

# **Structural investigations of the Id helix-loop-helix dimerization domain**

## **Dissertation**

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



vorgelegt von  
**Sebastian Donatus Kiewitz**  
aus Passau

**Regensburg 2007**



Diese Arbeit wurde angeleitet von:

Prof. Dr. A. Buschauer

Promotionsgesuch eingereicht am:

9. Mai 2007

Promotionskolloquium am:

1. Juni 2007

Prüfungsausschuß:

Vorsitzender: Prof. Dr. G. Schmeer

1. Gutachter: Prof. Dr. A. Buschauer

2. Gutachter: Prof. Dr. B. König

3. Prüfer: Prof. Dr. R. Gschwind



Diese Arbeit entstand in der Zeit von April 2004 bis April 2007 an der Fakultät für Chemie und Pharmazie der Universität Regensburg in der Arbeitsgruppe Prof. Dr. A. Buschauer unter Anleitung von Dr. C. Cabrele.



*meinen Eltern*





“If we knew what we were doing,  
it wouldn’t be called research, would it ?”

Albert Einstein



# Contents

I	Id proteins – key players in cellular processes	1
I.1	The HLH protein family	1
I.2	The Id proteins	3
I.2.1	Mode of action	3
I.2.2	Regulation of the Id proteins	6
I.3	Id in cellular pathways	9
I.4	Id in developmental processes	11
I.5	Id proteins and cancer	15
I.6	The Id proteins as therapeutic targets	19
I.7	References	20
II	Scope and objectives	29
III	Modulating the folding of the Id1 HLH domain by sequence modifications	31
III.1	Structural preferences of the HLH proteins	31
III.2	Peptide design	34
III.2.1	Loop surrogates and retro-sequences	34
III.2.2	Phenylalanine scan	36
III.3	Results	37
III.3.1	Loop surrogates	37
III.3.2	Phenylalanine scan	42
III.4	Discussion	45
III.5	References	47
IV	Structural preferences of the Id1 HLH dimer	49
IV.1	Introduction	49

IV.2	Peptide design	50
IV.3	Results	51
IV.3.1	Formation of the oxidized homodimers	51
IV.3.2	Thiol exchange assay	52
IV.4	Discussion	53
IV.5	References	55
V	Solution NMR studies on the Id2 HLH motif	57
V.1	Introduction	57
V.2	Results	58
V.2.1	CD measurements	58
V.2.2	NMR measurements	60
V.2.3	Chemical shift indices (CSIs) of the $\alpha$ H protons	62
V.2.4	NOE signals	62
V.3	Discussion	63
V.4	References	64
VI	Toward the recognition of the Id HLH motif by an artificial receptor	67
VI.1	Introduction	67
VI.2	Results	69
VI.3	Discussion	79
VI.4	References	81
VII	Summary	85
VIII	Experimental part	89
VIII.1	Materials	89
VIII.2	Methods	89

VIII.2.1	Peptide synthesis and purification	89
VIII.2.1.1	Solid-phase peptide synthesis (SPPS)	89
VIII.2.1.2	Experimental procedure	93
VIII.2.1.3	Preparative HPLC	94
VIII.2.2	Oxidation experiments	94
VIII.2.2.1	Monomer oxidation	94
VIII.2.2.2	Thiol exchange assay	94
VIII.2.3	Peptide characterization	95
VIII.2.3.1	Analytical HPLC	95
VIII.2.3.2	Mass spectrometry	95
VIII.2.3.3	UV spectroscopy	95
VIII.2.4	Structural characterization	96
VIII.2.4.1	CD spectroscopy	96
VIII.2.4.1.1	Introduction	96
VIII.2.4.1.2	Experimental procedure	98
VIII.2.4.2	NMR spectroscopy	98
VIII.2.4.2.1	Introduction	98
VIII.2.4.2.2	Sequential assignment	99
VIII.2.4.2.3	Chemical shift index (CSI)	100
VIII.2.4.2.4	NOE signals	101
VIII.2.4.2.5	Experimental procedure	102
VIII.2.5	Fluorescence spectroscopy	103
VIII.2.5.1	Experimental procedure	103
VIII.3	References	103
IX	Appendix	107
IX.1	Synthesized peptides – analytical data	107
IX.2	CONTIN analysis of the CD data for the Id2 HLH motif	108
IX.3	NMR data of the Id2 HLH motif	109

## List of Abbreviations

### Amino Acid Codes

One-Letter Code	Three-Letter Code	Amino Acid
A	Ala	Alanine
C	Cys	Cysteine
D	Asp	Aspartic acid
E	Glu	Glutamic acid
F	Phe	Phenylalanine
G	Gly	Glycine
H	His	Histidine
I	Ile	Isoleucine
K	Lys	Lysine
L	Leu	Leucine
M	Met	Methionine
N	Asn	Asparagine
P	Pro	Proline
Q	Gln	Glutamine
R	Arg	Arginine
S	Ser	Serine
T	Thr	Threonine
V	Val	Valine
W	Trp	Tryptophan
Y	Tyr	Tyrosine
-	β-Ala	β-Alanine
-	Ahx	Aminohexanoic acid
-	Orn	Ornithine

<b>Ac</b>	acetyl	<b>Mtt</b>	4-methyltrityl
<b>ACN</b>	acetonitrile	<b>MW</b>	molecular weight
<b>BMP</b>	bone morphogenic protein	<b>MyoD</b>	myogenic determination factor
<b>Boc</b>	tert-butyloxycarbonyl	<b>NES</b>	nuclear export signal
<b>Cbz</b>	benzyloxycarbonyl	<b>NLS</b>	nuclear localization signal
<b>CD</b>	circular dichroism	<b>NMP</b>	<i>N</i> -methyl-2-pyrrolidone
<b>CDK</b>	cyclin-dependent kinase	<b>NMR</b>	nuclear magnetic resonance
<b>CDKI</b>	cyclin-dependent kinase inhibitor	<b>NOESY</b>	nuclear Overhauser effect spectroscopy
<b>COSY</b>	correlation spectroscopy	<b>Pbf</b>	2,2,4,6,7-pentamethyldihydro-benzofuran-5-sulfonyl
<b>DCM</b>	dichloromethane	<b>PG</b>	protecting group
<b>DIC</b>	<i>N,N'</i> -diisopropylcarbodiimide	<b>pRb</b>	retinoblastoma protein
<b>DIEA</b>	diisopropylethylamine	<b>PS</b>	polystyrene
<b>DMF</b>	<i>N,N</i> -dimethylformamide	<b>PyBOP</b>	benzotriazol-1-yl-oxytri-pyrrolidino-phosphonium hexafluorophosphate
<b>EDT</b>	ethanedithiol	<b>RMSD</b>	root mean square deviation
<b>ESI</b>	electrospray ionization	<b>RP-HPLC</b>	reverse phase high performance liquid chromatography
<b>Fmoc</b>	fluorenyl-9-methoxycarbonyl	<b>SBE</b>	Smad-binding element
<b>HATU</b>	<i>O</i> -(7-azabenzotriazol-1-yl)- <i>N,N,N',N'</i> -tetramethyluronium hexafluorophosphate	<b>SPPS</b>	solid phase peptide synthesis
<b>HBTU</b>	<i>O</i> -(1-benzotriazolyl)- <i>N,N,N',N'</i> -tetramethyluroniumhexafluorophosphate	<b>TBTU</b>	<i>O</i> -(benzotriazol-1-yl)- <i>N,N,N',N'</i> -tetramethyluronium tetrafluoroborate
<b>HEPES</b>	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid	<b>TFA</b>	trifluoroacetic acid
<b>HLH</b>	helix-loop-helix	<b>TFE</b>	2,2,2-trifluoroethanol
<b>HOBt</b>	hydroxybenzotriazole	<b>TGF</b>	transforming growth factor
<b>Id</b>	inhibitor of DNA binding / inhibitor of differentiation	<b>TIS</b>	triisopropylsilane
<b>LZ</b>	leucine zipper	<b>TMSBr</b>	trimethylsilylbromide
<b>MALDI-ToF</b>	Matrix-Assisted-Laser-Desorption-Ionization Time of Flight	<b>TOCSY</b>	total correlation spectroscopy
<b>MAPK</b>	mitogen-activated protein kinase	<b>t<sub>R</sub></b>	retention time
<b>MBHA</b>	Methylbenzhydrylamine	<b>Trt</b>	trityl
<b>MMP</b>	matrix metalloproteinase	<b>UV</b>	ultraviolet
<b>MS</b>	mass spectrometry	<b>VEGF</b>	vascular endothelial growth factor





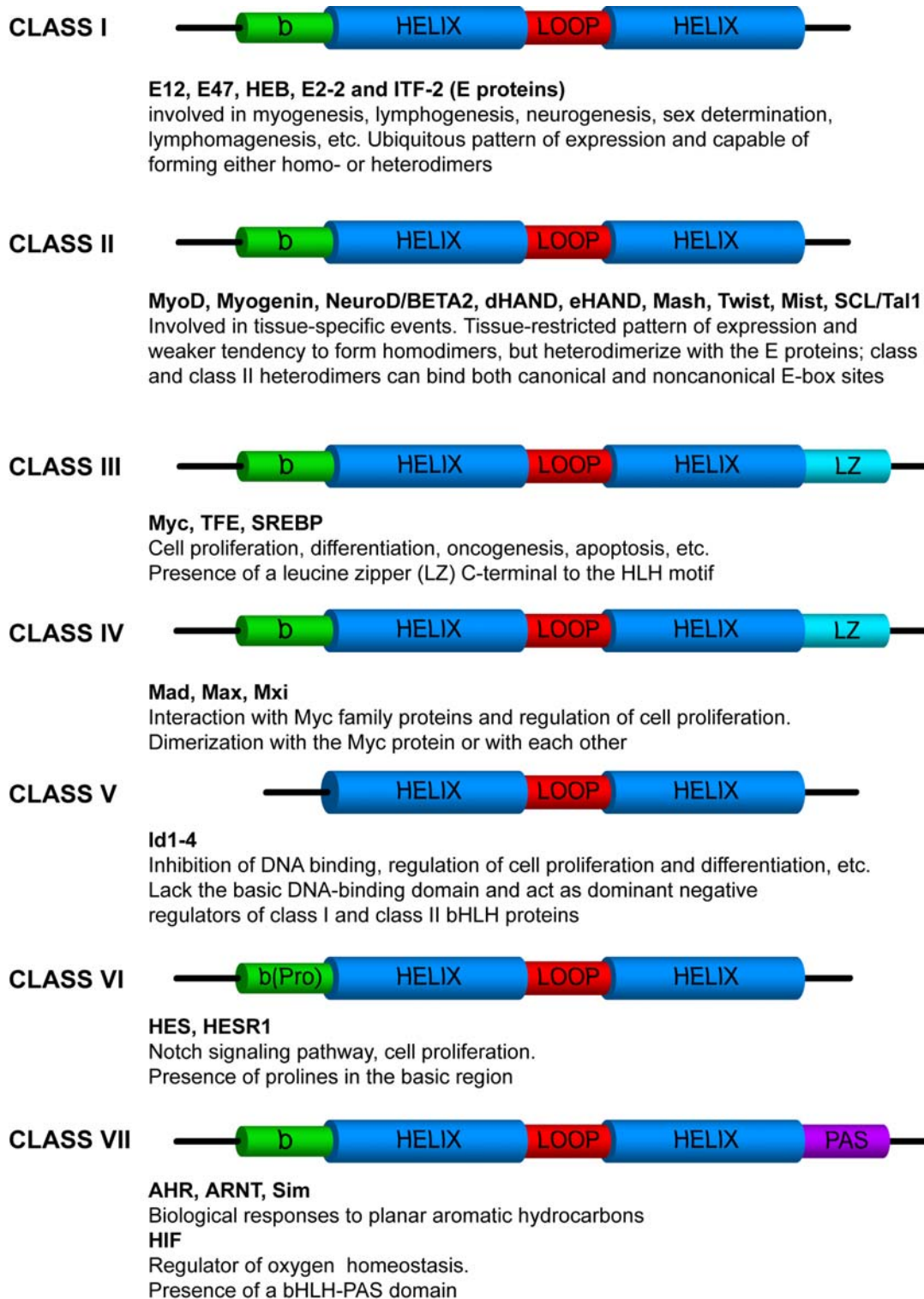
## I Id proteins – key players in cellular processes

### I.1 The HLH protein family

The first step in transferring the genetic information to the synthesis of proteins is the transcription of DNA to RNA. This process is initiated by proteins, so-called transcription factors, which bind to the DNA and thus trigger the transcriptional activation. Typically, these are classified on the basis of their DNA-binding domain. Common domains are the leucine zipper (LZ), the Helix-Turn-Helix (HTH) motif, the Helix-Loop-Helix (HLH) motif, zinc-coordinating binding domains and  $\beta$ -scaffolds for minor groove contacts. The more than 240 proteins containing a HLH motif are grouped in the HLH protein superfamily which is further divided into different classes depending on the presence of other domains in addition to the HLH domain (Scheme I.1). Class III and IV proteins present a C-terminal LZ, while others exhibit a proline-containing basic region (class VI) or a Per-Ant-Sim (PAS) domain (class VII). Alternatively, the HLH proteins can also be classified by their DNA binding preferences or their phylogenetic similarity<sup>1</sup>; however, these methods are less commonly used.

All HLH proteins except those of class V bear an additional basic region adjacent to the HLH motif, which enables the formed homo- or heterodimers to bind to specific hexameric DNA consensus sequences like the Ephrussi box (E-box) *CANNTG*, the N-box *CACNAG* or the Ets site *GGAA/T*, and thus to activate transcription. A special role is played by the class V proteins Id1-4 which are the only group lacking the additional basic region but, nevertheless, form heterodimers with proteins from the classes I and II, thus preventing the DNA binding of the latter. Using this inhibition mechanism, the Id proteins act as regulators of gene expression. The involvement of the HLH proteins in a wide range of developmental processes like cellular differentiation, lineage commitment, and sex determination renders them interesting subjects of research.

**Scheme I.1.** Classification of the HLH proteins by structural and functional characteristics. Adapted from Desprez et al.<sup>2</sup>. (LZ: leucine zipper; b: basic region; PAS: Per-Ant-Sim region).



## I.2 The Id proteins

The Id proteins were first identified in 1990 by Robert Benezra.<sup>3</sup> Thus far four mammalian Id proteins, Id1-4, as well as homolog proteins in *drosophila* (Emc) and in *xenopus* were identified.<sup>4-6</sup> Their lengths range from 119 amino acids for Id3 to 161 for Id4, which makes them rather small proteins. In humans the four Id proteins are located on chromosomes 20q11 (Id1)<sup>7,8</sup>, 2p25 (Id2)<sup>7</sup>, 1p36.1 (Id3)<sup>9,10</sup> and 6p21-22 (Id4)<sup>11</sup>. Also splicing variants like Id1<sup>8</sup> and Id3L<sup>12</sup> were observed, but reports about their functional significance are scarce. Sequence alignment reveals a potential HLH motif (Figure I.1), thus making them eligible to belong to the HLH protein family. As in the other HLH subfamilies, the HLH region is the most conserved throughout the Id proteins. Within the HLH motif helix-2 is higher conserved than helix-1. The loop region presents the lowest conservation except for the N-terminal *PxxP* sequence which is present in all four Id proteins and is assumed to be crucial for their activity.<sup>13</sup>

	HELIX 1															LOOP					HELIX 2																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																			
	1	5	10	15	20	25	30	35	40																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																																															</

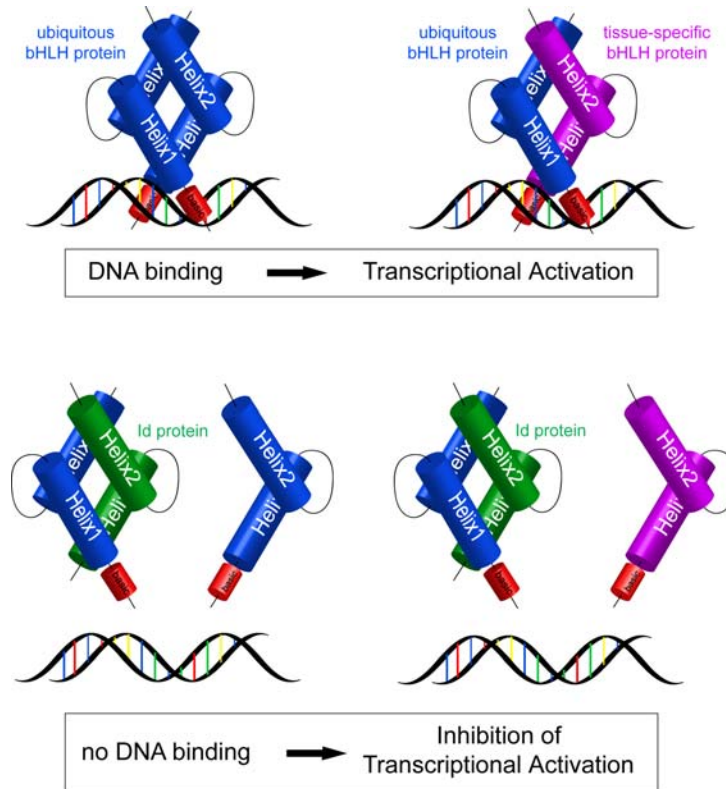
**Figure I.1.** Amino acid sequences of the human Id HLH motifs.

### I.2.1 Mode of action

The Id proteins participate in a wide range of biological events by forming heterodimers with other HLH family members. Preferred partners are the ubiquitously expressed E proteins (class I) and tissue-specific transcription factors like MyoD or myogenin (class II). As the resulting dimers are no longer able to bind to the DNA, the transcriptional activity of these bHLH factors is blocked (Scheme I.2). Hence, the Id proteins are referred to as *in*hibitors of DNA binding or, likewise, as *in*hibitors of differentiation describing the consequence of their interaction. As the interaction between the bHLH dimers and the DNA is highly energetic, it is unclear how the Id proteins overcome this energy. To disrupt the DNA-dimer complex, a widely accepted hypothesis suggests that the Id proteins sequester the fraction of transiently

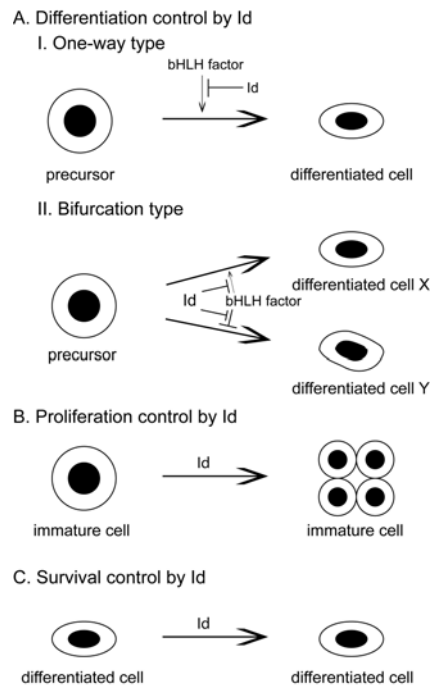
free bHLH proteins shuttling between different chromatin regions, thus preventing their reassociation to the chromatin.<sup>14</sup>

**Scheme I.2.** Inhibition of the activity of bHLH transcription factors by the Id proteins.



The function of the Id proteins as regulators of bHLH transcription factors can be divided into two categories (Scheme I.3). In the one-way type the Id proteins solely work as inhibitors, e.g. in neuronal differentiation<sup>15</sup>, while in the bifurcation type they act both as inhibitors and stimulators depending on the type of bHLH factor, e.g. in the natural killer (NK) cell development<sup>16</sup>.

In addition to bHLH proteins, also non-bHLH proteins were identified as binding partners of the Id proteins, thus expanding the range of potential biological implication: pRb family proteins, Ets-family proteins<sup>18</sup>, MIDA1<sup>19,20</sup>, Pax transcription factors<sup>21</sup>, adenovirus E1A protein<sup>22</sup>, ADD1/SREBP-1c<sup>23</sup> and S5a, a subunit of the 26 S proteasome<sup>24</sup> (Table I.1). As some of these interactions are specific for individual Id family members, it is suggested that they are not mediated by the highly conserved HLH motif, but rather by the less conserved N- and C-terminal regions.

**Scheme I.3.** Summary of the functions of the Id proteins *in vivo*.<sup>17</sup>**Table I.1.** Non-HLH protein binding partners of the Id proteins.

Protein	Description	Binding partners
<b>S5a</b>	subunit of the 26 S proteasome	Id1
<b>pRb family</b>	cell cycle regulation; as no direct interaction between pRb and Id2 could be observed, a multiprotein complex is likely to be involved in this interaction <sup>25</sup>	Id2
<b>Ets family (Elk-1, SAP-1, SAP-2)</b>	HTH transcription factors; regulation of expression of immediate early genes such as <i>c-fos</i> and <i>egr-1</i>	Id1-3
<b>MIDA1</b>	two thirds of the MIDA1 N-terminus resemble Zuotin, a Z-DNA binding protein; regulation of cell growth	Id1
<b>Pax</b>	paired-domain transcription factor, involved in several developmental processes	Id1 > Id2, Id3
<b>APC/C subunits Apc1, Apc5, Apc8/Cdc23</b>	degradation of cyclins and other key cell cycle regulators	Id2
<b>adenovirus E1A proteins</b>	induce cell proliferation and death, while inhibiting cell differentiation	Id1, Id2
<b>ADD1/SREBP-1c</b>	control the expression of adipocyte genes, involved in lipogenesis	Id2, Id3

### I.2.2 Regulation of the Id proteins

A prerequisite for the involvement of proteins in a wide range of diverse physiological processes is their precise regulation in terms of time and location. This is achieved by tissue-specific expression, a dynamic rate of protein synthesis, time-specific cellular localization and degradation. In the following, mechanisms regulating the Id proteins are reviewed.

#### *Expression of the Id proteins*

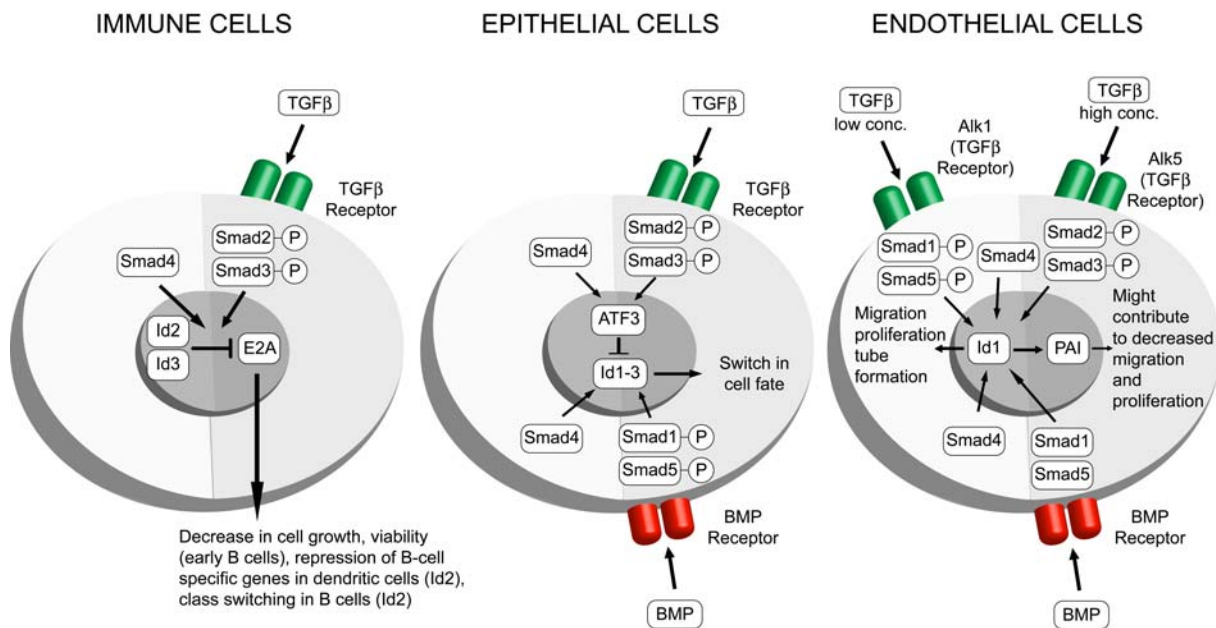
Tournay et al. identified Id1 as a physiological target for the early response gene *Egr-1*, which is rapidly induced in the absence of new protein synthesis by a variety of stimuli.<sup>26</sup> Downregulation of Id1 expression is mediated by a range of diverse signals like mitogen deprivation, vitamin D, granulocyte colony-stimulating factor, retinoic acid and TGF $\beta$ . A hormonal regulation of the Id proteins can be found in Sertoli cells, where Id1 expression is down-regulated by FSH.<sup>27</sup> Finally, also activin was shown to suppress Id expression in keratinocytes *in vivo*.<sup>28</sup>

#### *TGF $\beta$ /BMP regulation of Id expression*

Id protein expression was found to be mediated by the TGF $\beta$ /BMP signaling pathway. Proteins of the TGF $\beta$  (transforming growth factor  $\beta$ ) family are signaling molecules that are involved in the regulation of important cellular processes like cell growth, differentiation and fate determination. Their signals are transferred by the phosphorylated Smad transcription factors. A subclass of the TGF $\beta$  family is formed by the bone morphogenic proteins (BMPs), which repress neuronal and myogenic cell fate. BMPs<sup>29</sup>, as well as TGF $\beta$  itself, were found to participate in Id regulation in a cell-specific manner (Scheme I.4). In epithelial or endothelial cells BMP stimulation results in Id gene expression mediated by specific regions on the Id promoter: two SBE (Smad-binding elements) regions which bind Smad1 and Smad5, and a GC-rich region. Although this has only been reported for the Id1 gene<sup>31</sup>, it is assumed that Id2 and Id3 regulation are influenced by the same factors. In some epithelial cell lines the TGF $\beta$  pathway counteracts the BMP upregulation by downregulating the levels of the Id proteins. This downregulation is mediated by the synthesis of the transcriptional repressor ATF3 which binds to the ATF/CREB site on the Id promoter and thereby blocks Id expression.<sup>32</sup> In contrast thereto, the response upon TGF $\beta$  stimulation in endothelial cells is concentration dependent and mediated by the two TGF $\beta$  type I receptors Alk1 and Alk5.

**Scheme I.4.** Regulation of Id genes by members of the TGF $\beta$  family (adapted from Benezra and coworkers<sup>30</sup>).

PAI: Plasminogen activator inhibitor.



At low TGF $\beta$  concentrations, Alk1 signaling leads to the migration and proliferation of the cells, while at high TGF $\beta$  concentrations these processes are inhibited by Alk5 signaling.<sup>33</sup> Finally, in immune cells Id2 and Id3 expressions are upregulated by TGF $\beta$  resulting in sequestration of E2A.

#### Phosphorylation of the Id proteins

Another regulation mechanism of the Id protein function is the posttranslational phosphorylation. In the late G1 phase of the cell cycle Id2 and Id3 are phosphorylated by cyclin-dependent kinase 2 (CDK2), which modulates their activity and binding behavior and allows the progression to the next phase of the cell cycle.<sup>34,35</sup>

#### Degradation of the Id proteins

The very narrow time-frame of the Id protein functionality is mirrored in their short half life time ( $t_{1/2} = 0.5-1h$ ), which results from a coordinated interplay of expression and ubiquitin-dependent degradation<sup>36</sup>. While the Id proteins Id1-3 are degraded *via* the 26 S proteasome pathway, Id4 degradation is regulated differently.<sup>37</sup> In case of Id1 and Id3, the ubiquitin-dependent degradation was reported to be mediated by the recently identified COP9 signalosome (CSN)<sup>38, 39</sup>. This was confirmed by the observation of a direct physical interaction between the CSN complex and Id3, and the fact that inhibition of the CSN

associated kinases triggers the ubiquitin-dependent degradation of Id1 and Id3. Furthermore, the interaction with binding partners renders the Id proteins more stable towards degradation, as it was shown for Id1<sup>36</sup> and Id3<sup>37, 40</sup>. In addition, Iavarone et al. reported on a conserved destruction box (D-box) motif *RxxLxxxN* in all Id proteins but Id3 (Table I.2) which triggers their degradation by the anaphase-promoting complex (APC/C<sup>Cdh1</sup>).<sup>41</sup>

**Table I.2.** Alignment of D-boxes in the Id proteins.<sup>41</sup>

Protein	D-box	Location
<b>Id1</b>	<b>RAPLSTLN</b>	126-133
<b>Id2</b>	<b>RTPLTTLN</b>	100-107
<b>Id4</b>	<b>RTPLTALN</b>	137-144

### Localization of the Id proteins

Besides the temporal regulation, the function of the Id proteins also depends on the spatial distribution. Depending on the cell type and cell cycle state, the Id proteins are found in the nucleus or in the cytoplasm.<sup>40, 42, 43</sup> As the Id proteins are relatively small proteins with molecular masses in the range from 13 to 18 *kDa*, passive diffusion is favored as the principal way of localization. Nonetheless, also nuclear export signals (NES) were identified in the Id proteins which regulate the transport by nuclear pore complexes (Table I.3). While Id1 exhibits a NES in the second helix, the NES of Id2 is located in the C-terminal region.<sup>44</sup> Although the Id proteins lack defined canonical nuclear localization signals (NLS), it is likely that the basic residues present in the HLH domain fulfill the task of nuclear localization as it was shown for Id1<sup>36</sup>. For Id3, which lacks a NLS-like sequence, another localization mechanism was observed, which can also be imagined for the other Id proteins: by binding to the bHLH protein E47 that does possess a NLS region, Id3 is exclusively localized to the nucleus.<sup>40</sup>

**Table I.3.** NES- and NLS-like sequences of the Id proteins Id1 and Id2.

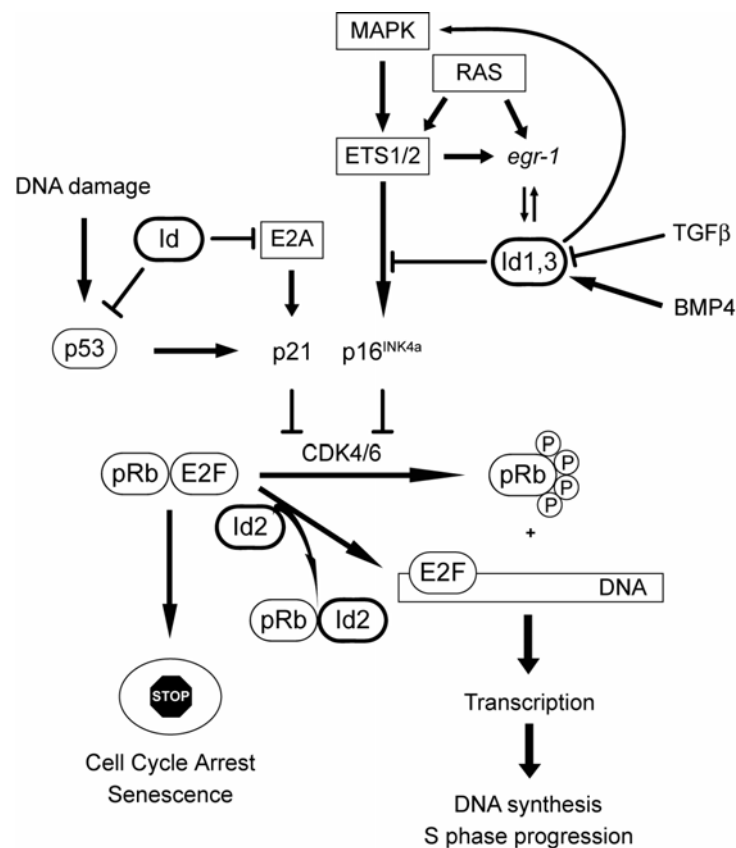
	NES	Putative NLS
<b>Id1</b>	98 99 100 101 102 103 104 105 106 107 108 109 V I D Y I R D L Q L E L	74 75 76 77 78 79 80 81 82 83 84 85 86 87 88 89 90 91 S R L K E L V P T L P Q N R K V S K
<b>Id2</b>	106 107 108 109 110 111 112 113 114 115 L N T D I S I L S L	



### I.3 Id in cellular pathways

The eukaryotic cell cycle is composed of a series of events between two cell divisions. An important phase is the G1 to S transition, in which the decision upon the further fate of the cell is made. As Id1 and Id2 were found to be upregulated during this cell cycle phase<sup>45</sup>, it was first speculated and later proven that they contribute to cell cycle regulation (Scheme I.5).

**Scheme I.5.** The role of the Id proteins in the cell cycle.



#### *pRb Pathway*

Iavarone et al. reported upon an interaction between Id2 and the tumor suppressor protein pRb in U2OS and SAOS-2 cell lines, which alters the pRb function of suppression of cell cycle progression.<sup>42, 46</sup> However, a direct protein-protein interaction between the small pocket domain of pRb and Id2 could not be observed *in vitro*.<sup>25</sup> The pRb activity is indirectly regulated by Id1 and Id3, as they lead to decreased expression of p16<sup>INK4a</sup> by binding to Ets-family proteins.<sup>47, 48</sup> p16<sup>INK4a</sup> belongs to the INK4a family of CDK inhibitors (CKI) and is a potent inhibitor of the CDK4/CDK6/CyclinD complex that mediates the phosphorylation of

pRb. The hyperphosphorylated form of pRb does no longer bind the transcription factor E2F, thus promoting cell growth. These findings were confirmed by Id1<sup>-/-</sup> primary mouse embryonic fibroblasts which prematurely enter senescence.<sup>47</sup>

#### *p53 Pathway*

Although *p53* is not required for cell cycle progression, it has the crucial function of deciding upon the cell fate once DNA damage occurred: either the cell is arrested in the G1 phase (senescence) or the cell is subjected to apoptosis. The *p53* gene is the most frequently mutated gene in human cancer (more than 50% of human cancers are associated with a *p53* mutation<sup>49</sup>), wherein it is responsible for uncontrolled proliferation. The Id proteins are also involved in this pathway, as Id1 transfected human keratinocytes were found to impair the DNA damage response of the *p53* pathway after exposure to a DNA damaging agent.<sup>50</sup> Moreover, Nakajima et al. reported that the suppression of E1A-induced apoptosis by mutated *p53* is overcome by coexpression with Id proteins.<sup>22</sup>

#### *MAPK Pathway*

The MAPK (mitogen-activated protein kinase) signaling pathway induces cell proliferation upon mitogen stimulation. Key regulators of this pathway are Raf and MEK1/2, which are found after Id1 expression.<sup>51</sup> In prostate cancer elevated MAPK levels and the expression of *egr-1* are connected to a poorer outcome. Ling et al. therefore suggested Id1 to be a positive upstream regulator of the MAPK signaling pathway.

#### *p21 Pathway*

The ubiquitous class I bHLH protein E2A possesses two inhibitory regions in addition to the HLH domain and blocks the G1-S phase transition by activating the transcription of the p21<sup>Cip1</sup> gene, another CDKI. This process was reported to be regulated by the Id proteins, which bind to the E2A protein and block its growth suppression action.<sup>52,53</sup>

#### *Immortalization*

The lifespan of eukaryotic cells is determined by the telomeric DNA ends which are shortened with each cell division. After reaching the end of their lifespan, normal cells enter a replicative senescence, in which they stop dividing but remain viable. Telomerase is able to counteract the telomere shortening.

In human keratinocytes Id1 expression was found to increase telomerase activity and thus to produce immortalized cells.<sup>50</sup> However, these results are challenged by other reports which could not establish a link between Id1 expression and increased telomerase activity. Instead of an immortalization only a postponed entry in senescence was observed.<sup>54</sup>

## I.4 Id in developmental processes

The role of the Id proteins in developmental processes is based on the cellular pathways described above. Several studies on animal knockout models elucidated the participation of Id proteins in tissue-specific events (Table I.4).

**Table I.4.** Results from studies on Id knockout animals.<sup>55</sup>

Deleted Gene	Phenotype
Id1 <sup>-/-</sup>	No major abnormalities, elevated TSP-1 expression
Id2 <sup>-/-</sup>	Lack of lymph nodes, NK cells and Peyer's patches; defects in spermatogenesis; loss of Langerhans cells
Id3 <sup>-/-</sup>	Abnormalities in humoral immunity and T cell development
Id1 <sup>-/-</sup> , Id3 <sup>-/-</sup>	Lethal E13.5, vascular defects in forebrain, premature maturation of neurons
Id1/Id3 loss of single or multiple alleles	Defects in vascularization and growth of tumor xenografts
Id2 <sup>-/-</sup> , pRb <sup>-/-</sup>	Compensatory resolution of pRb <sup>-/-</sup> defects in neurogenesis and hematopoiesis, rescue of pRb <sup>-/-</sup> embryonic lethality, but severe muscle loss leads to postnatal demise

Surprisingly Id1<sup>-/-</sup> mice did not produce severe abnormalities, while double knockout of Id1 and Id3 was lethal. This suggests that lack of Id1 can be compensated by Id3, and that Id1 is not essential for embryonic development. Furthermore, the Id proteins are involved in the terminal differentiation steps of many cell lineages as diverse as myoblasts, T and B cells, mammary epithelial cells, oligodendrocytes and dendritic cells.

### *Ids in neurogenesis*<sup>56</sup>

Astroglia, oligodendrocytes, ependymal and microglia are the four cell types in the central nervous system (CNS) responsible for the support and protection of neural cells. While Id1 and Id3 are abundantly expressed in proliferating neuroepithelial cells and undifferentiated CNS regions, their expression is decreased in the late stage of CNS development. This pattern suggests a role in the proliferation of neural precursor cells and in the inhibition of cell differentiation. Id2, in contrast, is still expressed in postnatal development and adulthood, pointing to a more complex role in neural tissue. Id4, however, is only detected in neural tissue during postgastrulational mouse development and was shown to be involved in the development of oligodendrocytes.<sup>57</sup> In combination with Id2 it is further responsible for the timing of oligodendrocyte regulation.<sup>17</sup> These processes are mediated by Olig1 and Olig2, two class II bHLH proteins which are sequestered by Id2 and Id4, resulting in inhibition of oligodendrocyte lineage commitment. As a consequence, astrocyte development, which is not dependent on bHLH factors, is enhanced.<sup>58</sup>

### *Ids in the digestive system*

Although no potential bHLH binding partners for the Id proteins have yet been identified in the oropharyngeal cavity and in the esophagus, it is suggested that the Id proteins are regulating the cell growth of these lineages. Evidence came from the observation of Id1 upregulation in esophageal squamous cell carcinoma (SCC). This involvement is assumed to be mediated by the APC/ $\beta$ -catenin pathway.<sup>59</sup> However, the Id proteins seem not to contribute to the homeostasis of the small intestine epithelium. In the pancreas an increased expression of Id1 and Id2 was found during the transformation of  $\beta$ -cells into  $\delta$ -cells associated with the activation of the insulin gene. Furthermore, Id expression was shown to inhibit insulin production in insulinoma cells. Suggested mechanisms are the inhibition of the *cis*-regulated insulin control element (ICE) and the sequestration of E2A gene products, which bind to HLH complexes and the insulin activator factor INSAT. Finally, in the liver Id1 is upregulated during the activation and proliferation of hepatocytes.

