Structural Investigations of the Tetramerization Domains of p53/p73-like Transcription Factors from the Tunicate Species *Ciona intestinalis*

Dissertation

zur Erlangung des Doktorgrades

der Naturwissenschaften

vorgelegt beim Fachbereich

Biochemie, Chemie und Pharmazie (FB 14)

der Johann Wolfgang Goethe-Universität

in Frankfurt am Main

von

Jan Peter Heering

aus Eckernförde

GOETHE UNIVERSITÄT FRANKFURT AM MAIN

Frankfurt am Main 2014

(D 30)

Vom Fachbereich Biochemie, Chemie und Pharmazie (FB 14) der
Johann Wolfgang Goethe-Universität als Dissertation angenommen.

Dekan:	Prof. Dr. Thomas Prisner
Gutachter:	Prof. Dr. Volker Dötsch
	Dr. Martin Vabulas

Datum der Disputation: 28.Mai.2015

“I have a theory that the truth is never told during the nine-to-five hours.”

Hunter S. Thompson

Eidesstattliche Versicherung

Ich erkläre hiermit an Eides Statt, dass ich die vorgelegte Dissertation über

„Structural investigations of the tetramerization domains of p53/p73-like proteins from the tunicate species *Ciona intestinalis*“

selbstständig angefertigt und mich anderer Hilfsmittel als der in ihr angegebenen nicht bedient habe, insbesondere, dass alle Entlehnungen aus anderen Schriften mit Angabe der betreffenden Schrift gekennzeichnet sind.
Ich versichere, die Grundsätze der guten wissenschaftlichen Praxis beachtet, und nicht die Hilfe einer kommerziellen Promotionsvermittlung in Anspruch genommen zu haben.

Frankfurt am Main, den 20.Nov.2014

Jan Peter Heering

Table of Contents

Table of Contents I
Abbreviations VI
Glossary on homology IX
Summary X
Zusammenfassung XV
1. Introduction 1
1.1. The vertebrate p53 family of transcription factors 1
1.1.1. The transcription factor p53 – the guardian of the genome 1
1.1.2. Regulation of p53 3
1.1.3. Function of the p53 TD and regulation of the oligomeric state 4
1.1.4. Structure of the tetramerization domain of *homo sapiens* p53 6
1.1.5. The transcription factors p63 and p73 are closely related to p53 7
1.1.6. Structures of the TDs of p63 and p73 in the tetrameric state 10
1.1.7. p63 ensures the genomic integrity of the female germ line 11
1.1.8. Guardian of the genome – the ancestral function of the p53 family 12
1.2. Aim of study 13
1.2.1. Superordinate research objective 13
1.2.2. Previous investigations 13
1.2.3. Objective of this thesis 15
1.3. Tunicates are the closest living relatives of vertebrates 16
1.3.1. The ascidians *Ciona intestinalis* and *Ciona savignyi* 17
1.3.2. Proteins from the p53 family of transcription factors in *Ciona intestinalis* 20
2. Materials 23
2.1. Laboratory Equipment 23
2.2. Chromatography 24
2.3. Membranes for ultrafiltration and dialysis 25
2.4. Kits 25
2.5. DNA Ladders 25
2.6. Protein standards 25
2.7. Enzymes 26
2.8. Plasmids 26

2.9. Bacterial Strains 27
2.10. Reagents 27
2.11. Common buffers, media and reagents 30
2.11.1. Antibiotics 30
2.11.2. Inducers (Induction of Expression) 30
2.11.3. Media for cultivation of bacteria 30
2.11.4. Supplement solutions for M9 medium 31
2.11.5. Nuclease stock solutions for lysis of bacterial pellets 31
2.11.6. General stocks for protein biochemistry 31
2.11.7. Buffers for DNA agarose gel electrophoresis 32
2.11.8. Buffers for Tris-Tricine-SDS PAGE 32
2.11.9. Buffers for cell lysis 33
2.11.10. Buffers for protein purification and analysis 33
2.11.11. Buffers used for condition screening of *C.int.* p53/p73-a 33
2.12. Software 34
2.13. Web server operated programs 34
3. Methods 36
3.1. Methods for microbiology and the manipulation of plasmid DNA 36
3.1.1. Polymerase Chain Reaction (PCR) 36
3.1.2. PCR Purification 36
3.1.3. Digestion of DNA with Restriction Endonucleases 36
3.1.4. Ligation of DNA fragments 37
3.1.5. PCR based vector preparation for background free cloning 37
3.1.6. Transformation of bacteria 39
3.1.7. Plasmid DNA preparation from bacterial cultures 39
3.1.8. Determination of DNA concentration 40
3.1.9. Site-directed mutagenesis 40
3.1.10. DNA sequencing 40
3.1.11. Agarose Gel Electrophoresis 40
3.2. Protein biochemical methods 41
3.2.1. Determination of protein concentration 41
3.2.2. Concentrating of protein samples 41
3.2.3. Heterologous expression in *E. coli* 41
3.2.4. Cell Lysis of *E. coli* 43

3.2.5. Protein purification by Immobilized Metal Ion Chromatography (IMAC) 43
3.2.6. Cleavage of fusion protein or purification tag with TEV protease 44
3.2.7. Reversed Immobilized Metal Ion Chromatography (Re-IMAC) 44
3.2.8. Size Exclusion Chromatography (SEC) 45
3.2.9. Dialysis 45
3.2.10. Tris-Tricine-SDS-PAGE (adapted from Schaegger) 45
3.2.10.1. Coomassie brilliant blue staining 46
3.3. Biophysical methods 47
3.3.1. Circular Dichroism spectroscopy 47
3.3.1.1. Curve fitting of thermal denaturation curves determined for homo-tetramers characterized by reversible coupled disassembly and unfolding 47
3.3.2. Nuclear Magnetic Resonance (NMR) in solution 53
3.3.2.1. Sample preparation for the C.int. p53/p73 project 53
3.3.3. Spectra analysis 53
3.3.4. Sequential assignments 53
3.3.5. Hetero NOE experiments 53
3.3.6. TALOS (torsion angle likeliness obtained from shift & sequence similarity) 54
3.3.7. Determination of secondary structure propensity based on chemical shifts 54
3.3.8. Calculation of chemical shift perturbation 56
3.3.9. Assignment of side chain resonances and pro-chiral methyl groups in *C.int.* p53/p73-b TD 56
3.3.10. Assignment of inter-monomer NOEs in the *C.int.* p53/p73-b TD 57
3.3.11. Calculation of the solution structure of the *C.int.* p53/p73-b TD 58
4. Results 60
4.1. Alignment of *Ciona intestinalis* and *Ciona savignyi* p53/p73 proteins towards sequences from other species 60
4.2. Structural analysis of the TD and C-terminal sections of *Ciona intestinalis* p53/p73-a 65
4.2.1. Detailed sequence comparisons of the TD and C-terminal sections of the *Ciona intestinalis* and *Ciona savignyi* p53/p73-a isoforms 65
4.2.2. C-terminus specific to *Ciona intestinalis* p53/p73-a isoform 2 is unstructured 67
4.2.3. Determination of section with secondary structure in C-termini of *Ciona int.* p53/p73-a isoform 1, 3 and 4 73
4.2.4. The truncated construct corresponding to aa 415-450 of *Ciona intestinalis* p53/p73-a forms tetramers 79
4.2.5. *Ciona int.* p53/p73-a TD of isoform 1 and 2 does not form a second helix 80

4.2.6. The core tetramerization domain of *C.int.* p53/p73-a isoform 1 and 2 is capable to support a second helix 82

4.2.7. Determination of minimal number of mutations causing a second helix 88

4.2.8. Formation of second helix upon mutations in hinge region 95

4.2.9. The second helix resulting from mutations in the hinge region adopts a position relative to the core TD like in human p73 98

4.2.10. Change of secondary structure propensity upon mutation of hinge region 99

4.2.11. Determination of stabilizing effect of hinge region mutations by CD 102

4.2.12. Determination of the gain in stability of the second helix by measurements of heteronuclear nuclear Overhauser effects 107

4.2.13. Lower temperatures result in stabilization of the second helix 109

4.3. Structural analysis of the TD of *Ciona intestinalis* p53/p73-b 112

4.3.1. Detailed sequence comparison of the TD sequences of *Ciona intestinalis* and *Ciona savignyi* p53/p73-b 112

4.3.2. Determination of section spanned by secondary structure elements within the C-terminus of *Ciona intestinalis* p53/p73-b 112

4.3.3. *Ciona intestinalis* p53/p73-b TD possesses a stably folded second helix 116

4.3.4. The second helix in the tetramerization domain of *Ciona intestinalis* p53/p73-b is not essential for the oligomeric state 117

4.3.5. Determination of the NMR solution structure of *C.int.* p53/p73-b TD 118

4.4. Structural comparison of the TD of *C.int.* p53/p73-b and TDs of p53 and p73-like proteins from vertebrate species 127

4.5. The TDs of *Ciona intestinalis* p53/p73-a and p53/p73-b do not form hetero oligomers 131

4.6. Prediction of a single genomic copy of a p53/p73 like protein in the genome of *Botryllus schlosserie* 133

5. Discussion 136

5.1. Phylogenetic relationship between p53-like transcription factors from vertebrate and tunicate species 136

5.2. Tunicate p53/p73 proteins do not possess a SAM domain 138

5.3. TD of *C.int.* p53/p73-a does not possess a stably folded second helix 140

5.4. The only partially folded second helix in *C.int.* p53/p73-a could function in regulating interactions or in fine tuning the K_d of tetramerization 142

5.5. The TD of *C.int.* p53/p73-b possesses a rigid second helix with an alternative protein sequence 144

5.6. The second helix in the course of deuterostome evolution 146

5.7. Outlook: the colonial ascidian *Botryllus schlosseri* 149

6. References 150

7. Appendix 165
7.1. Protein Sequences 165
7.1.1. *Ciona intestinalis* p53/p73-a isoforms with numbering according to T. Noda [140] 165
7.1.2. *Ciona intestinalis* p53/p73-b 168
7.1.3. *Ciona savignyi* p53/p73-a isoforms 169
7.1.4. *Ciona savignyi* p53/p73-b 171
7.1.5. Preliminary protein sequence of *Botryllus schlosserie* p53/p73 172
7.1.6. Reference numbers of protein sequences from other invertebrate species as listet in the uniprot database 172
7.1.6.1. Branchiostoma floridae (Florida lancelet; amphioxus) protein sequences 172
7.1.6.2. Drosophila melanogaster (fruit fly) protein sequence 172
7.1.6.3. Caenorhabditis elegans (roundworm) protein sequence 172
7.1.7. Reference numbers of protein sequences from vertebrate species as listet in the uniprot database 173
7.1.7.1. Callorhynchus milii (elephant shark) protein sequences 173
7.1.7.2. Danio rerio (zebrafish) protein sequences 173
7.1.7.3. Xenopus tropicalis (western clawed frog) protein sequences 173
7.1.7.4. Mus musculus (mouse) protein sequences 173
7.1.7.5. Homo sapiens (human) protein sequences 173
7.1.8. Alignments 174
7.1.8.1. Percentage amino acid identity within DBD between p53 family proteins of vertebrate and *Ciona* species 174
7.1.8.2. Percentage amino acid identity within TD between p53 family proteins of vertebrate and *Ciona* species 175
7.1.8.3. Percentage amino acid identity within SAM domain between vertebrate p63 and p73 proteins 176
7.2. Prediction of secondary structure propensity by SSP and ncSPC 177
Acknowledgements 179
Curriculum vitae 181

Abbreviations

^{13}C	Carbon-13
^{15}N	Nitrogen-15
^{1}H	Hydrogen-1 (proton)
AU	Absorbance units
BSA	Bovine serum albumin
CSI	Chemical Shift Index
CSP	Chemical shift perturbation
CT	constant time
Da	Dalton
DBD	DNA binding domain
DNA	Desoxyribonucleic acid
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDTA	Ethylendiaminetetraacetic acid
EST	Expressed sequence tag
EtBr	Ethidium bromide
EtOH	Ethanol
h	Hour
HECT	Homologous to the E6-AP Carboxyl Terminus (catalytic active domain of Itch)
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
hetNOE	heteronuclear nuclear Overhauser effect
6xHis	Hexa-histidine (affinity tag)
HMQC	Heteronuclear Multiple Quantum Coherence

HSQC	Heteronuclear Single Quantum Coherence
IPTG	Isopropyl β-D-1-thiogalactopyranoside
IMAC	Immobilized metal affinity chromatography
Itch	E3 ubiquitin-protein ligase Itchy homolog
kbp	Kilo base pair
L	Liter
LB	Lysogeny broth
MDM2	Mouse double minute 2 homolog
MDM4 also known as MDMX	Mouse double minute 4 protein (MDM2-like)
MES	2-(N-Morpholino)-ethanesulfonic acid
min	Minute
MS	Mass spectrometry
MW	Molecular weight
MWCO	Molecular weight cut off
NaPi	Sodium phosphate buffer
NMR	Nuclear magnetic resonance
NOE	Nuclear Overhauser effect
NOESY	Nuclear Overhauser effect spectroscopy
OD	Oligomerization domain
ON	Overnight
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PIC	Protease inhibitor cocktail
PPI	Protein protein interaction
RMSD	Root-mean-square deviation

RNA	Ribonucleic acid
rpm	Rounds per minute
RT	Room temperature
s	Second
SAM	Sterile alpha motif (small domain composed exclusively of α-helices)
SDS	Sodium dodecyl sulfate
SEC	Size exclusion chromatography
SSE	Secondary structure element
TALOS(+)	torsion angle likeliness obtained from shift & sequence similarity
TA	Transactivation domain
TBS	Tris buffered saline
τ_c	Molecular rotational correlation time
TD	Tetramerization domain
TEMED	N,N,N´,N´-tetramethylethylenediamine
TEV	Tobacco etch virus protease
TI	Transactivation inhibitory domain
TOCSY	Total Correlation Spectroscopy
Tris	Tris(hydroxymethyl)aminomethane
TROSY	Transverse relaxation optimized spectroscopy
TWEEN 20	Polyoxyethylene (20) sorbitan monolaurate
V	Volt

Amino acids are abbreviated using the common single and three letter code

Glossary on homology

Homologous	Genes related to one another by descent from a common ancestral DNA sequence.
Orthologous	A set of homologous genes that diverged from each other as a consequence of speciation.
Paralogous	A set of homologous genes that resulted from genetic duplication.
Retrotransposition	Describes a rare mechanism where an mRNA transcript is spontaneously reverse transcribed back into DNA and afterwards inserted into chromosomal DNA. Infections with retroviruses are considered as the root cause for gene duplications resulting from retrotransposition.

Summary

The central task of the transcription factor (TF) p53 is to prevent erroneous, somatic cells from duplicating themselves which otherwise could result in the formation of cancer. In response to cellular stresses as for instance DNA damage or increased oncogene activity p53 is activated by upstream signaling cascades. Through activating the expression of a variety of different target genes p53 then in turn can arrest the cell cycle, initiate DNA damage repair or in more severe cases also trigger elimination of the cell by apoptosis. Therefore p53 is often called the “guardian of the genome”. Its importance for human health is impressively illustrated by the observation that about half of all carcinosises can be correlated with mutations in p53. And also in the remaining tumor entities p53 is mostly just as functionally reduced, for instance as a result of the inhibition of upstream or downstream acting signaling cascades or as a result of increased expression of proteins, which down regulate the cellular concentration of p53.

Beside p53 vertebrates possess two additional p53-like TFs, p63 and p73. During evolution all three proteins developed from the same common ancestor as a result of two sequential gene duplications and therefore possess a high degree of structural homology and sequence identity. They have a modular architecture with several independent domains. p53 as well as all isoforms of p63 and p73, which are able to activate the expression of target genes, possess an N-terminal transactivation domain (TA) followed by a DNA binding domain (DBD) and a tetramerization domain (TD). The longest isoforms (TA α) of p63 and p73, respectively, comprise two additional domains, a sterile alpha motif (SAM) domain and a transcription inhibitory (TI) domain, which are located C-terminal to the TD.

TAp63α is exclusively expressed in primary oocytes, which are arrested in the prophase of meiose I, until they are recruited for ovulation. If, during this phase, the DNA is damaged, TAp63α initiates the elimination of the oocyte similar to the function of p53 in somatic cells. Therefore p63 is also called the “guardian of the female germ line”. Interestingly, also in *Caenorhabditis elegans*, a species from the phylum nematoda, this function of securing the genetic integrity of oocytes is fulfilled by the only p53-like TF of the species. But invertebrates, such as *Caenorhabditis elegans*, do not need a tumor suppressor for somatic cells as they do not possess renewable tissues and are therefore not threatened by tumors. Therefore it is nowadays believed that securing the genetic integrity of germ cells has been

the ancestral function of p53-like TFs. And in combination with insights from phylogenetic studies this puts the assumption close that among vertebrate p53-like TFs p63 represents the most ancestral form. Deciphering the functional history of the protein family of p53-like TFs could in combination with detailed structural investigations contribute to a better understanding of their present-day functions in vertebrates.

The formation of tetramers via their TDs is a common characteristic of all p53-like TFs from vertebrate species. This oligomerization allows for cooperative, sequence-specific binding of the DBDs to DNA responsive elements and is a prerequisite for the transactivation of target genes. The DBD is highly conserved in all p53-like TFs even over large phylogenetic distances. The fact that the TD is much less conserved in comparison reveals that its architecture has undergone substantial changes during evolution. Therefore the superordinate research objective, to which this thesis writing contributes, is to elucidate these changes by examining the TDs of p53-like proteins from invertebrate species.

