
September 2004–now	PhD studies, **University of Regensburg**, Germany Group of Prof. A. Buschauer, Supervisor Dr. C. Cabrele
September 1998–June 2004	MSc. studies, **Institute of Chemical Technology**, Prague, Czech Republic Group of Prof. I. Stibor and Prof. P. Lhoták Diploma Thesis: "Design and Synthesis of New Receptors Based on Calix[4]arenes"
September 2003	Study stay, **Universidad de Salamanca**, Spain Group of Prof. J. Rodríguez Morán
July 2002 and July 2003	Internship, Pharmaceutical company **Synthon CZ** Supervisor Dr. J. Píš

Acknowledgements

- I would like to thank *Prof. A. Buschauer* who gave me the opportunity to study the PhD in Regensburg.

- I would especially like to thank *Dr. C. Cabrele* who offered me the chance to work in her group, gave me an interesting research project in the world of peptides, and supported its development at any time.

- I am grateful to *Prof. G. Bernhardt* and especially to *E. Hofinger* who helped me with the SDS-PAGE technique.

- Special thanks to *Prof. H. Tschochner* for the MALDI-TOF measurement of full-length Id3 protein, *Dr. Kiermaier and his team* for their assistance during MALDI-TOF measurements and for the LC-ESI analyses; *Dr. R. Vasold* and *Mr. E. Lautenschlager* for their assistance during analytical and preparative HPLC.

- Additionally, I thank the working group of *Prof. O. Wolfbeis* for the use of the circular dichroism spectrometer and the working group of *Prof. B. König* for the use of the fluorimeter.

- For the financial support, I am grateful to the DFG and DAAD.

- Special thanks to my colleagues *Dr. Noemi Colombo* and *Sebastian Kiewitz* for the nice time we spent together in the lab and for their friendship. I would like to thank *Roberto*, *Christian* and *Andrea* who bring a piece of Italy to our lab; *Giovanni* for the delicious Italian dinners; *Kristina* and *Alex* for the Romanian evenings with board games.

- I thank *Šárka* whose friendship never faltered; *Katka*, *Radek*, *Hanka*, *Jarek*, *Petr* and *Marek* for the nice time spent not only during the weekends.

- I would especially like to thank my husband *Jiřík* for his never-ending support, understanding and helpfulness at any time.

- Finally, *my parents* and entire *my and Jiřík's family* receive a great thank from my entire heart for their support from distance, and encouragement they gave me at any time.