### *Ids and thyroid*

Stimulation of thyroid cell growth *in vitro* led to upregulation of Id1, whereas activation of the protein kinase C pathway downregulated Id1.<sup>60</sup> With the tissue-specific class II bHLH protein Ash-1 (achaete-scute homolog-1) a potential target of Id1 in thyroid was identified.<sup>61</sup>

*Ids and the immune system*

The major cellular components of the immune system are the lymphocytes which are divided into three major cell types: the natural killer (NK), T and B cells. All these cell types emanate from a common, multipotent precursor. Main participants in lymphocyte commitment are the ubiquitous E proteins E2A and HEB, and the Id proteins Id2 and Id3. While Id2 is involved in cell fate determination to NK cells, as demonstrated in Id2<sup>-/-</sup> mice<sup>16, 62</sup>, Id3 is crucial for the development of T cells.

Peyer's patches (PPs) and lymph nodes (LNs) constitute, together with the spleen, the secondary lymphoid system. Their development is dependent on the lymphotoxin (LT) signaling pathway through the LTβ receptor and is assumedly regulated by Id2, as mice deficient in Id2 do not develop PPs and peripheral LNs.<sup>62</sup> It is suggested that a yet unidentified bHLH factor negatively regulating the differentiation of lymphoid precursor cells to LT-expressing cells is repressed by Id2, which allows a normal development of secondary lymphoid organs.<sup>17</sup>

*Ids and mammary differentiation*

A contribution of the Id proteins to the differentiation of mammary epithelial cells is underlined by the finding that Id1 expression decreased when murine SCp2 cells were induced to differentiate in culture, and that constitutive Id1 expression blocked differentiation.<sup>63</sup> With ITF-2 (class I HLH protein) a potential binding partner for the Id proteins in mammary epithelial cells was identified.<sup>64</sup> Furthermore, it is suggested that during mammary differentiation Id1 acts as an activator rather than as an inhibitor. Studies on Id2 deficient female mice showed a severe lactation defect, and the mammary glands of Id2<sup>-/-</sup> pregnant mice failed to proliferate and remained immature.<sup>65</sup> These results stress the importance of Id2 in cell cycle progression of mammary epithelial cells and propose a coordinated interplay between Id1 and Id2 during mammary differentiation.

*Ids and spermatogenesis*

In the testis the Id family members possess a distinct expression pattern not only in spatio-temporal terms, but also in a dynamic subcellular distribution.<sup>66</sup> Closer studies have only been conducted with Id2 deficient adult mice which show malformed seminiferous tubules and a dramatic decrease in mature sperm. Id2 is mainly expressed in Sertoli cells, which are formed in fetal and neonatal periods and whose number is constant during adulthood, suggesting Id2 participation in their long-term survival.

### *Ids in skin development*

The keratinocytes are the predominant cell type encountered in the epidermis. As the Id proteins were found to be expressed in human epidermis and dysregulated in squamous cell carcinoma (SCC)<sup>67</sup>, an Id regulatory role in keratinocyte proliferation and differentiation was proposed. In fact, Id1, Id2, and Id3 are downregulated by activin, a member of the TGF- $\beta$  family, which is involved in skin morphogenesis and wound healing.<sup>28</sup> The contribution of the Id proteins to wound healing was demonstrated by Schaefer et al., who observed the upregulation of Id1 upon skin injury. After mediating the conversion from a sessile into a mobile keratinocyte alongside its proliferation, the Id1 level is downregulated again.<sup>68</sup> The influence of the Id proteins on cell motility can also be observed in tumorigenesis, where they facilitate tumor propagation. A pathogenic role of Id1 in psoriasis has also been suggested, as it was found to be upregulated in lesioned psoriatic skin.<sup>69</sup>

### *Ids in myogenesis*

In the course of myogenesis myoblasts are differentiated into myotubes. This process is regulated by the highly coordinated interplay of the muscle regulatory factors MyoD, Myf5, myogenin, and Mrf4, (class II bHLH proteins) and their binding partners, the ubiquitous E proteins (class I) and the Id proteins (class V). The transcription of muscle-specific target genes is activated by heterodimers formed between the muscle regulatory factors and the ubiquitous E proteins. As Id1 possesses a high binding affinity for the E proteins, it prevents the formation of transcription activating dimers by sequestering the E proteins. This inhibition is overcome by p204, a p200 family protein, which binds to the Id proteins and also triggers a decrease in their level, presumably by shuttling them from the nucleus to the cytoplasm and thus accelerating their degradation.<sup>70</sup> Indeed, it was reported that in proliferating C2C12 myoblasts MyoD and Id1 are co-localized in the nucleus, while in differentiated myotubes MyoD was located in the nucleus and Id1 in the cytoplasm.<sup>71</sup>

### *Ids and angiogenesis*

Angiogenesis, the formation of new blood vessels out of preexisting ones, is an important process during development and wound healing. First evidence for a role of the Id proteins in angiogenesis was provided by a study on Id1/Id3 double knockout mice which suffered from haemorrhage.<sup>72</sup> With thrombospondin-1 (TSP-1) a target gene for the transcriptional repression by Id1 was identified.<sup>73</sup> TSP-1 is a glycoprotein known to be a potent inhibitor of *in vivo* neovascularization and tumorigenesis. Id1 represses these two

events by inhibiting the transcription of TSP-1 *via* a mechanism yet to be identified. Another key player in angiogenic events is the vascular endothelial growth factor (VEGF), a heparin-binding glycoprotein produced by almost every cell type: loss of Id function has been reported to lead to decreased VEGF expression.<sup>15</sup>

## I.5 Id proteins and cancer

In comparison to normal cells, important cellular pathways are dysregulated in cancer cells, resulting in uncontrolled cell division and spreading to adjacent tissue (invasion) as well as distant sites (metastasis). As the Id proteins are involved in cellular pathways regulating proliferation and differentiation, it was not surprising to find them also contributing to tumor-related processes.

### *Ids in tumorigenesis*

Although the Id proteins do not strictly meet the definition of oncogenes and, up to now, no tumor-associated defects in the Id genes have been observed, the fact that overexpression of the Id genes is mediated by oncogenes like *myc*, *RAS* and *(EWS)-ETS* and affects oncogenic pathways emphasizes their importance in tumorigenic events. Elevated levels of the Id proteins have been found in a wide range of tumor types (Table I.5), in which they were often connected with heightened severity and poor prognosis.<sup>74-76</sup>

### *Ids and centrosomes*

Centrosomes are the primary microtubule organizing centers (MTOC) in mitotic and post-mitotic cells. They are situated adjacent to the nucleus and are regulators of cell cycle progression. Hasskarl et al. showed that a fraction of Id1, but not of the other Id proteins localizes to the centrosomes and mitotic spindle poles, and subsequently induces abnormal centrosome and centriole numbers.<sup>79</sup> These defects in the centrosome duplication presumably contribute to genomic instability and tumor formation, as they decrease the accuracy of mitotic replication. Furthermore, this property of the Id1 protein could be linked to the presence of its N-terminal and HLH regions.

**Table I.5.** Dysregulated expression of Id mRNAs and proteins in primary human tumors.<sup>77, 78</sup>

Dysregulated Id	Tumor type
Id1, Id3	Glioblastoma, medulloblastoma, neuroblastoma
Id1, Id2, Id3	Pancreatic cancer
Id1, Id2, Id3, Id4	Seminoma
Id1	Medullary thyroid cancer
Id1, Id2, Id3	Squamous cell carcinoma
Id1	Breast carcinoma
Id1	Endometrial cancer
Id1	Cervical cancer
Id1	Melanoma
Id2	Neuroblastoma
Id1	Hepatocellular carcinoma
Id1, Id2, Id3	Colorectal adenocarcinoma
Id1, Id2, Id3	Astrocytic tumor
Id2	Ewing's sarcoma
Id1, Id3	Ovarian tumor
Id1, Id2	Prostate cancer

### *Ids in cancer progression*

The pivotal events in tumor progression are angiogenesis, tissue invasion, anaplasia and proliferation. In all these processes the Id family members were found to be involved. As within the Id protein family Id1 contribution was most exhaustively reviewed, the following paragraphs will focus on this Id member.

### *Tumor progression*

The spreading of tumor cells into adjacent or remote regions requires the destruction of the basement membrane and the extracellular matrix (ECM) surrounding the cancer cell. *In vivo* studies revealed that high levels of Id are correlated with an upregulation of MMP2, one of the matrix metalloproteinases (MMPs) which are known to mediate membrane degradation. Indeed, downregulating Id1 expression resulted in a suppression of invasion and metastatic ability of breast cancer cells.<sup>80, 81</sup> Moreover Id1<sup>-/-</sup>Id3<sup>+/-</sup> mice failed to grow tumors, due to the poor vascularization of the xenografts.<sup>15</sup>



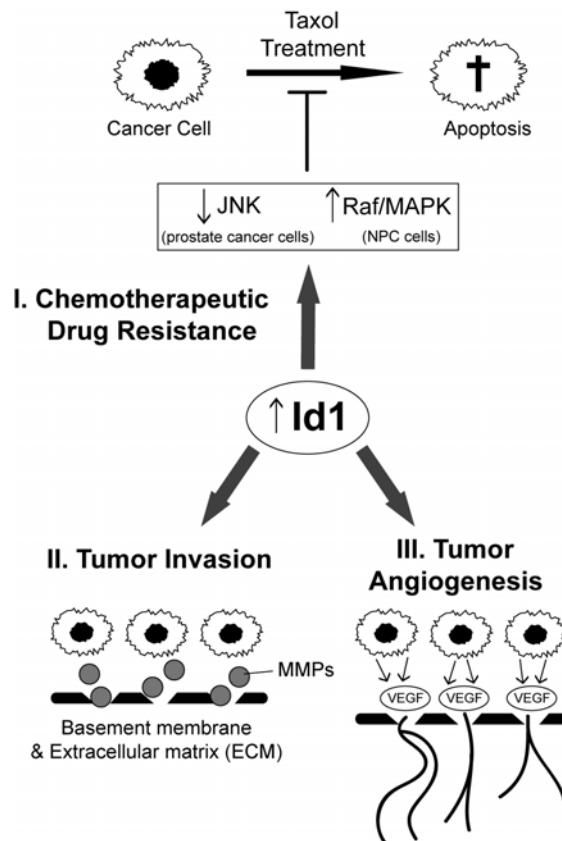
### *Tumor angiogenesis*

Prerequisite for tumor progression and metastasis is a sufficient blood supply maintained by the formation of new blood vessels in a process called tumor angiogenesis. The initiation of angiogenesis in tumors, the so-called “angiogenic switch”, is triggered by the upregulation of VEGF that promotes the exponential growth of the tumor. Like in angiogenic events during normal development, also here the Id proteins are involved.<sup>82</sup>

### *Development of chemotherapeutic drug resistance*

Chemotherapeutic drugs function by inducing apoptosis in cancer cells. A limitation of chemotherapy is the emergence of drug resistance which is associated with a more aggressive course of the cancer disease and the resistance to further chemotherapy. Responsible for the development of chemotherapeutic drug resistance is the deregulation of the Raf-1/MAPK, NF- $\kappa$ B and JNK pathways. Based on the fact that Id1 can activate the Raf-1/MAPK and NF- $\kappa$ B pathways<sup>51, 83</sup>, a role of Id1 in the development of drug resistance was anticipated. In fact, ectopic Id1 expression was able to induce resistance to taxol treatment in prostate and nasopharyngeal carcinoma (NPC) cells (Scheme I.6).<sup>84, 85</sup>

**Scheme I.6.** The role of the Id1 protein in cancer progression (adapted from Ling et al.<sup>88</sup>).



Furthermore, also an association to doxorubicin and cyclophosphamide resistance was reported<sup>86</sup> and, more recently, it was shown that high levels of the Id1 protein protect cancer cells against the apoptosis pathway initiated by nine different anticancer drugs, suggesting that Id1 promotes cell survival by acting as universal antiapoptotic factor.<sup>87</sup> These results deliver a hypothesis explaining the link between upregulation of the Id proteins and poor prognosis and severity of some human cancer types.

### *Neuroblastoma*

Neuroblastoma is an extracranial cancer striking infants and children. A characteristic of this cancer type is the amplification and overexpression of oncogenic *N-Myc* (neuronal *Myc*) which was shown to be linked to Id2 upregulation.<sup>89</sup> Although Perk et al. reported about Id2 expression as prognostic value for neuroblastoma patients, these results are disputed by other working groups.<sup>90</sup>

### *Breast cancer*<sup>91</sup>

High levels of Id1 expression in several breast cancer cell lines are associated with a high level of aggressiveness and invasiveness, therefore Id1 could be imagined as potential marker for this pathology. Contrarily to Id1, Id2 is downregulated in breast carcinogenesis.

### *Colon carcinogenesis*

Most colon cancers are caused by mutations in the APC (adenomatous polyposis coli) gene and/or  $\beta$ -catenin gene. In these adenomas a similar expression pattern of Id2 and  $\beta$ -catenin was observed.<sup>92</sup>

### *Pancreatic carcinoma*

The Id proteins are implicated in the tumorigenesis of human pancreatic cancer, a highly malignant and almost always fatal cancer type.<sup>93</sup> Id1-3 localize in tumor cells and duct-like cancer cells, while Id1 and Id2 are also found in the transformed cell cytoplasm. Like in other cancer types, overexpression of the Id proteins is associated with an enhanced proliferative activity and inhibition of differentiation and, therefore, Id1-3 could represent markers of pancreatic malignancy.

### *Epithelial cell derived tumors*

Overexpression of the Id protein in an epithelial cell tumor was firstly reported by Tang et al., who observed an upregulation of Id1 in Kaposi's sarcoma (KS), the most common neoplasm in AIDS patients. Up to now, the exact mechanism is unclear, but a contribution of the latency-associated nuclear antigen (LANA) is undisputed.<sup>94</sup>

### *Hormone refractory cancer*

A drawback in the treatment of patients suffering from breast or prostate cancer is that the cancers progress to a hormone refractory stage, where they are no longer dependent on hormonal stimulation. Hormonal ablation therapy as the main treatment option fails and a high mortality rate is associated with these cancer types. Studies conducted by Lin et al. showed a loss of estrogen dependency upon ectopic expression of the Id1 protein, suggesting that Id1 overexpression is, at least in part, responsible for the development of the hormone refractory stage.<sup>95</sup> Despite the exact mechanism being unknown, Id1 was identified as a downstream target of the androgen receptor.<sup>88</sup>

## I.6 The Id proteins as therapeutic targets

Cancer diseases as one of the main causes of death in developed countries fuel the constant need for researching novel, reliable anti-tumor therapeutics. The Id proteins represent interesting targets for such a purpose, as they are involved in key events not only related to tumorigenesis, but also to cancer progression. The main benefit of targeting the Id proteins is their differential expression in normal and cancer adult cells. Moreover, studies conducted on Id-knockout models showed that already a partial reduction of Id levels leads to reduced tumor invasion and metastasis.<sup>81</sup> Also, downregulating Id1 could provide an interesting strategy to overcome chemotherapeutic drug resistance, a major drawback in cancer treatment.<sup>87</sup> Another field of research has come up with the recent reports on a connection between Id2 expression and neuronal maturation, offering new therapeutic approaches for the treatment of patients suffering from neurological pathologies.

However, given the vast field of events, in which the Id proteins are participating, selectively addressing one key domain, while not tampering with other crucial processes, necessitates the precise deciphering of the individual physiological and structural preferences of the Id proteins.

## I.7 References

1. Atchley, W. R.; Fitch, W. M., A natural classification of the basic helix-loop-helix class of transcription factors. *Proc Natl Acad Sci USA* **1997**, 94, 5172-6.
2. Desprez, P. Y.; Sumida, T.; Coppe, J. P., Helix-loop-helix proteins in mammary gland development and breast cancer. *J Mammary Gland Biol Neoplasia* **2003**, 8, 225-39.
3. Benezra, R.; Davis, R. L.; Lockshon, D.; Turner, D. L.; Weintraub, H., The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. *Cell* **1990**, 61, 49-59.
4. Christy, B. A.; Sanders, L. K.; Lau, L. F.; Copeland, N. G.; Jenkins, N. A.; Nathans, D., An Id-related helix-loop-helix protein encoded by a growth factor-inducible gene. *Proc Natl Acad Sci USA* **1991**, 88, 1815-9.
5. Riechmann, V.; van Cruchten, I.; Sablitzky, F., The expression pattern of Id4, a novel dominant negative helix-loop-helix protein, is distinct from Id1, Id2 and Id3. *Nucleic Acids Res* **1994**, 22, 749-55.
6. Sun, X. H.; Copeland, N. G.; Jenkins, N. A.; Baltimore, D., Id proteins Id1 and Id2 selectively inhibit DNA binding by one class of helix-loop-helix proteins. *Mol Cell Biol* **1991**, 11, 5603-5611.
7. Mathew, S.; Chen, W.; Murty, V. V.; Benezra, R.; Chaganti, R. S., Chromosomal assignment of human ID1 and ID2 genes. *Genomics* **1995**, 30, 385-7.
8. Nehlin, J. O.; Hara, E.; Kuo, W. L.; Collins, C.; Campisi, J., Genomic organization, sequence, and chromosomal localization of the human helix-loop-helix Id1 gene. *Biochem Biophys Res Commun* **1997**, 231, 628-34.
9. Deed, R. W.; Hirose, T.; Mitchell, E. L.; Santibanez-Koref, M. F.; Norton, J. D., Structural organisation and chromosomal mapping of the human Id-3 gene. *Gene* **1994**, 151, 309-14.
10. Ellmeier, W.; Aguzzi, A.; Kleiner, E.; Kurzbauer, R.; Weith, A., Mutually exclusive expression of a helix-loop-helix gene and N-myc in human neuroblastomas and in normal development. *Embo J* **1992**, 11, 2563-71.
11. Pagliuca, A.; Bartoli, P. C.; Saccone, S.; Della Valle, G.; Lania, L., Molecular cloning of ID4, a novel dominant negative helix-loop-helix human gene on chromosome 6p21.3-p22. *Genomics* **1995**, 27, 200-3.
12. Deed, R. W.; Jasiok, M.; Norton, J. D., Attenuated function of a variant form of the helix-loop-helix protein, Id-3, generated by an alternative splicing mechanism. *FEBS Lett* **1996**, 393, 113-6.

13. Pesce, S.; Benezra, R., The loop region of the helix-loop-helix protein Id1 is critical for its dominant negative activity. *Mol Cell Biol* **1993**, 13, 7874-80.
14. O'Toole, P. J.; Inoue, T.; Emerson, L.; Morrison, I. E.; Mackie, A. R.; Cherry, R. J.; Norton, J. D., Id proteins negatively regulate basic helix-loop-helix transcription factor function by disrupting subnuclear compartmentalization. *J Biol Chem* **2003**, 278, 45770-6.
15. Lyden, D.; Young, A. Z.; Zagzag, D.; Yan, W.; Gerald, W.; O'Reilly, R.; Bader, B. L.; Hynes, R. O.; Zhuang, Y.; Manova, K.; Benezra, R., Id1 and Id3 are required for neurogenesis, angiogenesis and vascularization of tumour xenografts. *Nature* **1999**, 401, 670-7.
16. Ikawa, T.; Fujimoto, S.; Kawamoto, H.; Katsura, Y.; Yokota, Y., Commitment to natural killer cells requires the helix-loop-helix inhibitor Id2. *Proc Natl Acad Sci USA* **2001**, 98, 5164-9.
17. Yokota, Y., Id and development. *Oncogene* **2001**, 20, 8290-8.
18. Yates, P. R.; Atherton, G. T.; Deed, R. W.; Norton, J. D.; Sharrocks, A. D., Id helix-loop-helix proteins inhibit nucleoprotein complex formation by the TCF ETS-domain transcription factors. *Embo J* **1999**, 18, 968-76.
19. Shoji, W.; Inoue, T.; Yamamoto, T.; Obinata, M., MIDA1, a protein associated with Id, regulates cell growth. *J Biol Chem* **1995**, 270, 24818-25.
20. Inoue, T.; Shoji, W.; Obinata, M., MIDA1 is a sequence specific DNA binding protein with novel DNA binding properties. *Genes Cells* **2000**, 5, 699-709.
21. Roberts, E. C.; Deed, R. W.; Inoue, T.; Norton, J. D.; Sharrocks, A. D., Id helix-loop-helix proteins antagonize pax transcription factor activity by inhibiting DNA binding. *Mol Cell Biol* **2001**, 21, 524-33.
22. Nakajima, T.; Yageta, M.; Shiotsu, K.; Morita, K.; Suzuki, M.; Tomooka, Y.; Oda, K., Suppression of adenovirus E1A-induced apoptosis by mutated p53 is overcome by coexpression with Id proteins. *Proc Natl Acad Sci USA* **1998**, 95, 10590-5.
23. Moldes, M.; Boizard, M.; Liepvre, X. L.; Fève, B.; Dugail, I.; Pairault, J., Functional antagonism between inhibitor of DNA binding (Id) and adipocyte determination and differentiation factor 1/sterol regulatory element-binding protein-1c (ADD1/SREBP-1c) trans-factors for the regulation of fatty acid synthase promoter in adipocytes. *Biochem J* **1999**, 344 Pt 3, 873-80.

24. Anand, G.; Yin, X.; Shahidi, A. K.; Grove, L.; Prochownik, E. V., Novel regulation of the helix-loop-helix protein Id1 by S5a, a subunit of the 26 S proteasome. *J Biol Chem* **1997**, 272, 19140-51.
25. Smialowski, P.; Singh, M.; Mikolajka, A.; Majumdar, S.; Joy, J. K.; Nalabothula, N.; Krajewski, M.; Degenkolbe, R.; Bernard, H. U.; Holak, T. A., NMR and mass spectrometry studies of putative interactions of cell cycle proteins pRb and CDK6 with cell differentiation proteins MyoD and ID-2. *Biochim Biophys Acta* **2005**, 1750, 48-60.
26. Tournay, O.; Benezra, R., Transcription of the dominant-negative helix-loop-helix protein Id1 is regulated by a protein complex containing the immediate-early response gene Egr-1. *Mol Cell Biol* **1996**, 16, 2418-30.
27. Chaudhary, J.; Johnson, J.; Kim, G.; Skinner, M. K., Hormonal regulation and differential actions of the helix-loop-helix transcriptional inhibitors of differentiation (Id1, Id2, Id3, and Id4) in Sertoli cells. *Endocrinology* **2001**, 142, 1727-36.
28. Rotzer, D.; Krampert, M.; Sulyok, S.; Braun, S.; Stark, H. J.; Boukamp, P.; Werner, S., Id proteins: novel targets of activin action, which regulate epidermal homeostasis. *Oncogene* **2006**, 25, 2070-81.
29. Nakashima, K.; Takizawa, T.; Ochiai, W.; Yanagisawa, M.; Hisatsune, T.; Nakafuku, M.; Miyazono, K.; Kishimoto, T.; Kageyama, R.; Taga, T., BMP2-mediated alteration in the developmental pathway of fetal mouse brain cells from neurogenesis to astrocytogenesis. *Proc Natl Acad Sci USA* **2001**, 98, 5868-73.
30. Ruzinova, M. B.; Benezra, R., Id proteins in development, cell cycle and cancer. *Trends Cell Biol* **2003**, 13, 410-8.
31. Lopez-Rovira, T.; Chalaux, E.; Massague, J.; Rosa, J. L.; Ventura, F., Direct binding of Smad1 and Smad4 to two distinct motifs mediates bone morphogenetic protein-specific transcriptional activation of Id1 gene. *J Biol Chem* **2002**, 277, 3176-85.
32. Kang, Y.; Chen, C. R.; Massague, J., A self-enabling TGFbeta response coupled to stress signaling: Smad engages stress response factor ATF3 for Id1 repression in epithelial cells. *Mol Cell* **2003**, 11, 915-26.
33. Goumans, M. J.; Valdimarsdottir, G.; Itoh, S.; Rosendahl, A.; Sideras, P.; ten Dijke, P., Balancing the activation state of the endothelium via two distinct TGF-beta type I receptors. *Embo J* **2002**, 21, 1743-53.
34. Hara, E.; Hall, M.; Peters, G., Cdk2-dependent phosphorylation of Id2 modulates activity of E2A-related transcription factors. *Embo J* **1997**, 16, 332-42.

35. Deed, R. W.; Hara, E.; Atherton, G. T.; Peters, G.; Norton, J. D., Regulation of Id3 cell cycle function by Cdk-2-dependent phosphorylation. *Mol Cell Biol* **1997**, 17, 6815-21.
36. Trausch-Azar, J. S.; Lingbeck, J.; Ciechanover, A.; Schwartz, A. L., Ubiquitin-Proteasome-mediated degradation of Id1 is modulated by MyoD. *J Biol Chem* **2004**, 279, 32614-9.
37. Bounpheng, M. A.; Dimas, J. J.; Dodds, S. G.; Christy, B. A., Degradation of Id proteins by the ubiquitin-proteasome pathway. *Faseb J* **1999**, 13, 2257-64.
38. 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.; Chamovitz, D. A., Unified nomenclature for the COP9 signalosome and its subunits: an essential regulator of development. *Trends Genet* **2000**, 16, 202-3.
39. Berse, M.; Bounpheng, M.; Huang, X.; Christy, B.; Pollmann, C.; Dubiel, W., Ubiquitin-dependent degradation of Id1 and Id3 is mediated by the COP9 signalosome. *J Mol Biol* **2004**, 343, 361-70.
40. Deed, R. W.; Armitage, S.; Norton, J. D., Nuclear localization and regulation of Id protein through an E protein-mediated chaperone mechanism. *J Biol Chem* **1996**, 271, 23603-6.
41. Lasorella, A.; Stegmuller, J.; Guardavaccaro, D.; Liu, G.; Carro, M. S.; Rothschild, G.; de la Torre-Ubieta, L.; Pagano, M.; Bonni, A.; Iavarone, A., Degradation of Id2 by the anaphase-promoting complex couples cell cycle exit and axonal growth. *Nature* **2006**, 442, 471-4.
42. Iavarone, A.; Garg, P.; Lasorella, A.; Hsu, J.; Israel, M. A., The helix-loop-helix protein Id-2 enhances cell proliferation and binds to the retinoblastoma protein. *Genes Dev* **1994**, 8, 1270-84.
43. Jen, Y.; Weintraub, H.; Benezra, R., Overexpression of Id protein inhibits the muscle differentiation program: in vivo association of Id with E2A proteins. *Genes Dev* **1992**, 6, 1466-79.
44. Makita, J.; Kurooka, H.; Mori, K.; Akagi, Y.; Yokota, Y., Identification of the nuclear export signal in the helix-loop-helix inhibitor Id1. *FEBS Lett* **2006**, 580, 1812-6.
45. Hara, E.; Yamaguchi, T.; Nojima, H.; Ide, T.; Campisi, J.; Okayama, H.; Oda, K., Id-related genes encoding helix-loop-helix proteins are required for G1 progression and are repressed in senescent human fibroblasts. *J Biol Chem* **1994**, 269, 2139-45.

46. Lasorella, A.; Iavarone, A.; Israel, M. A., Id2 specifically alters regulation of the cell cycle by tumor suppressor proteins. *Mol Cell Biol* **1996**, 16, 2570-8.
47. Alani, R. M.; Young, A. Z.; Shifflett, C. B., Id1 regulation of cellular senescence through transcriptional repression of p16/Ink4a. *Proc Natl Acad Sci USA* **2001**, 98, 7812-6.
48. Ouyang, X. S.; Wang, X.; Ling, M. T.; Wong, H. L.; Tsao, S. W.; Wong, Y. C., Id-1 stimulates serum independent prostate cancer cell proliferation through inactivation of p16(INK4a)/pRB pathway. *Carcinogenesis* **2002**, 23, 721-5.
49. Sandal, T., Molecular aspects of the mammalian cell cycle and cancer. *Oncologist* **2002**, 7, 73-81.
50. Alani, R. M.; Hasskarl, J.; Grace, M.; Hernandez, M. C.; Israel, M. A.; Munger, K., Immortalization of primary human keratinocytes by the helix-loop-helix protein, Id-1. *Proc Natl Acad Sci USA* **1999**, 96, 9637-41.
51. Ling, M. T.; Wang, X.; Ouyang, X. S.; Lee, T. K.; Fan, T. Y.; Xu, K.; Tsao, S. W.; Wong, Y. C., Activation of MAPK signaling pathway is essential for Id-1 induced serum independent prostate cancer cell growth. *Oncogene* **2002**, 21, 8498-505.
52. Peverali, F. A.; Ramqvist, T.; Saffrich, R.; Pepperkok, R.; Barone, M. V.; Philipson, L., Regulation of G1 progression by E2A and Id helix-loop-helix proteins. *Embo J* **1994**, 13, 4291-301.
53. Prabhu, S.; Ignatova, A.; Park, S. T.; Sun, X. H., Regulation of the expression of cyclin-dependent kinase inhibitor p21 by E2A and Id proteins. *Mol Cell Biol* **1997**, 17, 5888-96.
54. Nickoloff, B. J.; Chaturvedi, V.; Bacon, P.; Qin, J. Z.; Denning, M. F.; Diaz, M. O., Id-1 delays senescence but does not immortalize keratinocytes. *J Biol Chem* **2000**, 275, 27501-4.
55. Sikder, H. A.; Devlin, M. K.; Dunlap, S.; Ryu, B.; Alani, R. M., Id proteins in cell growth and tumorigenesis. *Cancer Cell* **2003**, 3, 525-30.
56. Tzeng, S. F., Inhibitors of DNA binding in neural cell proliferation and differentiation. *Neurochem Res* **2003**, 28, 45-52.
57. Kondo, T.; Raff, M., The Id4 HLH protein and the timing of oligodendrocyte differentiation. *Embo J* **2000**, 19, 1998-2007.
58. Samanta, J.; Kessler, J. A., Interactions between ID and OLIG proteins mediate the inhibitory effects of BMP4 on oligodendroglial differentiation. *Development* **2004**, 131, 4131-42.



59. Kimura, Y.; Shiozaki, H.; Doki, Y.; Yamamoto, M.; Utsunomiya, T.; Kawanishi, K.; Fukuchi, N.; Inoue, M.; Tsujinaka, T.; Monden, M., Cytoplasmic beta-catenin in esophageal cancers. *Int J Cancer* **1999**, 84, 174-8.
60. Kebebew, E.; Treseler, P. A.; Duh, Q. Y.; Clark, O. H., The helix-loop-helix transcription factor, Id-1, is overexpressed in medullary thyroid cancer. *Surgery* **2000**, 128, 952-7.
61. Chen, H.; Carson-Walter, E. B.; Baylin, S. B.; Nelkin, B. D.; Ball, D. W., Differentiation of medullary thyroid cancer by C-Raf-1 silences expression of the neural transcription factor human achaete-scute homolog-1. *Surgery* **1996**, 120, 168-72; discussion 173.
62. Yokota, Y.; Mansouri, A.; Mori, S.; Sugawara, S.; Adachi, S.; Nishikawa, S.; Gruss, P., Development of peripheral lymphoid organs and natural killer cells depends on the helix-loop-helix inhibitor Id2. *Nature* **1999**, 397, 702-6.
63. Desprez, P. Y.; Hara, E.; Bissell, M. J.; Campisi, J., Suppression of mammary epithelial cell differentiation by the helix-loop-helix protein Id-1. *Mol Cell Biol* **1995**, 15, 3398-404.
64. Parrinello, S.; Lin, C. Q.; Murata, K.; Itahana, Y.; Singh, J.; Krtolica, A.; Campisi, J.; Desprez, P. Y., Id-1, ITF-2, and Id-2 comprise a network of helix-loop-helix proteins that regulate mammary epithelial cell proliferation, differentiation, and apoptosis. *J Biol Chem* **2001**, 276, 39213-9.
65. Mori, S.; Nishikawa, S. I.; Yokota, Y., Lactation defect in mice lacking the helix-loop-helix inhibitor Id2. *Embo J* **2000**, 19, 5772-81.
66. Sablitzky, F.; Moore, A.; Bromley, M.; Deed, R. W.; Newton, J. S.; Norton, J. D., Stage- and subcellular-specific expression of Id proteins in male germ and Sertoli cells implicates distinctive regulatory roles for Id proteins during meiosis, spermatogenesis, and Sertoli cell function. *Cell Growth Differ* **1998**, 9, 1015-24.
67. Langlands, K.; Down, G. A.; Kealey, T., Id proteins are dynamically expressed in normal epidermis and dysregulated in squamous cell carcinoma. *Cancer Res* **2000**, 60, 5929-33.
68. Schaefer, B. M.; Koch, J.; Wirzbach, A.; Kramer, M. D., Expression of the helix-loop-helix protein ID1 in keratinocytes is upregulated by loss of cell-matrix contact. *Exp Cell Res* **2001**, 266, 250-9.

69. Bjorntorp, E.; Parsa, R.; Thornemo, M.; Wennberg, A. M.; Lindahl, A., The helix-loop-helix transcription factor Id1 is highly expressed in psoriatic involved skin. *Acta Derm Venereol* **2003**, 83, 403-9.
70. Ding, B.; Liu, C. J.; Huang, Y.; Yu, J.; Kong, W.; Lengyel, P., p204 protein overcomes the inhibition of the differentiation of P19 murine embryonal carcinoma cells to beating cardiac myocytes by Id proteins. *J Biol Chem* **2006**, 281, 14893-906.
71. Sun, L.; Trausch-Azar, J. S.; Ciechanover, A.; Schwartz, A. L., Ubiquitin-proteasome-mediated degradation, intracellular localization, and protein synthesis of MyoD and Id1 during muscle differentiation. *J Biol Chem* **2005**, 280, 26448-56.
72. Benezra, R.; Rafii, S.; Lyden, D., The Id proteins and angiogenesis. *Oncogene* **2001**, 20, 8334-41.
73. Volpert, O. V.; Pili, R.; Sikder, H. A.; Nelius, T.; Zaichuk, T.; Morris, C.; Shiflett, C. B.; Devlin, M. K.; Conant, K.; Alani, R. M., Id1 regulates angiogenesis through transcriptional repression of thrombospondin-1. *Cancer Cell* **2002**, 2, 473-83.
74. Ouyang, X. S.; Wang, X.; Lee, D. T.; Tsao, S. W.; Wong, Y. C., Over expression of ID-1 in prostate cancer. *J Urol* **2002**, 167, 2598-602.
75. Schoppmann, S. F.; Schindl, M.; Bayer, G.; Aumayr, K.; Dienes, J.; Horvat, R.; Rudas, M.; Gnant, M.; Jakesz, R.; Birner, P., Overexpression of Id-1 is associated with poor clinical outcome in node negative breast cancer. *Int J Cancer* **2003**, 104, 677-82.
76. Schindl, M.; Oberhuber, G.; Obermair, A.; Schoppmann, S. F.; Karner, B.; Birner, P., Overexpression of Id-1 protein is a marker for unfavorable prognosis in early-stage cervical cancer. *Cancer Res* **2001**, 61, 5703-6.
77. Lasorella, A.; Uo, T.; Iavarone, A., Id proteins at the cross-road of development and cancer. *Oncogene* **2001**, 20, 8326-33.
78. Fong, S.; Debs, R. J.; Desprez, P. Y., Id genes and proteins as promising targets in cancer therapy. *Trends Mol Med* **2004**, 10, 387-92.
79. Hasskarl, J.; Duensing, S.; Manuel, E.; Munger, K., The helix-loop-helix protein ID1 localizes to centrosomes and rapidly induces abnormal centrosome numbers. *Oncogene* **2004**, 23, 1930-8.
80. Minn, A. J.; Gupta, G. P.; Siegel, P. M.; Bos, P. D.; Shu, W.; Giri, D. D.; Viale, A.; Olshen, A. B.; Gerald, W. L.; Massague, J., Genes that mediate breast cancer metastasis to lung. *Nature* **2005**, 436, 518-24.
81. Fong, S.; Itahana, Y.; Sumida, T.; Singh, J.; Coppe, J. P.; Liu, Y.; Richards, P. C.; Bennington, J. L.; Lee, N. M.; Debs, R. J.; Desprez, P. Y., Id-1 as a molecular target in

- therapy for breast cancer cell invasion and metastasis. *Proc Natl Acad Sci USA* **2003**, 100, 13543-8.
82. Benezra, R., Role of Id proteins in embryonic and tumor angiogenesis. *Trends Cardiovasc Med* **2001**, 11, 237-41.
83. Ling, M. T.; Wang, X.; Ouyang, X. S.; Xu, K.; Tsao, S. W.; Wong, Y. C., Id-1 expression promotes cell survival through activation of NF-kappaB signalling pathway in prostate cancer cells. *Oncogene* **2003**, 22, 4498-508.
84. Cheung, H. W.; Ling, M. T.; Tsao, S. W.; Wong, Y. C.; Wang, X., Id-1-induced Raf/MEK pathway activation is essential for its protective role against taxol-induced apoptosis in nasopharyngeal carcinoma cells. *Carcinogenesis* **2004**, 25, 881-7.
85. Zhang, X.; Ling, M. T.; Wang, X.; Wong, Y. C., Inactivation of Id-1 in prostate cancer cells: A potential therapeutic target in inducing chemosensitization to taxol through activation of JNK pathway. *Int J Cancer* **2006**, 118, 2072-81.
86. Lin, J. C.; Chang, S. Y.; Hsieh, D. S.; Lee, C. F.; Yu, D. S., The association of Id-1, MIF and GSTpi with acquired drug resistance in hormone independent prostate cancer cells. *Oncol Rep* **2005**, 13, 983-8.
87. Zhang, X.; Ling, M. T.; Wong, Y. C.; Wang, X., Evidence of a novel antiapoptotic factor: Role of inhibitor of differentiation or DNA binding (Id-1) in anticancer drug-induced apoptosis. *Cancer Sci* **2007**.
88. Ling, M. T.; Wang, X.; Zhang, X.; Wong, Y. C., The multiple roles of Id-1 in cancer progression. *Differentiation* **2006**, 74, 481-7.
89. Lasorella, A.; Nosedà, M.; Beyna, M.; Yokota, Y.; Iavarone, A., Id2 is a retinoblastoma protein target and mediates signalling by Myc oncoproteins. *Nature* **2000**, 407, 592-8.
90. Perk, J.; Iavarone, A.; Benezra, R., Id family of helix-loop-helix proteins in cancer. *Nat Rev Cancer* **2005**, 5, 603-14.
91. Swarbrick, A.; Akerfeldt, M. C.; Lee, C. S.; Sergio, C. M.; Caldon, C. E.; Hunter, L. J.; Sutherland, R. L.; Musgrove, E. A., Regulation of cyclin expression and cell cycle progression in breast epithelial cells by the helix-loop-helix protein Id1. *Oncogene* **2005**, 24, 381-9.
92. Rockman, S. P.; Currie, S. A.; Ciavarella, M.; Vincan, E.; Dow, C.; Thomas, R. J.; Phillips, W. A., Id2 is a target of the beta-catenin/T cell factor pathway in colon carcinoma. *J Biol Chem* **2001**, 276, 45113-9.