Ciona intestinalis (*C.int.*) is a species from the subphylum tunicata. Tunicates are the closest living relatives of vertebrates and *C.int.* is a popular experimental model system for studying embryonic development. Its genome encodes two p53-like TFs, which are named p53/p73-a and p53/p73-b. The structures of their TDs were examined by nuclear magnetic resonance (NMR) spectroscopy.

The TD of human p53 can best be described as a dimer of dimers. Each monomer contributes a β-strand and an α-helix, which are linked by a single highly conserved glycine residue allowing for the sharp turn between the two secondary structure elements (SSE). Within the primary dimer the β-strands form an intermolecular antiparallel β-sheet and the helices pack against each other lengthwise with opposing orientation. In the tetramer two such primary dimers then arrange in such a way that a bundle of four helices is formed, which is sandwiched between the two β-sheets. This minimal motif is highly conserved and is designated in the following also as the core TD, since it is in the TDs of other p53-like proteins often stabilized by additional SSEs. In the TDs of human p63 and p73 each monomer possesses an additional second helix at its C-terminus. The second helices from each primary dimer wrap around the other primary dimer in a clamp like fashion and in this way stabilize the tetramer. The stable binding to the core TD thereby relies on the two sequential, highly conserved residues, tyrosine and arginine (Tyr-Arg motif), within the center of the second helix, which form stabilizing contacts to the core TD.

Just this Tyr-Arg motif is also conserved in the TD of *C.int.* p53/p73-a, but the analysis of the secondary structure on the basis of NMR chemical shifts revealed that the TD of *C.int.* p53/p73-a does not form a second helix at 25°C. However, utilizing chimeric TD peptides, in which parts of the *Ciona* sequence were exchanged against the corresponding sections from human p73, it was shown that the core TD of *C.int.* p53/p73-a is able to stabilize a second helix and that therefore beside the Tyr-Arg motif also the section between first and second helix is decisive. Within this section the protein sequences of human p73 and *C.int.* p53/p73-a differ from one another in six sequential residues. Out of these, those four were identified, whose side chains contribute most to the stability of the TD of human p73, and in the TD of *C.int.* p53/p73-a the respective positions were then exchanged in different combinations for the corresponding residues from human p73. A minimum of three sequential substitutions turned out to be sufficient to cause the formation of a second helix and to increase the thermal stability of the TD with respect to the wild type protein. However, in contrast to the chimeric protein, in which starting at the end of helix one the complete C-terminus was exchanged for the respective sequence of human p73, none of the TD peptides with up to four stabilizing mutations did form a stably folded second helix. The second helix in each case folded only transiently and conformations with and without a second helix were in fast exchange with respect to the NMR time scale. This observation suggested examining a possible temperature dependency, as well. And indeed it could be shown that at lower temperatures also the wild type TD of *C.int.* p53/p73-a is able to form a transiently folding, second helix. At 10°C an estimated fraction of about 25% possesses a folded second helix in temporal means.

The stability of the tetrameric state of the TDs of human p63 and p73 highly depends on their stably folded second helices, respectively. If these are truncated they form primarily only dimers even at high µ-molar concentrations. In contrast it was shown by analytical ultracentrifugation that at 150 µM an accordingly truncated construct of the TD of *C.int.* p53/p73-a forms almost exclusively tetramers. However, *in vivo* TFs are generally present in by far lower concentrations and therefore it is to be assumed that the temperature-dependent stabilization of the second helix significantly strengthens the formation of tetramers at lower temperatures. But beside that the transiently folding helix could also be involved in regulating intra- or inter-molecular interactions.

The gene encoding *C.int.* p53/p73-b resulted from a relatively recent lineage-specific gene duplication caused by retrotransposition. According to phylogenetic data most likely an mRNA transcript for *C.int.* p53/p73-a, or more precisely the ancestral version of *C.int.* p53/p73-a existing at that time, was spontaneously reverse transcribed back into DNA and inserted into chromosomal DNA. Therefore a high degree of sequence identity between the paralogous proteins would be expected, but interestingly, within the TD residues with conservation in both proteins are restricted to the section spanning the β-strand and the first helix. Hence, *C.int.* p53/p73-b lacks the characteristic Tyr-Arg motif, and it had from there been predicted that *C.int.* p53/p73-b possessed only a minimal / core TD.

However, analysis of NMR chemical shifts revealed that in the TD of *C.int.* p53/p73-b every monomer possesses a rigidly folded second helix at its C-terminus. And since the TD of *C.int.* p53/p73-b therefore exhibits a so far unknown motif, its NMR solution structure was determined (pdb entry 2MW4). Surprisingly, although missing the Tyr-Arg motif the second helix adopts a position with respect to the core TD, closely resembling the overall architecture of the TD of human p73. The second helix is wrapping around the neighboring primary dimer in the same clamp like fashion and hence its formation depends on and stabilizes the tetrameric state. However, analytical ultracentrifugation experiments revealed that the second helix is albeit not required for tetramerization, and this was attributed to the observation that the core TD of *C.int.* p53/p73-b is stabilized by salt bridges involving residues from both primary dimers.

While the TD of human p53 is likewise stabilized by salt bridges spanning across the tetramerization interface, at suitable positions the core TDs of human p63 and p73 are lacking residues with accordingly oppositely charged side chains. Sequence comparisons and the recently published structure of the TD of zebrafish p53 put the conclusion close that originally also the TDs of all p53 proteins from vertebrate species possessed a second helix stabilized by a Tyr-Arg motif and that just relatively recently the second helix was lost on most of the p53 proteins due to the accumulation of degenerating amino acid substitutions. The loss of the second helix could thereby probably only be compensated since already before changes in the protein sequence of the core TD enabled the formation of inter-dimer salt bridges stabilizing the tetramer.

The observation that the TD of *C.int.* p53/p73-b possesses both stabilizing features is particularly interesting. Since the sequence of the second helix lacks similarity to the

respective sections in *C.int.* p53/p73-a or homologous proteins from vertebrate species, it could likely be an innovation, which is limited to the *Ciona* lineage. The tetramer is already stable without the second helix and its main function could instead be to prevent the formation of mixed tetramers with *C.int.* p53/p73-a.

Just recently the genome of another tunicate species, *Botryllus schlosserie* (*B.schl.*), has been sequenced, as well. However, gene models and protein sequences are not yet available. With the aim to improve the phylogenetic classification of the *Ciona* p53/p73 proteins as well as to gain some more insights into the structural changes the TD has undergone during tunicate evolution the genomic sequence of *B.schl.* was searched for sections encoding the DBD of a p53-like TF. Only a single gene was identified and for consistency the *B.schl.* p53 family protein was named *B.schl.* p53/p73. Based on sequence comparisons it was predicted with high confidence that the TD of *B.schl.* p53/p73 also possesses a stably folded second helix, which again deviates in its position relative to the core TD only little from the structure of the TD of human p73.

Taken together the results reveal that the architecture of the TD with a second helix had already been the prototype for the TDs of all p53-like TFs from vertebrates and tunicates. Therefore the second helix, which functions like a clamp in stabilizing the TD, must have developed prior to the divergence of the two phyla.

The NMR solution structure of the TD of *C.int.* p53/p73-b was submitted to the pdb and will be accessible upon release under the pdb entry code **2MW4**.

A manuscript for the publication of the data presented in this dissertation is in preparation and the publication will be authored as follows:
J. Heering, H. Jonker, F. Loehr, H. Schwalbe, V. Doetsch

Zusammenfassung

Die zentrale Aufgabe des Transkriptionsfaktors (TF) p53 ist es zu verhindern, dass sich entartete somatische Zellen vermehren und in der Folge Krebs entsteht. Als Reaktion auf zellulären Stress wie etwa Schädigungen der DNA, Hypoxie oder Onkogen-Aktivierung kann p53 die Zellteilung vorübergehend stoppen, die Reparatur von DNA Schäden initiieren oder in schweren Fällen Apoptose auslösen. Daher wird p53 auch oftmals als „Wächter des Genoms“ bezeichnet. Seine Bedeutung für die menschliche Gesundheit wird dadurch verdeutlicht, dass etwa die Hälfte aller Tumorerkrankungen im Menschen mit Mutationen von p53 korreliert werden können. Und auch in den übrigen Tumorentitäten ist p53 zumeist ebenso in seiner Funktion eingeschränkt, da etwa die vor- oder nachgeschalteten Signalkaskaden inhibiert sind oder Proteine vermehrt exprimiert werden, welche die Menge an p53 in den Zellen negativ regulieren. In Wirbeltieren gibt es neben p53 zwei weitere ähnliche TFs, welche mit p63 bzw. p73 bezeichnet werden. Alle drei Proteine sind durch Genduplikationen im Laufe der Evolution der Wirbeltiere aus einem gemeinsamen Vorläufer hervorgegangen und besitzen daher hohe strukturelle Homologie und Sequenzidentität. Ihr Aufbau ist modular mit mehreren voneinander unabhängig faltenden Domänen. p53 sowie alle Isoformen von p63 und p73, welche zur Aktivierung der Expression von Zielgenen befähigt sind, besitzen eine N-terminale Transaktivierunsdomäne (TA) gefolgt von einer DNA-Bindungsdomäne (DBD) und einer Tetramerisierungsdomäne (TD). In den jeweils längsten Isoformen (TA α) von p63 und p73 folgen auf die TD C-terminal jeweils noch ein ausschließlich aus α-Helices aufgebautes Motiv (engl. „sterile alpha motif“; SAM), welches eigenständig faltet und daher als SAM Domäne bezeichnet wird, sowie am C-terminus eine Transaktivierungsinhibitions- (TI) Domäne.

TAp63α wird ausschließlich in primären Oozyten exprimiert, welche in der Prophase der ersten Reifeteilung verharren, bis sie zur Ovulation rekrutiert werden. Kommt es während dieser Phase zu Schädigungen der DNA aktiviert TAp63α, analog zur Funktion von p53 in somatischen Zellen, die Eliminierung der Oozyte. p63 wird daher auch als „Wächter der weiblichen Keimbahn“ bezeichnet. Interessanter Weise wird diese Funktion zur Sicherstellung der genetischen Integrität der Oozyten auch in *Caenorhabditis elegans*, einer Spezies aus dem Stamm der Fadenwürmer, vom einzigen p53 ähnlichen TF dieser Spezies

wahrgenommen. Zudem verfügen auch primitive vielzellige Tiere über jeweils zumindest einen p53 ähnlichen TF, obgleich Wirbellose wie *Caenorhabditis elegans* keinen Tumorsuppressor für somatische Zellen benötigen. Denn sie sind nicht durch die Entstehung möglicher Tumore bedroht, da sie nur eine kurze Lebenszeit haben und nicht über sich erneuernde Gewebe verfügen. Es wird daher allgemein angenommen, dass nicht die Funktion eines Tumorsuppressors in somatischen Zellen, sondern die Sicherstellung der genetischen Integrität der Keimzellen, die ursprüngliche Funktion von p53 ähnlichen TFs war. Sowohl die Ausübung dieser Funktion als auch phylogenetische Studien legen nahe, dass p63 die evolutionär ursprünglichste Form der p53 TF Familie in Wirbeltieren darstellt.

Die evolutionären Ursprünge sowie die funktionelle Entwicklung der p53 Proteinfamilie aufzuklären könnte in Verbindung mit detaillierten Untersuchungen der Proteinstrukturen dazu beitragen ein besseres Verständnis der Funktionen zu erlangen, welche diese Proteine im Menschen ausüben. Dabei wären speziell neue Erkenntnisse in Bezug auf regulatorische Wechselwirkungen als auch direkte molekulare Interaktionen sehr nützlich, könnten sie doch dazu beitragen Krebsentstehung besser zu verstehen und möglicher Weise sogar neue Perspektiven für die Behandlung von Krebserkrankungen eröffnen.

Die DBD ist in allen Proteinen der p53 Familie auch über große phylogenetische Abstände hinweg hoch konserviert. Im Vergleich dazu variiert die Proteinsequenz der TD deutlich stärker was andeutet, dass ihre Struktur im Laufe der Evolution erhebliche Veränderungen durchlaufen hat. Die Art und Reihenfolge jener Veränderungen aufzuklären wird allerdings dadurch erschwert, dass naturgemäß die Proteinsequenzen direkter Vorläufer niemals verfügbar sind, da die Evolution nicht still steht. Das übergeordnete Forschungsvorhaben, zu welchem diese Dissertationsschrift beiträgt, ist es daher die TDs von p53 ähnlichen Proteinen aus wirbellosen Tieren zu untersuchen und diese mit den TDs von menschlichem p53, p63 und p73 zu vergleichen, um so Einblicke in die evolutionäre Vergangenheit und die strukturelle Entwicklung der TD zu erlangen.

Die über die TD vermittelte Bildung von Tetrameren ist eine gemeinsame Eigenschaft aller p53 ähnlichen TFs aus Wirbeltieren. Die Zusammenlagerung ermöglicht es den DBDs kooperativ, sequenzspezifisch an entsprechende DNA Sequenzmotive zu binden und ist eine essentielle Bedingung für die Rekrutierung von Co-Faktoren zur Aktivierung der Expression von Zielgenen.

Die Regulation des Oligomerisierungszustands und damit der Aktivierung von p53 erfolgt hauptsächlich indirekt über die Steuerung seiner Konzentration innerhalb der Zelle. p53 wird permanent neu exprimiert, aber in Abwesenheit von Stresssignalen durch in ihrer Aktivität fein regulierte E3 Ubiquitin Ligasen nach nur kurzer Zeit für den Abbau im Proteasom markiert. Dies bewirkt eine sehr kurze Halbwertszeit sowie niedrige Proteinkonzentrationen, sodass p53 hauptsächlich als Monomer und in Form von Dimeren vorliegt. In Reaktion auf zellulären Stress wird p53 posttranslational modifiziert, wodurch die Interaktion mit den E3 Ligasen unterbunden wird. In der Folge akkumuliert p53 und bildet vermehrt aktive Tetramere. Im Gegensatz dazu liegt TAp63α in Oozyten in hohen Konzentrationen vor und wird durch ein Netzwerk aus Wechselwirkungen zwischen der TI und der TA Domäne sowie der TD in einem auto-inhibierten, inaktiven, dimeren Zustand stabilisiert. Kommt es zu irreparablen DNA Doppelstrangbrüchen wird TAp63α phosphoryliert, wodurch die auto-inhibierenden Wechselwirkungen aufgebrochen werden und sich aktive Tetramere bilden können, was letzten Endes die Apoptose der geschädigten Oozyte einleitet.

Diese Doktorarbeit befasst sich mit der strukturellen Charakterisierung der TDs sowie der C-terminalen Sequenzabschnitte der zwei p53/p73 ähnlichen Proteine aus *Ciona intestinalis* (*C.int.*), einer Spezies aus dem Unterstamm der Manteltiere. Die Manteltiere sind die engsten lebenden Verwandten der Wirbeltiere und *C.int.* ist die mit Abstand am besten charakterisierte Spezies. Sie zählt zu den sogenannten Seescheiden, welche die artenreichste Klasse der Manteltiere bilden. Obgleich der erwachsene Organismus sehr primitiv wirkt, hat das Larvenstadium große Ähnlichkeit mit dem Grundbauplan der Wirbeltiere. Die Larven verfügen über einen von kräftigen Muskeln angetriebenen Ruderschwanz und entlang des Rückens verläuft eine stabförmige, elastische Stütze, die Chorda dorsalis, die einen zentralen Nervenstrang, das Neuralrohr, umschließt. Die große Ähnlichkeit dieser Struktur mit dem Aufbau der Wirbelsäule und damit das enge verwandtschaftliche Verhältnis von *C.int.* und Wirbeltieren wurde bereits Mitte des 19ten Jahrhunderts von Charles Darwin und Alexander Kowalevsky erkannt. Seither ist *C.int.* ein bedeutender Modelorganismus für die Erforschung der Embryonalentwicklung.

Das Genom von *C.int.* sowie auch das Genom der eng verwandten Spezies *Ciona savignyi* (*C.sav.*) beinhaltet zwei Gene, welche für p53 ähnliche TFs kodieren. Phylogenetischen Analysen zu Folge gehen diese zwei Gene auf eine vermutlich nicht sehr weit zurück liegende Genduplikation zurück, welche spezifisch nur in der Entwicklungslinie von *Ciona* stattfand.

Sie erfolgte unabhängig von den zwei aufeinander folgenden Genduplikationen, welche während der Evolution der Wirbeltiere zu den drei paralogen Genen für p53, p63 und p73 geführt haben. Die beiden paralogen *Ciona* Proteine bilden daher eine unabhängige monophyletische Gruppe. Sie sind beide gleichermaßen Co-Orthologe der drei p53 ähnlichen Proteine der Wirbeltiere und aus dieser gruppenbezogenen Homologie folgt, dass keines der Wirbeltier Proteine einzig auf nur eines der *Ciona* Proteine zurückgeführt werden kann. Um diesem Verwandtschaftsverhältnis Rechnung zu tragen, werden die zwei *Ciona* Proteine daher mit p53/p73-a und p53/p73-b bezeichnet. Nach aktuellem Kenntnisstand geht das Gen, welches für *C.int.* p53/p73-b kodiert, auf eine für *Ciona* spezifische durch Retrotransposition hervorgerufene Genduplikation zurück. Dabei wurde ein mRNA Transkript, welches für die zu jenem Zeitpunkt existierende Vorläuferversion von *C.int.* p53/p73-a kodierte, spontan zurück in DNA übersetzt und in die chromosomale DNA eingefügt. Im Anschluss an diese Duplizierung haben sich die beiden Kopien in Bezug auf die Aminosäuresequenz der TD dann allerdings stark auseinander entwickelt.