93. Kleeff, J.; Ishiwata, T.; Friess, H.; Buchler, M. W.; Israel, M. A.; Korc, M., The helix-loop-helix protein Id2 is overexpressed in human pancreatic cancer. *Cancer Res* **1998**, 58, 3769-72.
94. Tang, J.; Gordon, G. M.; Muller, M. G.; Dahiya, M.; Foreman, K. E., Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen induces expression of the helix-loop-helix protein Id-1 in human endothelial cells. *J Virol* **2003**, 77, 5975-84.
95. Lin, C. Q.; Singh, J.; Murata, K.; Itahana, Y.; Parrinello, S.; Liang, S. H.; Gillett, C. E.; Campisi, J.; Desprez, P. Y., A role for Id-1 in the aggressive phenotype and steroid hormone response of human breast cancer cells. *Cancer Res* **2000**, 60, 1332-40.

## II Scope and objectives

The last decade saw much effort in unraveling the role of the Id proteins in numerous physiological events, as diverse as neuronal differentiation, cell cycle control and angiogenesis. In contrast thereto, not much is known about the structural features of these biomolecules. Recently, we have reported that the HLH motifs of the Id proteins are, as predicted by a sequence alignment study, highly helical, while the N-terminal and C-terminal regions display no preferred intrinsic conformation.

The goal of the work presented in the next chapters has been to investigate the structural preferences of the Id HLH motifs by using, on one hand, analogs based on modifications of the native amino acid sequences or, on the other hand, peptides containing covalently linked HLH monomers. The peptides have been synthesized by solid-phase methodology and conformationally characterized by circular dichroism (CD) spectroscopy. In case of the Id2 HLH motif, NMR investigations have also been conducted to provide a more detailed insight into the secondary structure. Furthermore, the possibility of using artificial receptors for the molecular recognition of the Id HLH motifs has been investigated, providing first hints on accessible side chains on the surface of the HLH domain, which might turn out to be useful for the development of selective small-molecule modulators of Id homo- and heterodimerization.

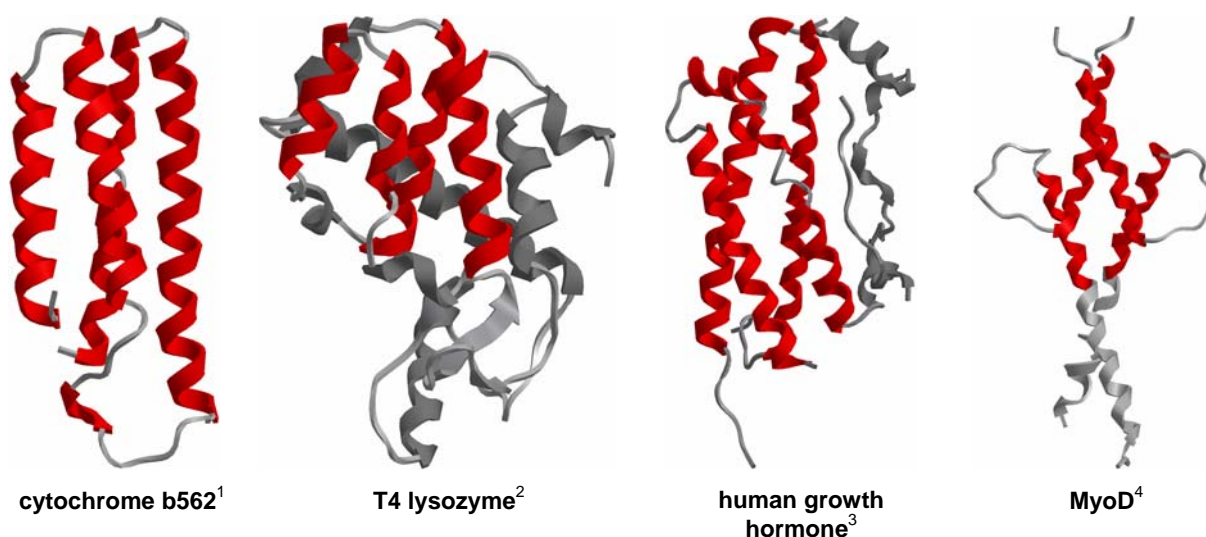


### III Modulating the folding of the Id1 HLH domain by sequence modifications

Many physiological events are based on protein-protein or protein-ligand interactions. The specificity of these processes has to be guaranteed by a uniquely defined protein structure which is based on the amino acid sequence, the so-called primary structure, on secondary structures like  $\alpha$ -helices,  $\beta$ -sheets and turns, and on a complex spatial arrangement of the latter elements *via* intra- or intermolecular interactions, leading to the tertiary or quaternary structure, respectively.

#### III.1 Structural preferences of the HLH proteins

A common tertiary/quaternary structural motif found in proteins is the four-helix bundle (Figure III.1), which consists of four helices packed together, either in a parallel or an anti-parallel orientation.



**Figure III.1.** Four-helix bundle containing proteins. Marked in red are the four-helix bundles.

The four-helix bundle occurs in nature either as isolated fold (cytochrome b562) or as a domain of a larger protein (T4 lysozyme), but also as the result of interacting polypeptide chains (MyoD).<sup>5</sup> Proteins containing four-helix folds take part in diverse functions like

As no structural information about the Id proteins is available so far, it remains unclear if they adopt a structure similar to the one shown by their parent HLH proteins. However, a first hint on the conformational properties of the Id proteins is given by the identification of heptad repeats (*abcdefg*) within the amino acid sequence corresponding to their dimerization domain. In these heptads, positions *a* and *d* are predominantly occupied by hydrophobic amino acids, while at the other positions polar residues prevail (Scheme III.1). This pattern is found in amphiphilic helices, where a hydrophilic surface is opposite to a hydrophobic one, thus allowing the formation of a hydrophobic core upon dimerization, which contributes to the overall stabilization of a protein. An indirect evidence for such a folding mechanism in the Id protein family came from the observation of Id2 homodimerization *in vivo*.<sup>10</sup> Crucial for this dimerization were shown to be the HLH domain, a cysteine residue in the first helix (Cys<sup>42</sup>) and the regions adjacent to the HLH domain.



The first steps towards a comprehensive, systematic structural characterization of the HLH proteins were taken by Chavali et al. who presented a comparative study of several HLH proteins, among them the Id proteins, based on the known crystal structures of E47 and MyoD.<sup>11</sup> A high similarity between all HLH proteins was observed, and the loop region was not found to contribute to the dimer formation. The loop region seems to be rather important to drive the appropriate spatial orientation of the two helices in the HLH motif.

Another interesting feature of the HLH proteins is that, despite the high similarity of their HLH domains, their individual aggregation behavior differs: the myogenic transcription factors myogenin and MyoD (class II) were reported to form both higher order complexes but with different average size, being significantly smaller in the case of MyoD (tetramers).<sup>12</sup> However, another report even describes the presence of MyoD micelles consisting of 100 monomers.<sup>13</sup> In contrast thereto, the ubiquitous factor E12 (class I) exists only as a dimer or a mixture of monomers and dimers<sup>12</sup>, while the Id proteins probably oligomerize, as the isolated HLH domains were found to form stable tetramers in the concentration range from 20 to 300  $\mu\text{M}$ .<sup>14</sup> The equilibrium between monomeric and oligomeric species might constitute an important regulation mechanism for the diverse biological functions of the HLH proteins, as only their monomeric forms are able to interact with their binding partners. The reason for the different behavior of similar HLH motifs might be found in the hydrophobic core: indeed, it was shown that the oligomeric state of a LZ protein can be controlled by mutating residues in its hydrophobic core.<sup>15</sup> As every HLH protein possesses a more or less unique functional profile, it is likely that despite a rather similar overall structure, the specificity of the diverse functionalities is governed by subtle, but crucial structural changes within this protein family, which also influence the oligomerization behavior. Therefore, to fully understand the implication of the structure on the physiological function, it would be necessary to investigate each member individually.

To evaluate the effect of primary structure alterations on the conformational preferences of the Id1 HLH domain, in the presented work we synthesized a series of peptide analogs of the Id1 HLH motif containing mutations in the loop region as well as the retro-sequence of helix-1 (Table III.1). Additionally, analogs bearing substitutions of fully conserved residues were prepared. All these analogs were then investigated by CD spectroscopy, providing information on the impact of each modification on the overall structure of the Id1 HLH region.

**Table III.1.** Sequences of the synthetic Id1 HLH analogs.

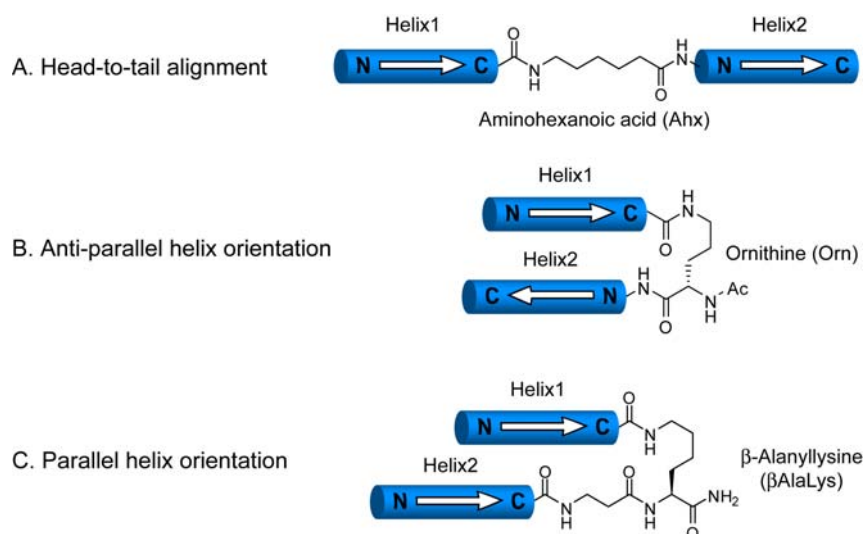
No.	Description	Sequence
III.1	Ac-(66-106)	Ac-LYDMNGCY SRLKELVPTLPQNRKVS KVEILQHVIDYIRDLQ-NH <sub>2</sub>
III.2	Ac-[retro <sup>66-80</sup> , des <sup>81</sup> ]- (66-106)	Ac-VLEKLRSYCGNMDYL-TLPQNRKVS-KVEILQHVIDYIRDLQ-NH <sub>2</sub>
III.3	Ac-[Ahx <sup>82-90</sup> ]- (66-106)	Ac-LYDMNGCY SRLKELVP-Ahx-KVEILQHVIDYIRDLQ-NH <sub>2</sub>
III.4	Ac-[retro <sup>66-80</sup> , des <sup>81</sup> , Ahx <sup>82-90</sup> ]- (66-106)	Ac-VLEKLRSYCGNMDYL-Ahx-KVEILQHVIDYIRDLQ-NH <sub>2</sub>
III.5	Ac-[Orn <sup>82-90</sup> ]- (66-106)	Ac-LYDMNGCY SRLKELVP-Orn (Ac) -KVEILQHVIDYIRDLQ-NH <sub>2</sub>
III.6	Ac-[retro <sup>66-80</sup> , des <sup>81</sup> , Orn <sup>82-90</sup> ]- (66-106)	Ac-VLEKLRSYCGNMDYL-Orn (Ac) -KVEILQHVIDYIRDLQ-NH <sub>2</sub>
III.7	Ac-(89-106)-βAlaLys(retro <sup>66-80</sup> )	Ac-VSKVEILQHVIDYIRDLQ-βAlaLys (Ac-VLEKLRSYCGNMDYL) -NH <sub>2</sub>
III.8	Ac-(89-106)-βAlaLys(66-83)	Ac-VSKVEILQHVIDYIRDLQ-βAlaLys (Ac-LYDMNGCY SRLKELVPTL) -NH <sub>2</sub>
III.9	Ac-(89-106)-βAlaLys	Ac-VSKVEILQHVIDYIRDLQ-βAlaLys-NH <sub>2</sub>
III.10	Ac-[retro <sup>69-80</sup> , ProGly <sup>81-90</sup> ]- (69-103)	Ac-VLEKLRSYCGNM-PG-KVEILQHVIDYIR-NH <sub>2</sub>
III.11	Ac-[F <sup>73, 101</sup> ]- (66-106)	Ac-LYDMNGC <b>F</b> SRLKELVPTLPQNRKVS KVEILQHVID <b>F</b> IRDLQ-NH <sub>2</sub>
III.12	Ac-[F <sup>101</sup> ]- (66-106)	Ac-LYDMNGCY SRLKELVPTLPQNRKVS KVEILQHVID <b>F</b> IRDLQ-NH <sub>2</sub>
III.13	Ac-[F <sup>73</sup> ]- (66-106)	Ac-LYDMNGC <b>F</b> SRLKELVPTLPQNRKVS KVEILQHVIDYIRDLQ-NH <sub>2</sub>

## III.2 Peptide design

### III.2.1 Loop surrogates and retro-sequences

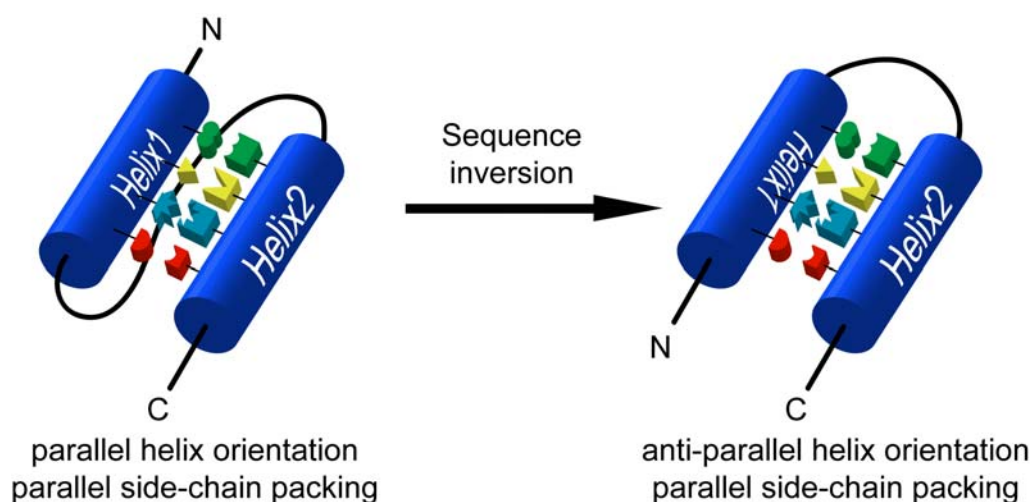
To elucidate the importance of interhelical contacts for the stabilization of the Id1 HLH fold, peptides containing the native Id1 HLH helices connected by different linkers were synthesized (Table III.1). The incorporation of rigid linkers in place of the native nine residue long loop should restrict the number of possible orientations of the two helices to each other (Scheme III.2). By comparison of the secondary structure elements composition calculated by the CONTIN algorithm for each analog out of the CD data with the one of the native Id HLH motif, insights into the preferred orientation of the two helices to each other can be gained. Another set of peptides containing the same loop surrogates were synthesized, but in these peptides the native sequence of helix-1 was inverted (retro-helix-1).

**Scheme III.2.** Loop substitutions and their putative helix-1/helix-2 orientations induced by the used loop surrogates.



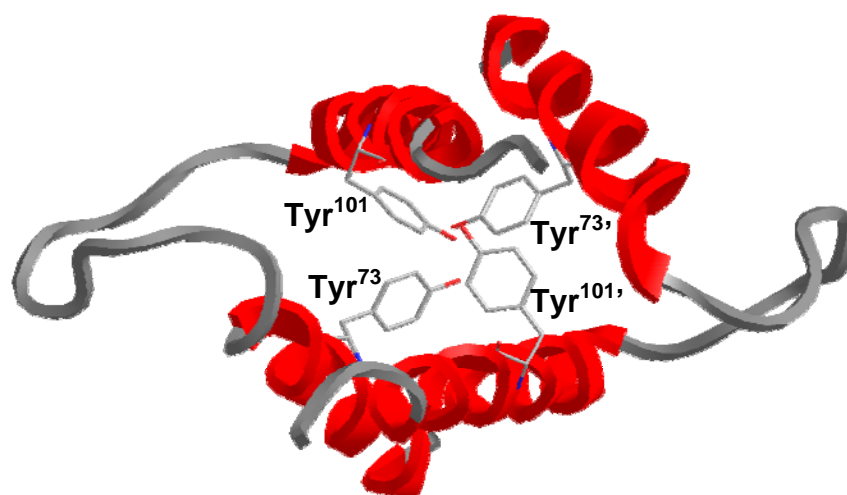
This approach allows the investigation of the influence of the side-chain packing on the overall structure. In order to maintain its favored side-chain packing, the peptide is forced to change its native helix orientation, as illustrated in Scheme III.3. Obviously, this change will be prevented by a rigid linker substituting the more flexible, native loop region.

**Scheme III.3.** Influence of a retro-sequence on the preference for an anti-parallel or parallel conformation.



### III.2.2 Phenylalanine scan

A number of different interactions are involved in the stabilization of the three-dimensional structure of peptides and proteins including hydrogen bonds, salt bridges,  $\pi$ - $\pi$  stacking and cation- $\pi$  interactions. The latter occur between a cationic group like the side-chains of lysine, arginine, or histidine and the  $\pi$ -face of the aromatic side chains of phenylalanine, tyrosine or tryptophan. As the primary sequences of the Id HLH motifs contain several basic residues as well as two conserved tyrosine residues, cation- $\pi$  interactions might contribute to the stabilization of both secondary and tertiary structures. The tyrosine side chains may be involved in contacts using their phenyl ring ( $\pi$ - $\pi$  and cation- $\pi$  interactions) as well as their OH group (H-bonds). To investigate the role of the aromatic residues on the folding, peptides reproducing the HLH motif of the Id1 protein, but containing substitutions of the native tyrosine residues by phenylalanine were synthesized and characterized by CD spectroscopy (Table III.1). For the substitutions, positions 73 and 101 were chosen, as these positions are assumed to be involved in self, as well as in partner interactions with the HLH proteins (Figure III.2).<sup>11</sup> Moreover, position 73 is occupied by phenylalanine in many other members of the HLH family, including the ubiquitous transcription factors, suggesting a role in the regulation of binding specificity.



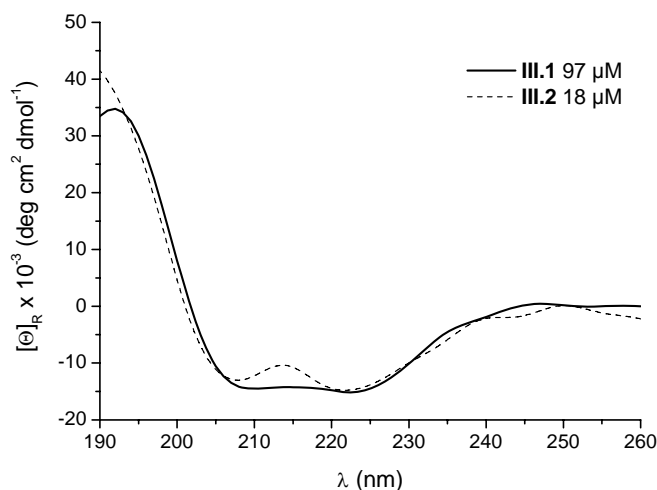
**Figure III.2.** Top view of the MyoD HLH dimer showing the potential location of the tyrosine positions in the hydrophobic core of the Id1 HLH motif.

### III.3 Results

#### III.3.1 Loop surrogates

##### *Native HLH motifs*

Peptides **III.1** and **III.2** contain both the native sequence of the Id1 HLH motif, but in peptide **III.2** the sequence of the N-terminal helix is inverted. As this modification led to a dramatically reduced solubility, probably due to strong aggregation, peptide **III.2** could only be measured at a maximum concentration of 18  $\mu\text{M}$ . Like the native HLH motif (peptide **III.1**), peptide **III.2** shows a CD spectrum characteristic for a helical structure, with two minima at 222 and 208 nm, respectively, and a maximum below 200 nm (Figure III.3). The ratio of the ellipticity of the two minima,  $R = [\Theta]_{222}/[\Theta]_{208}$ , is 1.14, which suggests the presence of interacting helices and is even higher than in the native HLH motif (1.07). Also the helix content of 54% (corresponding to 21 out of 40 helical amino acids) slightly surpasses the one of the native sequence.



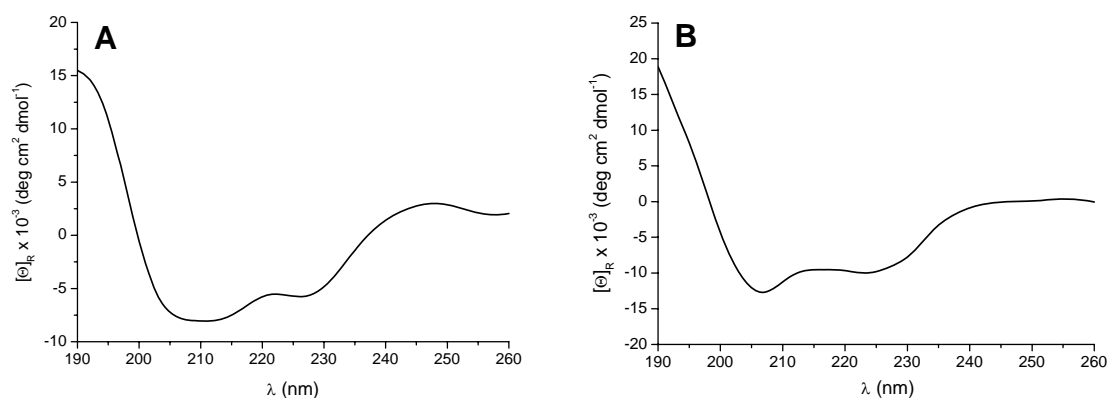
**Figure III.3.** CD spectra of peptides **III.1** and **III.2** in 0.1 M phosphate buffer (pH 7.3).

As the only variation between the two peptides is the retro-sequence of the N-terminal helix, which should induce an inversion of the helix orientation to guarantee an efficient side-chain packing, the fact that the helical character of peptide **III.2** not only is maintained, but even improved suggests that the peptide backbone underwent a switch to allow a native-like side-

chain packing, which additionally led to a more favorable orientation of the two helix dipoles (Scheme III.3).

### *Ahx loop containing peptides*

Peptides **III.3** and **III.4** contain an aminohexanoyl (Ahx) unit as a linker (spacer) between the two helices of the HLH motif. This loop surrogate should primarily reduce the intramolecular interactions between the two helical flanking regions. In peptide **III.4** the native sequence of the N-terminal helix is inverted, in order to control whether any interhelix contact has been suppressed.



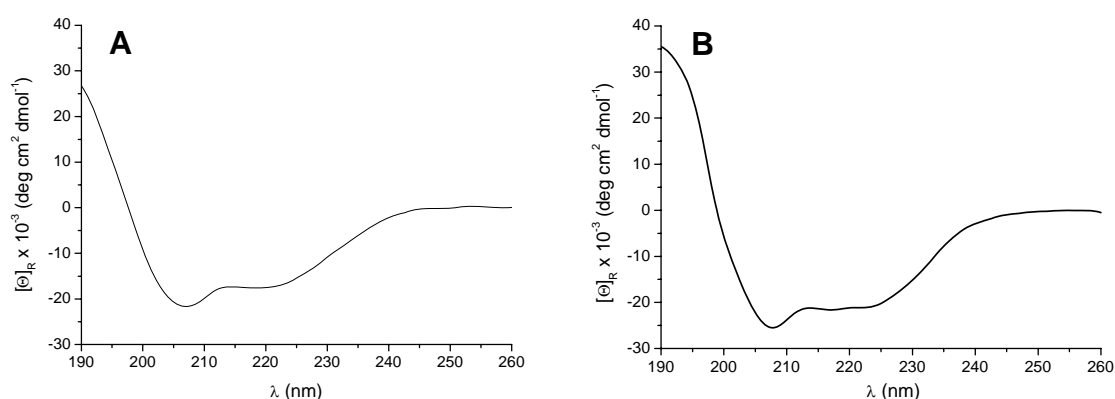
**Figure III.4.** CD spectra of peptide **III.3** (52  $\mu$ M, A) and peptide **III.4** (45  $\mu$ M, B) in 0.1 M phosphate buffer (pH 7.3).

Peptides **III.3** and **III.4** show CD spectra resembling those of a helix, but display some peculiarities (Figure III.4): in case of peptide **III.3**, the minimum at 208 nm is quite broad and the second minimum, usually located at 222 nm in case of a helical structure, is shifted to a longer wavelength (Figure III.4A). Moreover, both minima are only moderately intense. In case of peptide **III.4** the helical character of the CD curve is more pronounced, with the characteristic minima at 222 and 208 nm and a maximum below 200 nm. CONTIN analysis yields similar helix contents for both analogs: 30% for **III.3** and 33% for **III.4** (~10 helical residues). In addition thereto, also the  $\beta$ -strand, turns and disordered fractions are comparable. The poor helix content suggests that the two putative flanking helices are probably not in contact with each other, which is also supported by R values lower than 1 (0.7 for peptide **III.3** and 0.8 for peptide **III.4**). However, as the presence of the retro-sequence of helix-1 resulted in a slightly increased helical character in comparison to the native one, residual

helix-helix interactions can be presumed. This would correspond to the situation encountered for peptide **III.2**. Nevertheless, the Ahx containing peptide **III.4** features a very short and rather rigid linker, thus generating a sterical hindrance for the folding of the two helices towards each other. Accordingly, the small number of helical residues for both peptides supports the presence of a single helical segment, which is likely to be localized within the C-terminal region that has been previously shown to possess a high intrinsic helix propensity.<sup>16</sup>

### *Ornithine loop containing peptides*

Peptides **III.5** and **III.6** contain an ornithine loop in place of the native nine residue long loop sequence. Additionally, in peptide **III.6** the sequence of the first helix is inverted. The  $\delta$ -ornithine linker was chosen, as it could enable the formation of a mirror-image  $\beta$ -turn similar to the one observed with a D-prolylglycine unit<sup>17</sup>; the two helices should thus be able to fold towards each other, producing an anti-parallel orientation of the helix dipoles. Both CD spectra present two minima at 222 and 208 nm, accompanied by a maximum below 200 nm, which are characteristic for a helical structure (Figure III.5). The curve of peptide **III.6** containing the retro helix-1 is moderately superior in intensity, which is illustrated by the 60% helix content obtained by the CONTIN analysis opposed to 42% for peptide **III.5**. This implies that in peptide **III.5** 14 residues out of 32 are in a helical conformation, while in peptide **III.6** 19 helical residues are present. The latter analog displays nearly the same number of helical residues as the native HLH motif **III.1**. By contrast, in comparison to the Ahx containing peptides **III.3** and **III.4**, the ornithine containing peptides reveal a clearly superior amount of helical structure, which can be explained by the ability of  $\delta$ -ornithine to induce and stabilize a turn-like motif that improves the interaction of the two flanking helices.



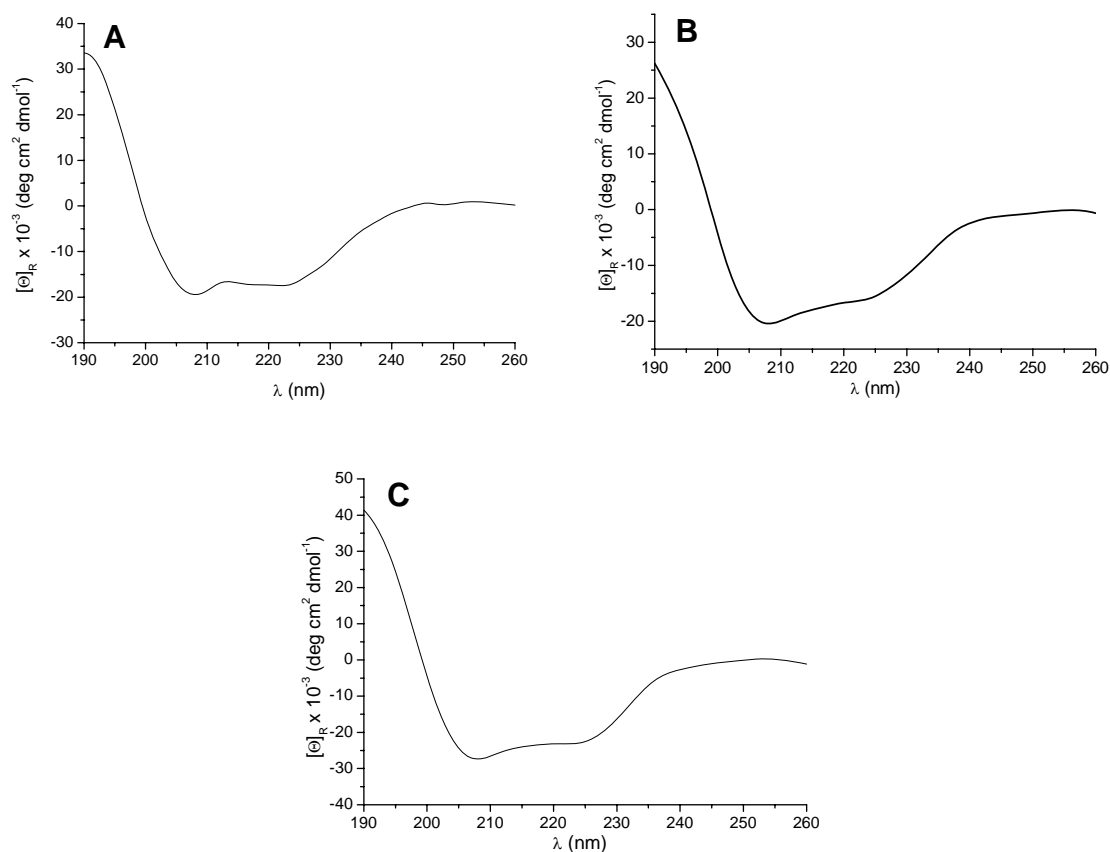
**Figure III.5.** CD spectra of peptide **III.5** (66  $\mu$ M, A) and peptide **III.6** (65  $\mu$ M, B) in 0.1 M phosphate buffer (pH 7.3).

Furthermore, the side-chain packing in peptide **III.6**, which corresponds to the one obtained for a parallel orientation of the two native helices, seems to be favored, as indicated by the additional five helical residues compared to peptide **III.5**. Another striking feature of analogs **III.5** and **III.6** is their reduced  $\beta$ -strand fraction (2%), compared to the one obtained for the native HLH motif (11%), which is likely to reflect a change in the oligomerization properties. This might also be the reason for the reduced R values found for peptides **III.5** and **III.6** (0.80 and 0.83, respectively) with respect to those of the native sequences **III.1** and **III.2**. However, the comparison between peptides **III.5** and **III.6** shows that the inter-helical side-chain packing can be assumed to play an important role in stabilizing the first helix, as the non optimal packing in peptide **III.5** results in a diminished number of helical residues.

### *$\beta$ -AlaLys loop containing peptides*

In peptides **III.7** and **III.8** the native loop sequence is replaced by a  $\beta$ -alanyllysine loop. This linker should provide the necessary proximity between the two helices to allow them to interact with each other. Moreover, a parallel orientation between the two helix dipoles should be favored (Scheme III.2C). In peptide **III.7** the retro sequence of helix-1 (retro<sup>66-80</sup>) has been introduced to investigate the importance of the side-chain helix packing. By contrast, in peptide **III.8** the native sequence 66-83 is present, which contains not only helix-1, but also residues located at the junction with the loop (P<sup>81</sup>TL<sup>83</sup>). Peptide **III.9** was obtained as side-product during the synthesis of peptide **III.8**, due to the incomplete deprotection of the  $\epsilon$ -amino group of the lysine residue (incomplete Mtt cleavage), and proved to be useful as a reference to study the C-terminal helix-2 alone. The CD curves of the three peptides are characteristic for a  $\alpha$ -helical structure (Figure III.6). CONTIN analysis provided a helical fraction of 48% for **III.7**, 46% for **III.8** and 63% for **III.9** corresponding to 17 out of 35 or 38 helical residues for **III.7** and **III.8**, respectively, and to 12 out of 20 helical residues for **III.9**. The R values are as follows: 0.90 for **III.7**, 0.80 for **III.8** and 0.85 for **III.9**. Interestingly, peptide **III.9** that lacks the N-terminal helix-1 is still exhibiting a highly helical structure, which again confirms the intrinsic helix propensity of the helix-2. However, conjugation of the retro or native sequence of helix-1 (peptide **III.7** and **III.8**, respectively) is accompanied by an increase in the number of helical residues of about five.



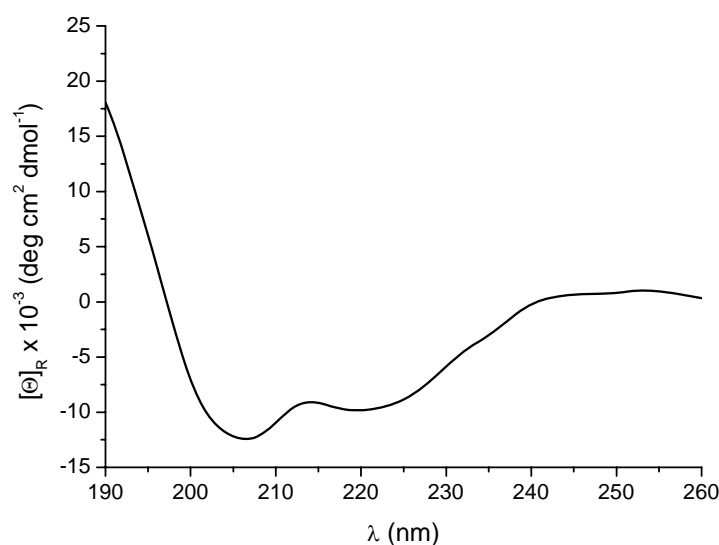


**Figure III.6.** CD spectra of peptide **III.7** (53  $\mu$ M, A), **III.8** (50  $\mu$ M, B) and **III.9** (44  $\mu$ M, C) in 0.1 M phosphate buffer (pH 7.3).

Based on the behavior of the HLH analogs **III.2**, **III.4** and **III.6** described above, a minor helicity was expected for the  $\beta$ -AlaLys loop containing peptide **III.7**, as it was designed to yield parallel helix dipoles with consequent anti-parallel side-chain packing, whereas in peptide **III.8** the helix dipoles as well as the side-chain packing should be parallel. Thereby, it should be noted that the two peptides do not contain exactly the same connecting motif between the two helices: while peptide **III.7** contains just the  $\beta$ -AlaLys moiety, in peptide **III.8** the first three N-terminal residues of the native loop were included to produce the larger motif  $\beta$ -AlaLys(P<sup>81</sup>TL<sup>83</sup>). In the last case the presence of a proline might result in an increased flexibility of the linker and, consequently, promote the dynamics of the folding. In the case of peptide **III.7** the loop is shorter but, nevertheless, might still be long enough (contrarily to the  $\delta$ -ornithine loop in peptide **III.5**) to allow a partially anti-parallel juxtaposition of the two helices, which would lead not only to a parallel side-chain packing like in peptide **III.8**, but also to a favorable anti-parallel helix dipole orientation.

### ***Prolylglycine loop containing peptide***

Peptide **III.10** contains a prolylglycine as a surrogate for the native loop sequence, which connects the retro sequence of helix-1 to the native sequence of helix-2 (both helices were shortened by three residues to evaluate the importance of the helix length). The CD curve exhibits the characteristic bands for a helical structure, but slightly blue-shifted, which might reflect an additional contribution of turns (Figure III.7).



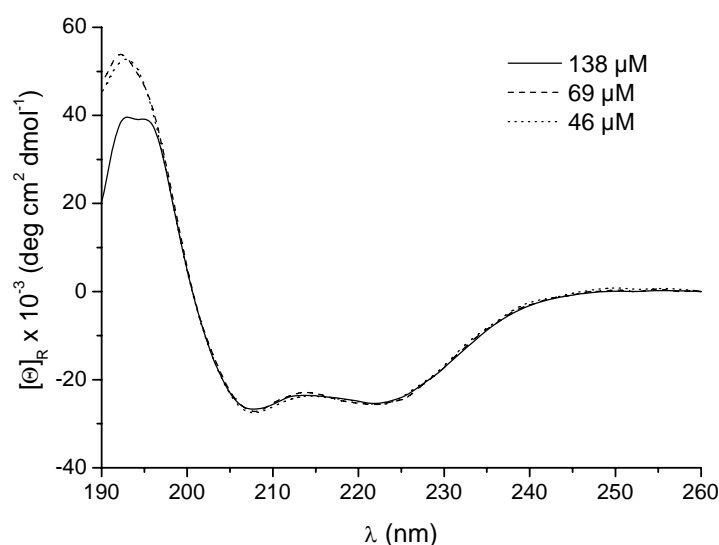
**Figure III.7.** CD spectrum of peptide **III.10** (84  $\mu$ M) in 0.1 M phosphate buffer (pH 7.3).

Accordingly, CONTIN analysis revealed only 25% helicity for peptide **III.10**, corresponding to 7 helical amino acids out of 27, and the ratio helix/turns was 2:1. The low helical content is probably due to a combination of negative effects, including the shortening of the helices as well as of the loop.

### **III.3.2 Phenylalanine scan**

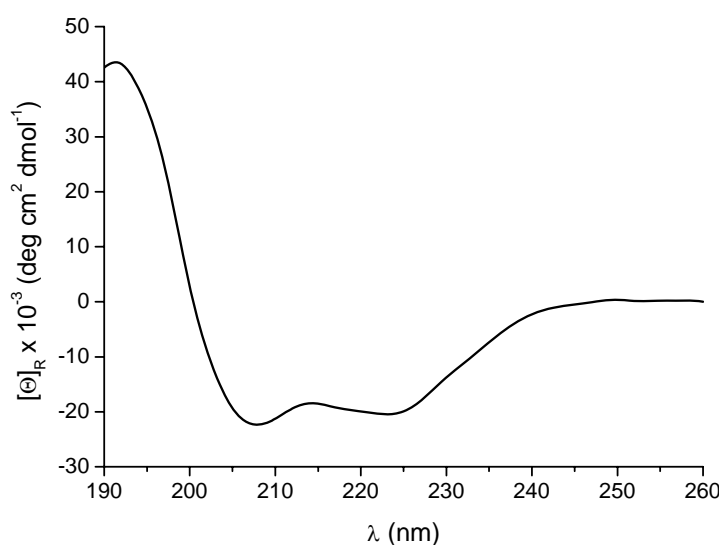
In peptide **III.11** the tyrosines at positions 73 and 101 were substituted by phenylalanine. The CD spectra recorded at different peptide concentrations from 46  $\mu$ M to 138  $\mu$ M are superimposable and characterized by minima at 208 nm and 222 nm, indicating the presence of a helical structure (Figure III.8). CONTIN analysis of the CD spectrum of the 46  $\mu$ M peptide solution reveals 78% helicity, corresponding to 32 helical residues and

matching the number of helical residues predicted for the HLH motif by sequence alignment and homology studies.<sup>11, 18</sup> The ratio R is 0.93.



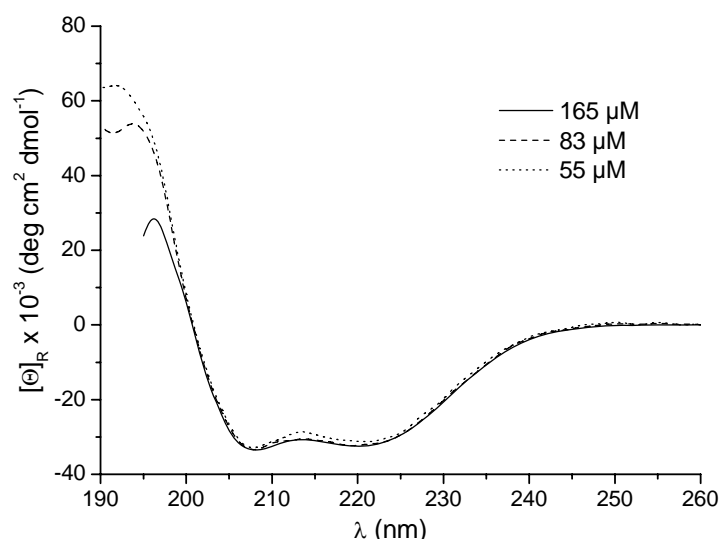
**Figure III.8.** CD spectra of peptide **III.11** at different concentrations in 0.1 M phosphate buffer (pH 7.3).

In peptide **III.12** only the tyrosine at position 101 in the second helix was replaced with phenylalanine. Due to low solubility in phosphate buffer, the peptide could only be measured at a concentration of 58  $\mu$ M. The obtained CD spectrum (Figure III.9) is indicative for a helical structure, as it shows the characteristic bands at 208 and 222 nm. CONTIN analysis calculated a helicity of 58%, which is clearly reduced compared to peptide **III.11**. Nonetheless, the ratio R is in the same range and accounts for 0.91.



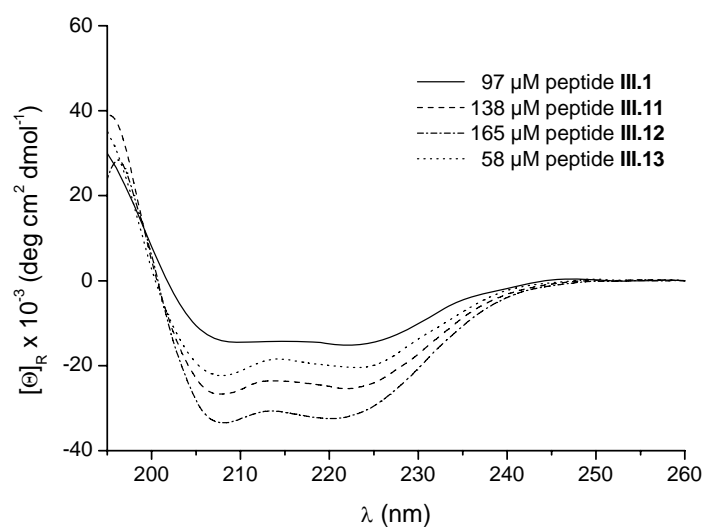
**Figure III.9.** CD spectrum of peptide **III.12** at the concentration of 58  $\mu$ M in 0.1 M phosphate buffer (pH 7.3).

Peptide **III.13** contains a tyrosine/phenylalanine substitution at position 73 in the first helix. Similarly to peptides **III.11** and **III.12**, the obtained CD spectrum is characteristic for a helical structure (Figure III.10). The overall helicity is 85%, as calculated by CONTIN analysis, representing the highest one obtained for the peptides presented in this study. The ratio R is similar to those obtained for peptides **III.11-12** and accounts for 0.95.



**Figure III.10.** CD spectra of peptide **III.13** at different concentrations in 0.1 M phosphate buffer (pH 7.3).

A superposition of the CD spectra of the three tyrosine-modified peptides **III.11-13** with the one obtained for the native sequence clearly shows that the replacement of a phenol group with a phenyl group leads to an increment in the  $\alpha$ -helix content (Figure III.11).



**Figure III.11.** Superposition of the CD spectra of peptides **III.1**, **III.11**, **III.12** and **III.13**.

### III.4 Discussion

Regarding the obtained CD data for the synthesized Id1 HLH peptides containing loop variations, it is evident that the overall helicity of the HLH fold does not only stem from contributions of the individual helices, but is also based on an interhelical stabilization. This stabilization is strongly dependent upon the correct orientation of the two helices towards each other, as shown by the conformational preferences of the synthesized loop analogs (Table III.2). Peptide **III.9**, the isolated helix-2, possesses already remarkable helicity, confirming our previous structural studies on the isolated helices of the Id HLH motif.<sup>16</sup> This underlines the hypothesis of helix-2 serving as nucleation seed for the folding of the HLH domain. The loop, on the other hand, presumably plays the important role to let the two flanking regions approach and juxtapose in a proper way for interactions, which consequently results in the formation and/or stabilization of helix-1.

**Table III.2.** Results of the CONTIN analysis for the secondary structure elements composition of the measured peptides.  $n$  – number of amino acids,  $n_{\text{helical}}$  – calculated number of helical amino acids,  $n_{\text{theor}}$  – theoretical number of helical residues corresponding to those predicted for the HLH motif by sequence alignment,  $R$  – ratio between the ellipticity minima at 222 nm and 208 nm.

No.	c in $\mu\text{M}$	n	helix	$\beta$ -strands	turns	unordered	$n_{\text{helical}}$	$n_{\text{theor}}$	R
<b>III.1</b>	97	41	49%	11%	14%	25%	20	32	1.07
<b>III.2</b>	18	40	54%	16%	7%	23%	21	31	1.14
<b>III.3</b>	52	33	30%	14%	21%	36%	10	33	0.70
<b>III.4</b>	45	32	33%	10%	17%	40%	11	31	0.80
<b>III.5</b>	66	33	42%	2%	9%	47%	14	32	0.80
<b>III.6</b>	65	32	60%	2%	11%	27%	19	31	0.83
<b>III.7</b>	53	35	48%	3%	11%	37%	17	33	0.90
<b>III.8</b>	50	38	46%	3%	16%	35%	17	36	0.80
<b>III.9</b>	44	20	63%	2%	10%	25%	12	18	0.85
<b>III.10</b>	84	27	25%	15%	12%	49%	7	25	0.79
<b>III.11</b>	46	41	77%	1%	10%	12%	32	32	0.94
<b>III.12</b>	58	41	64%	0%	8%	28%	26	32	0.91
<b>III.13</b>	55	41	85%	0%	9%	6%	35	32	0.95

The shortest Id HLH analog that showed a number of helical residues close to the one found in the native full-length HLH motif is peptide **III.6** containing an ornithine loop and the retro sequence of helix-1. These modifications should favor an anti-parallel orientation of the helix dipoles and a parallel packing of the side chains. Compared thereto, in peptide **III.5**, the

ornithine analog presenting the native sequence of helix-1, the amount of unordered structure is remarkably increased at the expense of the helix content. Taken together, these results suggest the preference for a parallel helix side-chain packing. Peptides **III.7** and **III.8**, the  $\beta$ -AlaLys-loop containing analogs, should prefer a parallel helix orientation, resulting in an anti-parallel side-chain packing for peptide **III.7**, while in a parallel one for peptide **III.8**. Unexpectedly, the two peptides displayed a similar structural composition, which can be explained by the fact that the putative parallel orientation could be converted into an anti-parallel one. A similar behavior was observed in case of peptides **III.1** and **III.2** (Scheme III.3), in which the modification of the native HLH sequence by inverting helix-1 showed almost no effect on the helix content. Interestingly, by comparing peptide **III.8** with peptide **III.6**, the superior helicity of the latter might reflect a positive contribution of the anti-parallel alignment of the helix dipoles.

The lowest amount of helicity was obtained for peptides **III.3-4**, containing the Ahx moiety, and for peptide **III.10**, containing the PG dipeptide. This suggests that the introduction of short elements which do not elicit the possibility of the two flanking regions to get into proximity is highly unfavorable, thus confirming the importance of interhelical contacts.

Most remarkably, the peptides containing tyrosine-phenylalanine substitutions showed an increase in helicity compared to the native HLH motif, suggesting that the phenylalanine residues are able to increase the stabilization of the structure, presumably by improving the packing of the hydrophobic core. As Chavali et al. reported, the modified positions 73 and 101 are found at the junction of the two helices and are assumedly involved in self as well as in partner interactions.<sup>11</sup> These interactions are likely to be mediated by  $\pi$ - $\pi$  or cation- $\pi$  interactions which are improved after substitution with phenylalanine, according to the reports by Pletneva et al., who determined a better cation- $\pi$  binding propensity for phenylalanine compared to tyrosine.<sup>19</sup> The highest stabilization was obtained for peptide **III.13**, which bears a Phe  $\rightarrow$  Tyr substitution at position 73. The substitution at position 101 does not seem to be equally favored, as both the Phe<sup>101</sup> containing peptides **III.11** and **III.12** showed a reduced helicity compared to analog **III.13**. This might suggest a superior role of position 73 in stabilizing the hydrophobic core of the fold. Moreover, all three Phe  $\rightarrow$  Tyr analogs feature reduced interhelical interactions, as indicated by R values  $< 1$ . This might be related to the modified hydrophobicity of the core, which might modulate the oligomerization behavior, as also observed by Harbury et al. for another protein system.<sup>15</sup>

Furthermore, by comparing the obtained R values it can be speculated that this dichroic parameter presents an indicator of aggregation tendency, as the property of analog **III.2** to

aggregate already at very low concentrations was reflected by the highest R value obtained (1.14). Most of the other analogs exhibited better solubility and were characterized by R values lower than 1.

### III.5 References

1. Lederer, F.; Glatigny, A.; Bethge, P. H.; Bellamy, H. D.; Matthew, F. S., Improvement of the 2.5 Å resolution model of cytochrome b562 by redetermining the primary structure and using molecular graphics. *J Mol Biol* **1981**, 148, 427-48.
2. Nicholson, H.; Tronrud, D. E.; Becktel, W. J.; Matthews, B. W., Analysis of the effectiveness of proline substitutions and glycine replacements in increasing the stability of phage T4 lysozyme. *Biopolymers* **1992**, 32, 1431-41.
3. Clackson, T.; Ultsch, M. H.; Wells, J. A.; de Vos, A. M., Structural and functional analysis of the 1:1 growth hormone:receptor complex reveals the molecular basis for receptor affinity. *J Mol Biol* **1998**, 277, 1111-28.
4. Ma, P. C.; Rould, M. A.; Weintraub, H.; Pabo, C. O., Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* **1994**, 77, 451-9.
5. Kamtekar, S.; Hecht, M. H., Protein Motifs. 7. The four-helix bundle: what determines a fold? *Faseb J* **1995**, 9, 1013-22.
6. Oakley, M. G.; Hollenbeck, J. J., The design of antiparallel coiled coils. *Curr Opin Struct Biol* **2001**, 11, 450-7.
7. Ellenberger, T.; Fass, D.; Arnaud, M.; Harrison, S. C., Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev* **1994**, 8, 970-80.
8. Ferre-D'Amare, A. R.; Prendergast, G. C.; Ziff, E. B.; Burley, S. K., Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **1993**, 363, 38-45.
9. Ferre-D'Amare, A. R.; Pognonec, P.; Roeder, R. G.; Burley, S. K., Structure and function of the b/HLH/Z domain of USF. *Embo J* **1994**, 13, 180-9.
10. Liu, J.; Shi, W.; Warburton, D., A cysteine residue in the helix-loop-helix domain of Id2 is critical for homodimerization and function. *Biochem Biophys Res Commun* **2000**, 273, 1042-7.

11. Chavali, G. B.; Vijayalakshmi, C.; Salunke, D. M., Analysis of sequence signature defining functional specificity and structural stability in helix-loop-helix proteins. *Proteins* **2001**, 42, 471-80.
12. Farmer, K.; Catala, F.; Wright, W. E., Alternative multimeric structures affect myogenin DNA binding activity. *J Biol Chem* **1992**, 267, 5631-6.
13. Laue, T. M.; Starovasnik, M. A.; Weintraub, H.; Sun, X. H.; Snider, L.; Klevit, R. E., MyoD forms micelles which can dissociate to form heterodimers with E47: implications of micellization on function. *Proc Natl Acad Sci USA* **1995**, 92, 11824-8.
14. Fairman, R.; Beran-Steed, R. K.; Anthony-Cahill, S. J.; Lear, J. D.; Stafford, W. F., 3rd; DeGrado, W. F.; Benfield, P. A.; Brenner, S. L., Multiple oligomeric states regulate the DNA binding of helix-loop-helix peptides. *Proc Natl Acad Sci USA* **1993**, 90, 10429-33.
15. Harbury, P. B.; Zhang, T.; Kim, P. S.; Alber, T., A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants. *Science* **1993**, 262, 1401-7.
16. Colombo, N.; Cabrele, C., Synthesis and conformational analysis of Id2 protein fragments: impact of chain length and point mutations on the structural HLH motif. *J Pept Sci* **2006**, 12, 550-8.
17. Nowick, J. S.; Brower, J. O., A new turn structure for the formation of beta-hairpins in peptides. *J Am Chem Soc* **2003**, 125, 876-7.
18. Wibley, J.; Deed, R.; Jasiok, M.; Douglas, K.; Norton, J., A homology model of the Id-3 helix-loop-helix domain as a basis for structure-function predictions. *Biochim Biophys Acta* **1996**, 1294, 138-46.
19. Pletneva, E. V.; Laederach, A. T.; Fulton, D. B.; Kostic, N. M., The role of cation-pi interactions in biomolecular association. Design of peptides favoring interactions between cationic and aromatic amino acid side chains. *J Am Chem Soc* **2001**, 123, 6232-45.



## IV Structural preferences of the Id1 HLH dimer

### IV.1 Introduction

Disulfide bonds play a crucial role in the tertiary and quaternary structure of proteins.<sup>1</sup> By restraining the conformational freedom of peptides and proteins, they contribute considerably to the stabilization of biologically active conformations. Furthermore, intramolecular disulfide bridges can be valuable tools for introducing a conformational constraint in chemically synthesized peptides. Another widely used application of the disulfide bond is the formation of covalent dimers: one example being the covalently linked MyoD homodimer, which was obtained by oxidation of the single, native cysteine residue located in the C-terminal helix-2. This oxidized dimer was then used to study the solution structure of MyoD by NMR spectroscopy<sup>2</sup>, as strong aggregation of the reduced form caused by formation of higher order oligomers hindered a reasonable NMR investigation. Intermolecular disulfide bridging has also been applied for the stabilization of peptidic  $\beta$ -helices<sup>3</sup> or coiled coils in the HIV gp41 protein<sup>4</sup>. Recently, Gellman and coworkers used cysteine bridges to elucidate the preferred helical orientation of a *de novo* designed peptide.<sup>5</sup> For this purpose, the peptides of interest were modified by introduction of an N-terminal cysteine containing moiety and subjected to heterodimerization by thiol oxidation. A following thiol exchange experiment between the heterodimers and different thiol containing peptides allowed the identification of the preferred helical coiled coil.

We considered a similar approach to be useful to gain information about the Id1 HLH domain, whose structural preferences are not fully understood yet. The approach is based on the synthesis of peptides containing GGC moieties attached either to the N-terminal (peptide **IV.1**) or the C-terminal end (peptide **IV.2**) of the native HLH motif. These peptides were subsequently oxidized to form disulfide bridged homodimers (Table IV.1).

**Table IV.1.** Amino acid sequences of the synthetic Id1 HLH analogs used for dimerization and thiol exchange studies.

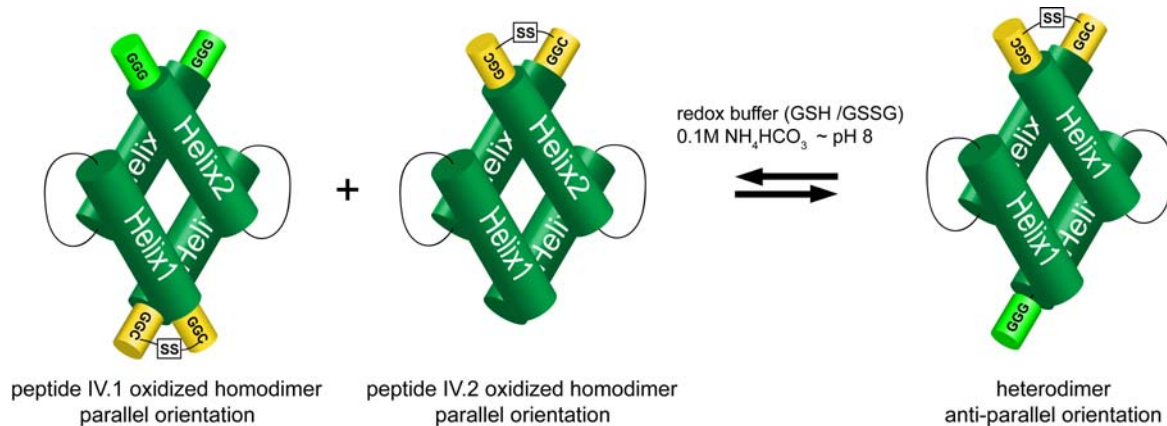
No.	Description	Sequence
<b>IV.1</b>	Ac-CGG-[S <sup>72</sup> ](66-106)-GGG	Ac-CGG-LYDMNGSYSRLKELV-PTLPQNRKVS-KVEILQHVIDYIRDLQ-GGG-NH <sub>2</sub>
<b>IV.2</b>	Ac-[S <sup>72</sup> ](66-106)-GGC	Ac-LYDMNGSYSRLKELV-PTLPQNRKVS-KVEILQHVIDYIRDLQ-GGC-NH <sub>2</sub>

A mixture of the two homodimers was then subjected to conditions allowing for a thiol exchange, through which both homo- and heterodimers might be formed. The preference for either a homo- or heterodimer should deliver insight into the helix packing and orientation of the Id1 HLH dimer.

## IV.2 Peptide design

Two analogs of the native Id1 HLH motif were synthesized by SPPS: peptide **IV.1** bears an additional CGG moiety for disulfide bridging at the N-terminus of the domain, whereas peptide **IV.2** bears this moiety at the C-terminus. Moreover, peptide **IV.1** contains a GGG moiety at the C-terminus to facilitate the identification of potential heterodimers with peptide **IV.2** by mass spectrometry. The cysteines to be oxidized were separated by two glycines from the respective helices to guarantee a sufficient flexibility of the covalent linkage in respect to the native motif and to prevent a strong structural impact on the folding of the HLH motif in the dimer. Moreover, in both peptides the native cysteine at position 72 was substituted by serine to prevent the formation of undesired oxidation products. Oxidation of these peptides should generate covalently linked homodimers, in which the helices adjacent to the disulfide bridge adopt a parallel orientation. The mixing of the two oxidized peptides **IV.1<sup>ox</sup>** and **IV.2<sup>ox</sup>** under conditions that allow a rearrangement of the disulfide bridge could, however, produce the mixed dimer displaying an antiparallel arrangement (Scheme IV.1), thus providing information about the preferred helix orientation.

**Scheme IV.1.** Thiol exchange assay. *GSH* – reduced glutathione; *GSSG* – oxidized glutathione.

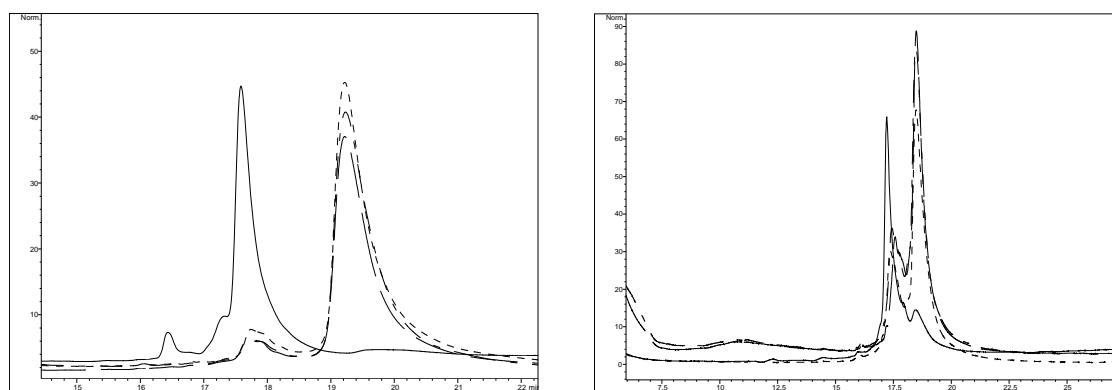


As appropriate mixture for the thiol exchange a solution containing a glutathione redox buffer (the oxidized and reduced species of glutathione) in a slightly basic buffered solution (0.1 M ammonium carbonate at pH ~ 8) was prepared. The oxidations of the monomeric peptides, as well as the thiol exchange assay, were monitored by analytical HPLC and mass spectrometry.

### IV.3 Results

#### IV.3.1 Formation of the oxidized homodimers

The synthetically prepared peptides **IV.1** and **IV.2** were dissolved in ammonium carbonate (pH ~ 8) at the concentration of 1 mg mL<sup>-1</sup> and subjected to air oxidation, in order to produce the corresponding covalently linked homodimers **IV.1<sup>ox</sup>** and **IV.2<sup>ox</sup>**. The disulfide bond formation was monitored by HPLC over 72 h. However, already after 24 h the HPLC peak of the reduced monomer disappeared and a new peak corresponding to the oxidized peptide emerged at a higher retention time (Figure IV.1). The fact that the oxidation reaction ran smoothly and almost to completeness suggests that the formed disulfide bridge does not impose an unfavorable constraint to the dimer.



**Figure IV.1.** Monitoring of the air oxidation of peptides **IV.1** (left) and **IV.2** (right) by HPLC. The solid line represents the HPLC of an aliquot taken immediately after dissolving the peptide in ammonium carbonate buffer.

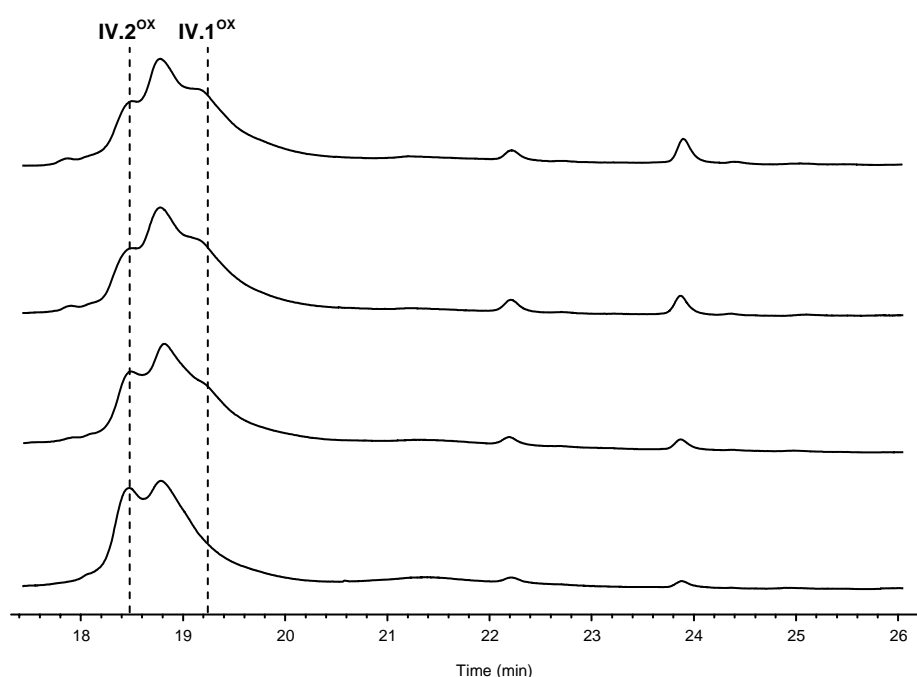
After 72 h, the peptides were lyophilized and the presence of the oxidized species was additionally confirmed by mass spectrometry (Table IV.2).

**Table IV.2.** Mass characterization of the lyophilized homodimers **IV.1<sup>OX</sup>** and **IV.2<sup>OX</sup>**.

	MW <sub>mon, calc</sub> (Da)	MW <sub>mon, found</sub> (Da)	MW <sub>dimer, calc</sub> (Da)	MW <sub>dimer, found</sub> (Da)
peptide <b>IV.1</b>	5303.1	5299.0	10604.3	10613
peptide <b>IV.2</b>	5132.0	5132.7	10262.0	10262

### IV.3.2 Thiol exchange assay

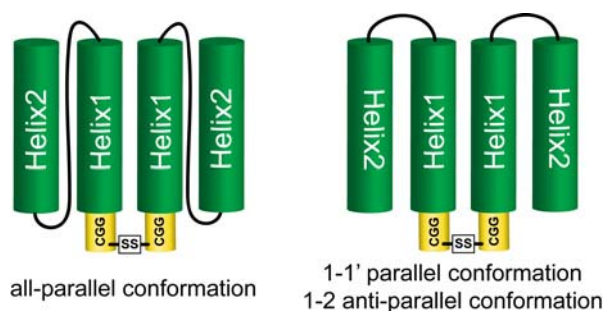
To probe for selectivity towards one distinct orientation, the two homodimers of peptides **IV.1** and **IV.2** were mixed under conditions that allow a thiol exchange assay. Therefore, an equimolar mixture of the two homodimers in ammonium carbonate buffer (pH ~ 8) containing reduced and oxidized glutathione in a 1:2 ratio was prepared. Already in the first aliquot, taken immediately after the start of the thiol exchange assay, mass measurements showed the reduced species of **IV.1** and **IV.2**, as well as their glutathione adducts, as indicated by the masses of the monomers plus 307 Da. The corresponding HPLC profile showed a broad peak at 18.75 min preceded and followed by shoulders at 18.5 min and 19.2 min, respectively. By comparison of the retention times, the two shoulders were assigned to the homodimers **IV.2<sup>OX</sup>** and **IV.1<sup>OX</sup>**, respectively, whereas the central peak presumably corresponds to the glutathione adducts of the reduced peptides (Figure IV.2).

**Figure IV.2.** HPLC runs of aliquots of the thiol exchange mixture taken at 0 h, 4.5 h, 22 h and 46 h (from top to bottom).

After 4.5 h and 22 h, the shoulder of **IV.1**<sup>OX</sup> decreased, while the peak attributed to **IV.2**<sup>OX</sup> became more pronounced. Finally, after 46 h the peaks at 18.5 min and 18.75 min were still distinguishable, suggesting that the main products were the homodimer of **IV.2** and mixed peptide-glutathione disulfides. This was also confirmed by mass spectrometry, which revealed not only the presence of a product with a mass of 10270 Da corresponding to **IV.2**<sup>OX</sup>, but also the masses of glutathione adducts of both monomers.

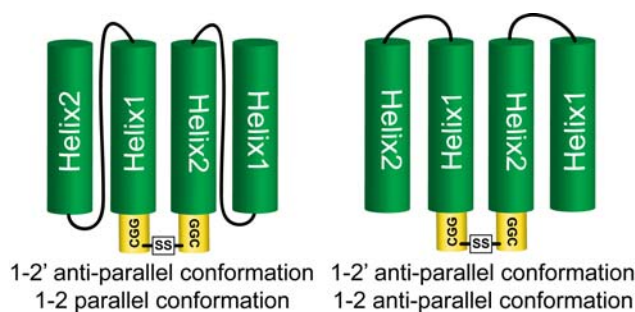
#### IV.4 Discussion

Two conceivable folding models for the covalent homodimer of peptide **IV.1** are shown in figure **IV.3**. In the all-parallel conformation the four helices adopt the same orientation to each other with the two helix-1 fragments being N-terminally linked, whereas in the other model the N-terminally linked helix-1 fragments are parallel to each other but anti-parallel to the helix-2 fragments. Thereby, the latter model resembles the one of the disulfide-bridged homodimer of MyoD<sup>2</sup>. In case of the **IV.2** homodimer, the same models are plausible, with the difference that the two helix-2 fragments are C-terminally linked. The fact that the oxidations of both peptides **IV.1** and **IV.2** proceeded smoothly and almost to completeness suggests that the formation of the corresponding homodimers was favorable in both cases, and that a parallel orientation of the two helix-1 fragments in **IV.1**<sup>OX</sup> or the two helix-2 fragments in **IV.2**<sup>OX</sup> is possible.



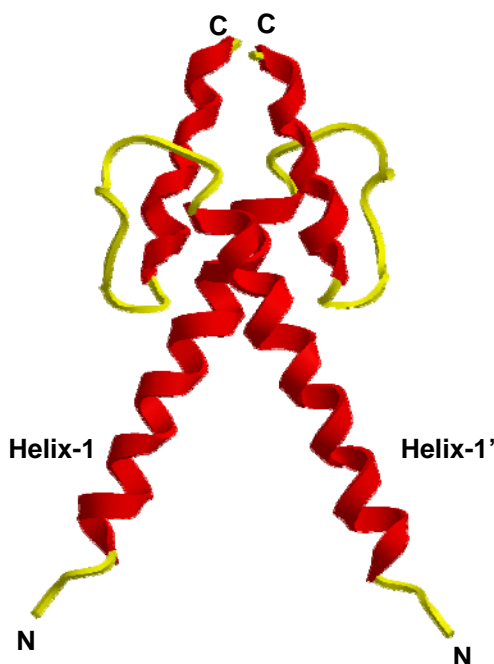
**Figure IV.3.** Schematic illustration of the possible folding models for the oxidized peptide **IV.1**.

By contrast, the generation of the mixed dimer of **IV.1** and **IV.2** would propose two different types of helix packing, which are schematically drawn in figure IV.4.



**Figure IV.4.** Schematic illustration of the possible folding models for the mixed dimer of peptides **IV.1** and **IV.2**.

In both possible dimeric arrangements helix-1 of **IV.1** and helix-2 of **IV.2** would adopt an anti-parallel orientation to each other. As the thiol exchange assay did not show the formation of covalent heterodimers, it can be assumed that an interchain anti-parallel arrangement of helix-1 and helix-2 is unfavored. Moreover, as the homodimer **IV.2<sup>OX</sup>** was still present in the glutathione containing buffer, whereas the homodimer **IV.1<sup>OX</sup>** was converted into mixed disulfides with the glutathione, it can be suggested that the covalently linked second helices experience a better stabilization in comparison to the covalently linked first helices. This may reflect the superior intrinsic helix propensity of helix-2 with respect to helix-1, which could also better promote helix interaction and dimerization.



**Figure IV.5.** Homodimer of the bHLH-LZ motif of Max obtained from a co-crystal with DNA.<sup>6</sup> The C-terminal LZ region, the N-terminal DNA-binding region as well as the DNA strands were omitted for reasons of clarity.

Taken together, the results suggest that the following helix arrangements are favored in the Id1 HLH dimer: helix-1 parallel to helix-1', helix-2 parallel to helix-2', and helix-1 parallel to helix-2'. These arrangements correspond to an all-parallel four-helix bundle, as it was also found in the crystal structures of the DNA-bound non-covalent homodimers of the related proteins MyoD<sup>7</sup>, E47<sup>8</sup> and Max<sup>6</sup>; the crystal structure of the latter is exemplarily shown in Figure IV.5.

## IV.5 References

1. Kadokura, H.; Katzen, F.; Beckwith, J., Protein disulfide bond formation in prokaryotes. *Annu Rev Biochem* **2003**, 72, 111-35.
2. Starovasnik, M. A.; Blackwell, T. K.; Laue, T. M.; Weintraub, H.; Klevit, R. E., Folding topology of the disulfide-bonded dimeric DNA-binding domain of the myogenic determination factor MyoD. *Biochemistry* **1992**, 31, 9891-903.
3. Albrecht Jacobi, D. S., How To Stabilize or Break beta-Peptidic Helices by Disulfide Bridges: Synthesis and CD Investigation of beta-Peptides with Cysteine and Homocysteine Side Chains. *Helvetica Chimica Acta* **1999**, 82, 1150-1172.
4. Bianchi, E.; Finotto, M.; Ingallinella, P.; Hrin, R.; Carella, A. V.; Hou, X. S.; Schleif, W. A.; Miller, M. D.; Geleziunas, R.; Pessi, A., Covalent stabilization of coiled coils of the HIV gp41 N region yields extremely potent and broad inhibitors of viral infection. *Proc Natl Acad Sci U S A* **2005**, 102, 12903-8.
5. Lai, J. R.; Fisk, J. D.; Weisblum, B.; Gellman, S. H., Hydrophobic core repacking in a coiled-coil dimer via phage display: insights into plasticity and specificity at a protein-protein interface. *J Am Chem Soc* **2004**, 126, 10514-5.
6. Ferre-D'Amare, A. R.; Prendergast, G. C.; Ziff, E. B.; Burley, S. K., Recognition by Max of its cognate DNA through a dimeric b/HLH/Z domain. *Nature* **1993**, 363, 38-45.
7. Ma, P. C.; Rould, M. A.; Weintraub, H.; Pabo, C. O., Crystal structure of MyoD bHLH domain-DNA complex: perspectives on DNA recognition and implications for transcriptional activation. *Cell* **1994**, 77, 451-9.
8. Ellenberger, T.; Fass, D.; Arnaud, M.; Harrison, S. C., Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev* **1994**, 8, 970-80.

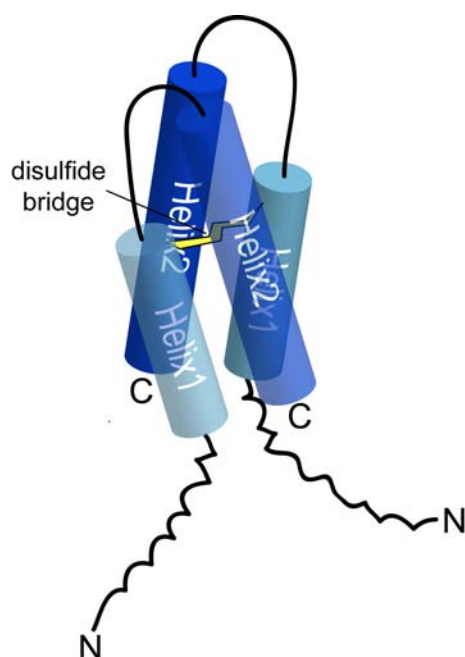




## V Solution NMR studies on the Id2 HLH motif

### V.1 Introduction

The biological functions of the HLH proteins are strongly dependent upon equilibria between their monomeric and homo-oligomeric forms, which regulate their interactions with other HLH proteins and/or with the DNA.<sup>1,2</sup> Due to their pronounced aggregation propensity at the concentration range necessary for NMR measurements, only a few examples of NMR studies on HLH proteins have been reported thus far. Successful NMR experiments were for example carried out for the bHLH motifs of MyoD, a class II HLH protein and the ubiquitous E47, a class I HLH protein.<sup>3,4</sup> The NMR structure of MyoD proved to be difficult to obtain, as its bHLH motif was found to form predominantly tetramers that were not suitable for a reasonable NMR evaluation. Therefore, a covalently linked dimer was prepared by intermolecular disulfide bridge formation using the native cysteine residue at position 135 (close to the C-terminal end of helix-1) and subsequently assigned by heteronuclear NMR



**Figure V.1.** Model of the Cys<sup>135</sup> disulfide bonded dimer of MyoD bHLH.<sup>3</sup>

experiments. The results suggest an anti-parallel four-helix bundle, in which the helix-1 of one monomer is roughly parallel to the helix-1 of the other monomer (Figure V.1). This NMR model contradicts the crystal structure of the DNA-bound MyoD dimer, in which all four helices are parallel. Another bHLH motif studied by NMR is the one from the ubiquitous E47 protein (class I HLH). This protein was found to be a suitable candidate for NMR measurements, as, contrarily to MyoD or the Id proteins, it does form stable dimers but no tetramers or higher-order oligomers. Helix-2 was found to be well-defined, as indicated by the presence of  $(i, i+3)$  and  $(i, i+4)$  NOE signals, whereas helix-1 turned out to be rather flexible

under the conditions used. Except for the unordered basic region, which becomes structured only upon binding to the DNA, the obtained NMR results are in good agreement with the crystal structure of the DNA-bound E47 dimer, which shows a parallel four-helix bundle.<sup>5</sup>

Moreover, the whole HLH domain of the E47 protein features a high degree of dynamics, as indicated by the rapid hydrogen-deuterium exchange.