Die TD von menschlichem p53 (hp53) kann am besten als ein Dimer aus Dimeren beschrieben werden. In jedem Monomer folgt auf einen β-Strang ein hochkonserviertes Glycin, welches einen scharfen Knick hin zu einer direkt folgenden α-Helix ermöglicht. Jeweils zwei Monomere lagern sich achsensymmetrisch zu einem primären Dimer in der Weise zusammen, dass sich ein antiparalleles β-Faltblatt bildet und die α-Helices mit entgegen gesetzter Orientierung der Länge nach aneinander packen. Zwei dieser Dimere lagern sich dann so zum Tetramer zusammen, dass die Helices circa im rechten Winkel aufeinander packen und damit zwischen den polständigen β-Faltblättern ein Bündel aus vier Helices entsteht. Dieses minimale Motiv ist hoch konserviert und wird im Folgenden auch als Kern-TD bezeichnet, da es in allen bisher strukturell aufgeklärten TDs mit Ausnahme der von p53 durch zusätzliche Sekundärstrukturelemente (SSEs) stabilisiert wird.

In den TDs von menschlichem p63 und p73 (hp63 und hp73) verfügt jedes Monomer noch über eine weitere C-terminale α-Helix. Die zusätzlichen Helices eines jeden Dimers greifen wie Klammern um das jeweils andere primäre Dimer und stabilisieren auf diese Weise das Tetramer. Von besonderer Bedeutung sind dabei die beiden aufeinander folgenden und hoch konservierten Aminosäuren (ASs) Tyr-Arg, welche in der Mitte der zweiten Helix liegen und stabilisierende Kontakte zur Kern-TD formen.

Eben jenes Tyr-Arg Motiv ist ebenso in der TD von *C.int.* p53/p73-a konserviert, jedoch ergab die Vorhersage der Sekundärstruktur mit TALOS+ auf Basis der mittels 3D NMR Experimenten ermittelten chemischen Verschiebungen von entsprechend sensitiven und unmittelbar das Proteinrückgrad bildenden Atomen, dass sich keine stabile zweite Helix faltet. Dies wurde anhand von chimären TD Peptiden genauer untersucht, in welchen der C-terminale Abschnitt der Primärsequenz von *C.int.* p53/p73-a von unterschiedlichen Positionen an beginnend bis zum C-terminus des jeweiligen Peptids gegen die entsprechende Sequenz von hp73 ausgetauscht wurde. Damit war es möglich zu zeigen, dass die minimale TD von *C.int.* p53/p73-a im Prinzip fähig ist eine zweite Helix zu stabilisieren. Mit weiteren Mutationsexperimenten wurde außerdem bestätigt, dass auch im chimären Protein für die Stabilisierung der zweiten Helix über Interaktionen mit der Kern-TD von *C.int.* p53/p73-a das Tyr-Arg Motiv in der zweiten Helix essentiell ist. Dies legte den Schluss nahe, dass die zweite Helix eine Position relativ zur Kern-TD einnimmt, welche jener in der TD von hp73 sehr nahe kommt, denn jene Reste der Kern-TD, welche mit der zweiten Helix wechselwirken, sind in beiden Proteinen ebenfalls konserviert. Der kurze Abschnitt zwischen erster und zweiter Helix fungiert in hp73 nicht nur als Gelenk zwischen den beiden SSEs, sondern bildet auch mit dem gleichen Abschnitt eines anderen Monomers eine Vielzahl von hydrophoben Wechselwirkungen aus. Chimäre Proteine, in welchen dieser Abschnitt der Sequenz von *C.int.* p53/p73-a entsprach, bildeten keine zweite Helix aus. Zwischen dem Ende der ersten und dem Anfang der zweiten Helix unterscheiden sich die Proteinsequenzen von hp73 und *C.int.* p53/p73-a in sechs aufeinander folgenden Resten. Von diesen wurden jene vier identifiziert, deren Seitenketten in der TD von hp73 die größten stabilisierenden Effekte ausüben. In der TD von *C.int.* p53/p73-a wurden dann in unterschiedlichen Kombinationen die entsprechenden Positionen gegen die Reste aus hp73 getauscht. Für die drei Mutanten, deren HSQC Spektren die größten Unterschiede zum Spektrum des Wildtyps aufwiesen, wurden die chemischen Verschiebungen des Proteinrückgrads ermittelt und die Sekundärstrukturvorhersagen mit TALOS+ bestätigten jeweils die Faltung einer zweiten Helix.

Zusätzlich wurden ebenfalls auf Basis der chemischen Verschiebungen die prozentualen Anteile von Konformationen mit und ohne zweite Helix berechnet. Dies ergab, dass im Gegensatz zum chimären Protein keine der Mutanten eine stabile, konstant gefaltete zweite Helix ausbildet, was zusätzlich mit hetNOE (engl. heteronuclear nuclear Overhauser effect)

Experimenten bestätigt wurde. Die Daten zeigten, dass die zweite Helix nicht dem häufig in Bezug auf Proteinfaltung postulierten Alles-oder-nichts Prinzip folgend, sondern nur transient, faltet. Wurden unterschiedliche Mutationen kombiniert war deren Effekt circa additiv. In allen Mutanten war der Austausch zwischen Konformationen mit und ohne zweite Helix schnell im Verhältnis zur NMR Zeitskala, wodurch jeweils nur ein Satz an Peaks in den 2D NMR Spektren detektiert wurde und es nicht zur Linienverbreiterung kam.

Mittels Circulardicroismus (CD) Spektroskopie wurde gezeigt, dass minimal die Sequenz VAA zu LVP ausgetauscht werden muss, um die Faltung der zweiten Helix zumindest soweit zu stabilisieren, dass die thermische Stabilität der gesamten TD im Vergleich zum Wildtyp erhöht wird. In diesem Motiv kommt dem Prolin eine besondere Bedeutung zu, denn es besetzt in der TD von hp73 die Position unmittelbar vor der zweiten Helix (N-cap) und bildet zahlreiche zumeist hydrophobe Wechselwirkungen aus, welche die erste Windung der zweiten Helix stabilisieren.

Mit hetNOE Experimenten konnte außerdem gezeigt werden, dass bei niedrigen Temperaturen auch die Wildtyp TD von *C.int.* p53/p73-a befähigt ist eine zweite Helix auszubilden, welche ebenfalls nur transient faltet. Im zeitlichen Mittel haben bei 10°C Konformationen mit gefalteter zweiter Helix einen Anteil von etwa 25%.

Es ist bekannt, dass die TDs von hp63 und hp73 ohne zweite Helix auch bei hohen µ-molaren Proteinkonzentrationen zu einem Großteil in Dimere zerfallen. Im Gegensatz dazu konnte für die TD von *C.int.* p53/p73-a mittels analytischer Ultrazentrifugation gezeigt werden, dass diese bei 150 µM auch dann noch hauptsächlich Tetramere bildet, wenn diese um den Sequenzabschnitt, welcher der zweiten Helix entspricht, verkürzt wird.

Nativ liegen TFs im Allgemeinen aber in weitaus geringeren Konzentrationen vor. Daher ist davon auszugehen, dass die temperaturabhängige Stabilisierung der zweiten Helix die Bildung von Tetrameren bei niedrigen Temperaturen signifikant verstärkt. Des Weiteren könnte die transient faltende Helix von *C.int.* p53/p73-a aber auch eine Funktion zur Regulation von intra- oder intermolekularen Interaktionen erfüllen.

Obgleich die Genduplikation, welche in *Ciona* zu p53/p73-a und –b geführt hat, nach evolutionären Maßstäben vermutlich noch nicht lange zurück liegt, haben sich die beiden paralogen Proteine in Bezug auf die Proteinsequenz der TD stark auseinander entwickelt. Innerhalb der TD finden sich Reste, welche in beiden Proteinen konserviert sind, ausschließlich im Bereich der Kern-TD, also in jenem Abschnitt, welcher dem β-Strang und

der ersten Helix entspricht. Daher fehlt *C.int.* p53/p73-b auch das Tyr-Arg Motiv, welches essentiell für die Bildung der zweiten Helix in den TDs von hp63 und hp73 sowie *C.int.* p53/p73-a ist. Die Annahme, dass die TD von *C.int.* p53/p73-b somit auch keine zweite Helix besitzen würde, wurde jedoch durch die Analyse der chemischen Verschiebungen mit TALOS+ widerlegt. Da die TD von *C.int.* p53/p73-b demnach ein bisher unbekanntes Faltungsmotiv aufweist, wurde im Rahmen der vorliegenden Arbeit die NMR Lösungsstruktur ermittelt.

Für die Strukturaufklärung mittels NMR stellt ein Homo-Tetramer, in welchem wie in diesem Fall alle Monomere durch Symmetrieoperationen ineinander überführbar sind, eine besondere Herausforderung dar. Denn die chemische Umgebung eines jeden Monomers ist identisch und dadurch können die Resonanzen von einzelnen Kernen nicht monomer-spezifisch zugeordnet werden. Bei der Zuordnung der NOE Kontakte und damit dem Definieren von Abstandsinformationen resultiert hieraus die Problematik, dass jeweils nicht eindeutig bestimmbar ist zu welchen Monomeren die beteiligten Protonen gehören. Daher wurden Proben präpariert in denen die einzelnen Monomere jeweils nur entweder mit dem NMR aktiven Kohlenstoff-Isotop ^{13}C oder dem NMR aktiven Stickstoff-Isotop ^{15}N angereichert waren. Mit Hilfe entsprechend gefilterter Experimente war es so möglich zumindest Kontakte innerhalb der gleichen Peptidkette von jenen zu anderen Monomeren unterscheiden zu können.

Insgesamt ist die Topologie der TD von *C.int.* p53/p73-b derer der TDs von hp63 und hp73 sehr ähnlich, obwohl die Proteinsequenz, welche die zweite Helix faltet, im Gegensatz zur Kern-TD keine in hp63 oder hp73 konservierten Reste enthält. Die TD von *C.int.* p53/p73-b ist auch als ein Dimer aus Dimeren aufgebaut und die zweiten Helices eines jeden primären Dimers greifen ebenso wie in hp63 und hp73 um das jeweils andere Dimer. Dabei nehmen sie interessanter Weise eine Position relativ zur Kern-TD ein, welche nur wenig von der in hp73 abweicht. Die zweiten Helices sind stabil gefaltet, sind aber für die Stabilität des Tetramers nicht notwendig, was mittels eines entsprechend verkürzten Konstruktes und analytischer Ultrazentrifugation gezeigt werden konnte. Für die hohe Stabilität des durch die Kern-TD gebildeten Tetramers sind vermutlich vor allem Salzbrücken verantwortlich, an welchen jeweils eine Seitenkette aus jedem der primären Dimeren beteiligt ist.

Während die TD von p53 ebenfalls durch das Tetramerisierungsinterface überspannende Salzbrücken stabilisiert wird, fehlen in den Kern-TDs von hp63 und hp73 entsprechend

entgegengesetzt geladene Seitenketten an passenden Positionen. Sequenzvergleiche und die vor kurzem veröffentlichte Struktur der TD von Zebrafisch p53 legen den Schluss nahe, dass ursprünglich auch die TDs sämtlicher p53 Proteine aus Wirbeltieren jeweils eine über ein Tyr-Arg Motiv stabilisierte, zweite Helix aufwiesen und diese erst spät im Laufe der Evolution durch Akkumulation von degenerierenden AS Austauschen verloren ging. Der Verlust der zweiten Helix konnte dabei vermutlich nur dadurch kompensiert werden, dass bereits zuvor Veränderungen in der Sequenz der Kern-TD die Ausprägung von das Tetramer stabilisierenden Salzbrücken ermöglichten.

Die Beobachtung, dass die TD von *C.int.* p53/p73-b über beide Eigenschaften verfügt ist insbesondere deshalb interessant, da die Sequenz der zweiten Helix keinerlei Sequenzhomologie zu anderen p53 ähnlichen Proteinen aufweist und vermutlich eine einzig auf die Entwicklungslinie von *Ciona* beschränkte Innovation darstellt. Das Tetramer ist bereits ohne diese zweite Helix stabil und ihre Funktion liegt vermutlich vor allem darin die Bildung von gemischten Tetrameren mit *C.int.* p53/p73-a zu verhindern, was in Experimenten gezeigt und in entsprechenden theoretischen Betrachtungen anhand der Proteinstrukturen diskutiert wurde.

Vor kurzem wurde das Genom von *Botryllus schlosserie* (*B.schl.*), einer weiteren Manteltierspezies, sequenziert, Genmodelle und Proteinsequenzen sind jedoch noch nicht verfügbar. Mit dem Ziel die phylogenetische Einordnung zu verbessern sowie die strukturelle Evolution der TDs der *Ciona* p53/p73 Proteine genauer einordnen zu können, wurde daher die genomische Sequenz nach Abschnitten durchsucht, welche eine p53 ähnliche DBD kodieren. Dabei wurde nur ein einziges Gen identifiziert, welches aus Gründen der Einheitlichkeit als *B.schl.* p53/p73 bezeichnet wird. Für dieses wurde referenziert auf die unterschiedlichen Isoformen von *Ciona* p53/p73-a eine Vorhersage der Spleißstellen und damit der Proteinsequenzen möglicher Isoformen gemacht. Auf Basis von Sequenzvergleichen wurde vorhergesagt, dass auch die TD von *B.schl.* p53/p73 mit hoher Wahrscheinlichkeit über eine stabil gefaltete zweite Helix verfügt, welche in ihrer Position relativ zur Kern-TD wiederum nur wenig von der Struktur der TD von hp73 abweicht.

Die Ergebnisse zeigen zusammengenommen, dass die Architektur der TD mit zweiter Helix bereits der Prototyp für die TDs aller p53 ähnlichen Proteine der Wirbeltiere und Manteltiere war und die als eine Art Klammer das Tetramer stabilisierende zweite Helix sich nicht erst während Evolution der Wirbeltiere entwickelt hat.

Die Domänenstruktur der TA α Isoformen von hp63 und hp73, welche durch eine C-terminal auf die TD folgende SAM Domäne charaktierisiert ist, findet sich teilweise bereits in den Homologen einiger tierischer Spezies, wie etwa Seeigeln, deren letzter Vorfahre, welchen sie mit Wirbel- und Manteltieren gemeinsam haben, deutlich weiter zurück datiert. Interessanterweise, konnte jedoch gezeigt werden, dass keine der Isoformen der *Ciona* p53/p73 Proteine eine SAM Domäne enthält und ferner wurde für *B.schl.* p53/p73 keine Isoform vorhergesagt, welche über eine ausreichend lange C-terminale Sequenz verfügen würde.

Daraus konnten zwei mögliche Szenarien für die phylogenetische Verwandtschaft der p53 ähnlichen Proteine von Wirbeltieren und Manteltieren abgeleitet werden. Entweder die Genduplikation, welche im Verlauf der Evolution der Wirbeltiere zu p53 und dem Vorläuferprotein p63/p73 geführt hat, fand vor der Spaltung der Entwicklungslinien von Wirbeltieren und Manteltieren statt. Dann muss das p63/p73 Protein später selektiv in der Manteltierlinie ausgelöscht worden sein. Oder die SAM Domäne ging unabhängig sowohl im p53 der Wirbeltiere als auch in den p53/p73 Proteinen der Manteltiere verloren.

Die NMR-Lösungsstruktur der TD von *C.int.* p53/p73-b wurde an die pdb übermittelt und wird ab Veröffentlichung abrufbar sein unter der pdb Kennung **2MW4**.

Ein Manuscript zur Veröffentlichung der in dieser Dissertationsschrift präsentierten Daten ist in Vorbereitung und die Veröffentlichung wird die folgenden Autoren haben:
J. Heering, H. Jonker, F. Loehr, H. Schwalbe, V. Doetsch

1. Introduction

1.1. The vertebrate p53 family of transcription factors

The p53 family of transcription factors (TFs) has a long and complex evolutionary history. Its origin reaches back about one billion years of evolution and a p53 ancestor gene is first clearly observed in the modern-day descendants of the early metazoan sea anemone. In man this protein family comprises three different proteins, p53, p63 and p73. All three are products of gene duplication events and subsequent specialization during vertebrate evolution starting from a single genomic copy.[1]

1.1.1. The transcription factor p53 – the guardian of the genome

After p53 was first discovered in 1979 in cells infected by the SV40 it took nearly ten years to elucidate its role in cancer biology.[2,3] First cDNA clones of the p53 gene from mouse or human origin encoded mutant p53 and might have been generated from tumor tissues. In consequence the cooperation of the mutant protein with several oncogenes in the transformation of cells in culture was misguiding the research on p53 for several years.[4,5] In 1989 it was discovered that the short arm of chromosome 17, which is frequently deleted in colorectal carcinoma, is the genomic locus coding for p53. As the remaining allele of the p53 gene was found to be mutated in several entities of such tumors a function of wild-type p53 as a tumor suppressor was suggested.[6] Short after within the same year cell culture experiments revealed that, indeed, wild-type p53 can block oncogene mediated transformation.[7] Finally p53 was assigned to be the guardian of the genome, when sequencing data showed that 50% of all human tumors had mutations in the p53 gene.[8]

By cell culture, biochemical, and structural studies p53 has been characterized as a multi-domain protein consisting of three domains spaced by flexible linkers. As shown in Figure 1-1 the transcription activation domain (TA) at the N-terminus is followed by a DNA binding domain (DBD) and the oligomerization domain (OD) located at the C-terminus, which is frequently referred to as the tetramerization domain (TD) in more recent publications.