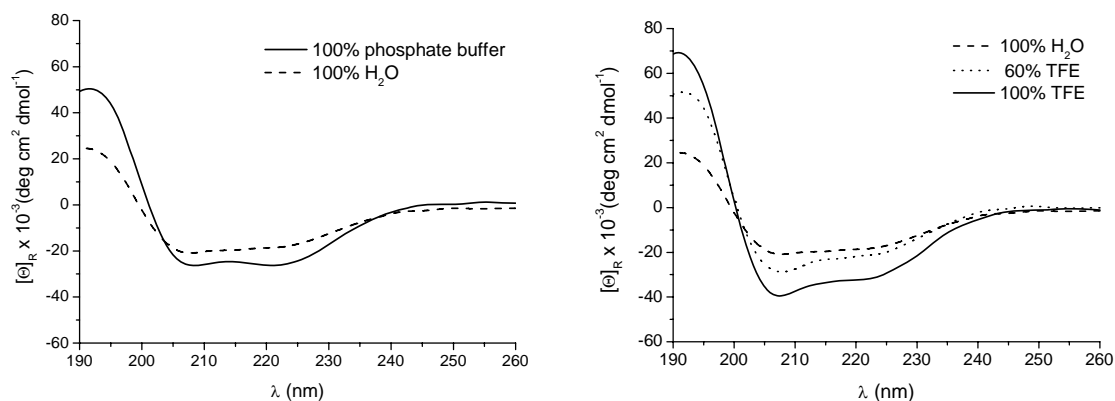
The tertiary/quaternary structure of the Id proteins is thus far not known, but, based on sequence alignment and homology models, it is proposed to be similar to the parallel four-helix bundle ascertained in the crystal structures of the related proteins MyoD and E47.<sup>6</sup> The conformational information available for the Id proteins is, up to now, limited to the secondary structure elements composition obtained by the analysis of CD data.<sup>7</sup> We envisioned that NMR experiments could provide a more detailed look on the structural preferences of the HLH motifs of the Id protein family. Like the other HLH proteins, also the Id proteins aggregate upon NMR conditions and present poor solubility; therefore, in order to avoid, or at least to reduce strong aggregation and to prevent precipitation during the NMR measurements, we decided to add the fluorinated alcohol trifluoroethanol (TFE) which is known to destroy tertiary and quaternary aggregations while, at the same time, it elicits the secondary structure propensity of a peptide chain. Although their use generates a non-physiological environment, the addition of fluorinated alcohols can still yield useful insights in the intrinsic structural features of a peptide. Interestingly, some reports also suggest the use of alcohols (MeOH or EtOH) and fluorinated alcohols like TFE to mimic the physiological environment near membranes or in the membrane, respectively.<sup>8</sup> Although it has to be kept in mind that the Id proteins are not membrane proteins, their biological functions require the crossing of the nuclear membrane in a receptor-independent manner and, thus, TFE conditions could also mirror physiologically relevant structural properties. Finally, the moderate length of 41 amino acids of the Id2 HLH domain benefits the acquisition of reasonable NMR data without isotope labeling, as it was already demonstrated for other peptides of similar lengths.<sup>9-12</sup> Thus, we assembled the Id2 HLH sequence 36-76 by using straight-forward SPPS methodology and undertook its NMR characterization.

## V.2 Results

### V.2.1 CD measurements

As the Id2 HLH motif is not soluble at a pH value of about 7 over the concentration range required for the NMR experiments ( $\geq 0.5$  mM), we dissolved the peptide in non-buffered water. Moreover, to further increase the solubility and to avoid precipitation, also TFE had to be added. To examine the influence of the pH value and the TFE addition on the

secondary structure of the Id2 HLH motif, CD measurements were performed before commencing the NMR experiments (Figure V.2).



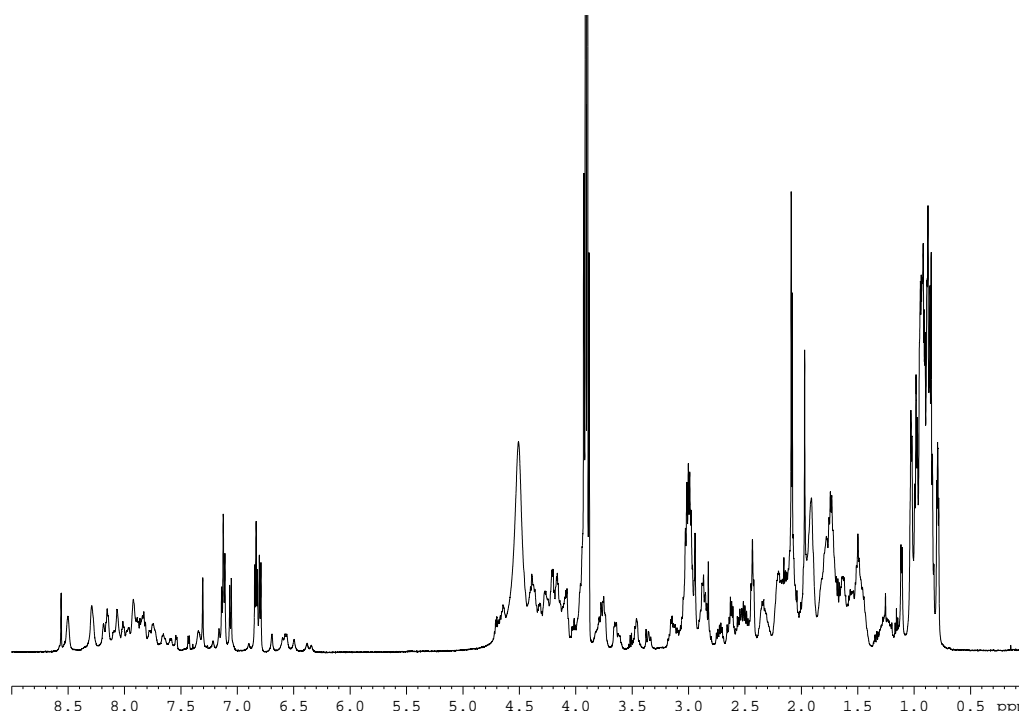
**Figure V.2.** CD spectra of the Id2 HLH motif (30  $\mu$ M). Left panel: peptide dissolved in 0.1 M phosphate buffer (pH 7.25) or in millipore water. Right panel: peptide dissolved in millipore water containing increasing amounts of TFE.

The obtained CD curves suggest a helical conformation for all measured samples, as indicated by the two negative bands at 208 nm and 222 nm, respectively, and by the maximum at 190 nm. Nevertheless, changing the pH from a neutral to an acidic value, as well as adding TFE affected the peptide structure. Indeed, the intensity of the CD bands decreased when going from neutral to acidic conditions (Figure V.2, left panel), indicating a slight loss of helicity. Moreover, the ratio between the two ellipticity minima at 222 nm and 208 nm,  $R = [\Theta]_{222}/[\Theta]_{208}$ , also decreased, reflecting changes in the interhelical contacts upon altering the charge distribution of the ionizable side chains. On the other hand, the loss of ellipticity observed when changing from a neutral to an acidic pH value could be compensated by the addition of TFE (Figure V.2, right panel). This increase in CD intensity was accompanied by a decrease in the ratio  $R$  from 0.88 in millipore water to 0.74 in a 60% TFE containing mixture. However, these changes were anticipated, as TFE is known not only for stabilizing the secondary structure preferences of a peptide, but also for destroying interhelical interactions. To quantify the obtained CD data, they were subjected to CONTIN analysis. For the peptide in millipore water a helical percentage of 49% was calculated, corresponding to 20 amino acids in a helical conformation. In comparison thereto, the percentage of helical secondary structure for the peptide in a 60% TFE solution was markedly increased: 72% helix content stemming from 30 helical amino acids. As the obtained CD spectrum of the peptide in

60% TFE supported the presence of a structured peptide, which is a prerequisite for a reasonable NMR investigation, we chose these conditions to record the NMR spectra.

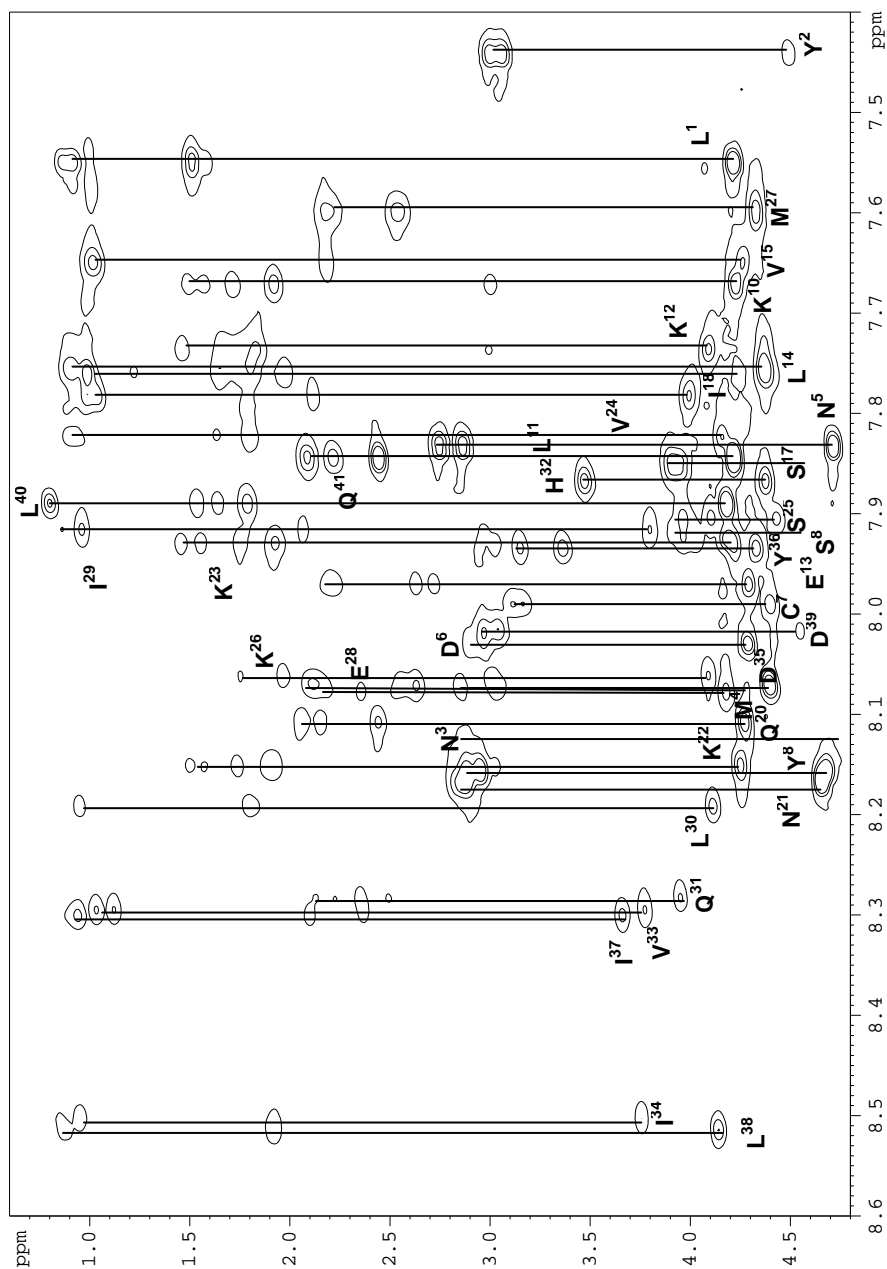
### V.2.2 NMR measurements

The NMR sample of the Id2 HLH peptide was prepared by dissolving it in a mixture of TFE- $d_2$ /water 60:40 (v/v) to a final peptide concentration of 0.5 mM. At first, 1D-NMR spectra were recorded at different temperatures in 6 K steps from 300 K to 318 K and evaluated in terms of minimal signal overlapping and maximal signal dispersion. As at 318 K (Figure V.3) the best results were obtained, the subsequent 2D-NMR experiments were performed at this temperature.



**Figure V.3.**  $^1\text{H}$ -NMR spectrum of the Id2 HLH motif in TFE- $d_2$ /water (60:40, v/v) recorded at 318 K.

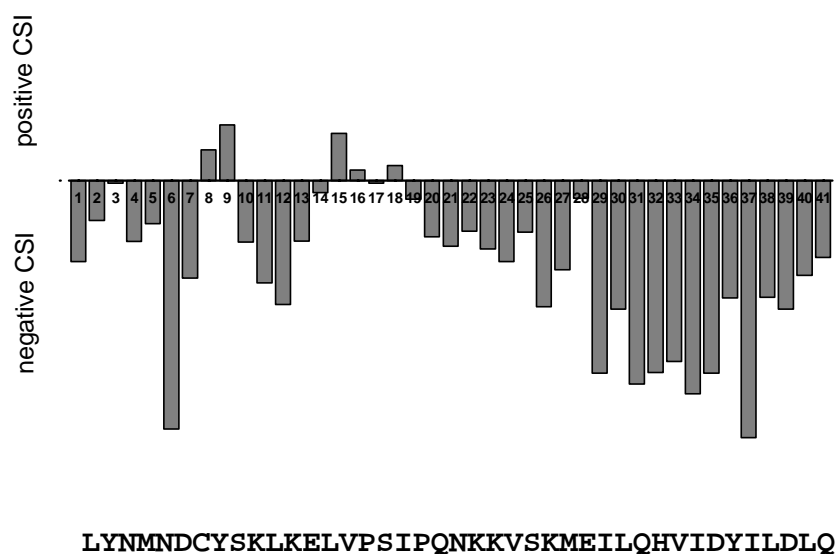
The fingerprint region of the COSY spectrum showed a good dispersion of the 39 non-proline NH- $\alpha$ H cross signals in the region from 8.6 to 7.4 ppm, indicating the presence of a structured peptide. Moreover, no major spectral overlaps could be detected, which facilitated the sequential assignment and, indeed, the obtained resolution of the two-dimensional spectra proved to be sufficient enough for the assignment of the 41 amino acids (Figure V.4. See Appendix IX for the complete list of the chemical shifts).



**Figure V.4.** Fingerprint region of the TOCSY spectrum of the Id2 HLH motif in TFE-d<sub>2</sub>/water (60:40) at 318 K. Resonance assignments are labeled using the one letter code for the amino acids combined with the residue number.

### V.2.3 Chemical shift indices (CSIs) of the $\alpha$ H protons

The deviations of the  $\alpha$ H chemical shifts obtained for the Id2 HLH peptide from those provided by Schwarzsinger et al. for a random coil protein in acidic 8 M urea<sup>13</sup> are reported in Figure V.5. The CSI pattern is dominated by negative values, indicating the predominance of a helical structure. More precisely, one large helical segment extends from residue 19 to the C-terminal end at position 41, whereas another one spans from the N-terminal end to residue 14 but is interrupted by two positive CSI values at positions 8 and 9. As clearly opposite CSI values of two adjacent residues are indicative for a transition from one secondary structure element to another<sup>14</sup>, it is likely that this short inversion of the CSI values reflects the presence of a turn or bent connecting two short helical elements (4-7 and 10-14). Furthermore, the central region 15-18 shows nearly neutral or positive CSI values and might correspond to the loop region predicted by sequence alignment.

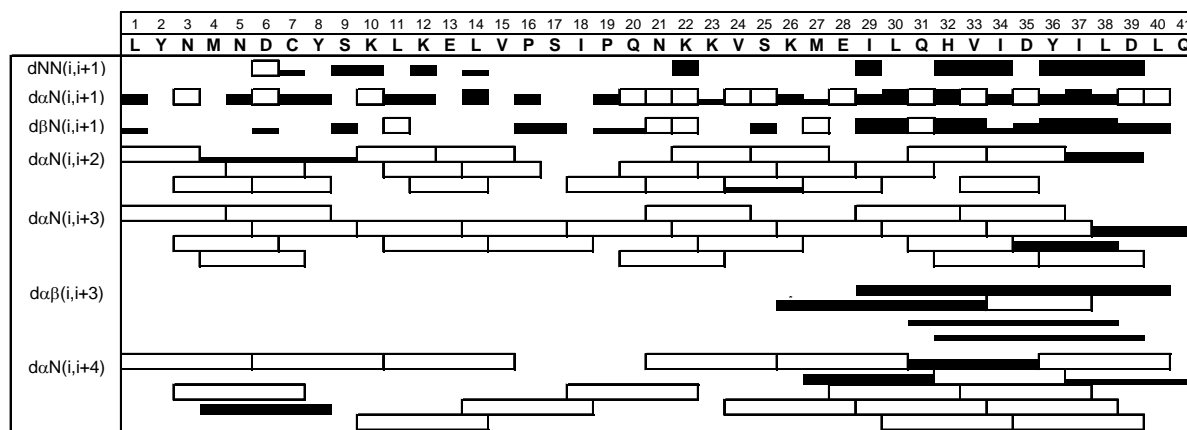


**Figure V.5.** CSIs for the  $\alpha$ H protons of the Id2 HLH motif in TFE- $d_2$ /water (60:40) at 318 K. The calculation was done using the reference values reported by Schwarzsinger et al.<sup>13</sup>

### V.2.4 NOE signals

The NOESY spectrum (recorded with a mixing time of 200 ms) was examined for NOE signals that are characteristic for the presence of secondary structure elements (Figure V.6). Unfortunately, the presence of some relevant NOE signals could not be evaluated, due to considerable overlap. Nevertheless, several medium to strong  $d_{NN}(i,i+1)$ ,  $d_{\alpha\beta}(i,i+3)$  and  $d_{\alpha N}(i,i+4)$  NOE signals were found in the C-terminal part of the sequence spanning the

residues 29-40, indicating the presence of a well-defined  $\alpha$ -helix in this region. Such NOE pattern agrees with the CSI values found for the same residues, which were remarkably more negative than those of the preceding residues 19-28, suggesting that this C-terminal helix is better defined at its C-terminus, while its N-terminal tail being probably more flexible.



**Figure V.6.** NOE cross signals obtained for the Id2 HLH motif. Missing bars indicate the absence of NOE signals. Open bars indicate the presence of overlapping that made it impossible to verify the presence of a NOE signal.

Contrarily to the C-terminal part of the sequence, the N-terminal part showed only few  $d_{NN}(i,i+1)$  and no  $d_{\alpha\beta}(i,i+3)$  NOE signals. However, two  $d_{\alpha N}(i,i+2)$  NOEs were present between residues 4-6 and 7-9, which might reflect the formation of turns or of a short  $3_{10}$ -helix.

### V.3 Discussion

The presented results provide the first NMR spectroscopic evaluation of the Id2 HLH motif. Despite its tendency to aggregate, the use of a non-buffered solution and the addition of TFE enabled the acquisition of reasonable NMR spectra. The spectral dispersion of the signals was sufficient enough to allow the sequential assignment and, subsequently, a first conformational analysis of this peptide. The CD study already anticipated the high helix content, which can be partly attributed to the addition of TFE. This result was confirmed by the analysis of the CSI values which suggest that the peptide is presumably adopting an  $\alpha$ -helical structure starting from Pro<sup>19</sup>, as supported also by the presence of several sequential and medium-range NOEs. The identified  $\alpha$ -helix spanning from Pro<sup>19</sup> to Gln<sup>41</sup> is apparently longer than the one predicted by sequence alignment studies which locate the start of the

helix-2 at Lys<sup>26</sup>. This might be explained by the strong influence of TFE on the stabilization of the secondary structure. However, it should be pointed out that the N-terminal tail of the helix-2 was not as rich in NOE connectivities as the C-terminal part. The loop region N-terminal to the helix-2 seems to be restricted to the residues 14-18, which would result to be much shorter than the predicted nine residues 17-25. In contrast to the remarkable helical character of the second half of the peptide, the predicted N-terminal helix-1 from residue 1 to residue 16 was poorly defined, as indicated by a scarce number of sequential and medium-range NOE signals. In particular, the lack of NOEs characteristic for  $\alpha$ -helices might be explained by non-classical helix elements of a highly dynamic conformation, as it was already observed for the helix-1 of the E47 HLH domain<sup>4</sup> and for the HLH domain of Max, a class IV protein<sup>15</sup>. In these studies the reduced number of sequential and medium-range NOEs was attributed to a poorly defined first helical region. Moreover, the NMR data presented here are consistent with our previous reports upon the superior helix propensity of the second helix of the Id HLH motif compared to the N-terminal region, as investigated by CD spectroscopy.<sup>16</sup> However, the majority of negative CSI values as well as the presence of some  $d_{\alpha N}(i,i+2)$  NOEs suggest the tendency to form local helical turns or a nascent helix, whose conversion into a stably folded  $\alpha$ -helix might correlate with the possibility to interact with the helix-2. The conditions used in the presented NMR study are not favorable to and rather prevent interhelical contacts, as TFE is known for destroying tertiary and quaternary folds.<sup>17</sup> Thus, it is reasonable that the peptide in the presence of a high content of TFE is not able to adopt a compact structure, but rather an opened conformation that might be also dictated by the N-terminal prolongation of the helix-2 at the expense of the central loop region. Finally, it can be suggested that the structural flexibility of the first helix constitutes a prevailing structural feature occurring in the whole HLH protein family, and is likely to provide the binding specificity to other proteins.

## V.4 References

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



2. Laue, T. M.; Starovasnik, M. A.; Weintraub, H.; Sun, X. H.; Snider, L.; Klevit, R. E., MyoD forms micelles which can dissociate to form heterodimers with E47: implications of micellization on function. *Proc Natl Acad Sci USA* **1995**, 92, 11824-8.
3. Starovasnik, M. A.; Blackwell, T. K.; Laue, T. M.; Weintraub, H.; Klevit, R. E., Folding topology of the disulfide-bonded dimeric DNA-binding domain of the myogenic determination factor MyoD. *Biochemistry* **1992**, 31, 9891-903.
4. Fairman, R.; Beran-Steed, R. K.; Handel, T. M., Heteronuclear (<sup>1</sup>H, <sup>13</sup>C, <sup>15</sup>N) NMR assignments and secondary structure of the basic region-helix-loop-helix domain of E47. *Protein Sci* **1997**, 6, 175-84.
5. Ellenberger, T.; Fass, D.; Arnaud, M.; Harrison, S. C., Crystal structure of transcription factor E47: E-box recognition by a basic region helix-loop-helix dimer. *Genes Dev* **1994**, 8, 970-80.
6. Chavali, G. B.; Vijayalakshmi, C.; Salunke, D. M., Analysis of sequence signature defining functional specificity and structural stability in helix-loop-helix proteins. *Proteins* **2001**, 42, 471-80.
7. Kiewitz, S. D.; Cabrele, C., Synthesis and conformational properties of protein fragments based on the Id family of DNA-binding and cell-differentiation inhibitors. *Biopolymers* **2005**, 80, 762-74.
8. Perham, M.; Liao, J.; Wittung-Stafshede, P., Differential effects of alcohols on conformational switchovers in alpha-helical and beta-sheet protein models. *Biochemistry* **2006**, 45, 7740-9.
9. Bansal, P. S.; Grieve, P. A.; Marschke, R. J.; Daly, N. L.; McGhie, E.; Craik, D. J.; Alewood, P. F., Chemical synthesis and structure elucidation of bovine kappa-casein (1-44). *Biochem Biophys Res Commun* **2006**, 340, 1098-103.
10. Hudson, F. M.; Andersen, N. H., Exenatide: NMR/CD evaluation of the medium dependence of conformation and aggregation state. *Biopolymers* **2004**, 76, 298-308.
11. Landon, C.; Meudal, H.; Boulanger, N.; Bulet, P.; Vovelle, F., Solution structures of stomoxyn and spinigerin, two insect antimicrobial peptides with an alpha-helical conformation. *Biopolymers* **2006**, 81, 92-103.
12. Schievano, E.; Mammi, S.; Monticelli, L.; Ciardella, M.; Peggion, E., Conformational studies of a bombolitin III-derived peptide mimicking the four-helix bundle structural motif of proteins. *J Am Chem Soc* **2003**, 125, 15314-23.

13. Schwarzing, S.; Kroon, G. J.; Foss, T. R.; Wright, P. E.; Dyson, H. J., Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J Biomol NMR* **2000**, 18, 43-8.
14. Wishart, D. S.; Sykes, B. D.; Richards, F. M., The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **1992**, 31, 1647-51.
15. Lavigne, P.; Kondejewski, L. H.; Houston, M. E., Jr.; Sonnichsen, F. D.; Lix, B.; Skyes, B. D.; Hodges, R. S.; Kay, C. M., Preferential heterodimeric parallel coiled-coil formation by synthetic Max and c-Myc leucine zippers: a description of putative electrostatic interactions responsible for the specificity of heterodimerization. *J Mol Biol* **1995**, 254, 505-20.
16. Colombo, N.; Cabrele, C., Synthesis and conformational analysis of Id2 protein fragments: impact of chain length and point mutations on the structural HLH motif. *J Pept Sci* **2006**, 12, 550-8.
17. Lau, S. Y.; Taneja, A. K.; Hodges, R. S., Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-coils. *J Biol Chem* **1984**, 259, 13253-61.

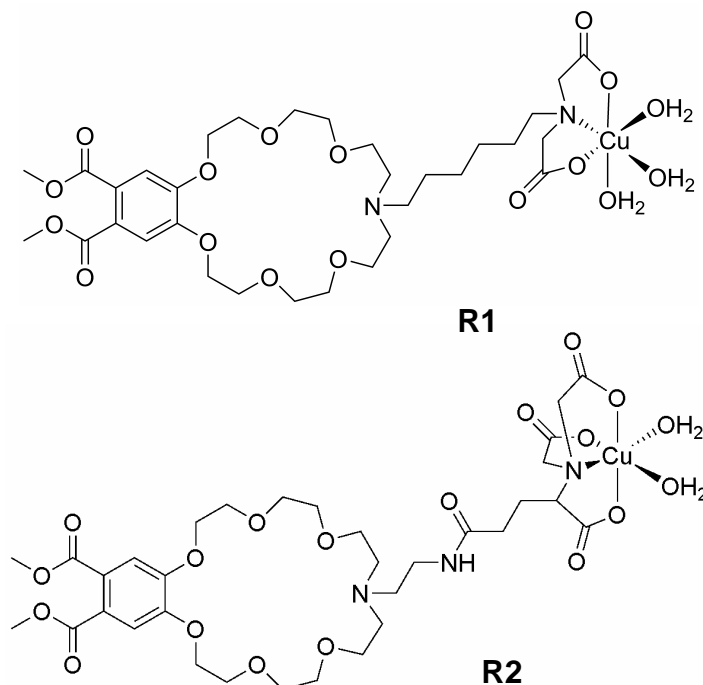
## VI Toward the recognition of the Id HLH motif by an artificial receptor

### VI.1 Introduction

Protein-protein interactions play a key role in many biological processes, including signal transduction, cell-cell contacts, immunitary response and DNA transcription. As they are involved both in normal and pathologic events, a major interest lies not only in a better understanding of underlying mechanisms of protein molecular recognition, but also in the development of new chemical tools to influence such mechanisms, especially in cancer pathways.<sup>1-4</sup> A range of artificial systems functioning as dimerizers, inhibitors of dimerization or inhibitors of protein-receptor interactions have successfully been developed.<sup>5</sup> The challenge of targeting protein interfaces arises from their large area (usually  $>600 \text{ \AA}^2$ ) and the highly specific distribution of a variety of functional groups (charged, polar, hydrophobic, H-bond acceptors/donors) on it. Nonetheless, recent approaches based on the design of protein-surface mimetics or protein-surface artificial receptors have been proven to be promising for the detection and/or inhibition of protein-protein activities. One of these approaches consists of the application of molecules incorporating transition metal ions with high affinity for specific amino acid side chains. For example, Mallik and co-workers have synthesized an artificial receptor bearing three  $\text{Cu}^{\text{II}}$ -IDA (iminodiacetic acid) units for the recognition and binding of three histidine side chains located at the N-terminus of the protein carbonic anhydrase.<sup>6</sup> Another interesting approach is to combine different functionalities into one receptor molecule to increase its multiplicity of interactions, including hydrophobic and electrostatic contacts. Such multivalent receptors have been used to selectively target peptide sequences in aqueous solutions: for example, Hossain and Schneider developed an artificial receptor containing an 18-crown-6 ether, a peralkylammonium group and a dansyl group for the sequence-selective binding of zwitterionic tripeptides bearing a central aromatic residue like phenylalanine or tryptophan.<sup>7</sup>

Recently, König and co-workers reported on the ditopic receptor **R1** featuring a  $\text{Cu}^{\text{II}}$ -IDA complex for the binding of a Lewis-base and a crown ether for binding to a Lewis-acid (Figure VI.1).<sup>8</sup> A phthalic diester unit incorporated in the crown ether allows the monitoring of cation coordination by measuring the increase in emission intensity at 394 nm. However, a fluorescence response is only observed when the ligand is able to simultaneously bind to the  $\text{Cu}^{\text{II}}$ -IDA, as the crown ether interaction for itself is generally too weak. Among several tested

di- and tripeptides, only those containing an N-terminal histidine residue adjacent to a lysine or a glycine were found to bind the receptor. The highest affinity was observed for the dipeptide H-His-Lys-OMe.



**Figure VI.1.** Structures of the receptors **R1** and **R2** developed by König and co-workers.<sup>8</sup>

The IDA unit, complexed with copper or nickel, has found wide application in immobilized metal affinity chromatography (IMAC), a technique which is used for enriching or purifying proteins. The proteins are separated on the basis of the density of electron donors on their surface, like the side chains of cysteine, histidine or tryptophan, which stick to the immobilized Lewis acidic copper complex. At neutral pH the affinity of amino acids towards the metal complex is as follows: Cys > His >> Asp, Glu > Trp, Tyr, Lys, Met, Asn, Gln, Arg.<sup>9</sup> Furthermore, Nomura et al. proved the application of Cu<sup>II</sup>-IDA as paramagnetic probe for the elucidation of long distance information: by NMR experiments they showed the localization of the Cu<sup>II</sup>-IDA to a histidine residue on the surface of ubiquitin.<sup>10</sup>

Another receptor molecule developed by König and co-workers presents a Cu<sup>II</sup>-NTA (nitrilotriacetic acid) unit in place of the Cu<sup>II</sup>-IDA (**R2**, Figure VI.1). The Cu<sup>II</sup>-NTA complex is also frequently used in IMAC and exhibits a higher metal binding affinity at the expense of the protein binding, which is lower due to the loss of one coordination site.<sup>11</sup>

The aim of the presented work was to apply the receptor **R1** for recognition and non-covalent fluorescence labeling of the native dimerization domain of the Id proteins. Also the

more recently developed receptor **R2** was tested to elucidate the influence of the metal complex on the binding properties.

## VI.2 Results

To investigate potential interactions of the receptor **R1** with the Id HLH regions, four 41-residue long peptides based on the HLH sequences of Id1-4 were synthesized by solid-phase methodology (Table VI.1).

**Table VI.1.** Synthesized peptides tested for receptor affinity. All peptides were N-terminally acetylated and C-terminally amidated.

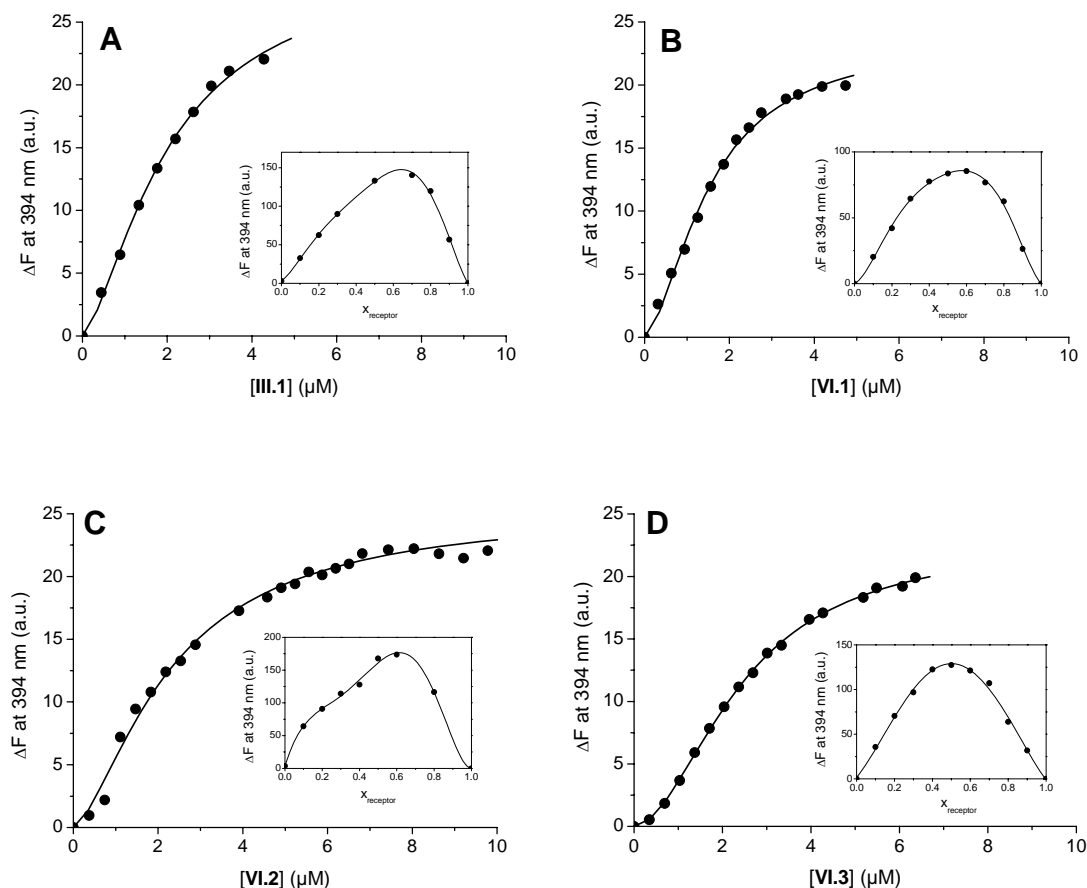
No.	Description	Helix-1	Loop	Helix-2
<b>III.1</b>	(66-106)-Id1	LYDMNGCYSRLKELVP	TLPQNRKVS	KVEILQHVIDYIRDLQ
<b>VI.1</b>	(36-76)-Id2	LYNMNDCYSKLKELVP	SIPQNKKVS	KMEILQHVIDYILDLQ
<b>VI.2</b>	(41-81)-Id3	LDDMNH CYSRLRELVP	GVPRGTQLS	QVEILQRVIDYILDLQ
<b>VI.3</b>	(65-105)-Id4	QCDMNDCYSRLRRLVP	TIPPNNKKVS	KVEILQHVIDYILDLQ
<b>VI.4</b>	(66-82)-Id1	LYDMNGCYSRLKELVP	T	
<b>VI.5</b>	(41-57)-Id3	LDDMNH CYSRLRELVP	G	
<b>VI.6</b>	(65-81)-Id4	QCDMNDCYSRLRRLVP	T	
<b>VI.7</b>	(90-105)-Id4			KVEILQHVIDYILDLQ
<b>VI.8</b>	(72-75)-Id1	CYSR		
<b>VI.9</b>	[A <sup>73</sup> ]- (72-75)-Id1	CASR		
<b>VI.10</b>	[A <sup>72</sup> ]- (72-75)-Id1	AYSR		
<b>VI.11</b>	(42-45)-Id2	CYSK		
<b>VI.12</b>	(47-52)-Id3	CYSRLR		
<b>VI.13</b>	(122-160)-MyoD	LSKVNEAFETLKRCTS	SNPNQRLP	KVEILRNAI RYIEGL
<b>VI.14</b>	(36-74)-Max	RDHIKDSFHSLRDSVP	SLQGEKAS	RAQILDKATEYIQYM

Formation of the receptor-peptide complex in a HEPES buffered solution (50 mM, pH 7.5) at 25 °C was monitored by fluorescence spectroscopy, following the change of the receptor emission intensity at 394 nm ( $\lambda_{\text{ex}} = 305$  nm) upon addition of the peptides. As they feature several groups which are prone to interact with the copper complex (imidazole, thiol, carboxylic groups and, to a minor extent, backbone amides) and with the crown ether (arginine and lysine side chains), the Id HLH motifs could be potential ligands of **R1**. Indeed,

an interaction between the receptor and the Id sequences was detected. In order to identify the region responsible for the binding, also the single helices and parts of them were synthesized and tested for receptor affinity. The results of these experiments are described in the following.

### *The Id HLH domains*

The fluorescence intensity of a solution of **R1** was monitored upon titration with each of the four peptides reproducing the 41-residue long HLH sequences of Id1-4 (peptides **III.1** and **VI.1-VI.3**; Figure VI.2 and Table VI.1).



**Figure VI.2.** Titration curves of **R1** (9.5  $\mu\text{M}$  in 50 mM HEPES buffer, pH 7.5) with peptides **III.1** (A), **VI.1** (B), **VI.2** (C) and **VI.3** (D). Solid lines represent the data fit obtained using the Hill equation. Insets: Job's plot analyses.

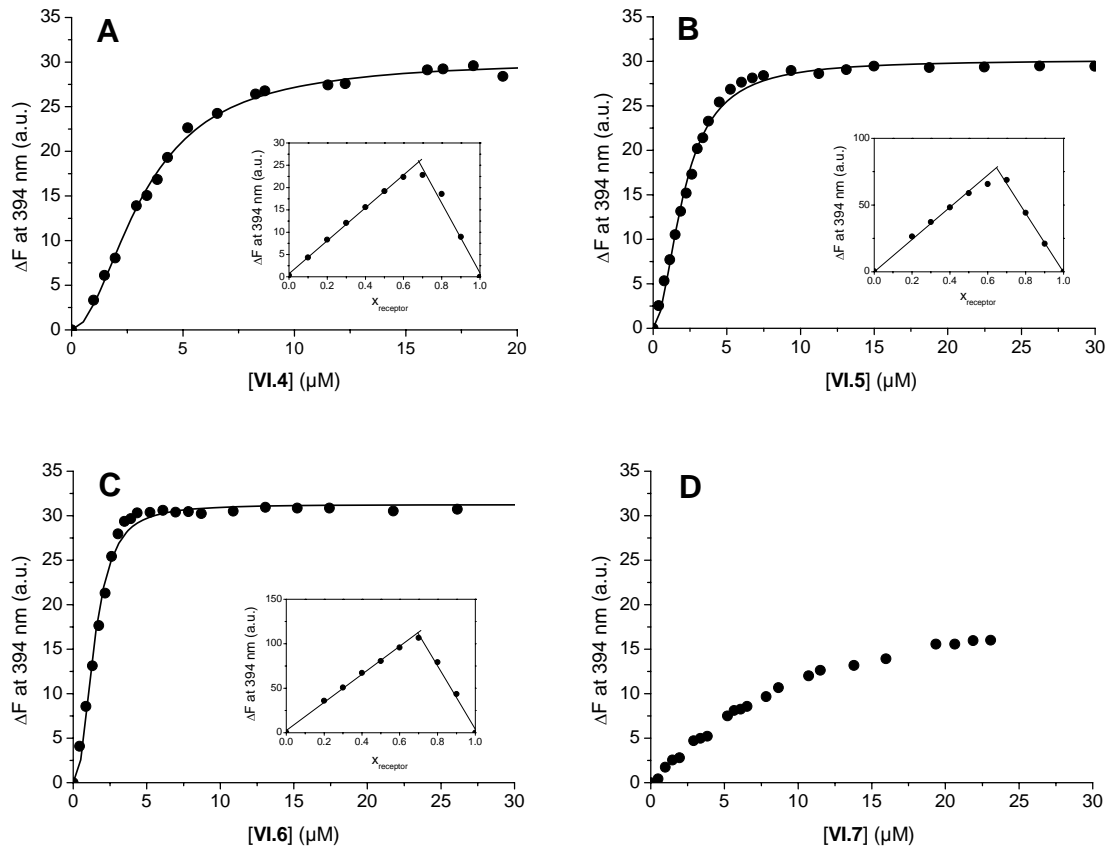
Upon increasing peptide concentration, the emission intensity of **R1** also increased, thus indicating cation coordination to the crown ether and, consequently, an interaction between **R1** and the peptide. In case of peptides **III.1** and **VI.1**, the fluorescence signal reached its maximum at ~0.5 equiv. peptide, suggesting a 2:1 **R1**/peptide ratio. Further increase in peptide concentration led to a decrease in fluorescence, which was accompanied by the appearance of a white precipitate (see Appendix, Figure IX.5). As the peptides themselves are completely soluble in the concentration range used, the turbidity might be attributed to a change in the solubility properties of the peptides in the titration mixture, presumably as a consequence of a conformational transition which elicited their self-association.

Similar titration curves were obtained for peptides **VI.2** and **VI.3**, for which, though, the fluorescence maximum was shifted to ~0.8 equiv. peptide, and the decrease in fluorescence intensity above this value was less pronounced, especially in case of **VI.2**. Moreover, no turbidity was observed during the titration of these two HLH motifs. To examine the binding stoichiometries, Job's plot analyses were conducted. For peptides **III.1**, **VI.1** and **VI.2** an ambiguous stoichiometry of ~1.6:1 (**R1**/peptide) was obtained, whereas peptide **VI.3** showed a 1:1 stoichiometry. As the shape of the curves was quite broad for all four peptides, the coexistence of different complex species in equilibrium can be assumed. By fitting the data preceding the fluorescence decrease with the Hill equation,  $K_{0.5}$  values in the range of 1.5-2.5  $\mu\text{M}$  were determined, with a Hill coefficient  $h$  increasing from 1.5 for **III.1** to 1.9 for **VI.3**.

### *Isolated Id helices*

To investigate the role of each single helix in the interaction with **R1**, we prepared peptides **VI.4-VI.7** (Table VI.1), reproducing the N-terminal helix (helix-1) of Id1 (**VI.4**), Id3 (**VI.5**) and Id4 (**VI.6**), and the C-terminal helix (helix-2) of Id4 (**VI.7**). All these peptides but **VI.7** showed a fluorescence response comparable to the one obtained for the full-length HLH motifs (Figure VI.3), indicating that helix-1 contains the site responsible for receptor binding. Peptide **VI.6**, representing the first helix of the Id4 HLH motif, exhibits the smallest  $K_{0.5}$  value among the isolated helices with 1.4  $\mu\text{M}$ . The corresponding helices of Id1 and Id3 (peptides **VI.4** and **VI.5**) show a moderately reduced binding with  $K_{0.5}$  values of 3.2  $\mu\text{M}$  and 2.1  $\mu\text{M}$ , respectively. The putative binding motif of the helix-1 of the Id HLH domains might correspond to the conserved tetrapeptide *CYSR(K)*, which is characterized by the presence of a cysteine residue at position  $i$  and of a basic residue (lysine for Id2 and arginine for the other Id proteins) at position  $i + 3$ . The moderately different  $K_{0.5}$  values shown by the three helix-1

segments probably reflects the less conserved amino acid positions in the first half of the sequences: indeed, the first two residues in peptide **VI.6** are glutamine and cysteine, which are found only in Id4, whereas peptides **VI.4** and **VI.5** start with leucine-tyrosine and leucine-aspartic acid, respectively. Moreover, the *CYSR(K)* motif is preceded by aspartic acid in peptide **VI.6**, by glycine in peptide **VI.4** and by histidine in peptide **VI.5**. These changes are likely to influence the binding and seem to have the most positive impact on peptide **VI.6**.



**Figure VI.3.** Titration curves of **R1** (9.5 μM in 50 mM HEPES buffer, pH 7.5) with peptides **VI.4** (A), **VI.5** (B), **VI.6** (C) and **VI.7** (D). Solid lines represent the data fit obtained using the Hill equation. Insets in panels A, B and C: Job's plot analyses.

In contrast to the Id helix-1 sequence, the helix-2 (only the Id4 helix-2, peptide **VI.7**, was used for this study, as all second helices of the Id family are highly identical) shows low affinity to the receptor. Although the *CYSR(K)* sequence is not present, other residues potentially binding to the receptor can be found, including lysine and histidine at positions  $i$  and  $i + 6$ , respectively. One reason for the weak fluorescence response might be an inappropriate side-chain distribution along the sequence, which does not allow simultaneous



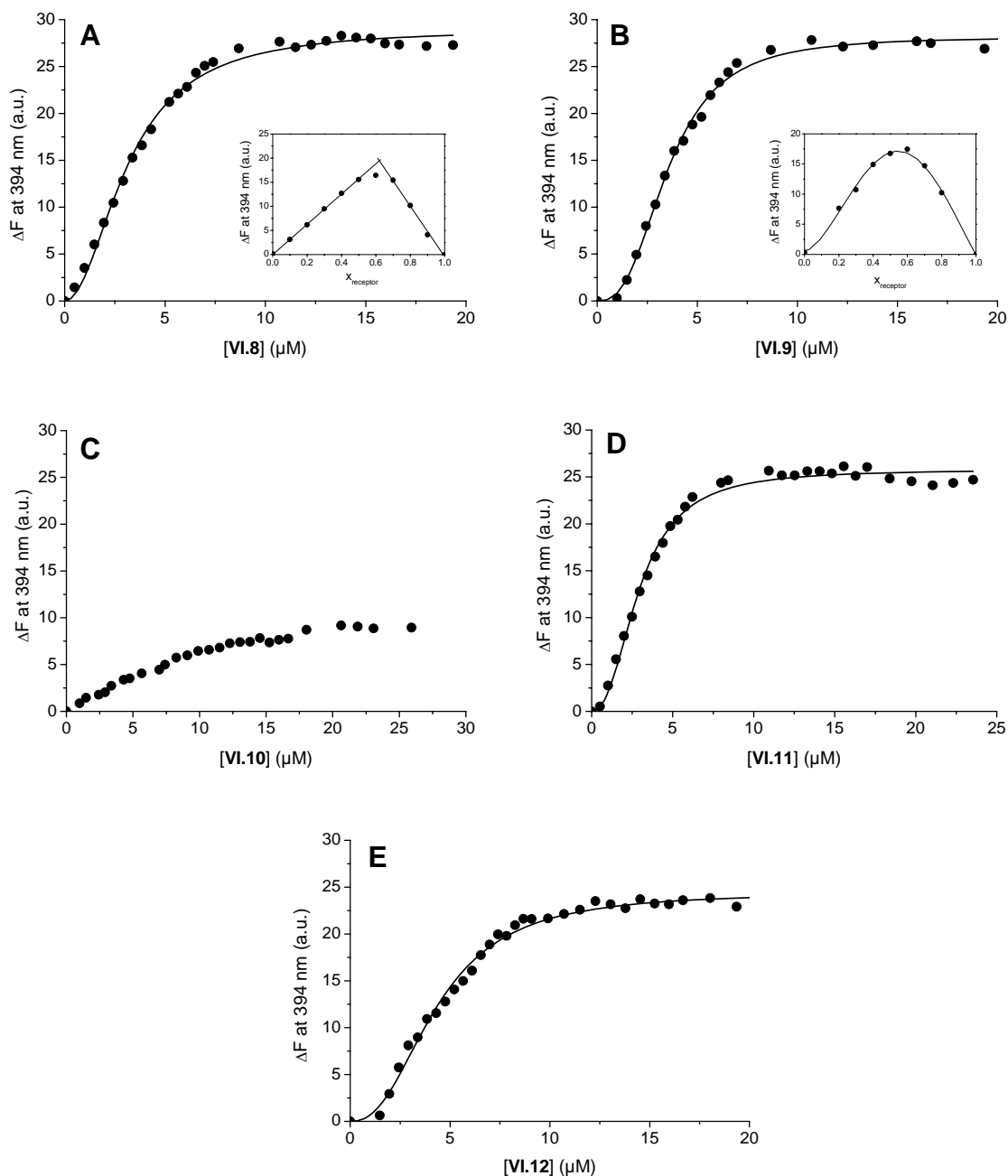
interactions of the peptide with the two receptor binding sites. Furthermore, it must be noted that the Id helix-2 shows a high intrinsic propensity to fold into stable amphiphilic  $\alpha$ -helices which can undergo self-association. This results in a well-defined side-chain arrangement on the helical surface, in which most of the potential  $\text{Cu}^{\text{II}}$ -IDA interacting residues, including histidine, aspartic and glutamic acid, and glutamine, would be located on the opposite helical face with respect to lysine. By contrast, the Id helix-1 possesses a very low intrinsic helix propensity and is, therefore, assumed to be rather flexible as isolated fragment. Moreover, the putative receptor binding site *CYSR(K)* features a cysteine residue, which is a strong  $\text{Cu}^{\text{II}}$ -IDA binder, and, two residues farther, a guanidinium or ammonium ion, which are potential ligands for the crown ether.

### ***The CYSR(K) motif***

To collect further evidence for the proposed receptor binding motif within the helix-1, we used a minimalist approach and examined the binding properties of several small peptides containing the *CYSR(K)* sequence (peptides **VI.8** and **VI.11**), and variations thereof: CASR (**VI.9**), AYSR (**VI.10**) and CYSRLR (**VI.12**). Peptides **VI.8**, **VI.11** and **VI.12**, featuring the native sequences, bind to **R1** with  $K_{0.5}$  values of 3.2  $\mu\text{M}$ , 2.9  $\mu\text{M}$  and 4.3  $\mu\text{M}$ , respectively, and with a receptor/peptide stoichiometry of 2:1, as extrapolated by the Job's plots (Figure VI.4). These  $K_{0.5}$  values suggest that the *CYSR(K)* motif is sufficient to bind the receptor in the low-micromolar range. Interestingly, the isolated *CYSR* motif showed an affinity comparable to the complete Id1 helix-1 (peptide **VI.4**), which points to a negligible contribution of the flanking amino acids in the longer peptide to the stabilization of the receptor/ligand complex. By contrast, the neighboring residues present in the Id3 and, especially, Id4 helix-1 (peptides **VI.5** and **VI.6**) seem to positively affect the binding, resulting in lower  $K_{0.5}$  values (2.1  $\mu\text{M}$  and 1.4  $\mu\text{M}$ , respectively). Strikingly, C-terminal elongation of the *CYSR* motif with leucine and an additional arginine (peptide **VI.12**) slightly impairs the binding. This might be caused by a competition between the two arginine residues for the coordination to the crown ether.

Substitution of the tyrosine residue adjacent to cysteine with an alanine (peptide **VI.9**) is accompanied by a slight increase in the  $K_{0.5}$  value (from 3.2  $\mu\text{M}$  for **VI.8** to 3.5  $\mu\text{M}$  for **VI.9**), leading to the assumption that the phenol ring positively influences the binding. Finally, the

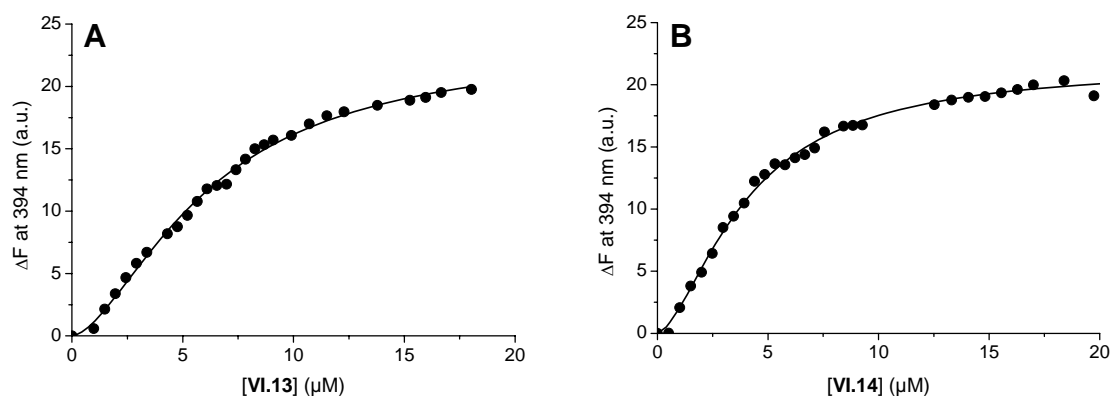
replacement of cysteine with alanine (peptide **VI.10**) leads to a dramatic loss of fluorescence response, indicating the need for a thiol group to obtain a strong binding to **R1**.



**Figure VI.4.** Titration curves of **R1** (9.5  $\mu M$  in 50 mM HEPES buffer, pH 7.5) with peptides **VI.8** (A), **VI.9** (B), **VI.10** (C), **VI.11** (D) and **VI.12** (E). Solid lines represent the data fit obtained using the Hill equation. Insets in panels A and B: Job's plot analyses.

### HLH domains of related proteins

As within the HLH protein family the *CxxR/K* motif is only found in the helix-1 of the Id proteins, we reckoned that the receptor **R1** might show some selectivity towards this subgroup of the HLH superfamily. Therefore, we tested the HLH motifs of MyoD (**VI.13**) and Max (**VI.14**), which are both natural dimerization partners of the Id proteins. Like the Id HLH motifs, also the MyoD HLH motif contains a cysteine residue preceded by two adjacent basic residues (*KRC*) in proximity of the C-end of helix-1. Nevertheless, peptide **VI.13** shows a reduced binding affinity to **R1** compared to the Id HLH motifs (Figure VI.5). A similar binding behavior is observed for the Max peptide **VI.14**, despite the presence of two histidine residues in the first helix at positions 3 and 9 close to basic residues like Lys<sup>5</sup> and Arg<sup>12</sup>.



**Figure VI.5.** Titration curves of **R1** (9.5  $\mu$ M in 50 mM HEPES buffer, pH 7.5) with peptides **VI.13** (A) and **VI.14** (B). Solid lines represent the data fit obtained using the Hill equation.

### Summary

For all measured peptides the  $K_{0.5}$  and  $h$  values were obtained from the Hill fitting of the curves (Table **VI.2**). The  $K_{0.5}$  values for the Id sequences increased in the order: Id2 HLH motif, Id4 helix-1 ( $\sim 1.5 \mu$ M) < Id1 HLH motif, Id3 helix-1 ( $\sim 2 \mu$ M) < Id3 HLH motif (2.31  $\mu$ M) < Id4 HLH motif (2.45  $\mu$ M) < *CYSR(K)*, Id1 helix-1 ( $\sim 3 \mu$ M) < *CASR* (3.46  $\mu$ M) < *CYSRLR* (4.33  $\mu$ M). Higher  $K_{0.5}$  values (5.2-6.0  $\mu$ M) were found for the peptides **VI.13** and **VI.14** representing the HLH domains of MyoD and Max. For peptides **VI.7** and **VI.10**, corresponding to the Id helix-2 and to the cysteine-free tetrapeptide AYSR, respectively, no Hill fitting was carried out, as only a weak fluorescence response was obtained.

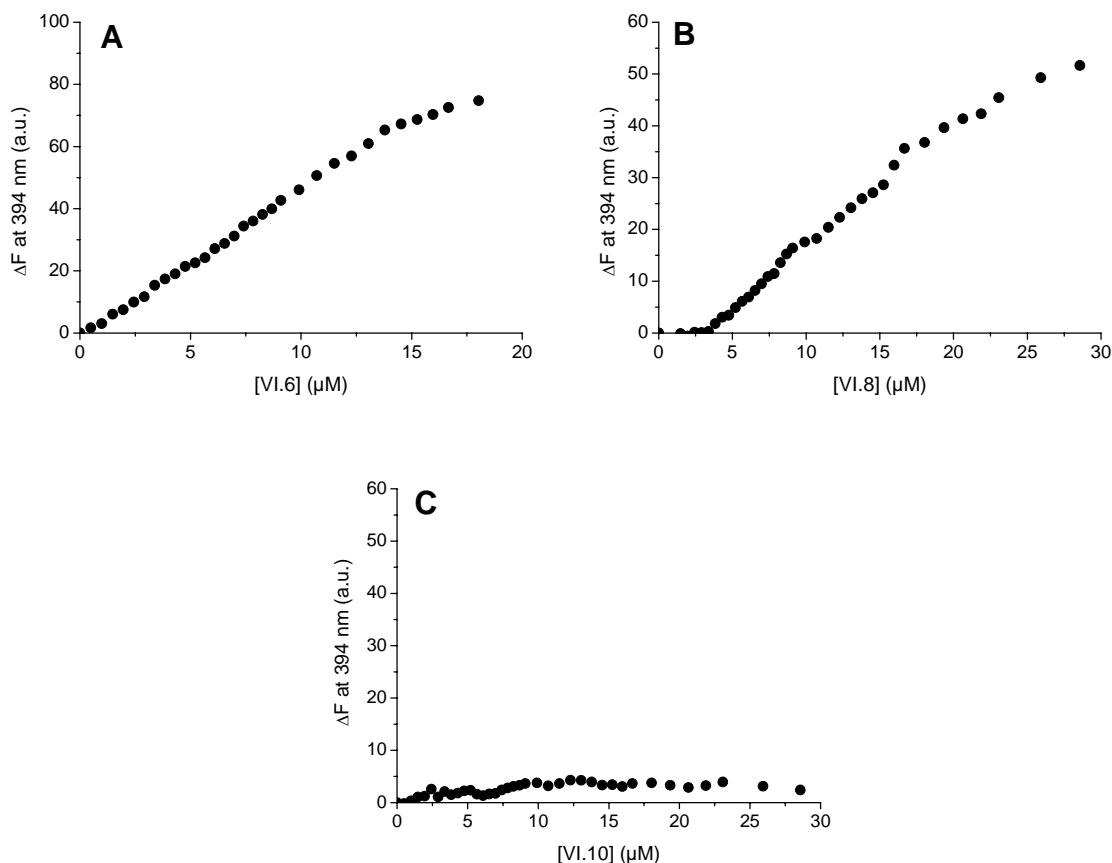
**Table VI.2.** Binding data for the synthetic peptides towards the receptor **R1**.  $K_{0.5}$  and  $h$  were obtained by fitting the experimental curves with the Hill equation.

Peptide	$K_{0.5}$ ( $\mu\text{M}$ )	$h$
<b>III.1</b>	$2.03 \pm 0.23$	$1.48 \pm 0.12$
<b>VI.1</b>	$1.52 \pm 0.13$	$1.62 \pm 0.13$
<b>VI.2</b>	$2.31 \pm 0.17$	$1.52 \pm 0.12$
<b>VI.3</b>	$2.45 \pm 0.05$	$1.89 \pm 0.05$
<b>VI.4</b>	$3.22 \pm 0.07$	$1.93 \pm 0.09$
<b>VI.5</b>	$2.06 \pm 0.05$	$1.91 \pm 0.09$
<b>VI.6</b>	$1.43 \pm 0.05$	$2.36 \pm 0.16$
<b>VI.7</b>	weak fluorescence response observed	
<b>VI.8</b>	$3.15 \pm 0.07$	$2.03 \pm 0.09$
<b>VI.9</b>	$3.46 \pm 0.06$	$2.91 \pm 0.14$
<b>VI.10</b>	weak fluorescence response observed	
<b>VI.11</b>	$2.93 \pm 0.06$	$2.33 \pm 0.12$
<b>VI.12</b>	$4.33 \pm 0.09$	$2.42 \pm 0.12$
<b>VI.13</b>	$5.98 \pm 0.20$	$1.65 \pm 0.06$
<b>VI.14</b>	$5.18 \pm 0.15$	$1.92 \pm 0.10$

### *The NTA-containing receptor*

To examine the influence of the metal complex on the binding, the two peptides **VI.6** and **VI.8**, which showed a good to moderate binding to receptor **R1**, and peptide **VI.10**, showing a weak fluorescence response of **R1**, were tested for affinity to receptor **R2** that contains, besides a modified linker bearing an amide bond, the  $\text{Cu}^{\text{II}}$ -NTA moiety in place of the  $\text{Cu}^{\text{II}}$ -IDA one (Figure VI.6). In the case of the cysteine-free tetrapeptide **VI.10** (AYSR), also the receptor **R2**, like **R1**, showed very low increase in the emission intensity upon increasing peptide concentration. This result was expected, as **VI.10** does not contain any residues which are able to strongly bind to the  $\text{Cu}^{\text{II}}$ -NTA complex. Instead, increasing concentrations of peptides **VI.6** and **VI.8** were accompanied by a linear increase in fluorescence, which might be explained by an unspecific peptide binding to the receptor. One reason for the different behavior of the two receptors might be the reduced coordination sites at the  $\text{Cu}^{\text{II}}$ -NTA complex, which prevent a successful binding to the peptide. However, also

the linker of receptor **R2** differs from the one of **R1**. While featuring the same chain length as the linker of **R1**, the linker of **R2** is rigidified by the presence of a central amide bond. Thus, it can not be excluded that the different binding behavior towards the same peptide ligands might also stem from different structural properties of the receptor itself.

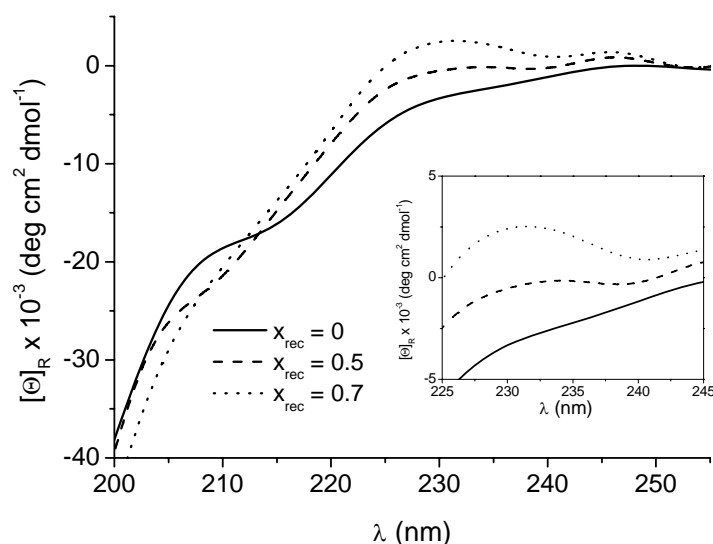


**Figure VI.6.** Titration curves of **R2** (9.5 μM in 50 mM HEPES buffer, pH 7.5) with peptides **VI.6** (A), **VI.8** (B) and **VI.10** (C).

### CD experiments

As the interaction of an artificial receptor with a peptide ligand might result in a conformational change of the latter, we carried out CD measurements of the tetrapeptide **VI.8** in the absence and presence of different amounts of receptor **R1**, and analyzed the far-UV region that is commonly used to monitor changes in the backbone conformation (Figure VI.7). The tetrapeptide is flexible in water, as shown by the weak negative ellipticity between 215 nm and 240 nm, and the strong negative band below 200 nm. With increasing concentrations of **R1** the negative contribution to the CD curve above 215 nm diminishes and two new

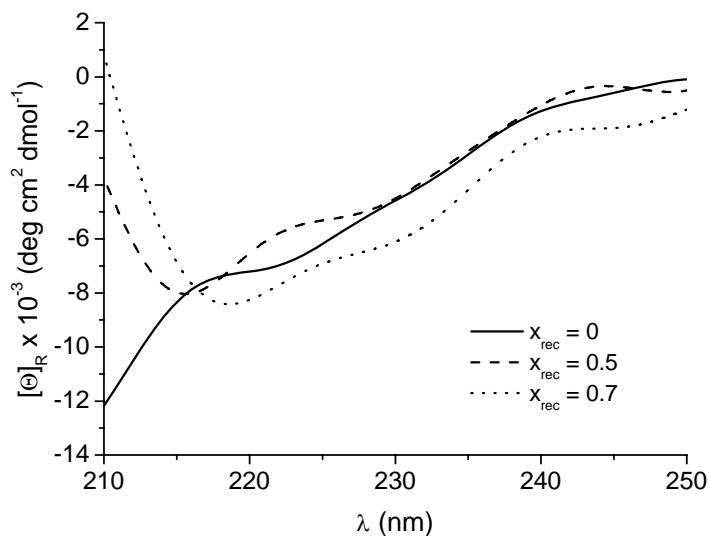
positive bands appear at 230 nm and 245 nm. This region is usually regarded to be sensitive to an interaction between the electronic transition of the peptide bond and the transition of an aromatic amino acid.<sup>12</sup> Thus, it can be speculated that a structural change in the peptide is induced upon binding to the receptor that alters the orientation of the tyrosine side chain to the main chain.



**Figure VI.7.** CD spectra of **R1/VI.8** mixtures in water. The concentration of **VI.8** was kept constant at 305  $\mu\text{M}$ , while the receptor concentration was varied to get receptor molar fractions  $x_{\text{rec}}$  between 0 and 0.7.

Like the Job's plots, also this CD experiment seems to indicate a receptor/peptide ratio higher than 1, as the maximum change in the CD spectra is observed at a receptor molar fraction higher than 0.5.

In addition, we also analyzed the Id4 helix-2 (peptide **VI.6**) in HEPES buffer (50 mM, pH 7.5) in the absence and presence of the receptor **R1** by CD spectroscopy (Figure VI.8). Unfortunately, technical reasons prevented the acquisition of the CD spectra below 210 nm. However, above this region some conformational changes could be noticed upon receptor addition: for example, the negative shoulder at 220 nm of the peptide alone disappeared and two new ones emerged at 230 nm and 245 nm, which might reflect an elicited contribution of the aromatic residue to the CD spectrum. Moreover, the decrease in the negative CD contribution below 215 nm and the concomitant appearance of a minimum around 216-218 nm suggest that the aperiodic, irregular peptide structure no longer prevails.



**Figure VI.8.** CD spectra of **R1/VI.6** mixtures in HEPES buffer (50 mM, pH 7.5).

Interestingly, the changes in the dichroic properties of the two peptides in the presence of the copper-containing receptor show some similarities to those observed for other flexible  $\text{Cu}^{2+}$ -binding peptides upon metal addition: in particular, the CD spectra of the tyrosine-containing amyloid  $\beta$  peptides (1-16)-A $\beta$  and (1-28)-A $\beta$  in water at pH 7.5-7.8 were characteristic of disordered structures in the absence of the metal, but displayed a positive band close to 225 nm and a decrease in the negative band below 215 nm in the presence of the metal.<sup>13</sup>

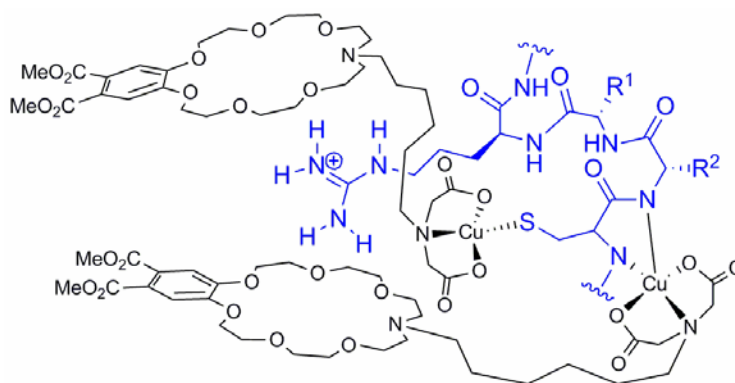
### VI.3 Discussion

The artificial receptor **R1** developed by König and co-workers for the combined recognition of Lewis-acidic and Lewis-basic amino acid residues was shown to bind to the HLH domain of the biologically relevant Id proteins. By using a minimalist approach, the binding motif was identified to be *CxxR/K*, a sequence located in the N-terminal helix of the Id HLH motif and unique for the Id family. Therein, the accommodation of the receptor might be facilitated by the rather flexible conformation observed in this region. Accordingly, the isolated C-terminal Id helix-2 showed weak interaction with the receptor, which might not only mirror the lack of a strong  $\text{Cu}^{\text{II}}$ -IDA binder in proximity of a basic residue, but also the tendency of the helix-2 to adopt a stable helical conformation with a well-defined spatial distribution of the side chains.

Although the Id1 helix-1 is sufficient to bind the receptor in the low-micromolar concentration range, the increase in the  $K_{0.5}$  value compared to the full-length HLH motif suggests that additional structural and functional elements arising from the loop and/or helix-2 can benefit the receptor binding. Contrary thereto, both the first helices of Id3 and Id4 exhibit smaller  $K_{0.5}$  values compared to those of the corresponding full-length HLH motifs, rather proposing a negative structural influence of the additional regions. Compared to the other parent Id proteins, both the helix-1 and the HLH sequence of Id4 showed another deviating feature: for example, the Id4 HLH motif was the only one showing a 1:1 receptor/peptide ratio. Presumably, the second cysteine residue in the helix-1, exclusively present in Id4, compensates the absence of the loop and helix-2 regions and even increases the ability of the isolated Id4 helix-1 to interact with the receptor, whereas this benefit might be less significant in the full-length Id4 HLH motif due to a more rigid helix folding.

With the exception of the Id4 HLH peptide, all other Id peptides were characterized by a receptor/peptide ratio higher than one. Also for the short tetrapeptide CYSR this stoichiometry was observed by Job's plot analysis and supported by CD measurements. The 2:1 stoichiometry might result from two binding processes: a first receptor molecule specifically binds to the epitope  $CxxR/K$ , while a second one non-specifically binds to the peptide backbone (Scheme VI.1). In fact, the ability of backbone amides to coordinate  $\text{Cu}^{2+}$  has been described for some peptide sequences, primarily for the amyloid peptide  $\text{A}\beta$ .<sup>14, 15</sup>

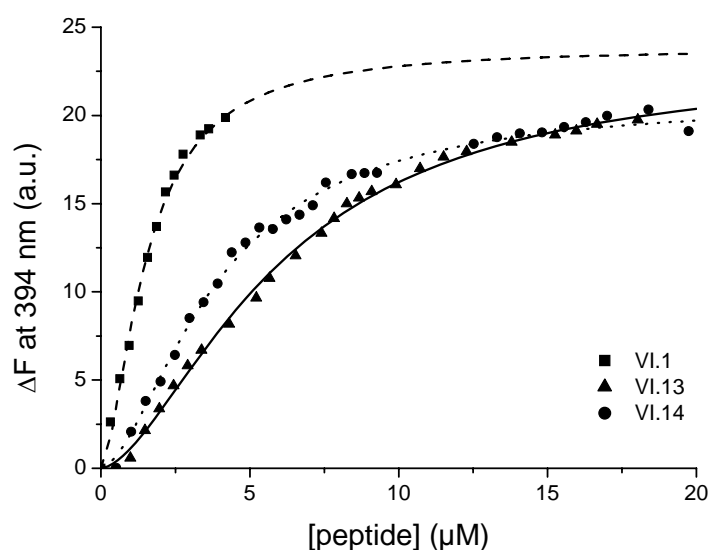
**Scheme VI.1.** A possible 2:1 binding model for the **R1**/ $CxxR$  complex. Peptide in blue, receptor molecules in black. Vacant coordination sites may be occupied by water.



Also for the peptides that do not contain the  $CxxR/K$  motif, an increase in peptide concentration was accompanied by a slight, but nevertheless detectable increment of the fluorescence response, which might stem from a receptor molecule anchored to the peptide backbone by its  $\text{Cu}^{\text{II}}$ -IDA complex.



The identified CxxR/K sequence might represent a target for the selective recognition of the Id proteins, as it is not found in the related HLH proteins.<sup>16</sup> The selectivity of the receptor towards this binding motif was tested with measurements on the HLH motifs of the Max and MyoD proteins, which, indeed, showed a moderately reduced binding compared to the Id HLH domains (Figure VI.9). Obviously, clearly superior affinity and selectivity are needed for an artificial receptor to be applied as a tool for the Id protein targeting. However, this study is the first attempt of molecular recognition of the dimerization domains of the Id proteins and may represent a starting point for the development of chemical devices able to modulate the protein-protein interactions involving the Id family.



**Figure VI.9.** Superposition of the titration curves of receptor **R1** with the HLH motifs of Id2 (**VI.1**), MyoD (**VI.13**) and Max (**VI.14**).

## VI.4 References

1. Harris, C. C., Protein-protein interactions for cancer therapy. *Proc Natl Acad Sci USA* **2006**, 103, 1659-60.
2. Berg, T., Modulation of protein-protein interactions with small organic molecules. *Angew Chem Int Ed Engl* **2003**, 42, 2462-81.
3. Arkin, M., Protein-protein interactions and cancer: small molecules going in for the kill. *Curr Opin Chem Biol* **2005**, 9, 317-24.
4. Yin, H.; Hamilton, A. D., Strategies for targeting protein-protein interactions with synthetic agents. *Angew Chem Int Ed Engl* **2005**, 44, 4130-63.

5. Veselovsky, A. V.; Ivanov, Y. D.; Ivanov, A. S.; Archakov, A. I.; Lewi, P.; Janssen, P., Protein-protein interactions: mechanisms and modification by drugs. *J Mol Recognit* **2002**, 15, 405-22.
6. Fazal, M. A.; Roy, B. C.; Sun, S.; Mallik, S.; Rodgers, K. R., Surface recognition of a protein using designed transition metal complexes. *J Am Chem Soc* **2001**, 123, 6283-90.
7. Hossain, M. A.; Schneider, H. J., Sequence-Selective Evaluation of Peptide Side-Chain Interaction. New Artificial Receptors for Selective Recognition in Water. *J Am Chem Soc* **1998**, 120, 11208-11209.
8. Kruppa, M.; Mandl, C.; Miltschitzky, S.; Konig, B., A luminescent receptor with affinity for N-terminal histidine in peptides in aqueous solution. *J Am Chem Soc* **2005**, 127, 3362-5.
9. Rosenberg, I. M., Protein Analysis and Purification: Benchtop Techniques. *Birkhäuser, Basel* **2004**.
10. Nomura, M.; Kobayashi, T.; Kohno, T.; Fujiwara, K.; Tenno, T.; Shirakawa, M.; Ishizaki, I.; Yamamoto, K.; Matsuyama, T.; Mishima, M.; Kojima, C., Paramagnetic NMR study of Cu(2+)-IDA complex localization on a protein surface and its application to elucidate long distance information. *FEBS Lett* **2004**, 566, 157-61.
11. Gaberc-Porekar, V.; Menart, V., Perspectives of immobilized-metal affinity chromatography. *J Biochem Biophys Methods* **2001**, 49, 335-60.
12. Liang, J. N.; Chakrabarti, B., Spectroscopic investigations of bovine lens crystallins. 1. Circular dichroism and intrinsic fluorescence. *Biochemistry* **1982**, 21, 1847-52.
13. Syme, C. D.; Nadal, R. C.; Rigby, S. E.; Viles, J. H., Copper binding to the amyloid-beta (A $\beta$ ) peptide associated with Alzheimer's disease: folding, coordination geometry, pH dependence, stoichiometry, and affinity of A $\beta$ -(1-28): insights from a range of complementary spectroscopic techniques. *J Biol Chem* **2004**, 279, 18169-77.
14. Hou, L.; Zagorski, M. G., NMR reveals anomalous copper(II) binding to the amyloid A $\beta$  peptide of Alzheimer's disease. *J Am Chem Soc* **2006**, 128, 9260-1.
15. Miura, T.; Suzuki, K.; Kohata, N.; Takeuchi, H., Metal binding modes of Alzheimer's amyloid beta-peptide in insoluble aggregates and soluble complexes. *Biochemistry* **2000**, 39, 7024-31.
16. Chavali, G. B.; Vijayalakshmi, C.; Salunke, D. M., Analysis of sequence signature defining functional specificity and structural stability in helix-loop-helix proteins. *Proteins* **2001**, 42, 471-80.