Figure 1-1: The domain architecture of p53. The transcription factor protein p53 is composed of an N-terminal transcription activation domain (TA), a DNA binding (DBD) and an oligomerization domain (OD) at the C-terminus. The OD is frequently referred to as the tetramerization domain (TD).

In response to various cellular stresses, such as DNA damage, p53 is modified. This results in increased cellular concentrations of the p53 protein and the formation of active tetramers. In this form p53 functions as a TF and binds via its DBD sequence-specific to DNA responsive elements in the sequences upstream of a variety of different target genes. The activation of the p53 pathway then in turn leads to apoptosis, cell cycle arrest or cell senescence.[9] In this way p53 prevents erroneous cells with mutations from duplicating themselves which otherwise could result in the formation of cancer. That p53 is not essential for development was demonstrated by p53 homozygous knock-out mice, which undergo normal development but are susceptible to tumorigenesis with early onset, within 6 months after birth.[10] Thus, it provides a possibility that extraneous intervention targeted at restoring abnormal p53 activity or protein concentration could be a promising therapeutic concept as normal development should not be affected by p53 gene therapy. In about half of all human tumors p53 is inactivated by point mutations within the DBD.[11] And it was found that such tumor-derived p53 mutants not only fail to suppress tumor growth,[7,12] but also inhibit the function of wild-type p53.[13–16] This dominant negative effect is believed to be the root cause for tumorigenesis derived from cells, which still possess one non-mutated p53 wild type allele. The effect can be explained by the sequestration of wild type p53 into mutant/wild type hetero-tetramers, which are inactive due to the partially defective DBDs. Only such p53 mutants with an intact TD are able to trans-dominantly inhibit wild type p53, and chimeric p53 proteins, which comprise heterologous TDs instead of the wild type p53 TD, are not trans-dominantly inhibited.[17,18] This shows that for a potential p53 gene therapy the TD would be of crucial importance, and therefore this domain received a lot of attention with hundreds of publications focusing on the molecular structure, stability and regulation of the oligomeric state as well as potential post-translational modifications and interaction partners.

1.1.2. Regulation of p53

Since p53 triggers life or death decisions, an exquisite control mechanism had to be evolved to prevent errant activation, while still enabling rapid responses to various stresses. The main regulatory mechanism for p53 is its interaction with the ubiquitin E3 ligase MDM2 (Figure 1-2), which results in a short half-life of the protein between 20 min and 2 hours.[9,19,20]

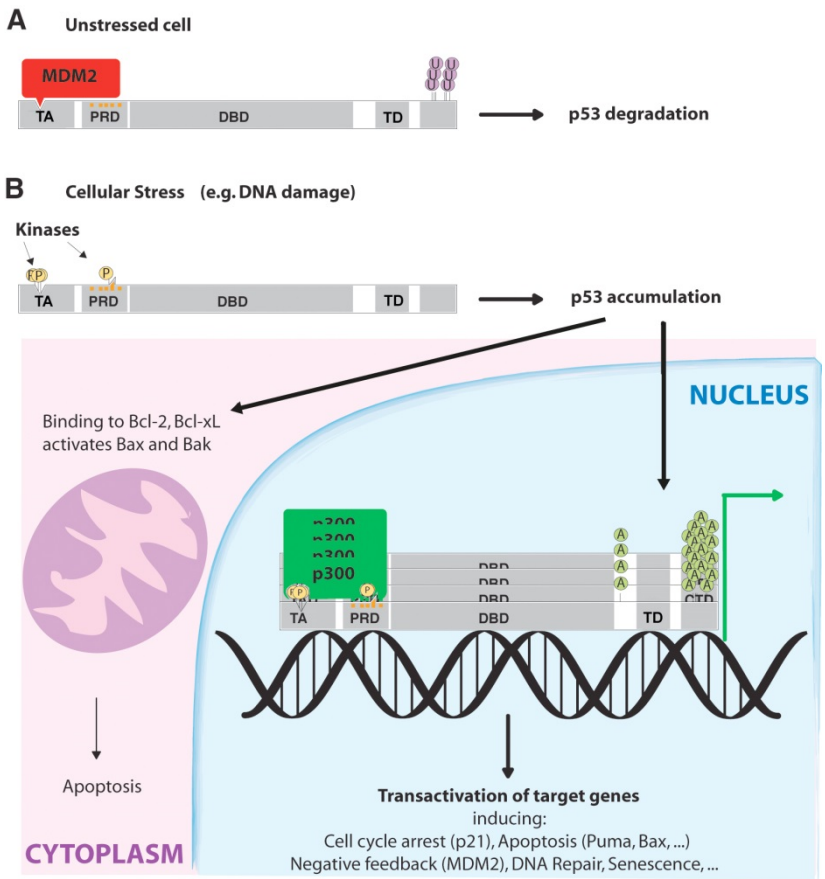


Figure 1-2 Regulation of p53 by MDM2 and post-translational modifications [modified from ref. [19]] (A) In unstressed cells MDM2 can bind to the transactivation domain (TA) of unmodified p53. This leads to the poly-ubiquitination at lysine residues in the section C-terminal to the tetramerization domain (TD) and targets p53 for proteasomal degradation. (B) Upon cellular stress p53 is phosphorylated which abrogates the interaction with MDM2 and in turn leads to the accumulation of p53.

MDM2 binds to the TA of p53 and covalently links poly-ubiquitin chains to lysine residues in its C-terminus which in turn targets p53 for degradation in the proteasomal pathway.[19–22] MDM2 and p53 are connected in a negative feedback loop, since MDM2 is itself a target gene of p53 and highly expressed upon activation of p53.[23,24] This allows for a finer tuning of the p53 response and establishes an additional security check reducing the chance of errant p53 activation.[25] As p53 is in this manner constantly expressed but kept at low levels, the

inhibition of the MDM2-mediated ubiquitination is a prerequisite for the activation of p53. The interaction between MDM2 and p53 can be disturbed by phosphorylation of residues in the TA of p53 by several different kinases, such as for example Chk1, Chk2, ATM and ATR, which are activated upon genotoxic stress.[26] And similarly, the phosphorylation of the p53 binding site in MDM2 by ATM or DNA-dependent protein kinase prevents p53 degradation.[27–29]

But phosphorylation is not always necessary for p53 stabilization and the oncogenes ras and myc use yet another mechanism to interfere with MDM2 regulation. They stimulate an increase of the tumor suppressor $p14^{ARF}$, which is also able to inactivate ubiquitination of p53.[30] $p14^{ARF}$ binds to the RING finger domain of MDM2 and thereby directly inhibits its E3 ligase activity.[31,32] MDMX, which is frequently also named MDM4, is a homolog of MDM2 with a similar structure but without intrinsic E3 ligase activity and is hence not capable to directly promote p53 degradation.[33] However, like MDM2, MDMX can bind to the TA of p53 and thereby inhibiting p53 transactivation by steric occlusion of the binding sites specific for other factors.[34] Furthermore, MDMX and MDM2 can form heterodimers through interaction of their C-terminal RING domains, and in this way MDMX stabilizes MDM2 by inhibiting its self-ubiquitination.[35] Similar to p53, MDMX can also be ubiquitinated by MDM2, and is thereby targeted for proteasomal degradation.[36–38] This system of multiple feedback circuits ultimately generates a steady state level of MDM2, MDMX and p53.[39]

The phosphorylation of p53 in the TA domain not only results in the release from MDM2 or MDMX but also in the recruitment of histone acetyl transferases like CBP and p300.[40,41] These act as co-activators and recruit a variety of other factors. Furthermore, they also subsequently acetylate lysine residues in the C-terminus of p53 which increases the ability of p53 to bind sequence-specific to DNA via the DBD.[42,43]

The numerous phosphorylations, acetylations and combinations thereof as well as interactions with modulators and inhibitors result in a highly complex network regulating and tuning the activity of p53.

1.1.3. Function of the p53 TD and regulation of the oligomeric state

The formation of tetramers is required for the direct binding of p53 to site-specific DNA responsive elements,[44–46] and mutations within the TD (aa 326-356) result in a reduction or

loss of the ability of p53 to activate transcription in cultured cells.[47] It was shown that such hindered formation of active tetramers causes predisposition for the development of cancers with early onset, known as the Li-Fraumeni syndrome.[48,49]

In *in vitro* studies, peptides resembling the TD assemble in a first step into homo-dimers with a K_d of ~1 nM.[50,51] These primary dimers then form tetramers with a K_d of ~100 nM,[50,52–54] which can be lowered by specific post-translational modifications.[52,53,55] As the p53 protein has an estimated mean cellular concentration of 140 nM,[56] it had already long been believed that at rest unmodified p53 is primarily dimeric.[57] This was recently proven by Gaglia et al.[51] who utilized fluorescence correlation spectroscopy to quantify the monomeric, dimeric, and tetrameric fractions of p53 contained in single living cells. Under basal conditions 58% of the p53 protein was detected to be bound in dimers, with 13% bound in tetramers and the remaining 29% being monomeric. The measured mean cellular concentration of p53 was up to ~400 nM and, thus, fairly higher than the dimeric dissociation constant of ~1 nM determined *in vitro* for TD peptides. Hence, the oligomeric state of p53 within cells is not solely dependent on the concentration of the protein, but a process highly regulated by additional factors.

In response to DNA damage the stoichiometric distribution is dramatically shifted towards tetramers leaving only minor fractions being monomeric or dimeric.[51] This tetramerization is largely independent of preceding basal p53 protein levels and does not require a prior increase in concentration.[51,58] The observed increase in cellular concentrations following the activation of p53 is, on the contrary, attributed to the stabilization of the protein rather than an increased expression. Tetrameric p53, however, does have a reduced half-life compared to monomeric or dimeric p53, and thus, cannot account for an increase in stability. The rapid surge of p53 tetramers can, therefore, best be explained by a combination of decreased degradation and active induction of tetramerization.[51]

In *in vitro* studies several cofactors have already been shown to regulate the oligomerization of p53. For instance, S100 proteins do preferentially bind to monomeric p53 and inhibit its oligomerization,[59] and 14-3-3σ does bind to phosphorylated p53, thereby reducing its tetrameric K_d.[52] The sequence-specific binding of p53 to DNA was also suggested to enhance the formation of tetramers, or rather, prevent their dissociation into dimers, once they have been formed. This implicates that modifications resulting in increased binding of p53 to DNA could indirectly effect the oligomeric state of p53.[54]

1.1.4. Structure of the tetramerization domain of *homo sapiens* p53

The topology of the p53 TD can be described as a dimer of dimers. As shown in Figure 1-3 each monomer contributes a β-strand and an α-helix, which are linked by a single highly conserved glycine residue allowing for the sharp turn between the two secondary structure elements (SSEs). Two monomers form a symmetrical primary dimer stabilized via an intermolecular anti-parallel β-sheet and antiparallel hydrophobic helix packing. The tetramer is composed of two such primary dimers, which pack against each other with their helices, resulting in a close to orthogonal arrangement with D_2 symmetry. Therefore, the tetramerization interface is solely formed by residues within the α-helices, and due to the large hydrophobic core, which is formed upon tetramerization, the tetrameric state is considered to be the most stable conformation of p53.[60–62] This is supported by the finding that the simultaneous mutation of Met340Gln and Leu344Arg leads to the abrogation of tetramerization and the formation of dimers only.[63]

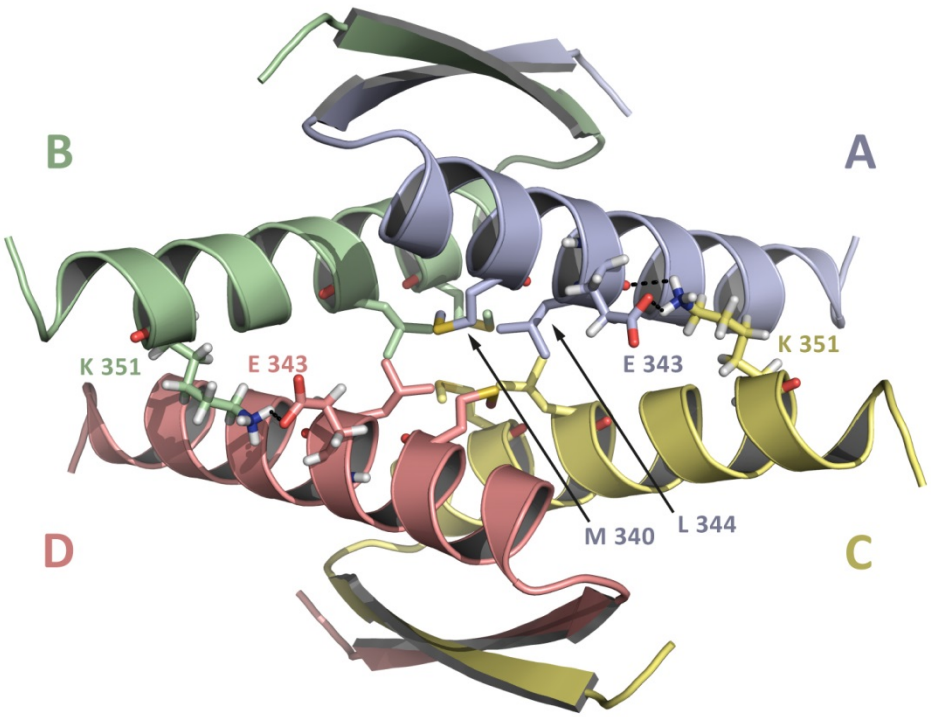


Figure 1-3 Crystal structure of the tetramerization domain of human p53 (pdb entry 1C26). Each monomer, colored blue, green, yellow and red, contributes a β-strand and an α-helix to the structure. Two highly intertwined monomers form a primary dimer in which the β-strands form an antiparallel β-sheet and the α-helices form an antiparallel helix-helix interface. Two such primary dimers pack against each other with their helices whereby the tetramerization interface is formed. Residues involved in inter-dimer salt bridges are depicted as sticks, and the hydrophobic residues Met 340 and Leu 344 are as well depicted as sticks but for clarity without the hydrogens. In reverse in each case the residues Glu 343 and Lys 351 from two adjacent monomers form salt bridges across the tetramerization interface, hence, a total of four. For clarity always only one of these interactions is shown per monomer pair, respectively.

As depicted in Figure 1-3 the tetramer is further stabilized by inter-dimer salt bridges, which are respectively formed between the side chain carboxylate group of Glu 343 and the primary amine from the side chain of Lys 351.[64,65] The importance of the stabilizing inter-dimer salt bridges is emphasized by the finding that the mutation of Lys351 to Glu results in a mutant whose transactivation activity is drastically reduced due to its ability to form dimers but not tetramers.[66,67] Furthermore the mutation Lys351Asn was identified in a cisplatin-resistant ovarian carcinoma cell line and was shown to significantly reduce the thermodynamic stability of the p53 tetramer, although not completely abrogating the formation of tetramers.[68]

1.1.5. The transcription factors p63 and p73 are closely related to p53

In 1997 and 1999, about twenty years after the discovery of p53, two paralogous proteins were discovered, named p63 and p73, whose functions diverge from p53.[69,70] Knock-out mice studies revealed that p63 is involved in limb and epidermal development,[71,72] and p73 being crucial for neuronal development.[73] In contrast to p53, which is not essential for normal development, the homozygous knock-out of p63 or p73 in mice resulted in a lethal phenotype. Mutations in p63 or p73 occur only rarely in human cancers (<1%) what excludes them as classic Knudson-type tumor suppressor genes,[74,75] but both proteins are expressed in a variety of different isoforms, which partially interplay with p53 mediated transcriptional activation. Like p53, both paralogs, p63 and p73, are modular proteins and all isoforms possess a DBD and a TD with high degrees of sequence identity to p53. As a result of different promoters at the N-terminus p63 and p73 can both be expressed as isoforms either containing the TA (TA isoforms) or not containing a complete N-terminus (ΔN), with the latter being capable to exert a dominant negative effect over the former. In addition, the N-terminal variants of p63 are alternatively spliced at different positions within the C-terminus resulting in a total of six different isoforms of p63. For p73 the complexity is even higher as at least seven different C-terminal splice variants were identified, although it is still a subject of debate how many of these truly exist as proteins.[76]

The TA alpha isoforms of p63 and p73 have the longest C-terminus, respectively, and as shown in Figure 1-4 possess two additional domains, a sterile alpha motif (SAM) domain and

a transcription inhibitory (TI) domain, which are located C-terminal to the TD and which are not found in p53.

Figure 1-4 The general domain architecture of the TA alpha isoforms of p63 and p73. The transcription factor proteins p63 and p73 possess an N-terminal transcription activation domain (TA) followed by a DNA binding (DBD) and a tetramerization domain (TD), as well as a sterile alpha motif (SAM) domain and a transcription inhibitory (TI) domain at the C-terminus.