## VII Summary

The Id proteins take part in many fundamental physiological processes during development and tumor-related events. A characteristic of these rather small proteins (119 up to 161 amino acids long) is a highly conserved dimerization domain, the helix-loop-helix (HLH) motif, which allows them to homo- or heterodimerize and makes them a subfamily of the large HLH protein family. Their preferred binding partners are transcription factors from the same family, whose homo- or heterodimers are able to bind to DNA by basic residues adjacent to the N-terminal end of the HLH fold. As the Id proteins lack such a basic region, dimers formed under their participation lose their ability to bind to DNA and, thus, transcriptional activation is blocked. By using this mechanism of inhibition of DNA binding, the *Id* family members regulate cellular processes like proliferation and differentiation. Although recent works also reported on non-HLH proteins able to interact with the Id proteins, dimerization with related HLH proteins seems to remain the major mode of action of the Id proteins. Structural information on the Id proteins is scarce and was so far deduced by sequence alignment and homology studies based on known HLH structures, disregarding their unique position within the HLH protein family.

The goal of the presented work was the elucidation of the conformational features underlying the Id HLH fold. For this purpose, the native Id sequences as well as a series of analogs based on amino acid substitution/truncation/deletion, sequence inversion (retro sequence) and disulfide bond formation were synthesized by solid-phase methodology and characterized by CD and NMR spectroscopy. In addition, the ability of small molecule receptors to bind to the Id HLH motifs was investigated by fluorescence spectroscopy.

A first approach was based on the design of Id1 HLH analogs displaying modified loop regions and/or the retro sequence of the first helix. As shown by CD experiments, the synthesized peptides possessed a broad diversity in terms of secondary structure elements composition. An analog containing the HLH domain with the retro helix-1 showed superior propensity to aggregate compared to the native HLH motif, as indicated by poor solubility in aqueous solutions. Nevertheless, this analog turned out to have an increased helical content, proving the importance not only of the interhelical side-chain packing, but also of the helix dipole arrangement for the HLH fold stabilization. The shortest analog that maintained a comparable helicity to the full length HLH motif was the one containing a  $\delta$ -ornithine residue in place of the native loop region, which should restrict the backbone conformational freedom

to an anti-parallel orientation of the helix dipoles, and the retro sequence of the first helix, which should allow for parallel side-chain orientation.

Further, by using tyrosine/phenylalanine replacement Id1 HLH analogs were obtained which showed a superior helicity compared to the native sequence, thus underlining the fact that the natural tyrosine side chains are predominantly participating in hydrophobic interactions, which result in the formation and stabilization of the hydrophobic core of the protein fold. Also the presence of cation- $\pi$  interactions could explain the positive effect of the tyrosine/phenylalanine substitution, as phenylalanine stabilizes such contacts better than tyrosine.

In another approach, covalently linked Id1 HLH dimers were formed by cysteine oxidation of analogs that contained GGC units attached to either the N-terminal or the C-terminal ends of the HLH motif, respectively. The oxidation proceeded fast and almost to completion, as monitored by mass spectrometry and analytical HPLC, indicating that a parallel orientation of the first and second helices, respectively, should be favored in the HLH dimer. A thiol exchange assay did not produce the heterodimer, in which the helix-1 and the helix-2 of the corresponding monomer would adopt an anti-parallel conformation. This observation supports a preferred parallel orientation of the helices in the Id1 HLH dimer, as it was also shown in the crystal structures of E47 and MyoD.

To further characterize the structural properties of the Id HLH fold, the HLH domain of the Id2 protein was investigated by NMR spectroscopy. Due to its pronounced aggregation tendency, the fluorinated alcohol TFE had to be added to improve the peptide solubility to a level, at which reasonable NMR spectra could be recorded. Indeed, the quality of the obtained NMR spectra proved to be sufficient to achieve the sequential assignment of the peptide. The chemical shift index (CSI) and the NOE patterns suggest the presence of a poorly defined N-terminal helix and of a stable C-terminal  $\alpha$ -helix, connected by a short flexible region. These results are in accordance with NMR spectroscopic investigations on the related HLH domains of E47 and Max.

Finally, the ability of a ditopic receptor to bind to the Id HLH motifs was investigated by fluorescence spectroscopy. The receptor consists of a fluorescent crown ether unit linked to a  $\text{Cu}^{\text{II}}$ -IDA complex by an alkyl chain. Only upon simultaneous binding to both moieties a fluorescence response is triggered. The receptor was shown to bind to peptides representing the four Id HLH domains with  $K_{0.5}$  values ranging from 1.5  $\mu\text{M}$  to 2.5  $\mu\text{M}$ . Further binding experiments on shorter Id peptide fragments allowed the identification of the tetrapeptide CxxR/K as strong binding motif located in the first helix of the Id HLH domains. The receptor

showed also moderate selectivity towards the Id HLH motifs compared to the ones of the related HLH proteins MyoD and Max, which showed  $K_{0.5}$  values of 6.0  $\mu\text{M}$  and 5.2  $\mu\text{M}$ , respectively. A similar receptor, containing a  $\text{Cu}^{\text{II}}$ -NTA in place of the  $\text{Cu}^{\text{II}}$ -IDA unit and a modified linker between the crown ether and the metal complex, did not succeed to bind the *CxxR* motif.

In conclusion, the presented work has provided additional information on the structural properties of the Id proteins. The oxidized Id1 HLH dimer is likely to prefer a parallel helix arrangement. Previous reports on the increased flexibility of helix-1 compared to helix-2 could be supported by the NMR data obtained for the Id2 HLH domain. By using a fluorescent artificial receptor, a unique tetrapeptide motif within the helix-1 could be identified, which might represent a starting point for the development of Id specific markers and small molecules able to interfere with the Id cellular pathways.





## VIII Experimental part

### VIII.1 Materials

The N<sup>α</sup>-Fmoc protected amino acids and HBTU were purchased from MultiSynTech (Witten, Germany); Fmoc-ε-aminohexanoic acid, Fmoc-β-alanine, Fmoc-ornithine(Mtt) and Fmoc-lysine(Mtt) were obtained from Novabiochem (Darmstadt, Germany). Peptide-synthesis grade DMF, NMP, piperidine and diethylether, ACN and TFA for spectroscopy were purchased from Biosolve (Valkenswaard, The Netherlands). HOBt, DIEA, TFA, DCM, DMSO, HEPES, reduced L-glutathione, oxidized L-glutathione and α-cyano-4-hydroxycinnamic acid for mass spectrometry were bought from Fluka (Taufkirchen, Germany). Methanol was obtained from J.T. Baker (Deventer, The Netherlands), TIS from Aldrich (Steinheim, Germany) and TFE from Acros (Geel, Belgium). EDT, acetic anhydride, acetic acid, NaH<sub>2</sub>PO<sub>4</sub> and Na<sub>2</sub>HPO<sub>4</sub> were purchased from Merck (Darmstadt, Germany). TFE-d<sub>2</sub> and D<sub>2</sub>O were from Deutero (Kastellaun, Germany).

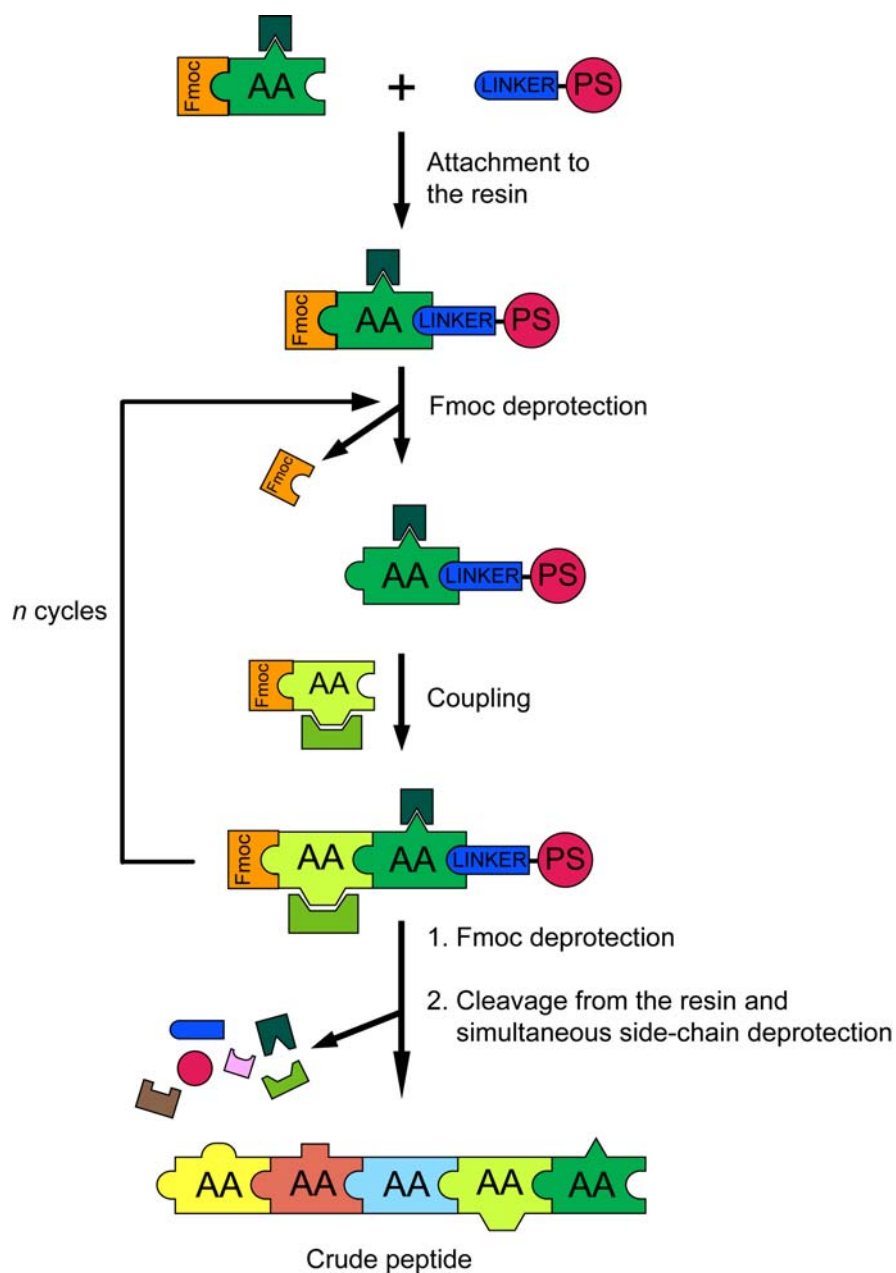
### VIII.2 Methods

#### VIII.2.1 Peptide synthesis and purification

##### VIII.2.1.1 Solid-phase peptide synthesis (SPPS)

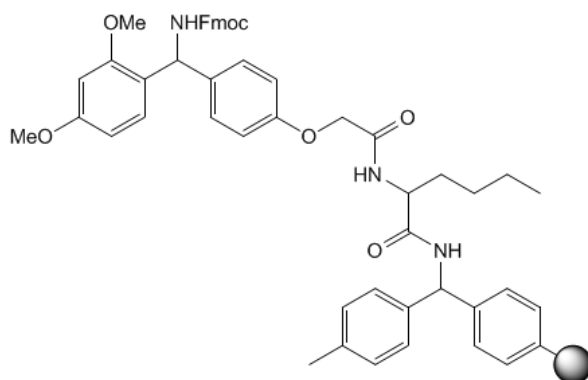
Developed in 1963 by the late Bruce Merrifield (Nobel Prize 1984) and steadily being improved since then, the solid-phase methodology is an indispensable tool for the synthesis of peptides and proteins, whose preparation by biochemical methods is challenging or impossible.<sup>1</sup> The C-terminal amino acid of the peptide to be synthesized is attached to an insoluble solid support *via* a so-called linker and, step by step, elongated with the successive amino acids (Scheme VIII.1). As excess reagents are washed away, tedious purification steps after each reaction, as in solution chemistry, are avoided. Moreover, this strategy is suitable for an automated process, which led to the development of sophisticated peptide synthesizers.

**Scheme VIII.1.** Fmoc-strategy based solid-phase peptide synthesis. *PS*: polymer matrix copoly(styrene-1% DVB); *AA*: amino acid.



#### *Solid support and linker*

The polystyrene beads used as solid support are functionalized with a linker, which serves as point of attachment for the first amino acid. The different commercially available resins distinguish themselves by the type of linker, which allows the preparation of peptides with different C-terminal ends, including acids, amides, hydrazines and aldehydes. The peptides presented in this work were synthesized on Rink amide MBHA resin (Scheme VIII.2) to obtain them as peptide amides.

**Scheme VIII.2.** Rink amide MBHA resin.

### Amino acid coupling

The reaction between an amine and a carboxylic acid to obtain a peptide bond requires the use of activation reagents. Widespread used in SPPS are phosphonium (PyBOP) or uronium salts (HBTU, TBTU, HATU) which all allow an *in situ* active ester formation.

### Protecting groups

The synthesis of peptides necessitates the use of appropriate protecting groups. While the side-chain functionalities have to be protected permanently during the synthesis to prevent the formation of undesired side products (Table VIII.1), the  $\alpha$ -amino group of the amino acid to be acylated has to be masked only temporarily to allow further chain elongation.

Frequently used temporary  $\alpha$ -amino protecting groups in peptide chemistry are, in the order of their development, benzyloxycarbonyl (Cbz)<sup>2</sup>, tert-butyloxycarbonyl (Boc)<sup>3, 4</sup> and fluorenyl-9-methoxycarbonyl (Fmoc)<sup>5, 6</sup>, the latter two being commonly applied in the automated SPPS. The Boc and Fmoc strategies differ in the removal of the respective  $\alpha$ -amino protecting group: while the Boc-group requires 50% TFA to be cleaved, the Fmoc-group is cleaved under basic conditions (20-40% piperidine). The advantages of the Fmoc-strategy compared to the Boc-strategy are: (i) exceptional acid stability, (ii) rapid nonhydrolytic deblocking by simple amines, (iii) compatibility with tert-butyl- and benzyl-based systems, (iv) direct liberation of protected amine in the free base form and (v) facile UV or fluorescence monitoring of deblocking.

**Table VIII.1.** Recommended side-chain protecting groups in the Boc- and Fmoc-strategy.<sup>7</sup> Those used in the presented work are underlined.

Amino acid	Suitable protecting groups	
	Boc-strategy	Fmoc-strategy
Cys	Acm, 4-MeBzl, 4-OMeBzl	Acm, tBu, <u>Trt</u>
Asp	Bzl, cHx	<u>tBu</u> , MPe
Glu	Bzl, cHx	<u>tBu</u>
His	Dnp, Bom, Tos, Cbz	<u>Trt</u>
Lys	2-Cl-Cbz	<u>Boc</u> , ivDde, Mmt
Asn	--	<u>Trt</u>
Gln	--	<u>Trt</u>
Arg	Tos, Mts, NO <sub>2</sub>	Pmc, <u>Pbf</u>
Ser	Bzl	<u>tBu</u>
Thr	Bzl	<u>tBu</u> , Trt
Trp	For	<u>Boc</u>
Tyr	2-Br-Cbz	<u>tBu</u> , 2-ClTrt
Ala, Phe, Gly, Ile, Leu, Met, Pro, Val	unprotected	unprotected

#### *Temporary side-chain protecting groups*

Besides the permanent protection of the amino acid side chains mentioned above, also temporary side-chain protecting groups can be applied, which are stable both at the conditions used for the removal of the other side-chain and  $\alpha$ -amino protecting groups. One example is the methyltrityl (Mtt) protecting group for lysine which can be selectively removed with 1% TFA in the presence of 1% triisopropylsilane (TIS) and allows the preparation of branched peptides or the introduction of fluorescent labels.

#### *Total cleavage*

During the repetitive deprotection and coupling steps of the peptide synthesis, the linkage of the peptide to the resin has to be stable, while it has to be cleaved in the last synthesis step to release the fully synthesized peptide. The final cleavage usually requires harsh conditions like anhydrous HF or other strong acids in Boc-strategy and trifluoroacetic acid (TFA) in Fmoc-strategy. Conveniently, the side-chain protecting groups are chosen to be also cleaved under these conditions. During the cleavage highly-reactive electrophilic species are formed, including tert-butyl and trityl cations and the resin-bound linker itself. In order to

prevent the reattachment of these species to the peptide chain, nucleophiles like TIS, triethylsilane (TES) or water have to be added to the cleavage mixture as scavengers.

### VIII.2.1.2 Experimental procedure

#### *General procedure*

All peptides were prepared on a 0.018 mmol scale using Fmoc chemistry on Rink amide MBHA resin (loading 0.6 mmol g<sup>-1</sup>) on a Syro-I peptide synthesizer from MultiSynTech. The following side-chain protections were applied: tBu (Asp, Glu, Ser, Thr), Boc (Lys), Trt (Cys, His, Asn, Gln) and Pbf (Arg). Chain assembly was accomplished by a double coupling procedure (2 x 40 min) using Fmoc-amino acid/HOBt/HBTU/DIEA (5:5:5:10 equiv.) in DMF/NMP (70:30, v/v), followed by Fmoc removal with piperidine in DMF/NMP (80:20, v/v) (3 min with a 40% base solution plus 10 min with a 20% base solution).

To monitor the progress of the synthesis, test cleavages using small amounts of resin were conducted and the resulting samples were analyzed by MALDI-ToF-MS and analytical HPLC. Final cleavage of the peptides from the resin and simultaneous side-chain deprotection was achieved by treatment of the peptidyl resin with 1.5 mL of a TFA/water/TIS mixture for three hours. For Met/Cys-containing peptides EDT was added to the cleavage cocktail. In all cases the ratio TFA/scavengers was 9:1 (v/v). The peptides were then precipitated from ice-cold ether, centrifuged and subjected to several ether-washing/centrifugation cycles to remove the scavengers.

#### *Mtt cleavage*

The Mtt protecting group used for the orthogonal side-chain protection of lysine and ornithine in the peptides **III.5-III.9** was cleaved by washing the resin with a solution containing 1% TFA and 1% TIS in DCM (ten cycles). After each cycle the resin was washed with DCM. Finally, the resin was washed with DCM (5 x) and DMF (5 x). After washing with a 1.2 M DIEA solution, the resin was washed again with DMF, DCM and diethyl ether. During the preparation of peptide **III.8** the Mtt cleavage resulted to be incomplete, which led to a truncated analog as a side-product (peptide **III.9**).

### *Methionine reduction*

Peptides that showed an oxidation of the methionine residue were treated with a reduction mixture consisting of TFA/TMSBr/EDT (95:2.5:2.5, v/v) for 1 h. Then, the peptides were precipitated from ice-cold ether and washed as described above.

#### **VIII.2.1.3 Preparative HPLC**

The synthesized peptides were purified by preparative RP-HPLC on a Phenomenex Luna C-18(2) column (10  $\mu$ m, 250  $\times$  21.2 mm). The crude products were dissolved in water or 20% acetic acid and filtered through polytetrafluoroethylene or polyvinylidene difluoride syringe filters. 0.004% (v/v) TFA in water (solvent A) and ACN (solvent B) were used as binary elution system. The gradient was 15% B for 5 min followed by 15-75% B over 66 min; the flow rate was 21 mL min<sup>-1</sup>. UV detection was done at 220 nm.

After collecting the fractions and evaporating the ACN, the peptides were lyophilized. The pure products were characterized by analytical HPLC and mass spectrometry, as reported in the following. All peptides were obtained with purities higher than 93%.

## **VIII.2.2 Oxidation experiments**

### **VIII.2.2.1 Monomer oxidation**

To produce the disulfide linked dimers **IV.1<sup>OX</sup>** and **IV.2<sup>OX</sup>**, the cysteine containing peptides **IV.1** and **IV.2** (1 mg each) were dissolved in 1 mL of degassed 0.1 M ammonium carbonate buffer (pH  $\sim$  8) and allowed to oxidize under air atmosphere in an opened Eppendorf vial. To follow the course of the oxidation reaction, small aliquots were taken after 0 h, 24 h, 48 h and 72 h, respectively, acidified by addition of 10  $\mu$ L of 20% acetic acid and subjected to analytical HPLC and MS measurements.

### **VIII.2.2.2 Thiol exchange assay**

Peptides **IV.1** and **IV.2** (1 mg each) were dissolved in 1 mL 0.1 M ammonium carbonate buffer (pH  $\sim$  8) containing 20 mM reduced glutathione and 40 mM oxidized glutathione.

After 4.5 h, 22 h and 46 h small aliquots were taken, acidified with 10  $\mu$ L 20% acetic acid and subjected to analytical HPLC and MS measurements.

### VIII.2.3 Peptide characterization

#### VIII.2.3.1 Analytical HPLC

For the analytical HPLC measurements a small amount of the peptide was dissolved in millipore water containing 0.1% TFA. The analytical RP-HPLC instrument consisted of the following parts: an L-6200A Intelligent pump from Merck, a HP detector series 1050 from Agilent and a Phenomenex Luna C-18(2) column (90 Å, 3  $\mu$ m, 150  $\times$  4.60 mm). The binary elution system consisted of 0.012% (v/v) TFA in water (eluent A) and 0.01% (v/v) TFA in ACN (eluent B). The following gradients were applied: 5% B for 3 min, 5-40% B over 40 min (peptides **VI.8-VI.12**) and 10% B for 3 min, 10-70% B over 40 min (remaining peptides).

#### VIII.2.3.2 Mass spectrometry

For the mass spectrometry measurements a small sample of the peptide was dissolved in 0.1% TFA in millipore water and mixed with a solution of  $\alpha$ -cyano-4-hydroxycinnamic acid in MeOH/ACN (1:1, v/v). The mass spectra were then recorded on a Future GSG spectrometer (Bruchsal, Germany) for MALDI-ToF-MS analysis.

#### VIII.2.3.3 UV spectroscopy

The concentration of the peptide samples prepared for CD and NMR measurements were determined UV spectrophotometrically on a Cary 100 Conc from Varian (Darmstadt, Germany) using 1 cm quartz cuvettes obtained from Hellma (Müllheim, Germany). By measuring the tyrosine absorption at 280 nm the concentration was calculated according to the Lambert-Beer law using an extinction coefficient of 1480 M<sup>-1</sup> cm<sup>-1</sup> per tyrosine residue.

The ratio between the peptide concentration values obtained by UV and by weight gave the peptide content. In case of peptides lacking tyrosine or tryptophan residues the peptide contents of similar peptides containing these residues were applied to calculate the concentration.

## VIII.2.4 Structural characterization

As the biological action of peptides and proteins is determined by their structural properties, the conformational analysis of these biomolecules is a major goal in peptide and protein research. The most powerful methods for the structural characterization of peptides and proteins are FT-IR spectroscopy, CD spectroscopy, NMR spectroscopy and X-ray crystallography. While FT-IR and CD data provide insights in the presence or absence of secondary structure elements, the latter two allow a detailed description of the localization and spatial arrangement of such elements within the amino acid sequence. However, the advantage of the CD and FT-IR spectroscopy lie in the fast and simple way experiments can be conducted. Moreover, the influence of solvent, pH, ionic strength and temperature on the protein structure can be easily monitored. In the presented work the synthesized peptides have been characterized by CD and NMR spectroscopy.

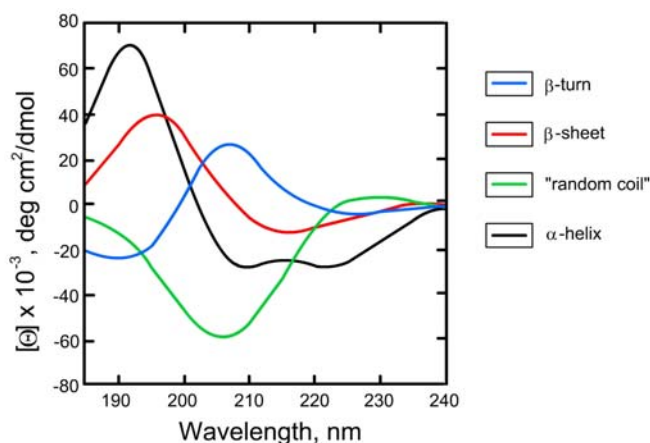
### VIII.2.4.1 CD spectroscopy

#### VIII.2.4.1.1 Introduction

CD spectroscopy is based on the differential absorption of left and right circularly polarized light. To deliver a CD signal, a compound has to possess a chromophore, which is either inherent chiral or surrounded by a chiral environment and absorbs in the observed wavelength range. These prerequisites are met by the amide bond in peptides and proteins, rendering the CD spectroscopy a suitable methodology for the elucidation of their structural preferences. The observed bands in the CD spectra of peptides and proteins stem from the  $n \rightarrow \pi^*$  and the  $\pi \rightarrow \pi^*$  transitions of the amide bonds.<sup>8</sup> Every secondary structure gives rise to a specific CD spectrum that represents a survey upon the averaged overall structure of a peptide or protein and it is impossible to assign structural preferences to an individual residue (Figure VIII.1). However, algorithms based on protein and peptide reference data sets enable the calculation of the secondary structure elements composition of a peptide/protein from its CD spectrum.

In addition to the far-UV region from 190-250 nm, also the near-UV region from 250 to 320 nm can deliver useful information, as it shows the contribution of aromatic residues like tyrosine or tryptophan.<sup>9, 10</sup>





**Figure VIII.1.** CD spectra characteristic for common secondary structure elements (adapted from Greenfield<sup>11</sup>).

The molar ellipticity  $[\Theta]_{\lambda}$  at the wavelength  $\lambda$  is defined as follows:

$$[\Theta]_{\lambda} = \frac{100 \cdot \Psi}{l \cdot c} \left[ \frac{\text{Deg} \cdot \text{cm}^2}{\text{dmol}} \right] \quad \text{Eq VIII.1}$$

where  $\Psi$  is the deviance of left- and right-circularly polarized light in degree,  $l$  the cell length in cm and  $c$  the peptide concentration in M. To compare peptides with different chain lengths, the molar ellipticity per residue is used, which is obtained by dividing  $[\Theta]_{\lambda}$  by the number of amino acids.

#### *Identification of interacting helices*

The CD bands characteristic for helical peptides at 190, 208 and 222 nm stem from the  $\pi \rightarrow \pi^*$  transition (190 nm and 208 nm are the components perpendicular and parallel to the helix axis, respectively) and  $n \rightarrow \pi^*$  transition (222 nm), with the latter being sensitive to hydrogen bond formation.<sup>12</sup> The ratio  $R$  between the two minima at 222 and 208 nm can be used as an indicator for the presence of interacting helices, as the band at 208 nm, corresponding to the parallel component of the  $\pi \rightarrow \pi^*$  transition, changes its energy upon an interhelical interaction. Ratios larger than 1 are commonly regarded as the result of interacting helices, as found in coiled coil structures, while ratios lower than 1 indicate the presence of non-interacting helices.<sup>13</sup>

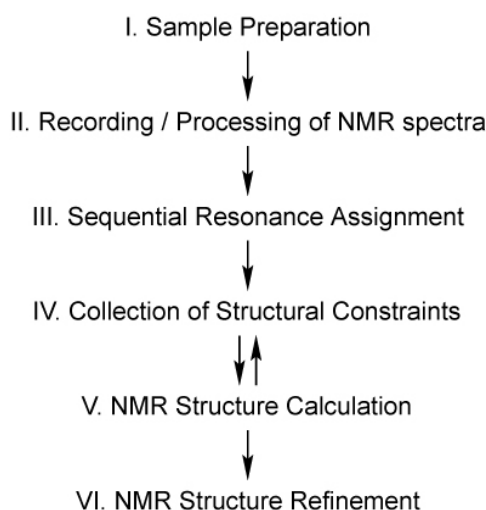
#### VIII.2.4.1.2 Experimental procedure

All CD spectra were recorded at room temperature on a JASCO J710 spectropolarimeter using a quartz cuvette with a path length of 0.02 cm. The peptide stock solutions were prepared in 0.1 M phosphate buffer (pH 7.3), and the concentrations were then determined UV spectrophotometrically. For each CD spectrum ten scans were accumulated using a step resolution of 1 nm, a bandwidth of 1 nm, a response time of 2 s, a scan speed of 20 nm min<sup>-1</sup> and a sensitivity of 10 mdeg, 20 mdeg or 50 mdeg, depending on the peptide concentration. To eliminate interferences which stem from cell, solvent and optical equipment, the CD spectrum of the buffer was subtracted from that of the peptide. The noise reduction was obtained by a Fourier transform filter using the program Origin (OriginLab Corporation, Northampton, USA). The obtained dichroic data were evaluated by using DICHROWEB<sup>14, 15</sup>, an online service for the evaluation of CD data. DICHROWEB relies on the five algorithms SELCON3, K2d, CDSSTR, VARSLC and CONTINLL<sup>16, 17</sup>; the latter one was used in combination with the reference dataset 7<sup>18</sup> for the CD analysis in the presented work. In all cases the analysis delivered a goodness-of-fit parameter RMSD that was below 0.13.

#### VIII.2.4.2 NMR spectroscopy

##### VIII.2.4.2.1 Introduction

Together with the X-ray crystallography, the NMR spectroscopy is the method of choice for determining three-dimensional protein structures. Drawbacks in the applicability of NMR experiments are the required solubility of the samples in the necessary concentration range (~0.5-1 mM) as well as the stability of the sample during the measurements. Moreover, the structure determination can be hindered by the lack of constraints. Following the sample preparation, the recording of 2D-NMR experiments represents the second step towards the protein structure determination, as outlined in Scheme VIII.3.

**Scheme VIII.3.** Protein structure determination by NMR.<sup>19</sup>

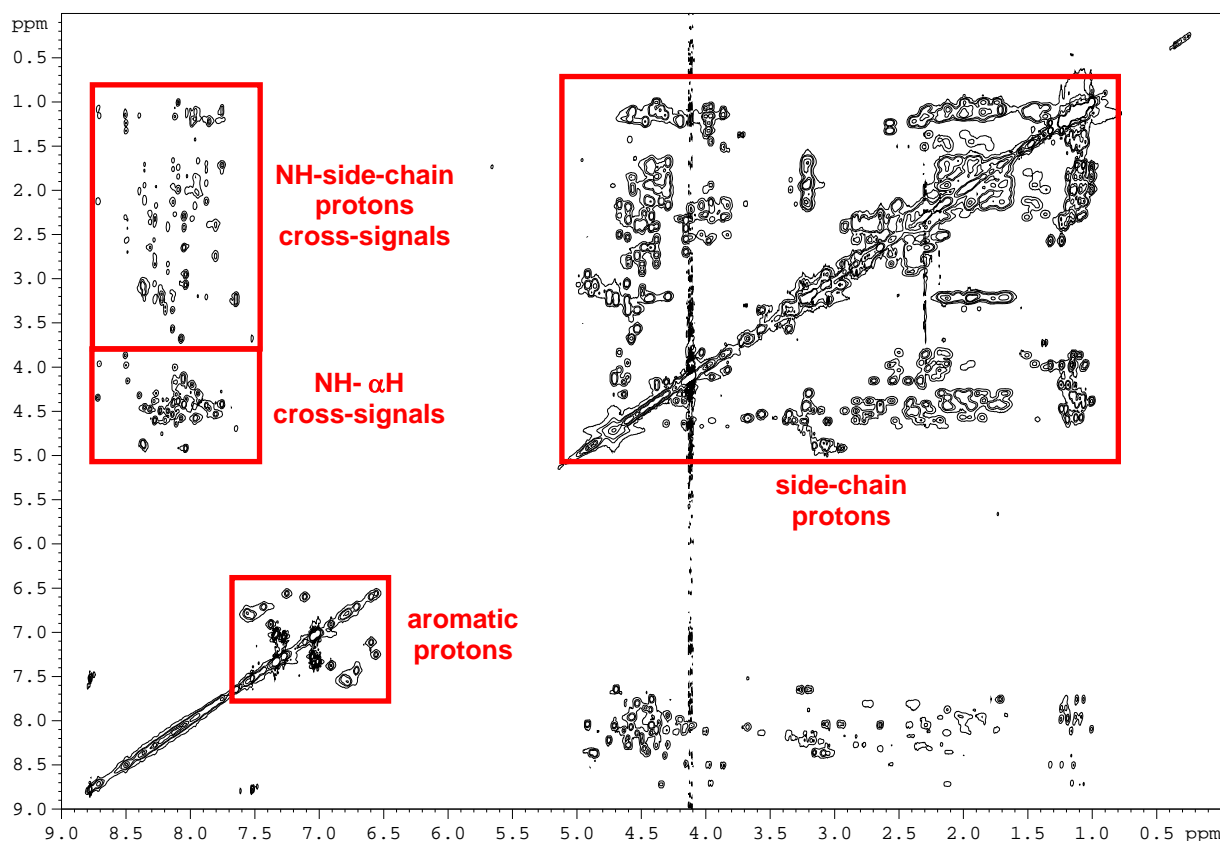
#### VIII.2.4.2.2 Sequential assignment

The sequential assignment of an unlabeled peptide or small protein requires the combined information of the following 2D-NMR experiments:

- i) the cross peaks observed in the COSY spectrum correspond to proton-proton correlations due to scalar (through-bond) couplings, though the respective protons may only be separated by a maximum of three covalent bonds within the amino acid;
- ii) the TOCSY spectrum contains all correlations between protons of one spin system (Figure VIII.2). As there is no scalar coupling across the amide bond, every amino acid represents a single spin system;
- iii) the NOESY spectrum provides cross peaks from protons which are spatially close. As the TOCSY spectrum is limited to the observation of the single peptide units, the sequential connectivities have to be elucidated from the NOESY experiment.

The sequential assignment usually starts by regarding the backbone NH- $\alpha$ H region of the COSY spectrum, which is considered a fingerprint of the corresponding protein. In this region every amino acid gives rise to a single cross signal. Exceptions are glycine, which shows two signals, and proline, which does not show any signal due to the absence of the NH proton. Counting the available COSY signals in this region rapidly assesses the quality of the NMR experiment in terms of spectral resolution and appearance. The next step is then the identification of the different spin systems in the TOCSY spectrum, which are lined up

vertically and correspond to the respective amino acids. The number of peaks per spin system, as well as their distribution is characteristic for every amino acid and can be used for the assignment. Finally, the necessary information about the sequential connectivities of amino acids is delivered by the NOESY experiment: a NOE signal from the NH of one amino acid to the  $\alpha$ H of the following one indicates the connection of the two amino acids.



**Figure VIII.2.** Exemplary TOCSY NMR spectrum of a peptide. Important regions are marked in red.

#### VIII.2.4.2.3 Chemical shift index (CSI)

Once the sequential assignment of the protein has been obtained, the chemical shift index (CSI) method offers a rapid evaluation of the secondary structure present in the sample. The CSI approach was introduced 1992 by Wishart et al. and is based on the dependence of the chemical shifts of the  $\alpha$ -protons on their surrounding environment, i.e. the secondary structure of the peptide or protein.<sup>20</sup> The CSI value for every amino acid is obtained by subtracting reference values (obtained from a randomly structured protein) from the experimental  $\alpha$ -proton chemical shifts. Four or more negative differences in a row indicate the presence of an  $\alpha$ -helix, while a grouping of three or more positive differences are

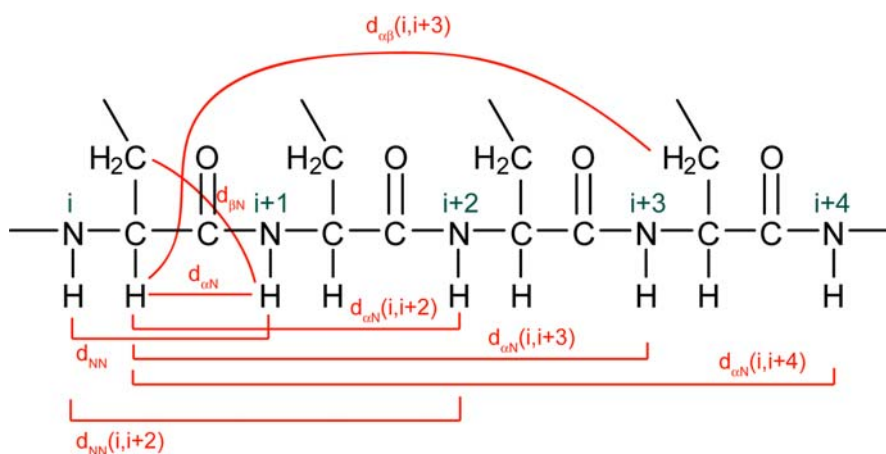
commonly regarded as evidence for a  $\beta$ -strand. The remaining residues are designated to be randomly ordered.

A comparison of the secondary structure content extrapolated by the CSI pattern with the one obtained from crystal structures of proteins showed that the results are strongly dependent upon the used reference values and only helices are estimated properly.<sup>21</sup> Nevertheless, the data set by Schwarzingner et al.<sup>22</sup> was found to be a good estimator of both helical and  $\beta$ -sheet structure content.

#### VIII.2.4.2.4 NOE signals

The 2D-NOESY (nuclear Overhauser effect spectroscopy) spectrum shows peaks stemming from nuclei that are close within the molecule, but not necessarily directly connected to each other. Therefore, the NOESY spectrum exhibits most or all peaks observed in the COSY spectrum and additionally shows cross signals from protons which are not neighboring in the sequence but close to each other in space ( $< 5 \text{ \AA}$ ). The obtained strength of the NOE cross signals then directly relates to the distance  $r$  of the participating nuclei and is inversely proportional to  $r^6$ . As the repetitive character of secondary structures in peptides and proteins gives rise to defined distances between the involved atoms, every structure element possesses a distinctive NOE “signature” and the NOESY spectrum can be used to identify the presence of these elements in the sample (Figure VIII.3 and Table VIII.2).

**Figure VIII.3.** Observable NOE signals in polypeptides.



**Table VIII.2.** Characteristic NOE patterns in polypeptides.<sup>23</sup> “–” – no cross signals, “+” – weak to medium strength cross signal, “++” – strong cross signals observable.

	$\beta$ -sheet	$\alpha$ -helix	$3_{10}$ -helix
$d_{\alpha N}(i, i+4)$	–	+	–
$d_{\alpha\beta}(i, i+3)$	–	++	+
$d_{\alpha N}(i, i+3)$	–	++	++
$d_{NN}(i, i+2)$	–	+	+
$d_{\alpha N}(i, i+2)$	–	–	+
$d_{NN}$	+	++	++
$d_{\alpha N}$	++	+	+

In practice, the NOE signals not corresponding to sequential amino acids (i.e. the ones not used for the sequential assignment) are assigned and classified according to their strength in weak, medium and strong signals. By calibrating the NOE signals with signals of known distance (e.g. geminal, diastereotopic protons or aromatic protons positioned ortho to each other), useful constraints for structural calculations can be obtained.

#### VIII.2.4.2.5 Experimental procedure

The peptide reproducing the HLH motif of Id2 was dissolved in a mixture of H<sub>2</sub>O/TFE-d<sub>2</sub> in a ratio of 40:60 (v/v) at a concentration of 0.5 mM. The NMR experiments were then recorded on a Bruker Avance 600 (600.13 MHz). A temperature scan from 300 K to 318 K in 6 K steps revealed the temperature of 318 K to be optimal for the experiments, as the least overlaps and sharper signals were observed in the NH region at this temperature. Consequently, all 2D measurements were conducted at this temperature. In the DQF-COSY, CLEAN-TOCSY and NOESY experiments the water resonance was suppressed by using the presaturation method. The mixing time accounted to 80 ms for the TOCSY and 200 ms for the NOESY experiment. The obtained data were processed using the TOPSPIN software purchased from Bruker (version 1.2). After the sequential assignment, the CSI pattern was calculated using the procedure developed by Wishart et al.<sup>20</sup>

## VIII.2.5 Fluorescence spectroscopy

### VIII.2.5.1 Experimental procedure

The fluorescence spectra were measured on a Cary Eclipse Spectrophotometer from Varian using a 1 cm path length quartz cuvette containing 1 mL of the receptor solution at a concentration of 9.5  $\mu\text{M}$  in 50 mM HEPES buffer (pH 7.5). Small aliquots (5  $\mu\text{L}$  up to 50  $\mu\text{L}$ ) of the peptide stock solution (64-142  $\mu\text{M}$  in HEPES buffer) were added manually to the receptor solution. The excitation wavelength was 305 nm. The emission spectra were recorded over the range 310-500 nm using a bandwidth of 5 nm at a temperature of 25 °C. Fitting of the titration data was done by using the Hill equation<sup>24</sup>:

$$\Delta F = \frac{\Delta F_{\max} \cdot c^h}{K_{0.5}^h + c^h} \quad \text{Eq VIII.2}$$

where  $h$  is the Hill coefficient,  $K_{0.5}$  the peptide concentration at which half of the receptor is bound,  $\Delta F_{\max}$  the fluorescence increment observed at saturation of the receptor and  $\Delta F$  the observed fluorescence increment at the peptide concentration  $c$ . For the Job's plot analysis<sup>25</sup>,<sup>26</sup> two solutions with equal concentrations of the receptor and the peptide ( $c$  in the range from 10  $\mu\text{M}$  to 41  $\mu\text{M}$ ) were prepared and mixed to produce eleven different solutions corresponding to a receptor molar ratio increasing from 0 to 1.

## VIII.3 References

1. Merrifield, R. B., The automatic synthesis of proteins. *Sci Am* **1968**, 218, 56-62.
2. Max Bergmann, L. Z., Über ein allgemeines Verfahren der Peptid-Synthese. *Berichte der deutschen chemischen Gesellschaft (A and B Series)* **1932**, 65, 1192-1201.
3. Zhou, Q.; Wang, S.; Anderson, D. J., Identification of a novel family of oligodendrocyte lineage-specific basic helix-loop-helix transcription factors. *Neuron* **2000**, 25, 331-43.
4. McKay, F. C.; Albertson, N. F., New Amine-masking Groups for Peptide Synthesis. *J Am Chem Soc* **1957**, 79, 4686-4690.
5. Carpino, L. A., The 9-fluorenylmethyloxycarbonyl family of base-sensitive amino-protecting groups. *Acc Chem Res* **1987**, 20, 401-407.

6. Carpino, L. A.; Han, G. Y., 9-Fluorenylmethoxycarbonyl function, a new base-sensitive amino-protecting group. *J Am Chem Soc* **1970**, 92, 5748-5749.
7. Novabiochem Catalog 2006/2007. pp 3.16 - 3.20.
8. Woody, R. W.; Koslowski, A., Recent developments in the electronic spectroscopy of amides and alpha-helical polypeptides. *Biophys Chem* **2002**, 101-102, 535-51.
9. Chakrabartty, A.; Kortemme, T.; Padmanabhan, S.; Baldwin, R. L., Aromatic side-chain contribution to far-ultraviolet circular dichroism of helical peptides and its effect on measurement of helix propensities. *Biochemistry* **1993**, 32, 5560-5.
10. Lisowski, M.; Olczak, J.; Zabrocki, J., Circular dichroic properties of the tyrosine residues in tetrazole analogues of opioid peptides. *J Pept Sci* **2006**, 12, 297-302.
11. Greenfield, N. J., Methods to estimate the conformation of proteins and polypeptides from circular dichroism data. *Anal Biochem* **1996**, 235, 1-10.
12. Sewald, N. J., H.-D., Peptides: Chemistry and Biology. Wiley, New York **2002**.
13. Lau, S. Y.; Taneja, A. K.; Hodges, R. S., Synthesis of a model protein of defined secondary and quaternary structure. Effect of chain length on the stabilization and formation of two-stranded alpha-helical coiled-coils. *J Biol Chem* **1984**, 259, 13253-61.
14. Lobley, A.; Whitmore, L.; Wallace, B. A., DICHROWEB: an interactive website for the analysis of protein secondary structure from circular dichroism spectra. *Bioinformatics* **2002**, 18, 211-2.
15. Whitmore, L.; Wallace, B. A., DICHROWEB, an online server for protein secondary structure analyses from circular dichroism spectroscopic data. *Nucleic Acids Res* **2004**, 32, (Web Server issue), W668-73.
16. Provencher, S. W.; Glockner, J., Estimation of globular protein secondary structure from circular dichroism. *Biochemistry* **1981**, 20, 33-7.
17. van Stokkum, I. H.; Spoelder, H. J.; Bloemendal, M.; van Grondelle, R.; Groen, F. C., Estimation of protein secondary structure and error analysis from circular dichroism spectra. *Anal Biochem* **1990**, 191, 110-8.
18. Sreerama, N.; Venyaminov, S. Y.; Woody, R. W., Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Anal Biochem* **2000**, 287, 243-51.
19. Montelione, G. T.; Zheng, D.; Huang, Y. J.; Gunsalus, K. C.; Szyperski, T., Protein NMR spectroscopy in structural genomics. *Nat Struct Biol* **2000**, 7 Suppl, 982-5.



20. Wishart, D. S.; Sykes, B. D.; Richards, F. M., The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy. *Biochemistry* **1992**, 31, 1647-51.
21. Mielke, S. P.; Krishnan, V. V., An evaluation of chemical shift index-based secondary structure determination in proteins: influence of random coil chemical shifts. *J Biomol NMR* **2004**, 30, 143-53.
22. Schwarzingher, S.; Kroon, G. J.; Foss, T. R.; Wright, P. E.; Dyson, H. J., Random coil chemical shifts in acidic 8 M urea: implementation of random coil shift data in NMRView. *J Biomol NMR* **2000**, 18, 43-8.
23. Wüthrich, K., NMR of Proteins and Nucleic Acids. *Wiley, New York* **1986**.
24. Weiss, J. N., The Hill equation revisited: uses and misuses. *Faseb J* **1997**, 11, 835-41.
25. Vosburgh, W. C.; Cooper, G. R., Complex Ions. I. The Identification of Complex Ions in Solution by Spectrophotometric Measurements. *J Am Chem Soc* **1941**, 63, 437-442.
26. Connors, K. A., Binding Constants: The Measurement of Molecular Complex Stability. *Wiley, New York* **2004**.



## IX Appendix

### IX.1 Synthesized peptides – analytical data

**Table IX.1.** Analytical data of the synthesized peptides. All peptides were N-terminally acetylated and C-terminally amidated. All retention times were measured using the gradient 10% ACN for 3 min, 10-70% ACN over 40 min, except the ones marked with an asterisk which were measured using the gradient 5% ACN for 3 min, 5-40% ACN over 40 min.

No.	Peptide	MW <sub>calc.</sub> (Da)	MW <sub>found</sub> (Da)	<i>t<sub>r</sub></i> (min)
III.1	Ac-(66-106)-Id1	4930.61	4932.7	18
III.2	Ac-[retro66-80, des81]-(66-106)-Id1	4833.00	4829.6	18
III.3	Ac-[Ahx82-90]-(66-106)-Id1	4037.85	4036.8	19
III.4	Ac-[retro66-80, des81, Ahx82-90]-(66-106)-Id1	3922.66	3922.1	18
III.5	Ac-[Orn82-90]-(66-106)-Id1	4062.80	4066.2	18
III.6	Ac-[retro66-80, des81, Orn82-90]-(66-106)-Id1	3964.65	3962.4	19.5
III.7	Ac-(89-106)-βAlaLys(retro66-80)-Id1	4237.02	4239.7	20
III.8	Ac-(89-106)-βAlaLys(66-83)-Id1	4549.40	4547.3	20
III.9	Ac-(89-106)-βAlaLys-Id1	2408.85	2410	17
III.10	Ac-[retro69-80, ProGly81-90]-(69-103)-Id1	3212.70	3215.2	18
III.11	Ac-[F73, 101]-(66-106)-Id1	4898.81	4905	19.5
III.12	Ac-[F101]-(66-106)-Id1	4914.81	4917.6	18
III.13	Ac-[F73]-(66-106)-Id1	4914.81	4910.1	17.5
IV.1	Ac-CGG-(66-106)-S72-GGG-Id1	5303.15	5299.5	16
IV.2	Ac-(66-106)-S72-GGC-Id1	5132.00	5132.7	15.7
VI.1	Ac-(36-76)-Id2	4906.85	4907.7	20.5
VI.2	Ac-(41-81)-Id3	4838.60	4839.6	22.1
VI.3	Ac-(65-105)-Id4	4896.82	4893.3	18
VI.4	Ac-(66-82)-Id1	2042.40	2042.1	16.4
VI.5	Ac-(41-56)-Id3	2058.36	2058.4	15.2
VI.6	Ac-(65-80)-Id4	2109.45	2110.5	14.2
VI.7	Ac-(89-105)-Id4	2067.43	2066.8	22.4
VI.8	Ac-(72-75)-Id1	568.66	569.3	5 (*)
VI.9	Ac-[A73]-(72-75)-Id1	476.56	477.0	3.5 (*)
VI.10	Ac-[A72]-(72-75)-Id1	536.59	537.3	3.5 (*)
VI.11	Ac-(42-45)-Id2	540.64	541.1	4 (*)
VI.12	Ac-(67-72)-Id3	838.01	838.3	10.5 (*)
VI.13	Ac-(122-160)-MyoD	4545.30	4531.4	19
VI.14	Ac-(36-74)-Max	4576.17	4575.8	13.4

## IX.2 CONTIN analysis of the CD data for the Id2 HLH motif

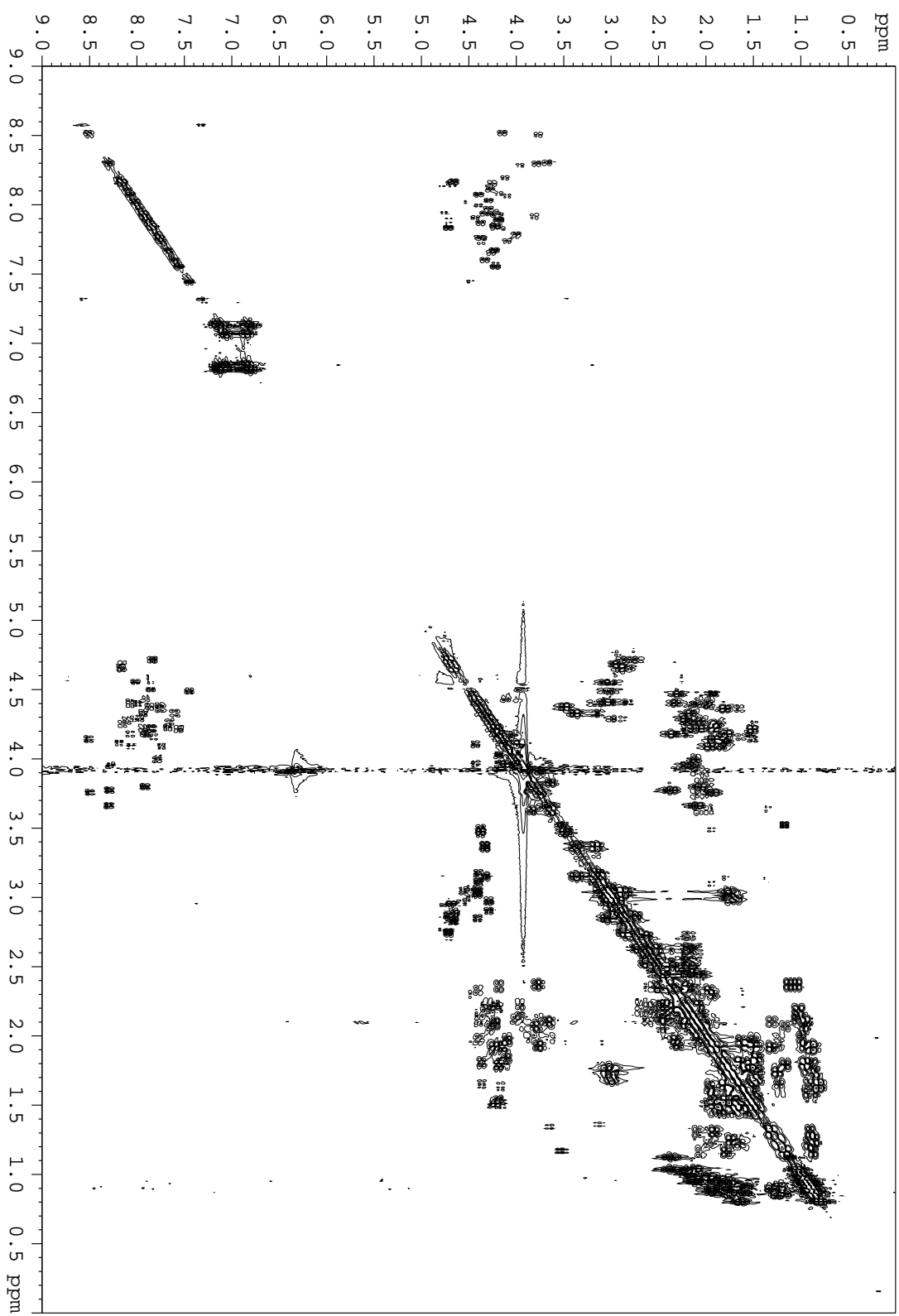
**Table IX.2.** Results of the CONTIN analysis for the structural composition of the Id2 HLH motif in different solvent mixtures ( $c = 30 \mu\text{M}$ ).  $n_{\text{helical}}$  – calculated number of helical amino acids,  $R$  – ratio between the ellipticity minima at 222 nm and 208 nm.

Solvent	helix	$\beta$ -strands	turns	unordered	$n_{\text{helical}}$	$R$
water	49%	3%	13%	35%	20	0.88
phosphate buffer 0.1M pH 7.25	60%	8%	13%	19%	24	1.04
water/TFE 40:60 (v/v)	72%	1%	12%	15%	30	0.74
100 % TFE	91%	0%	9 %	0%	37	0.81

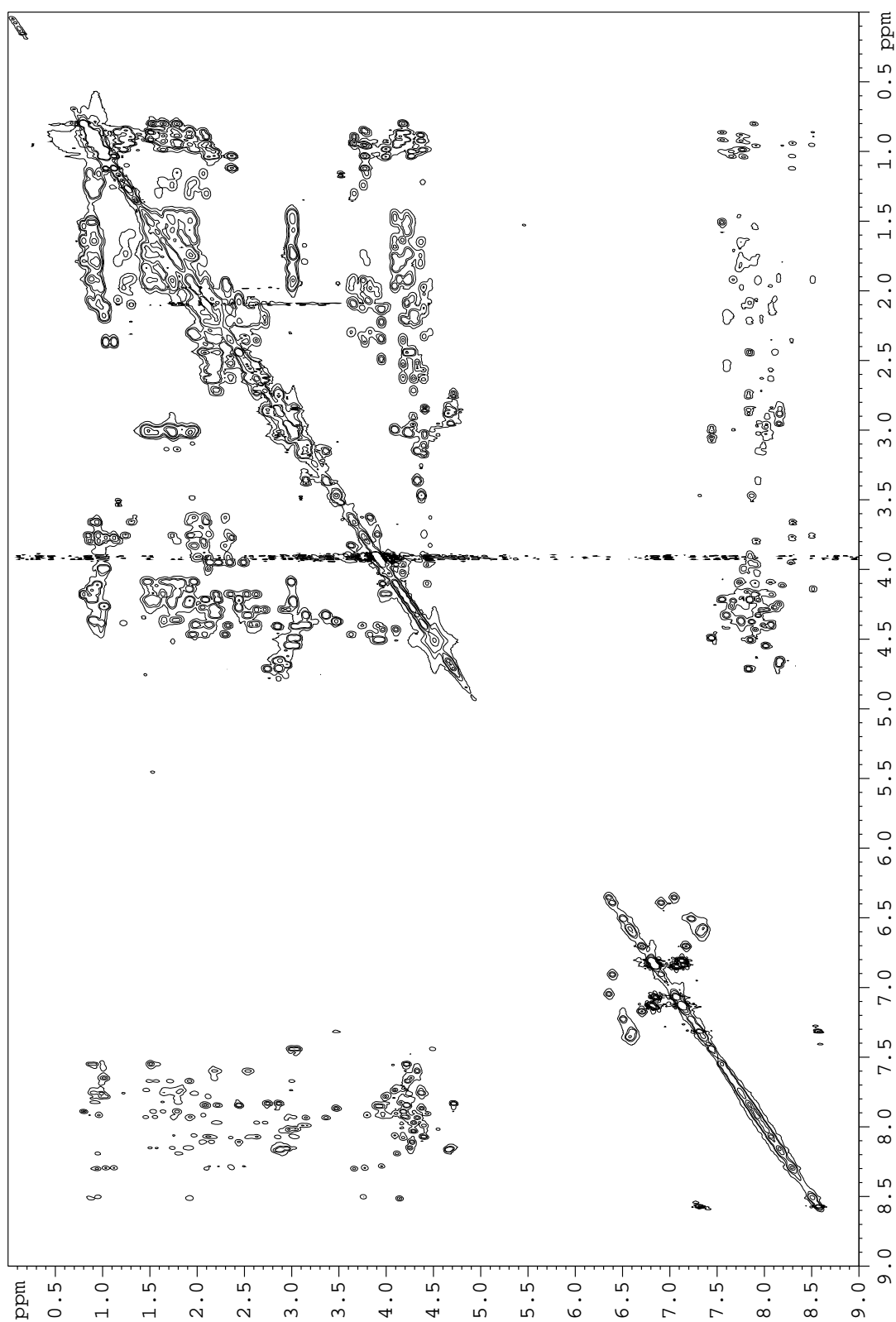
## IX.3 NMR data of the Id2 HLH motif

**Table IX.3.** The  $^1\text{H}$  chemical shifts (ppm) for the Id2 HLH motif (0.5 mM in water/TFE- $\text{d}_2$  40:60 at 318 K).

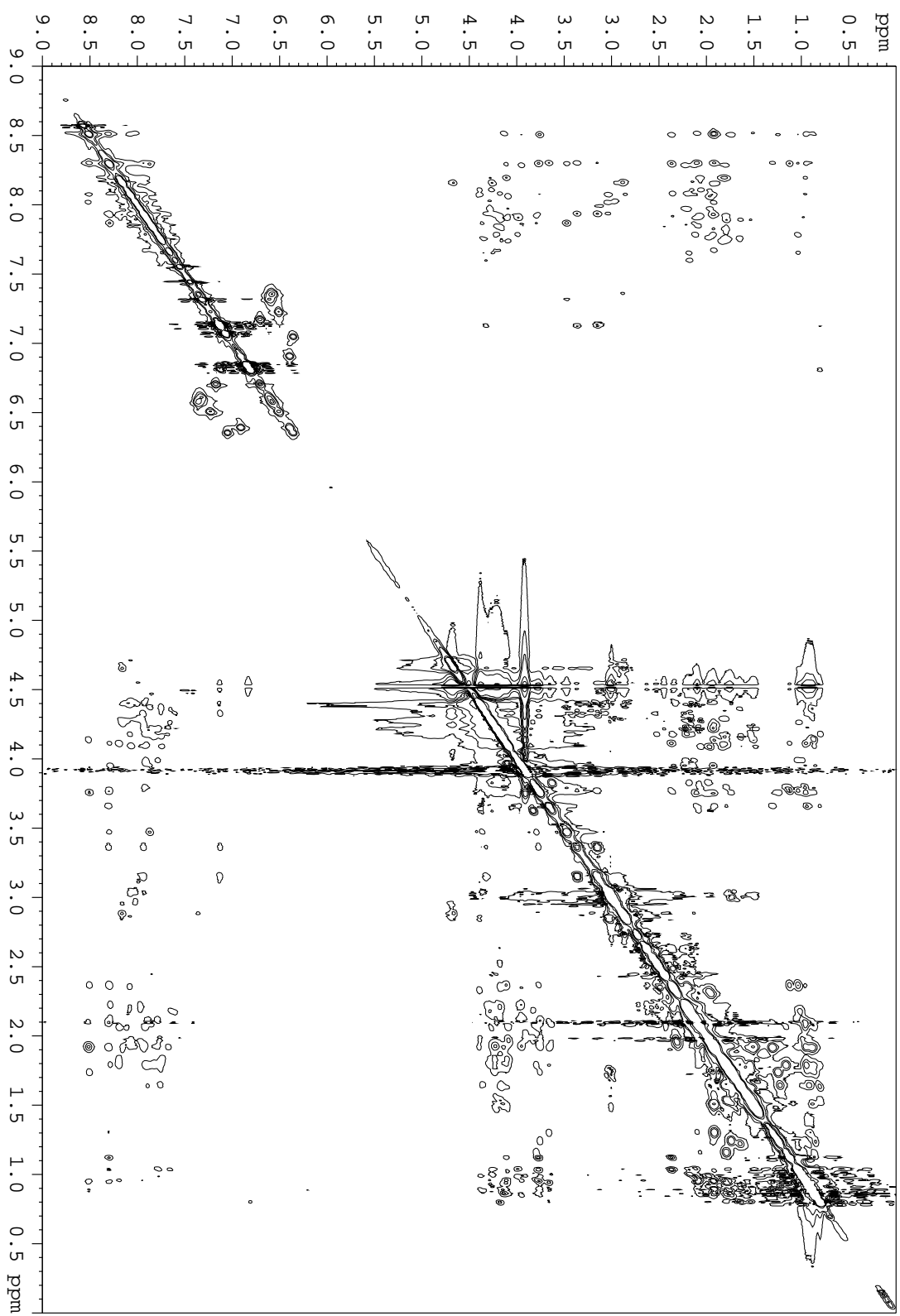
	NH	$\alpha\text{H}$	$\beta(\beta')$	$\gamma(\gamma')$	$\delta(\delta')$	other
Leu <sup>1</sup>	7.55	4.21	1.57	0.92	0.86	
Tyr <sup>2</sup>	7.44	4.49	3.06 (2.89)			Ring protons – 7.07, 6.85
Asn <sup>3</sup>	8.12	4.77	2.88 (2.77)			
Met <sup>4</sup>	8.07	4.39	2.13	2.64 (2.57)		
Asn <sup>5</sup>	7.83	4.70	2.86 (2.73)			
Asp <sup>6</sup>	8.03	4.29	2.90 (2.97)			
Cys <sup>7</sup>	7.99	4.38	3.12 (3.16)			
Tyr <sup>8</sup>	8.16	4.64	2.88			Ring protons – 7.36, 6.59
Ser <sup>9</sup>	7.92	4.63	4.04 (4.02)			
Lys <sup>10</sup>	7.67	4.23	1.92	1.48 (1.57)	1.73	$\epsilon$ – 3.01
Leu <sup>11</sup>	7.83	4.16	1.80	1.64	0.91	
Lys <sup>12</sup>	7.74	4.10	1.83	1.46	1.68 (1.75)	$\epsilon$ – 3.00
Glu <sup>13</sup>	7.97	4.29	2.22	2.72 (2.63)		
Leu <sup>14</sup>	7.76	4.35	1.98	1.80 (1.65)	0.91	
Val <sup>15</sup>	7.65	4.26	2.18	1.01		
Pro <sup>16</sup>		4.47	2.30 (1.96)	2.08	3.62 (3.81)	
Ser <sup>17</sup>	7.86	4.50	3.94	2.01 (2.33)	2.12	
Ile <sup>18</sup>	7.76	4.24	1.80	1.65	0.99	$\gamma_2$ – 1.22
Pro <sup>19</sup>		4.41	2.33 (2.01)	2.12	3.75 (3.91)	
Gln <sup>20</sup>	8.11	4.26	2.06 (2.16)	2.44		
Asn <sup>21</sup>	8.17	4.65	2.81 (2.90)			
Lys <sup>22</sup>	8.16	4.25	1.92	1.48 (1.57)	1.73	$\epsilon$ – 3.01
Lys <sup>23</sup>	7.93	4.21	1.93	1.46 (1.56)	1.76	$\epsilon$ – 3.03
Val <sup>24</sup>	7.78	3.99	2.11	1.00 (1.03)		
Ser <sup>25</sup>	7.91	4.40	3.97 (4.10)			
Lys <sup>26</sup>	8.06	4.09	1.97	1.50 (1.63)	1.75	$\epsilon$ – 2.98
Met <sup>27</sup>	7.60	4.33	2.20	2.54		
Glu <sup>28</sup>	7.07	4.38	2.20 (2.35)	2.52 (2.63)		
Ile <sup>29</sup>	7.91	3.80	2.07	1.16	0.86	$\gamma_2$ – 0.96
Leu <sup>30</sup>	8.19	4.11	1.84	1.80	0.96 (0.94)	
Gln <sup>31</sup>	8.28	3.95	2.12 (2.23)	2.35 (2.49)		
His <sup>32</sup>	7.87	4.38	3.45			NH(Imidazole) – 9.65 $\epsilon$ – 8.57 $\delta$ – 7.31
Val <sup>33</sup>	8.29	3.77	2.37	1.03 (1.12)		
Ile <sup>34</sup>	8.50	3.76	1.93	1.25	0.85	$\gamma_2$ – 0.95
Asp <sup>35</sup>	8.08	4.41	3.03 (2.85)			
Tyr <sup>36</sup>	7.94	4.33	3.37 (3.20)			Ring protons – 7.13, 6.81
Ile <sup>37</sup>	8.30	3.66	2.10	1.30	0.89	$\gamma_2$ – 0.94
Leu <sup>38</sup>	8.51	4.13	1.92	1.51	0.88	
Asp <sup>39</sup>	8.01	4.55	2.97 (3.05)			
Leu <sup>40</sup>	7.89	4.18	1.77	1.54 (1.65)	0.80	
Gln <sup>41</sup>	7.84	4.22	2.09 (2.21)	2.45		



**Figure IX.1.** DQF-COSY spectrum of the Id2 HLH motif in H<sub>2</sub>O/TFE-d<sub>2</sub> (40:60, v/v) at 318 K.

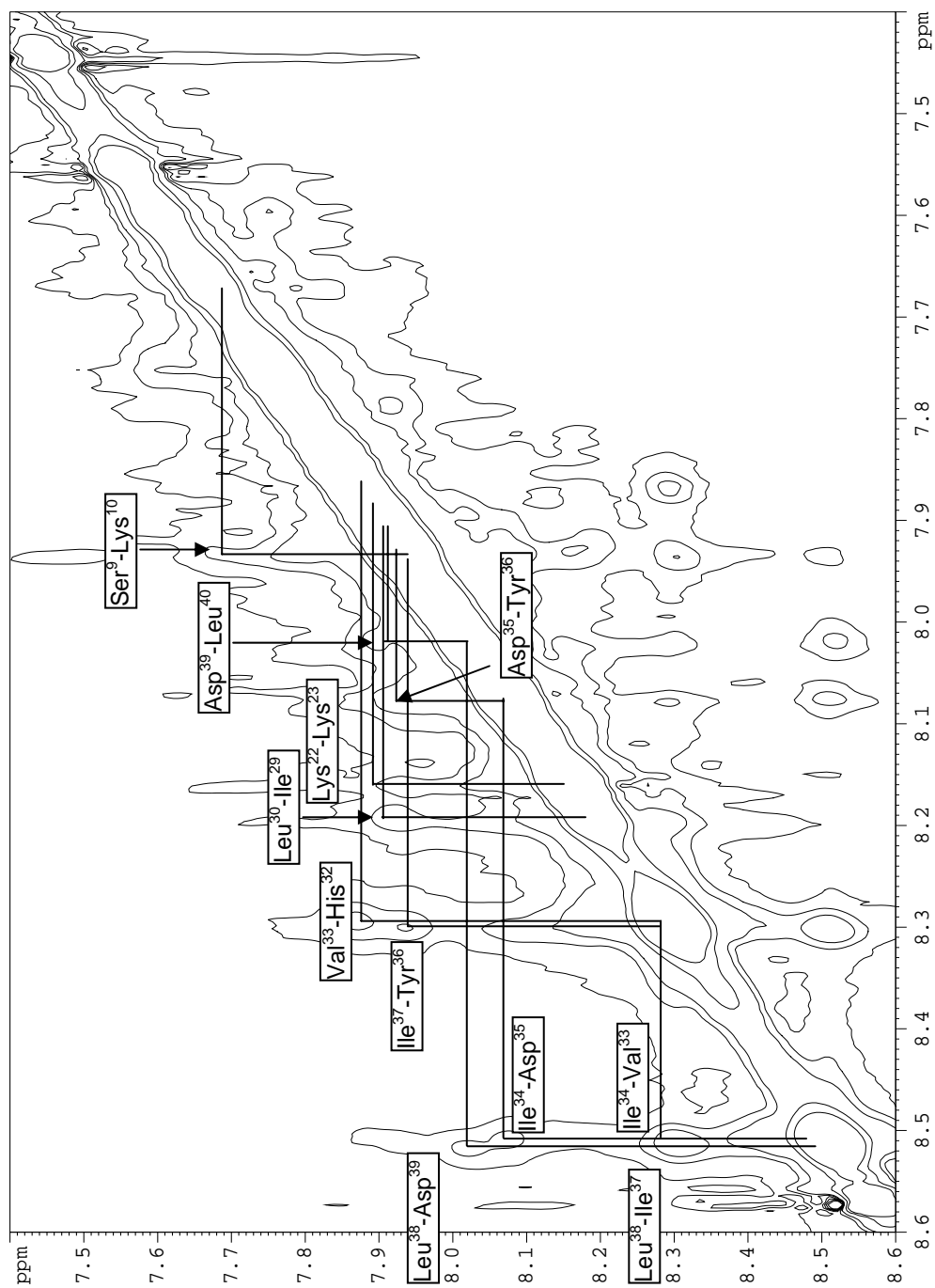


**Figure IX.2.** Clean-TOCSY spectrum of the Id2 HLH motif in H<sub>2</sub>O/TFE-d<sub>2</sub> (40:60, v/v) at 318 K.



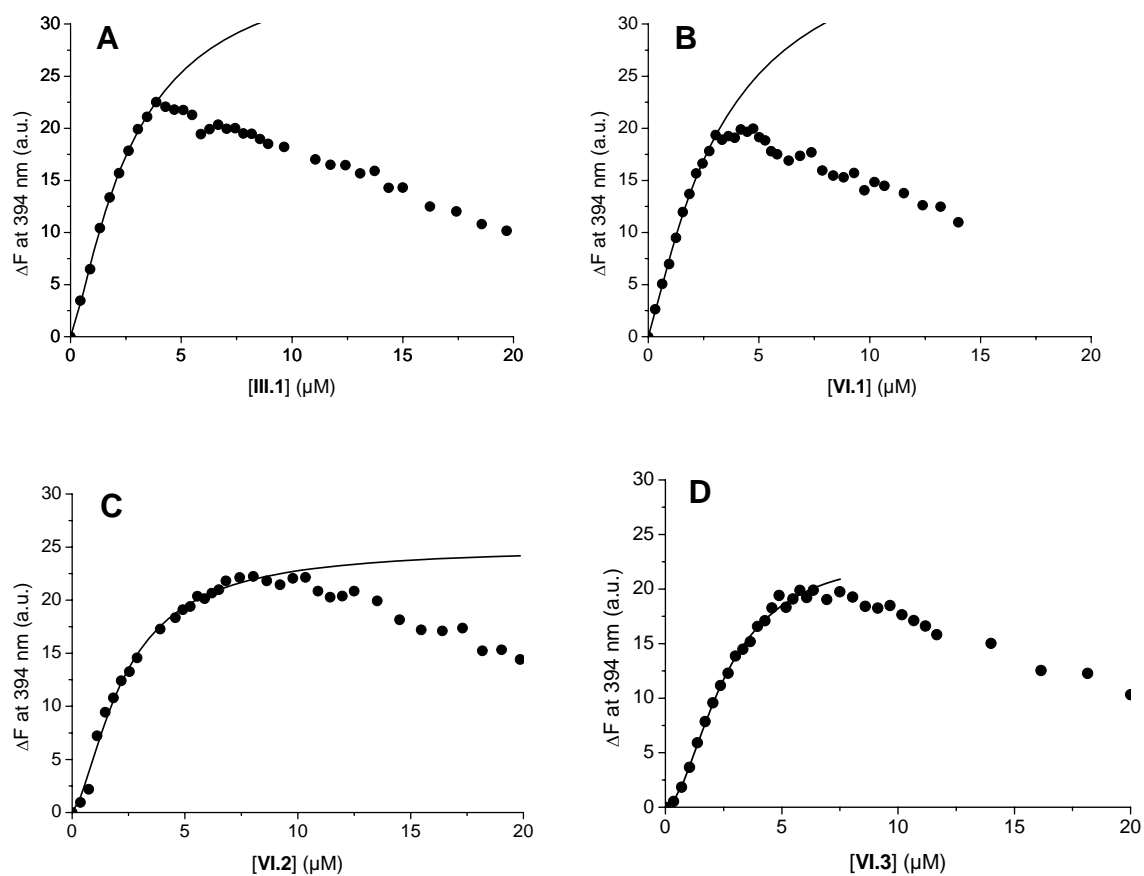
**Figure IX.3.** NOESY spectrum of the Id2 HLH motif in H<sub>2</sub>O/TFE-d<sub>2</sub> (40:60, v/v) at 318 K.





**Figure IX.4.** NOE connectivities observed for the NH-NH backbone of the Id2 HLH motif.

## IX.2 Fluorescence titrations



**Figure IX.5.** Titration curves of **R1** (9.5  $\mu M$  in 50 mM HEPES buffer, pH 7.5) with peptides **III.1** (A), **VI.1** (B), **VI.2** (C) and **VI.3** (D). Solid lines represent the fit of the experimental points preceding the loss of emission intensity.





## **Publications**

**Kiewitz S.**, 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.

Cabrele C., **Kiewitz S.**, Havrdová J., Colombo N. (2005) Studies on the chemical synthesis of the Id proteins and mutated analogs. *230<sup>th</sup> ACS National Meeting, Abstracts of Papers*.

## **Poster presentations**

Kiewitz S., Cabrele C. (2003) Id proteins. *Tag der Naturwissenschaften 2003, Regensburg (Germany)*.

Kiewitz S., Cabrele C. (2005) Synthesis and Characterization of Peptides Based on the HLH Domain of the Id1 Protein. *Summer School of Medicinal Chemistry, Shanghai (China)*.

## Curriculum Vitae

**Place of birth** Passau (Germany), 16<sup>th</sup> March 1978

### Education

05/2004 – 05/2007 PhD thesis at the faculty of Chemistry and Pharmacy, University of Regensburg, in the research group of Prof. A. Buschauer, under the supervision of Dr. Chiara Cabrele.

04/2003 – 04/2004 Diploma thesis at the faculty of Chemistry and Pharmacy, University of Regensburg, in the research group of Prof. A. Buschauer, under the supervision of Dr. Chiara Cabrele.

11/1998 – 03/2004 Chemistry studies at the University of Regensburg.

### Research Experience

09/2004 – 11/2004 Research stay at the University of Milan, Italy, under the supervision of Prof. Gelmi. Topic: “Synthesis of 2,3-Diamino-norbornane-2-carboxylic acids”.

01/2002 – 04/2002 Research stay at the Kansas University in Lawrence, USA, in the working group of Prof. Georg. Topic: “Synthesis of a small library of indol derivatives by combinatorial chemistry”.

### Fellowship

06/2007 Fellowship for the FY2007 JSPS Summer Program of the Japan Society for the Promotion of Science (JSPS).

## Acknowledgements

I would like to thank the following people:

*Prof. Dr. Buschauer*, for giving me the opportunity to carry out my PhD work at his institute.

*Dr. Chiara Cabrele*, under whose supervision this work was done, for taking me on a four-year journey into the world of peptide chemistry and her continued guidance during that time.

*Dr. Burgemeister and his team* for recording the NMR spectra, *Dr. Kiermaier and his team* for the assistance during MALDI-TOF measurements, *Dr. Vasold* and *Mr. Lautenschlager* for providing assistance during the preparative HPLC measurements.

The working groups of *Prof. Wolfbeis* and *Prof. König* for the possibility to use the CD spectrometer and the fluorimeter, respectively.

*Prof. Gelmi* for the interesting opportunity to conduct a research stay at the University of Milan, and *Dr. Sara Pellegrino* for the assistance during that time.

*Dr. Daniela Fischer* for useful advice on the evaluation of NMR spectra.

*Dr. Michael Kruppa* and *Stefan Stadlbauer* for providing the fluorescent receptors.

My lab colleagues *Dr. Noemi Colombo*, *Dr. Jaroslava Svobodová*, *Annette Snyder* and *Dominik Grögel* for the nice time and the support during the diploma and PhD years.

The Italian guest scientists *Cristian “E stasera ?” Cattaneo*, *Roberto “Pröst” Fanelli* and *Andrea “it’s incredible” Caporale* for creating a Mediterranean lab atmosphere, which is not often encountered north of the Alps.

*Timo Huber*, *Dominik Kohr* and *Harald Seger* for the nice time spent together during our chemistry studies.

My fellow *SHADO* members, *Martin Mayer* and *Stephanie Peters*, as well as *Florian Lechner* for providing the sorely needed distraction from chemistry.

*Melanie Schwarz* for her never-ending support, understanding and patience.