Similar to p53 the TA domains of p63 and p73 are rather unstructured but able to recruit regulatory proteins or co-activators.[77,78] Interestingly, the MDM2 binding motif within the TA domain is respectively conserved and MDM2 as well as MDMX have been shown to bind the TA domains of p63 and p73 *in vitro*.[79] Thereby the affinities of both MDM2 and MDMX for p73 were found to be of the same order as those for p53, which suggests that the proteins truly interact in cells. However, for p63 much lower affinities were determined and *in vivo* studies revealed that neither MDM2 nor MDMX are able to inhibit activation of transcription by p63 or to affect its half-life.[80] Instead, the E3 ubiquitin ligase Itch has been shown to target p63 as well as p73 for proteasomal degradation and to be the main regulator of endogenous p63 protein levels.[81–83]

The TA isoforms of p63 and p73 generally behave similarly to p53 in terms of biological functions and overlapping target promoters, which can be attributed to the fact that almost all residues within the DBD of p53 that directly interact with DNA are strictly conserved in p63 and p73. Only Arg283 of human p53 is exchanged to Lys but the respective position accounts only for an unspecific contact with the phosphate backbone of the DNA. Consequently, both p63 and p73 can transactivate p53-responsive genes, and are as well capable to mediate cell cycle arrest, cellular senescence or apoptosis in response to DNA damage.[70,84–87]

Such as p53, p63 and p73 can only activate the transcription of target genes when they form tetramers. Interestingly *in vitro* studies with the isolated TDs revealed that the TDs of p63 and p73 are able to form heterotetramers with those consisting of homodimers being the thermodynamically most stable form.[88] Thereby the dissociation of the p73 homotetramer, which is more stable than the p63 homotetramer, is the rate-limiting step.[89]

It was further shown that neither p63 nor p73 forms heterotetramers with p53, hence, the functional crosstalk via direct protein-protein interactions (PPIs) is restricted to p63 and p73.

The ΔN isoforms lack the N-terminal TA domain and act as dominant negative inhibitors of the TA proteins through the formation of mixed tetramers.[90,91] Therefore this inhibitory mechanism is dependent on the relative amounts of the isoforms. And it is elevated to an autoregulatory feedback loop by the fact that the transcription of the ΔN isoforms is directly activated by the TA proteins.[90,92] In addition to the dominant negative effect on the TA proteins the ΔN isoforms can also indirectly inhibit the activity of p53 by competing for the same DNA binding sites.[91] And p53 has also been shown to induce the expression of its antagonist ΔN p73 what establishes an comparable feedback loop.[93]

The specific loss of TA p73 has been shown to cause predisposition to spontaneous or carcinogen-induced tumorigenesis *in vivo*.[94] Therefore it is believed that the imbalance between the TA and ΔN isoforms of the p53 family, which was observed in several different tumor entities, could prevent the TA proteins from engaging in tumor-suppressive functions, and hence, account for tumorigenesis.[95–97]

The SAM domain is an 8 kDa protein domain, which is present in the sequences of diverse eukaryotic proteins and which is often involved in PPIs.[98] Mutations in the SAM domain of p63, which lead to a reduction in stability and finally to the degradation of the protein, have been implicated in several syndromes caused by partial loss of p63.[99–101] Immediately preceding the first SSE of the SAM domain, respectively, both proteins, p63 and p73, contain a highly conserved PPPY motif, to which the E3 ubiquitin ligase Itch has been shown to directly bind.[81,83] And this physical interaction results in the ubiquitination of the TA alpha and ΔN alpha isoforms and consequently promotes their proteasomal degradation. However, it was not yet investigated whether the PPPY in the absence of the SAM domain would still be sufficient for the interaction with Itch. For p63 it was shown that degradation can be mediated in a similar fashion by Nedd4 and WWP1, which both are Itch analogues and as well bind the PPPY motif.[102,103] Interestingly it was shown that all isoforms of p63, independently of whether or not they possess the PPPY motif and the SAM domain, are equally targeted by ITCH to degradation. And in the absence of the SAM domain the region encompassing residues 109 to 120 in TAp63 (residues 15 to 62 in ΔNp63) was identified to be required for the association with Itch. The respective region is not conserved in p73 which is likely the reason for the observation that the p73 isoforms, which lack the PPPY motif and

the SAM domain, are not targeted by Itch for proteasomal degradation.[82] So far no other direct interaction partners have been described for the region comprising the SAM domain, and hence, it is possible that its only function is to serve as an interaction module for the named ubiquitin ligases.

The last C-terminal domain of the alpha isoforms of p63 and p73 is the TI domain, which is natively unstructured, such as the TA domain. For p63 it was shown that the TI domain can bind to the N-terminal TA domain, thereby decreasing its ability to activate transcription by enforcing an auto-inhibitory closed conformation.[104,105] By mutational analysis three residues within the TI were identified to be important for the inhibitory mechanism, Phe605, Thr606 and Leu607. The simultaneous mutation to Ala in TAp63α results in a 7-fold increase in transcriptional activity compared to wild type TAp63α.[105]

Taken together a complicated interplay between the p53 family proteins is caused by the reciprocal regulatory networks and the structural complexity of the protein family in terms of different isoforms.

1.1.6. Structures of the TDs of p63 and p73 in the tetrameric state

In Figure 1-5 the NMR solution structure of the TD of human p73 and the crystal structure of the TD of human p63 are depicted. Although the TDs of p63 and p73 are as well composed of a dimer of primary dimers, there is a major difference in terms of their overall architecture in comparison to the TD of p53, since in p63 and p73 each monomer contains an additional α-helix. In the tetramer this second helix is wrapping around the neighboring dimer like a clamp, thereby stabilizing the overall structure.[88,106] Truncation of the second helix or the mutation of the strictly conserved motif of Tyr-Arg within its sequence destabilizes the TDs of p63 and p73. For p63 it was shown that upon truncation of the second helix the structure of the primary dimers remains mainly unchanged while the packing of the primary dimers within the tetramer changes from orthogonal to anti-parallel.[89] In p63 and p73 the tetramerization interface, which is formed between the first helices of two opposing dimers, is characterized only by hydrophobic interactions and lacks inter-dimer salt bridges. And this lack in polar interactions granting specificity is probably the reason for the observed fluidity of the interface.

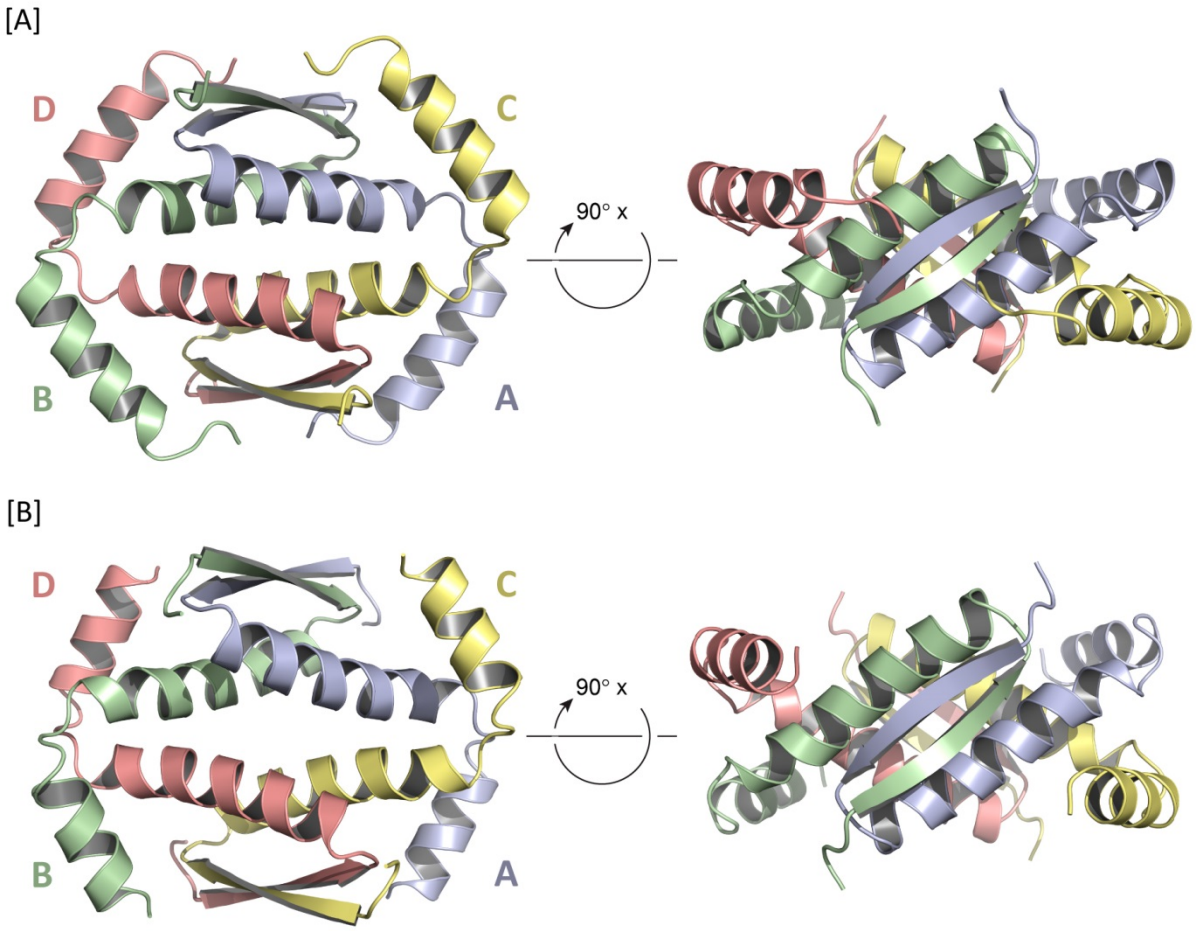


Figure 1-5 Comparison of the molecular structures of the tetramerization domains of human p73 and p63. [A] NMR solution structure of the tetrameriaztion domain (TD) of human p73 (pdb entry 2KBY) in comparison with [B] the crystal structure of the TD of human p63 (pdb entry 4A9Z).

1.1.7. p63 ensures the genomic integrity of the female germ line

TAp63α is highly expressed in undamaged oocytes and initiates oocyte death in response to DNA damage independently of p53.[107] Therefore, its destructive potential and its high protein level necessitate a tight control mechanism to prevent errant activation and loss of reproductive potential. This is achieved by TAp63α being stored in an inactive compact dimeric structure in which the TA and TI interact with one another and at the same time block the tetramerization interface of the TD.[108] In response to DNA damage p63 gets phosphorylated and this in turn results in the breakup of the autoinhibitory conformation. TA and TI detach from the TD and the protein forms tetramers with a ~20-fold higher DNA affinity. Once the lock is released the formation of the active tetramer is irreversible and it was shown *in vivo* that dephosphorylation does not convert the protein back to dimers. The

fact that p63 isoforms with shortened N- or C-termini are constitutive open tetramers demonstrates that TA and TI are both necessary for the formation of the closed, dimeric conformation. It was further shown *in vitro* with the isolated p63 TD and peptides from the TA that the TA binds onto the core TD, composed of the β-sheets and the first helices, thereby competing with the second helices of the TD for the same binding site.[108] Therefore, the structural motif of the second helices of the TD seems to have two functions at once, first to increase the stability of the tetrameric TD and second to block the interface to which the TA would bind within the autoinhibitory dimeric complex.

Despite the high degree of sequence identity this internal inhibitory mechanism could not be confirmed for p73, and recently it was shown that *in vivo* TAp73α is a constitutive open tetramer.[109] This implies that the regulation of p73's transcriptional activity might be more closely related to p53 than to p63.

1.1.8. Guardian of the genome – the ancestral function of the p53 family

The emergence of the p53 superfamily of TFs seems to predate the evolutionary origin of metazoans and at least one p53 superfamily member is present in all invertebrate species. This indicates that these TFs play fundamental roles in animals. Detailed functional analysis of the functions of the single p53 proteins of the two important model organisms, *Caenorhabditis elegans* (*C. elegans*; p53 homolog: Cep-1) and *Drosophila melanogaster* (*D. melanogaster*; p53 homolog: Dmp53) have revealed that they are required for the activation of apoptosis of germ cells in response to DNA damage.[110–112] Invertebrates, such as *C. elegans* have a short life span, do not possess renewable tissues and are therefore not threatened by tumors. But they have to ensure the genetic stability of their germ cells and have to eliminate those with damaged DNA. It is therefore nowadays believed that securing the genetic integrity of germ cells has been the ancestral function of the p53-like protein, and hence that in vertebrates p63 is the most ancient form.[107] Later in evolution the invention of renewable tissues allowed for an average life-time exceeding those of a cell but at the same making constant proliferation necessary. Consequently a more general surveillance system for somatic cells became mandatory and p53 appeared as a tumor suppressor in somatic cells.

1.2. Aim of study

Deciphering the functional history and evolutionary origins of the p53 protein family combined with detailed structural investigations could contribute to a better understanding of the present-day functions in vertebrates. And especially further insights into the network of regulatory as well as molecular interactions, in which the members of this protein family are engaged in, will have significant implications for our understanding of tumorigenesis, potentially offering new ways for cancer treatment.

The formation of oligomers via their oligomerization domains (ODs) is a common feature of all members of the protein family of p53-like transcription factors (TFs). This oligomerization allows for cooperative, sequence-specific binding of the DNA binding domains (DBDs) to DNA responsive elements and is a prerequisite for the transactivation of target genes. Therefore the activity of p53-like TFs strongly depends on their oligomeric state and in the consequence the molecular structure of the OD has for long attracted much attention.

1.2.1. Superordinate research objective

In comparison to the DBD the protein sequence of the OD is much less conserved revealing that its architecture has undergone substantial changes during evolution. Elucidating their kind and sequence is however hampered by the fact that the protein sequences from direct ancestors are naturally never obtainable. It is therefore believed that investigating the structures of the ODs of p53-like proteins from invertebrate species is the best way to gain insights into the evolutionary history of the OD of vertebrate p53-like TFs.

1.2.2. Previous investigations

Caenorhabditis elegans (*C. elegans*) and *Drosophila melanogaster* (*D. melanogaster*) are the two most popular model organisms from the superphylum Protostomia and are equally far distantly related to man; Figure 1-7. Both species respectively possess only a single protein belonging to the super family of p53-like TFs. And in a previous study from our group the NMR solution structures of the isolated ODs of the p53-like proteins from *C.elegans* (p53 homolog: Cep-1) and *D. melanogaster* (p53 homolog: Dmp53) were solved; Figure 1-6. [113]

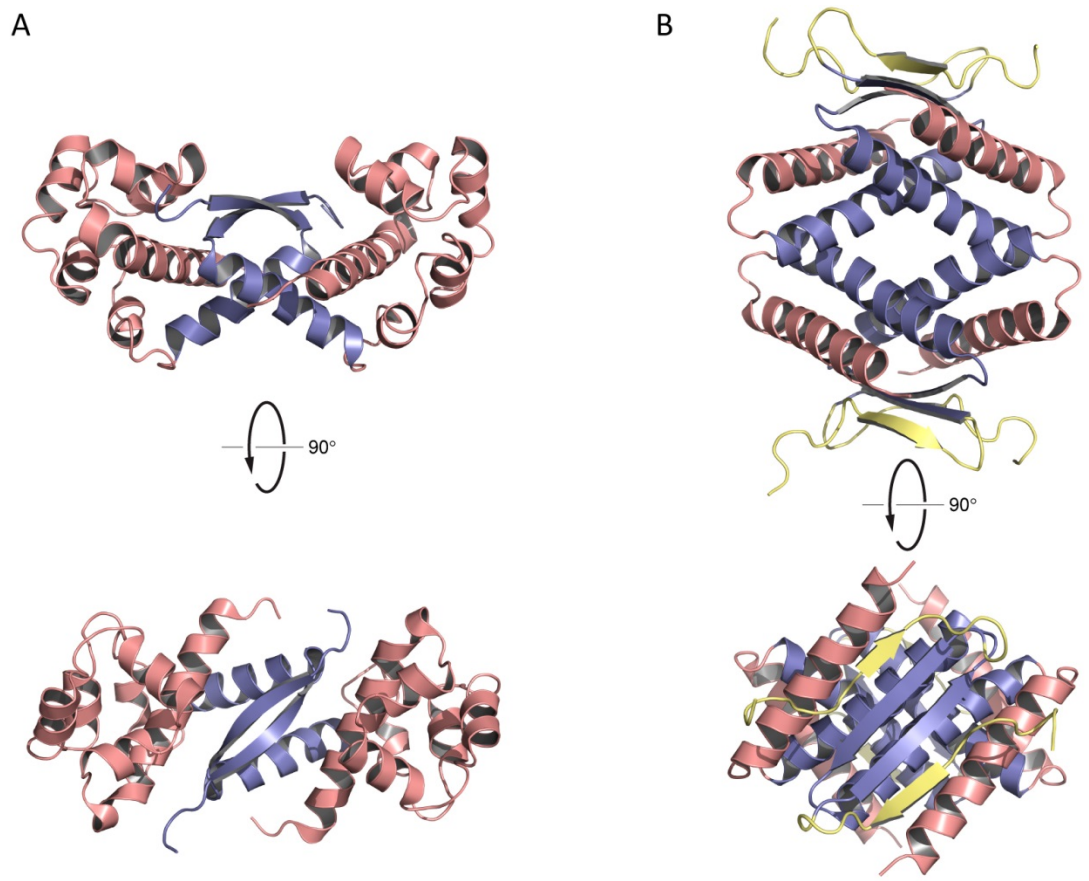


Figure 1-6 Comparison of the oligomerization domains of Cep-1 and Dmp53. (A) NMR solution structure of the oligomerization domain (OD) of Cep-1 (pdb entry 2RP5). (B) NMR solution structure of the OD of Dmp53 (pdb entry 2RP4). Both structures contain a motif, which closely resembles the primary dimer of human p53. These motifs depicted in blue are composed of the most central located α-helices from each monomer and the β-strands, respectively, immediately preceding these in the primary sequence. Additional α-helices with no equivalent in the ODs of human p53 or human p73 are depicted in red and additional β-strands are depicted in yellow.

Both ODs differ substantially from the architecture of any of the vertebrate p53 family proteins. The OD of Cep-1 is the only one known so far, which is forming only dimers. The N-terminal sections of the two monomers form a β-sheet and antiparallel packing α-helices, which together closely resemble the structure of the primary dimer in the OD of human p53. But each monomer additionally contributes a SAM domain folded from its C-terminal section. And these SAM domains were found to be an integral part of the OD of Cep-1 and essential for its stability. The interface, through which the first helices form inter-dimer contacts in human p53, is surface exposed but charged residues at crucial positions prevent the formation tetramers. The OD of Dmp53 forms tetramers as the ODs of the vertebrate p53-like proteins, but in comparison to the OD of human p53 each SSE is doubled. In the monomer two β-strands are followed by two α-helices. The respective second β-strands and first α-helices of two monomers form the core of the primary dimer, which closely resembles the primary dimer in the OD of human p53. In each monomer the first β-strand parallels the

second β-strand on the outer side resulting in a four bladed antiparallel β-sheet. The second helix is packing antiparallel against the first helix of the same monomer, is not wrapping around the neighboring dimer and in consequence the residues within the second helix barely interact with those from the other primary dimer. Taken together, the motif corresponding to the primary dimer within the OD of human p53 is conserved in the ODs of Cep-1 and Dmp53, but the overall topologies of the ODs differ fundamentally from those of human p53, p63 and p73. Furthermore, the additional SSEs, which in the two proteins respectively stabilize the primary dimer, are of different types, fold from different sections of the protein sequence and adopt different relative positions.

On the basis of these structural investigations two theories with respect to the evolution of the OD were formulated. The first one suggested that the stabilization by additional SSEs might have been a common feature of the ODs of ancestral forms of p53-like TFs. And resulting from the observation that Cep-1 forms only dimers the second theory asked whether the formation of higher order oligomers was only once or several times innovated during evolution.

1.2.3. Objective of this thesis

The fundamental differences in the topologies of the ODs of Cep-1 and Dmp-53 had so far prevented detailed elucidation of the developmental path at which end the more compact ODs of the orthologous human proteins stand. Therefore it was chosen as the objective of this thesis to investigate the domain architecture of the ODs of the p53-like TFs from an invertebrate species, being much more closely related to vertebrates than *C. elegans* or *D. melanogaster*. Since a lower number of p53-like proteins compared to vertebrates would be advantageous, the species should preferentially belong to a phylum, which diverged from vertebrates prior to the extensive genome duplications during their evolution. Other criteria have been that the genome had to be sequenced and completely annotated and that all p53-like proteins had to be validated at least on transcript level. All these criteria were matched by the tunicate species *Ciona intestinalis* (*C.int.*), which is a popular model organism in embryology and developmental biology.

1.3. Tunicates are the closest living relatives of vertebrates

More than 550 million years ago, at the time of the Cambrian explosion, the three major deuterostome phyla (echinoderms, hemichordates and chordates) arose from a common ancestor.[114–116] Subsequently the chordates diverged into three subphyla: vertebrates, chephalochordates and tunicates (also known as urochordates). The denotations tunicates and urochordates, both widely used in the literature, designate the same subphylum and are exchangeable against each other.

In 1866 Alexander Kowalevsky, a Russian embryologist, informed Charles Darwin about the extensive developmental and morphological similarities he had observed between vertebrates and the larval stage of ascidiacea, a class of tunicates. This prompted Darwin to hypothesize that ascidians might be useful to understand the origin of the vertebrate phylum as both lineages must have a common ancestry.[117]

The urochordate lineage was long generally believed to be the most basal among chordates, and the cephalochordates were considered to be the closest sister group to vertebrates. This was mainly justified by overall morphological similarities and the with respect to ascidians less primitive appearing body of adult cephalochordates. But recently this relationship was disproven. Based on the phylogenetic analysis of a dataset of nuclear genes Delsuc et al. [118] have shown that, indeed, tunicates and not cephalochordates are the closest living relatives of vertebrates urging for the taxonomy tree to be updated as shown in Figure 1-7.

The planktonic larval stage of tunicates shares prominent structural characteristics typical for chordates: a dorsal neural tube, a notochord, segmented musculature and gill slits;[117,118] Figure 1-8. This makes their tadpole larva a plausible approximation of ancestral chordates.

Species from the class ascidiacea within the subphylum tunicata have a long history as model organisms and are among the most suitable animals to investigate the genetic control mechanisms regulating development in early embryogenesis.

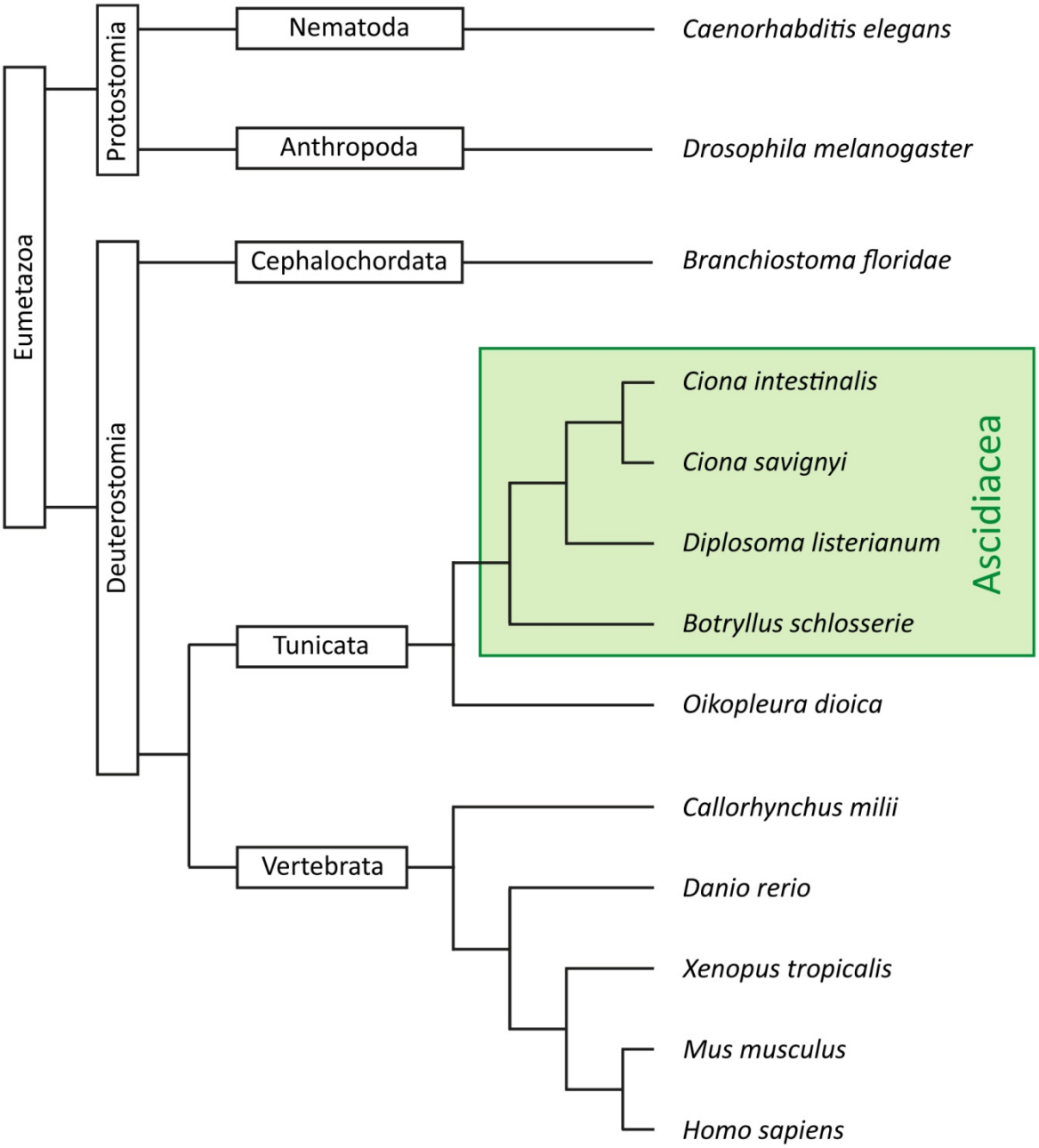


Figure 1-7 Taxonomy tree of the subkingdom eumetazoa within the animal kingdom with some sample species from different subphyla. Subkingdom (Eumetazoa) and superphyla (Deuterostomia, Protostomia) are written into boxes with vertical orientation. The subphyla are written into boxes with horizontal orientation. The taxonomic class of the ascidians (Ascidiacea) within the subphylum tunicata is emphasized with a green box. Species names are listed with italic lettering.

1.3.1. The ascidians *Ciona intestinalis* and *Ciona savignyi*

Among modern urochordates, or respectively, tunicates the ascidians are the most prevalent. Ascidians are familiarly known as sea squirts and are sessile, marine filter feeders that live in shallow ocean waters all around the world. As shown in Figure 1-8 the adult body

is encased in a thick, fibrous tunic, from which the name of the subphylum, tunicata, was derived.

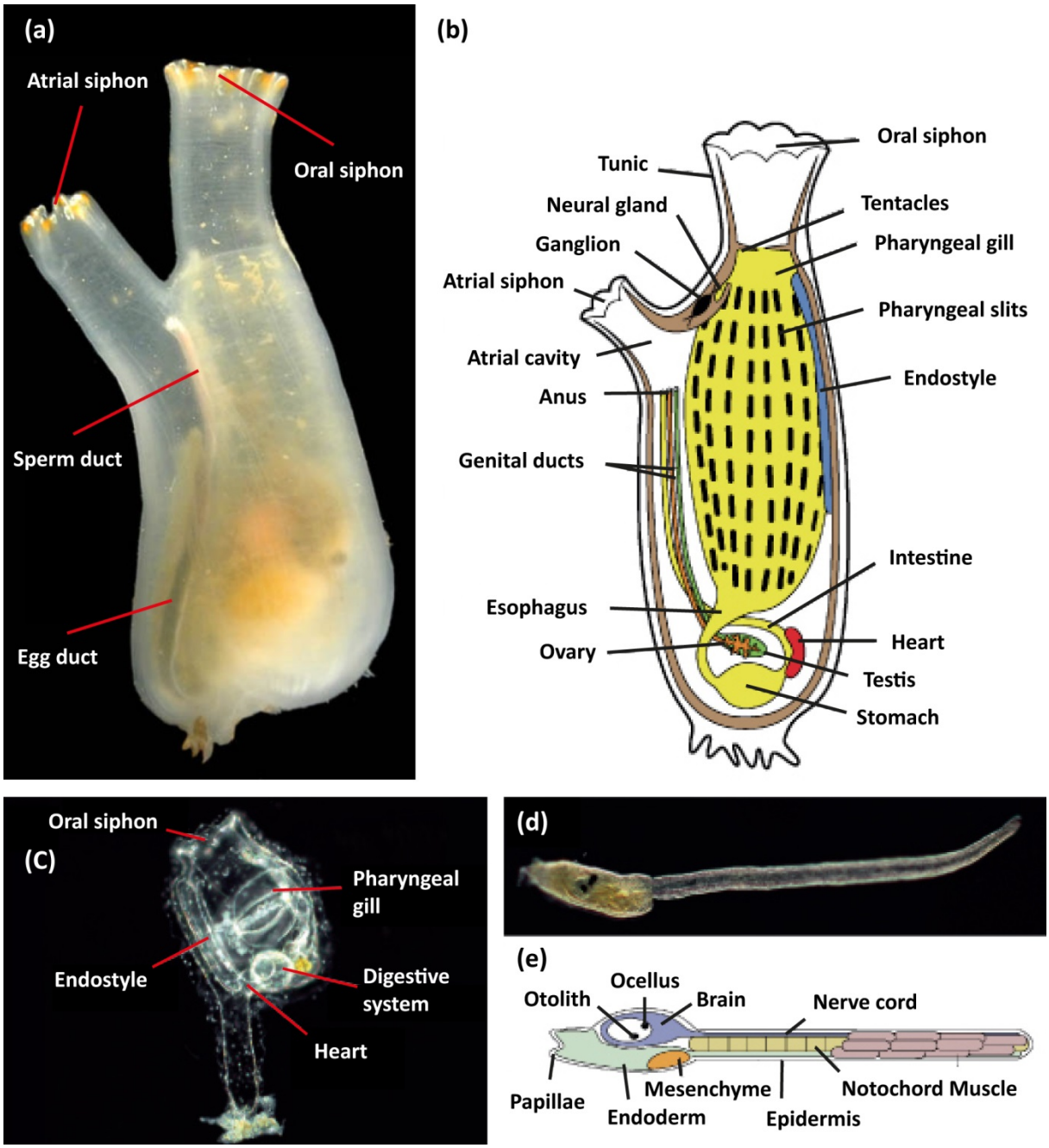


Figure 1-8 The body plan of the ascidian *Ciona intestinalis*. (a) An adult with incurrent (oral) and outcurrent (atrial) siphons. The sperm duct appears white and the egg duct, which is located parallel to it, appears orange. (b) Graphical illustration of the adult body, with organs and tissues labeled. (c) A juvenile a few days after metamorphosis, with most important organs labeled. (d) Tadpole larva before metamorphosis. (e) Graphical illustration of the tadpole body plan, with organs and tissues labeled. | Graphical illustrations (b) and (e) and pictures (c) and (d) as well as the text of the captions were taken over from the book Developmental Genomics of Ascidians by Noriyuki Satoh, 2013.[119]

This covering outer structure supports an incurrent oral siphon (branchial) and an outcurrent atrial siphon. Particles of plankton and other nutrients are passed through the

mouth into a large pharynx. Numerous gill slits along the dorsoventral side of this chamber function like a fine filter, and at the ventral margin a specialized organ, the endostyle, secretes large quantities of mucus for the capture of food particles. The digestive tract is build up by a stomach and an intestine, which terminates at the anus. The dejection is released into the outgoing current through the atrial siphon.

Embryogenesis is rapid and from fertilization to the free-swimming tadpole it takes only about 18 hours at 18°C. The larva consists of about 2600 cells, which form a small number of organs including the characteristic notochord. In response to light the larvae metamorphose into sessile individuals and lose most of their chordate phenotypes, but maintain a simple tube-like heart, an endostyle, and a neural complex; Figure 1-8.[117,118] The entire life cycle takes less than 3 months, facilitating experimental work.[120] As early as 1905, Edwin G. Conklin described the lineage-specificity of embryonic cells and suggested that maternal factors regulate their later differentiation. This was finally proven in 1973 by Richard Whittaker and as a result ascidians became a prominent experimental model system for molecular developmental biology, particularly with respect to investigations of the mechanisms underlying embryonic cell fate specification.

In 2002 *Ciona intestinalis* (*C.int.*) was the seventh animal whose genome was sequenced.[121] Among experimentally accessible chordates the genome of *C.int.* is one of the smallest. With 160 million base pairs it is about 20 times smaller compared to the human genome and contains an estimated 15,852 protein-coding genes.[122,121] This is similar to the number in other invertebrates, but only half the number found in vertebrates. After the divergence of the tunicate and vertebrate lineage, numerous gene duplications have occurred during evolution of the vertebrate lineage.[123–126] The extensive gain in genome size and gene number is often attributed to one or two rounds of genome-wide duplications. However, this polyploidization theory is now under debate for about 40 years and final evidence is still missing.[127] Instead several studies carried out on large-scale phylogenetic datasets have revealed that even the duplication of some exemplary complete gene clusters can better be explained with an en bloc mode of duplication of comparably small scale.[128,129]

Ascidians can be considered to have retained a basic, non-duplicated set of an ancestral chordate-type genome, and often a family of paralogous genes in vertebrates is represented by only a single gene in *C.int.*. However, certain single genes and as well complete clusters have also been subject of duplications.[130,131] Furthermore, in adaptation to their ecological

niche and probably as a result of their simplified life style tunicates lost ancestral genes, as well.[121,132] And in comparison to vertebrates the loss of entire ancestral gene families or single paralogous genes from multi-gene families occurred much more frequently in *C.int.*. Based on an estimated consensus ancestral genome it was calculated that *C.int.* lost 44% more ancestral gene families than human, and that of ancient gene duplicates, which arose prior to the last common ancestor of chordates and tunicates, losing one of the duplicates was four times as likely in *C.int.* than in vertebrates.[132] Overall about 77% of human genes can be traced back to protochordates with reasonable homology (e-value ≤ e^{-10}) which indicates that the common ancestral genome of tunicates and vertebrates had a gene repertoire with at least 77% homology to the human genome.[133]

The extent to which modern ascidians represent the last common ancestor of urochordates, cephalochordates and vertebrates in terms of gene function depends partly on how unambigously single *Ciona* genes can be assigned to individual genes or gene families in vertebrates. The genome of *C.int.* encodes a basic set of the developmentally relevant genes that are used by vertebrates, including TFs.[134] And *in silico* analysis revealed an unexpectedly high number of apoptosis-related genes. *C.int.* is equipped with as many caspases as higher vertebrates including mammals and interestingly, as well with as many anti-apoptotic proteins such as IAPs.

Notably it was proven that the cell death machinery controlled by Bax and Bcl-XL is conserved in *C.int.*.[135] This is of special importance since in vertebrates expression of the pro-apoptotic Bcl-2 protein Bax is induced by p53. An excess of Bax over its inhibitor Bcl-XL leads to the permeabilization of the outer mitochondrial membrane resulting in the release of cytochrome-C and the activation of the initiator caspase-9 by the apoptosome.

1.3.2. Proteins from the p53 family of transcription factors in *Ciona intestinalis*

Until today only a small number of genomes from tunicate species have been sequenced. For many of those with available genomic data a sufficient annotation is still missing. And most gene models have no evidence at the transcript level due to a lacking cDNA library. The only exception is the genome of *C.int.*, which is by far the best annotated genome for any tunicate and the only one with gene models for every gene. [121,136–138] And since *C.int.* is a popular model organism for many genes, espressed sequence tags (ESTs) are available as

well. The genome of *C.int.* and also the genome of the very closely related species *Ciona savignyi* (*C.sav.*) contains two genes encoding proteins belonging to the family of p53-like TFs. Phylogenetic analysis revealed that they resulted from a duplication event independent of the duplications, which lead to the p53 family in vertebrates.[139] Consequently, they form an independent monophyletic group and this makes the vertebrate p53, p63 and p73 proteins being co-orthologous to any of the two *Ciona* homologs, which are themselves paralogous to one another.[139] It is important to note that as a consequence of this group wise homology none of the vertebrate p53 family proteins can be traced back to a single *Ciona* protein. According to this relationship, the *Ciona* proteins were named p53/p73-a and p53/p73-b. The observation that the gene encoding *C.int.* p53/p73-b lacks any introns suggests retro-transposition as the mechanism most likely associated with the lineage-specific duplication. While in consequence only a single *C.int.* p53/p73-b protein is expressed, several different isoforms of *C.int.* p53/p73-a have been identified on transcript level. These are presented in Figure 1-9.

Until today only a single publication explicitly focuses on the function of *C.int.* p53/p73-a and –b. In his work T. Noda has shown that in eggs and early embryos both proteins are ubiquitously expressed from maternal mRNA, and that during embryogenesis they regulate the differentiation of the notochord.[140] It was shown that the knock down mediated by antisense morpholinos against either of the genes as well as the combined knock down results in morphological defects after the 64-cell stage. In the affected embryos the gastrulation movement during the gastrula stage does not occur, and in consequence the blastopore does not close. Furthermore, in the majority of the embryos expression of the notochord marker genes *Noto1* and *talin* is diminished, putting the suggestion close that *C.int.* p53/p73-a and *C.int.* p53/p73-b are both involved in the differentiation of the notochord. The TF Brachyury is a key regulator of notochord differentiation in *Ciona* embryos,[141] and is directly activated by ZicL.[142] Interestingly it was shown that in response to the knockdown of *C.int.* p53/p73-a or –b the expression of ZicL as well as Brachyury was down regulated. And since the region upstream of the gene encoding the *C.int.* ZicL homolog contains a putative p53-binding site this suggests that *C.int.* p53/p73-a and –b regulate the differentiation of the notochord in a ZicL dependent manner.

This observation is in agreement with many different studies, which have shown that in vertebrate species each of the three p53 family members is indispensable for certain

developmental processes.[143 (review)] For example, it was shown that during embryogenesis p53 contributes to the closure of the neural tube in mice and mesoderm specification in frogs.[144–146]

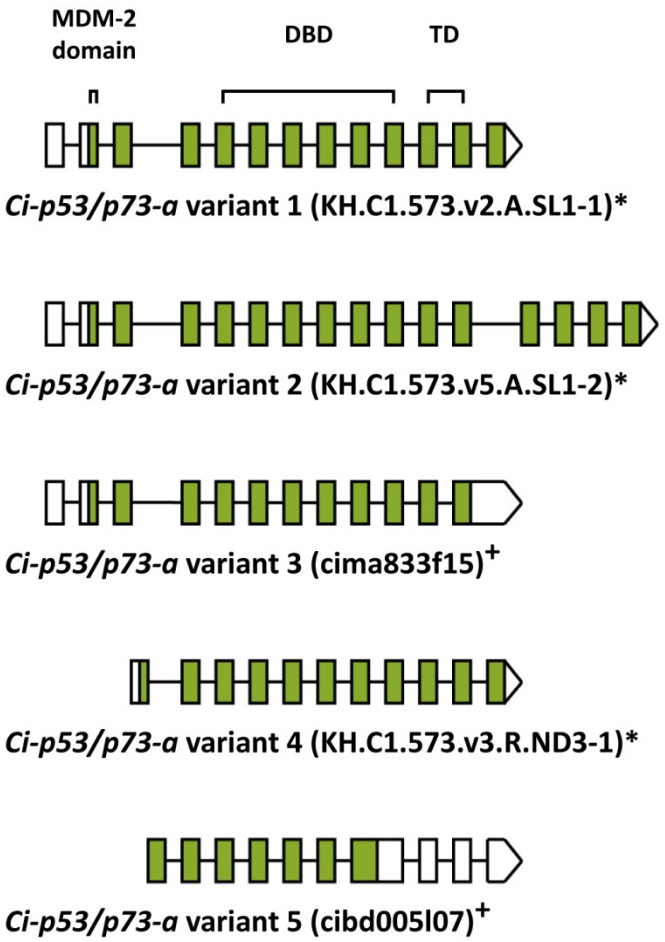


Figure 1-9 The splice variants of *Ciona intestinalis* p53/p73-a. These five splice variants of the gene were detected as mRNA by T. Noda in eggs and early embryos.[140] Untranslated regions are marked in white and the coding regions in green. The identifications of the transcripts are in parenthesis. The figure was captured with changes from T. Noda.[140]
(*) transcript model for gene KH2012:KH.C1.573 as annotated by the aniseed database. [137,147]
(+) corresponding cDNA sequence was submitted to the NCBI gene bank by the *Ciona intestinalis* cDNA project at the university of Kyoto, japan and referenced to Satou et al. 2002, [136]

2. Materials

2.1. Laboratory Equipment

Agar plate incubator	Memmert GmbH & Co. KG, Germany
ÄKTAprime plus FPLC system	GE Healthcare Europe GmbH, Germany
ÄKTApurifier FPLC system	GE Healthcare Europe GmbH, Germany
Analytical balance CPA124S-OCE	Sartorius AG, Germany
Analytical balance PB3002 DeltaRange	Mettler-Toledo GmbH, Germany
Autoclave GE 446 EC-1	Getinge AB, Sweden
Centrifuge Centricon H-401	Kontron-Hermle, Germany
Centrifuge Megafuge 16R	Heraeus, Germany
Centrifuge rotor A 6.9	Kontron-Hermle, Germany
Centrifuge rotor F34-6-38	Eppendorf AG, Germany
Centrifuge rotor GS-3	Sorvall Instruments, Germany
Centrifuge rotor GSA	Sorvall Instruments, Germany
Centrifuge rotor SS-34	Sorvall Instruments, Germany
Centrifuge rotor TX-400	Heraeus, Germany
Centrifuge Sorvall Evolution RC	Sorvall Instruments, Germany
Centrifuge Sorvall RC-5B	Sorvall Instruments, Germany
Centrifuge Sorvall RC-5C	Sorvall Instruments, Germany
Digital camera Powershot G3	Canon GmbH, Germany
Heating block	VWR International GmbH, Germany
Incubation shaker Innova 4330	New Brunswick Scientific GmbH, Germany
Incubation shaker Multitron	Infors-HT AG, Suisse
Magnetic Stirrer RTC basic	IKA-Werke GmbH & Co. KG, Germany
Milli-Q Academic ultrapure water system	Millipore GmbH, Germany
Mini-Protean Tetra Cell SDS-PAGE system	Bio-Rad Laboratories GmbH, Germany
NanoDrop 1000 UV-/Vis Spectrophotometer	PEQLAB Biotechnology GmbH, Germany
NMR spectometers with room temperature probes at proton frequencies: 499 MHz, 500 MHz	Bruker, Germany

NMR spectrometers with cryo probes at proton Frequencies: 599 MHz, 600 MHz, 700 MHz, 800 MHz, 900 MHz, 950 Mhz	Bruker, Germany
Peristaltic pump with VP drive	ISMATEC, Switzerland
PCR Cycler TPersonal	Biometra GmbH, Germany
PCR Gradient cycler peqSTAR	PEQLAB Biotechnology GmbH, Germany
pH-meter PHM210	Radiometer-Analytical SAS, France
Pipettes Research and Research Plus	Eppendorf AG, Germany
Pipetus, pipette (electr.) for large volumes	Hirschmann Laborgeräte, Germany
Power supply EPS 300	Pharmacia Biotech, Germany
Power supply Power PAC 3000	Bio-Rad Laboratories GmbH, Germany
Power supply Power Pack P25 T	Biometra GmbH, Germany
Refrigerated centrifuge 5810 R	Eppendorf AG, Germany
Refrigerated table top centrifuge Micro 22R	Andreas Hettich GmbH & Co. KG, Germany
Shaker unit Promax 2020	Heidolph Instr. GmbH & Co. KG, Germany
Spectropolarimeter J-810	Jasco Labortechnik, Germany
Sonifier Labsonic U	B. Braun Biotech International, Germany
Table top centrifuge 5415 D	Eppendorf AG, Germany
Table top centrifuge Biofuge 13	Heraeus, Germany
UV-/Vis-spetrometer U-1100	Hitachi Europe GmbH, Germany
Vortex MS2 Minishaker	IKA-Werke GmbH & Co. KG, Germany
Water bath TW12	JULABO Labortechnik GmbH, Germany

2.2. Chromatography

Material/ Column	Vendor / Producer
IMAC Sepharose 6 fast flow	GE Healthcare Europe GmbH, Germany
Glutathion Sepharose 4 Fast Flow	GE Healthcare Europe GmbH, Germany
OmniFit 25/100 column	msscientific, Germany
HiLoad 1660 Superdex 200	GE Healthcare Europe GmbH, Germany
HiLoad 1660 Superdex 75	GE Healthcare Europe GmbH, Germany
HiTrap IMAC FF 5 ml prepacked column	GE Healthcare Europe GmbH, Germany
Superdex 200 10/300 column	GE Healthcare Europe GmbH, Germany
Superdex 75 10/300 column	GE Healthcare Europe GmbH, Germany

2.3. Membranes for ultrafiltration and dialysis

Name	Vendor / Producer
Spectra/Por® 6 dialysis membranes (diff. MWCO)	Spectrum Europe B. V., Netherlands
Slide-A-Lyzer Dialysis Cassettes (diff. MWCO)	Pierce / Thermo Scientific, Germany
D-tube Dialyzer Midi, MWCO 3,5 and 6-8 kDa	Merck KGaA, Germany
Ultrafiltration disk membrane type YM1, MWCO 1000	Millipore GmbH, Germany
Ultrafiltration disk membrane type PLBC, MWCO 3000	Millipore GmbH, Germany
Ultrafiltration disk membrane type YM10, MWCO 10000	Millipore GmbH, Germany
Centriprep YM-3 / 10 and 30 concentrators	Millipore GmbH, Germany

2.4. Kits

Name	Producer
Gel Filtration Calibration Kits for low and high molecular weight range	GE Healthcare Europe GmbH, Germany
NucleoSpin Plasmid	Macherey-Nagel GmbH & Co. KG, Germany
NucleoBond PC 100 Midi Kit	Macherey-Nagel GmbH & Co. KG, Germany
Protease Inhibitor Cocktail Kit (Inhibitor and EDTA)	Pierce / Thermo Scientific, Germany
QIAquick PCR Purification Kit	Qiagen GmbH, Germany

2.5. DNA Ladders

Name	Vendor / Producer
GeneRuler 1 kb DNA Ladder	Thermo Scientific, Germany
GeneRuler 100 bp Plus DNA Ladder	Thermo Scientific, Germany
GeneRuler Low Range DNA Ladder	Thermo Scientific, Germany

2.6. Protein standards

Name	Vendor / Producer
Protein Marker Broad Range (2-212 kDa)	New England Biolabs GmbH, Germany
Precision Plus Unstained Standards (10-250 kDa)	New England Biolabs GmbH, Germany

2.7. Enzymes

Name	Vendor / Producer
Antarctic Phosphatase	New England Biolabs GmbH, Germany
BamHI-HF Restriction Endonuclease	New England Biolabs GmbH, Germany
DNase I from bovine pancreas type IV	Sigma-Aldrich Chemie GmbH, Germany
DpnI Endonuclease	New England Biolabs GmbH, Germany
Lysozyme from chicken egg white	Sigma-Aldrich Chemie GmbH, Germany
Phusion DNA Polymerase	Finnzymes, Finland
PfuTurbo DNA Polymerase	Stratagene, Agilent Technologies, Germany
PfuUltra High Fidelity DNA Polymerase	Stratagene, Agilent Technologies, Germany
Ribonuclease A from bovine pancreas type I-AS	Sigma-Aldrich Chemie GmbH, Germany
T4 DNA Ligase	New England Biolabs GmbH, Germany
TEV Protease	self-produced
XhoI Restriction Endonuclease	New England Biolabs GmbH, Germany

2.8. Plasmids

Name (Vendor / Source)	Features
pBH4 (Gift from Wendell Lim Laborat., UCSF)	pET15 derivative;, N-terminal 6xHis-tag followed by a TEV cleavage site; procures resistance towards ampicillin; expression induced by IPTG; T7 promoter origin of replication: pBR322 / ColE1
pRARE (co-plasmid from Rosetta™ strains; Novagen)	pACYC derivative; for supplementation with tRNAs for codons with low abundance in *E.coli* mRNA; procures resistance towards chloramphenicol; transcription induced by IPTG; origin of replication: p15A; compatible with pET vectors

2.9. Bacterial Strains

Strain name (Vendor)	Features
NEB 5-alpha competent *E. coli* (New England Biolabs, Germany)	K12 strain, DH5α™ derivative genotype: *fhuA2 Δ(argF-lacZ)U169 phoA* glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17
NEB T7 Express Competent *E.coli* (New England Biolabs, Germany)	B strain, BL21 derivative, T1 phage resistant deficient in proteases Lon and OmpT genotype: fhuA2 lacZ::T7 gene1 [lon] ompT gal sulA11 R(mcr-73::miniTn10--TetS)2 [dcm] R(zgb-210::Tn10--TetS) endA1 Δ(mcrC-mrr)114::IS10
NEB T7 Express Competent *E.coli* + pRARE plasmid	all characteristics from NEB T7 Express *E.coli* in addition chloramphenicol resistant

2.10. Reagents

Name	Vendor / Producer
Acetic acid, ROTIPURAN 100% p. a.	Carl-Roth GmbH & Co. KG, Germany
Agar-Agar- Kobe I	Carl-Roth GmbH & Co. KG, Germany
Agarose LE	Biozym Scientific GmbH, Germany
Ammonium molybdate tetrahydrate	Sigma-Aldrich Chemie GmbH, Germany
Ammoniumperoxid sulfate (APS)	Sigma-Aldrich Chemie GmbH, Germany
Ammonium chloride, NH_4Cl	Carl-Roth GmbH & Co. KG, Germany
^{15}N-NH_4Cl, isotopically enriched	Cambridge Isotope Laboratories, Inc., UK
Ampicillin sodium salt	Carl-Roth GmbH & Co. KG, Germany
Antifoam Y-30 emulsion	Sigma-Aldrich Chemie GmbH, Germany
Boric acid	Sigma-Aldrich Chemie GmbH, Germany
Bovine serum albumin 10 mg/mL solution	New England Biolabs GmbH, Germany
Bovine serum albumin Fraction V	Sigma-Aldrich Chemie GmbH, Germany
Chloramphenicol	Carl-Roth GmbH & Co. KG, Germany
Choline chloride	Sigma-Aldrich Chemie GmbH, Germany
Cobalt(II) chloride	Sigma-Aldrich Chemie GmbH, Germany
Complete Protease Inhibitor Cocktail (+/- EDTA)	Roche Diagnostics GmbH, Germany

Coomassie brilliant blue G250	Carl-Roth GmbH & Co. KG, Germany
Copper(II) sulfate	Sigma-Aldrich Chemie GmbH, Germany
Crystal violet	Carl-Roth GmbH & Co. KG, Germany
Dimethyl sulfoxide	Carl-Roth GmbH & Co. KG, Germany
1,4-Dithiotreitol (DTT)	Carl-Roth GmbH & Co. KG, Germany
DSS (4,4-dimethyl-4-silapentane-1-sulphonate)	Sigma-Aldrich Chemie GmbH, Germany
Ethanol	Carl-Roth GmbH & Co. KG, Germany
Ethidiumbromide	Carl-Roth GmbH & Co. KG, Germany
Ethylenediaminetetraacetic acid (EDTA)	Carl-Roth GmbH & Co. KG, Germany
Folic acid	Sigma-Aldrich Chemie GmbH, Germany
Glucose monohydrate	Carl-Roth GmbH & Co. KG, Germany
U-$^{13}C_6$-Glucose	Cambridge Isotope Laboratories, Inc., UK
Glycerol, 99% p. a.	Carl-Roth GmbH & Co. KG, Germany
HEPES	Carl-Roth GmbH & Co. KG, Germany
Hydrochloric acid, 37%	Carl-Roth GmbH & Co. KG, Germany
Hydrogen peroxide, 30%	Carl-Roth GmbH & Co. KG, Germany
Imidazole	Fluka Chemie GmbH, Germany
Isopropanol	Carl-Roth GmbH & Co. KG, Germany
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Carl-Roth GmbH & Co. KG, Germany
Kanamycinsulfat, ≥ 750 I.U./mg	Carl-Roth GmbH & Co. KG, Germany
Leupeptin hemisulfate salt	Sigma-Aldrich Chemie GmbH, Germany
Magnesium chloride	Carl-Roth GmbH & Co. KG, Germany
Manganese(II) chloride	Sigma-Aldrich Chemie GmbH, Germany
2-Mercaptoethanol (β-ME)	Carl-Roth GmbH & Co. KG, Germany
MES	Carl-Roth GmbH & Co. KG, Germany
Methanol	Carl-Roth GmbH & Co. KG, Germany
Myo-Inositol	Sigma-Aldrich Chemie GmbH, Germany
Nickel(II) sulfate hexahydrate	Carl-Roth GmbH & Co. KG, Germany
Nicotinamide	Sigma-Aldrich Chemie GmbH, Germany
DL-Pantothenic acid hemicalcium salt	Sigma-Aldrich Chemie GmbH, Germany
Pepstatin A	Sigma-Aldrich Chemie GmbH, Germany
Peptone from casein, tryptic digest	Carl-Roth GmbH & Co. KG, Germany
Potassium chloride	Carl-Roth GmbH & Co. KG, Germany
Potassium phosphate dibasic	Carl-Roth GmbH & Co. KG, Germany

Potassium phosphate monobasic	Carl-Roth GmbH & Co. KG, Germany
Pyridoxal hydrochloride	Sigma-Aldrich Chemie GmbH, Germany
Riboflavin	Sigma-Aldrich Chemie GmbH, Germany
Rotiphorese Gel 30 (37.5:1)	Carl Roth GmbH & Co. KG, Germany
Rotiphorese Gel 40 (19:1)	Carl Roth GmbH & Co. KG, Germany
Sodium azide, NaN_3	Carl Roth GmbH & Co. KG, Germany
Sodium chloride	Carl-Roth GmbH & Co. KG, Germany
Sodium dodecyl sulfate (SDS), pellets	Carl-Roth GmbH & Co. KG, Germany
Sodium hydroxide	Carl-Roth GmbH & Co. KG, Germany
Sodium phosphate dibasic dihydrate	Carl-Roth GmbH & Co. KG, Germany
Sodium phosphate monobasic dihydrate	Carl-Roth GmbH & Co. KG, Germany
TCEP Bond-Breaker, 0.5 M solution	Thermo Scientific, Germany
N,N,N’,N’-Tetramethylethylenediamine	Carl-Roth GmbH & Co. KG, Germany
Thiamine hydrochloride	Sigma-Aldrich Chemie GmbH, Germany
Tris(hydroxymethyl)aminomethane	Carl-Roth GmbH & Co. KG, Germany
Trypsin/EDTA	Life Technologies GmbH, Germany
TWEEN 20	Sigma-Aldrich Chemie GmbH, Germany
Yeast extract	Carl-Roth GmbH & Co. KG, Germany
Zinc sulfate monohydrate	Sigma-Aldrich Chemie GmbH, Germany

2.11. Common buffers, media and reagents

2.11.1. Antibiotics

Ampicillin (stock, 1000x):	100 mg/ml Na+-ampicillin in 50% (v/v) EtOH
Kanamycin (stock, 1000x):	35 mg/ml kanamycin sulfate in H_2O
Chloramphenicol (stock, 1000x):	34 mg/ml in EtOH

2.11.2. Inducers (Induction of Expression)

IPTG (stock, 1000x):	1 M, 11.9 mg/50 ml in H_2O

2.11.3. Media for cultivation of bacteria

LB plates (per liter):	10 g peptone, 5 g yeast extract, 10 g NaCl, 15 g agar-agar Adjust pH to 7.2 – 7.4 (NaOH) and sterilize by autoclaving.
LB medium (per liter):	10 g peptone, 5 g yeast extract, 10 g NaCl Adjust pH to 7.2 – 7.4 (NaOH) and sterilize by autoclaving.
2XYT (per liter):	16 g peptone, 10 g yeast extract, 5 g NaCl Adjust pH to 7.2 – 7.4 (NaOH) and sterilize by autoclaving.
SOC medium (per liter):	20 g peptone, 5 g yeast extract, 0.5 g NaCl Adjust pH to 7.2 – 7.4 (NaOH) and sterilize by autoclaving. Add 20 mM glucose solution, 25 mM KCl and 10 mM $MgCl_2$ and adjust volume with H_2O (all sterile filtered).
M9 medium (per liter):	6 g Na_2HPO_4, 3 g KH_2PO_4, 0.5 g NaCl Adjust pH to 7.0 – 7.4 (NaOH) and sterilize by autoclaving. Afterwards add (all sterile filtered): 2 ml 1 M $MgSO_4$ (use MgCl instead, if NH_4SO_4 is used as nitrogen source) 10 µl 1 M $CaCl_2$ 1 ml M9 vitamin mix 200 µl micronutrient mix 1 ml 10 mM $FeSO_4$ (fresh solution, 33 mg in 10 ml) 1 g NH_4Cl or 1 g ^{15}N-NH_4Cl 4 g glucose or 2 g ^{13}C-glucose

2.11.4. Supplement solutions for M9 medium

M9 vitamin mix (1000x, per 50 ml):
0.02 g choline chloride
0.025 g folic acid
0.025 g DL-pantothenic acid
0.025 g nicotinamide
0.05 g myo-inositol
0.025 g pyridoxal HCl
0.025 g thiamine HCl
0.0025 g Riboflavin

M9 Micronutrients (5000x):
$1.5 \cdot 10^{-5}$ M ammonium molybdate tetrahydrate, $(NH_4)_6Mo_7O_{24} \cdot 4H_2O$
2 mM boric acid, H_3BO_3
$1.5 \cdot 10^{-4}$ cobalt(II) chloride, $CoCl_2$
$5 \cdot 10^{-5}$ M copper(II) sulfate, $CuSO_4$
$4 \cdot 10^{-4}$ M manganese(II) chloride, $MnCl_2$
$5 \cdot 10^{-5}$ M zinc sulfate monohydrate, $ZnSO_4 \cdot H_2O$

2.11.5. Nuclease stock solutions for lysis of bacterial pellets

Deoxyribonuclease stock solution:
3000 Kunitz/ml DNAse I from bovine pancreas, type IV
20 mM Tris pH 8.0, 100 mM $MgSO_4$
store at -20°C

Ribonuclease stock solution:
500 Kunitz/ml RNAse A from bovine pancreas, type I-AS
20 mM Tris pH 8.0, 100 $MgSO_4$
store at – 20°C

2.11.6. General stocks for protein biochemistry

Protease inhibitor cocktail:

1.197 mg/ml AEBSF	[250 mM]
0.238 mg/ml leupeptin	[25 mM]
0.154 mg/ml bestatin	[25 mM]
0.098 mg/ml aprotinin	[750 µM]
0.089 mg/ml E-64	[12.5 mM]
0.034 mg/ml pepstatin	[2.5 mM]

Completely dissolved the ingredients in 50% v/v methanol at 4°C, evaporate solvent at RT under mild vacuum, and store at -20°C.
Alternatively one EDTA-free cOmplete™ protease inhibitor cocktail tablet can be dissolved in 1 ml MilliQ H2O and used as a 50x stock solution.

Sodium azide, NaN_3:
10 % (w/v) stock in H_2O

2.11.7. Buffers for DNA agarose gel electrophoresis

Agarose gel:	e.g. 1% (w/v) agarose boiled in 1x TAE buffer until completely dissolved. After cooling to ~ 60°C cast into chamber and wait until gel is getting solid.
Crystal violet stock solution (1000x):	10 mg/ml in milli-Q H_2O. Store at RT in the dark.
EtBr-stock solution (1000x):	10 mg/ml in milli-Q H_2O. Store at RT in the dark.
TAE buffer (50x stock solution):	Dissolve 2 M Tris, 1 M acetic acid and 50 mM EDTA in milli-Q H_2O. Adjust pH to 8.0. Store at RT.
6x DNA loading buffer (analytical):	10 mM Tris-HCl pH 7.6, 0.03% bromophenol blue, 0.03% xylene cyanol FF, 60 mM EDTA, 60% glycerol in milli-Q H_2O. Store at -20°C for long term.
6x orange DNA loading buffer:	10 mM Tris-HCl pH 7.6, 0.15% orange G, 0.03% xylene cyanol FF, 60 mM EDTA, 60% glycerol in milli-Q H_2O. Store at -20°C for long term.
4x DNA loading buffer (preparative)	0.2% crystal violet, 50% glycerol in milli-Q H_2O Store at RT.

2.11.8. Buffers for Tris-Tricine-SDS PAGE

reducing loading buffer:	150 mM Tris-HCl pH 7.0, 12% (w/v) SDS, 30% (w/v) glycerol, 0.05% (w/v) Coomassie brilliant blue G250 and 6% (v/v) β-mercaptoethanol in milli-Q H_2O. Store at -20°C
Tris-Tricine gel buffer (3x):	3 M Tris, 1 M HCl, pH 8.45 in milli-Q H_2O. Store at RT.
Anode buffer (10x):	1 M Tris, 0.225 M HCl, pH 8.9 in milli-Q H_2O. Store at RT.
Cathode buffer (10x):	1 M Tris, 1 M Tricine, 1% SDS (w/v), pH 8.25 in milli-Q H_2O. Store at RT.
APS stock solution:	10% (w/v) ammonium-peroxid sulfate in milli-Q H_2O. Store at 4°C.
Fixing solution:	50% Methanol, 10% Acetic acid, 100 mM ammonium acetate in milli-Q H_2O. Store at RT.
Gel staining solution:	10% Acetic acid, 0.025% Coomassie brilliant blue G250 in milli-Q H_2O. Store at RT in a dark bottle to avoid exposure to light.
Destaining solution:	10% Acetic acid in milli-Q H_2O. Store at RT.

2.11.9. Buffers for cell lysis

$MgSO_4$ stock solution:	1 M $MgSO_4$ in milli-Q H_2O. Store at RT.

2.11.10. Buffers for protein purification and analysis

Ni-NTA/IDA Stripping buffer:	500 mM NaCl, 100 mM EDTA, 20 mM Tris, pH 8.0 dissolve in milli-Q H_2O and degas. Store at RT.
Ni(II) charging solution:	100 mM NiSO4 dissolve in milli-Q H_2O and degas. Store at RT.
Lysis buffer (10 x):	4 M NaCl, 250 mM Tris, pH 7.8 dissolved in milli-Q H_2O. Store at 4°C.
IMAC buffer A1:	400 mM NaCl, 25 mM Tris, pH 7.8 dissolve in milli-Q H_2O, filter and degas. Store at 4°C.
IMAC buffer B1:	500 mM Imidazole, 400 mM NaCl, 25 mM Tris, pH 7.8 dissolve in milli-Q H_2O, filter and degase. Store at 4°C.
IMAC buffer A2 (reducing):	400 mM NaCl, 25 mM Tris, pH 7.8, 10 mM β-Mercaptoethanol, dissolve in milli-Q H_2O, filter and degas. Store at 4°C. Use for up to two weeks.
IMAC buffer B2 (reducing):	500 mM Imidazole, 400 mM NaCl, 25 mM Tris, pH 7.8, 10 mM β-Mercaptoethanol, dissolve in milli-Q H_2O, filter and degas. Store at 4°C. Use for up to two weeks.
CD (Circular dichroism) buffer:	Dissolve 50 mM NaH_2PO_4 in milli-Q H2O, adjust to pH 7.0 with NaOH, filter and degas. Store at RT.
D2O/DSS for NMR samples:	3 mM DSS, 0.02% (w/v) NaN_3 dissolved in D_2O. Store at 4°C.
RE buffer (NMR buffer, non-reducing):	50 mM L-arginine, 50 mM L-glutamate, dissolve in milli-Q H_2O, adjust to pH 7.0, filter and degas. Store at 4°C.
Reducing RE buffer (NMR buffer):	50 mM L-arginine, 50 mM L-glutamate, 1 mM DTT, dissolve in milli-Q H_2O, adjust to pH 7.0, filter and degas. Store at 4°C. Use for up to two months.

2.11.11. Buffers used for condition screening of *C.int.* p53/p73-a

Screening buffer 1:	50 mM NaPi, pH 7.0
Screening buffer 2:	50 mM NaPi, 100 mM KCl, pH 7.8
Screening buffer 3:	25 mM MES, 100 mM MES, 5.7
Screening buffer 4:	50 mM L-arginine, 50 mM L-glutamate, 50 mM NaCl, pH 7.0

2.12. Software

Adobe Creative Suite 5.5	Adobe Systems, Germany
Jalview	Barton Group, University of Dundee
MOLMOL 2K.2	Institute of Molecular Biology and Biophysik, ETH Zurich
MS Office 2010	Microsoft GmbH, Germany
pDRAW32	AcaClone, Denmark
PyMOL 1.6.0.0.	Schrödinger, LLC, USA
SPARKY 3.114	T. D. Goddard and D. G. Kneller, UCSF
Spectra Manager Version 2	JASCO Corporation
SSP (secondary structure propensity)	Joseph Marsh (ref.)
Swiss-PdbViewer 4.1.0 (DeepView)	Swiss Institute of Bioinformatics
Topspin	Bruker BioSpin GmbH, Germany
UNICORN 5.11	GE Healthcare Europe GmbH, Germany

2.13. Web server operated programs

ANISEED genome browser and blast engine	Ascidian Network for InSitu Expression and Embryological Data http://www.aniseed.cnrs.fr/
ClustalW2	multiple sequence alignment http://www.ebi.ac.uk/Tools/msa/clustalw2/
DIanna 1.1	software for cysteine state and disulfide bond partner prediction; developed by F. Ferre and P. Clote [148–150] clavius.bc.edu/~clotelab/DiANNA/
δ2D	server operated program for the determination of secondary structure populations based on NMR chemical shifts http://www-vendruscolo.ch.cam.ac.uk/d2D/
Genepyramid	genome browser at the Stanford University, USA genepyramid.stanford.edu/botryllusgenome/
graphical codon usage analyzer	Dr. Thomas Schödl, Universität Regensburg, Deutschland; http://gcua.schoedl.de/
PDBePISA	EMBL-European Bioinformatics Institute, England ebi.ac.uk/msd-srv/prot_int/pistart.html

primerX	automated design of mutagenic primers for site-directed mutagenesis http://www.bioinformatics.org/primerx/
PROPKA 3.0	prediction of protein pKa values and pH dependent stability based on pdb structure http://propka.ki.ku.dk/
ProtScale	calculation of relative hydrophobicity along a primary protein sequence http://web.expasy.org/protscale/
SPALN	Exon finding by genomic sequence comparison guided by automated mapping and alignment of user supplied cDNA or protein sequences. genome.ist.i.kyoto-u.ac.jp/~aln_user/index.html
SWISS-MODEL	SIB Swiss Institute of Bioinformatics, Switzerland http://swissmodel.expasy.org/
TALOS+ and viewer Java RAMA+	web server run by the workgroup of Ad Bax prediction of phi and psi angles based on secondary chemical shift data spin.niddk.nih.gov/bax/nmrserver/talos/ spin.niddk.nih.gov/bax/software/TALOS+/JRAMA+/

3. Methods

3.1. Methods for microbiology and the manipulation of plasmid DNA

3.1.1. Polymerase Chain Reaction (PCR)

Conventional PCR[151,152] was used for the amplification of specific DNA fragments. 20-50 ng template DNA, 5 µM of forward and reverse primer as well as 1 µl of pHusion DNA polymerase, 1 x polymerase buffer and 1-1.5 µl of 10 mM dNTP mix were used in a total volume of 100 µl. The melting temperature of the primers was calculated using pDraw32 and the temperature for annealing was set about 3°C below, generally between 54-68°C. Primers were ordered from Biospring (Biospring, Frankfurt). The temperature cycling protocol was set according to the manufacturer's instructions.

3.1.2. PCR Purification

PCR products and products of digestion with restriction endonucleases were purified using the QIAquick PCR purification kit according to the manufacturer's instructions. dsDNA fragments with a length smaller than 150 bp cannot be purified with this kit as they are not effectively immobilized by the cartridge. They were instead dialyzed against water in D-tube dialyzers with a MWCO of 6-8 kDa and afterwards dried in a speedvac.

3.1.3. Digestion of DNA with Restriction Endonucleases

DNA was digested with restriction enzymes from NEB in 20-60 µL reactions for 3-5 h at 37°C. Buffer conditions as well as enzyme concentrations were used according to the manufacturer's recommendations.

3.1.4. Ligation of DNA fragments

DNA fragments with sticky ends were ligated with T4-DNA-ligase, which was purchased from NEB. Reactions were performed in a total volume of 10 µl containing 1 µl 10 x T4-ligase buffers, 1 µl T4-ligase, 10 ng digested plasmid backbone and a fivefold molar excess of digested insert. After incubation overnight at 16°C the reactions were used to transform bacteria.

3.1.5. PCR based vector preparation for background free cloning

In conventional cloning a plasmid vector is digested by two different restriction enzymes within the multiple cloning site to allow for an orientation specific insertion of the insert. When some vector molecules are only incompletely digested they can easily get religated. After transformation they might than give rise to multiple colonies containing the old vector, a phenomenon often referred to as background. To prevent that incompletely digested vectors get religated or that two vector molecules get ligated with one another the 5' phosphates are routinely removed. But the respective phosphatase has to be either denatured by heat or removed from the vector before the insert can be added as otherwise the insert would as well lose its 5' phosphate making ligation impossible. After digestion and dephosphorylation the vector DNA is therefore normally applied to agarose gel electrophoresis and extracted from the gel slice, which corresponds to the expected size. But frequently a big portion of the DNA is lost in the process. And the ligation and transformation efficiencies are as well reduced as due to the missing phosphate group only one chain can be ligated.

To circumvent this problem I conceived an alternative PCR based protocol for the preparation of plasmid backbones with high ligation and transformation efficiencies but no background. It is schematically depicted in Figure 3-1. An important advantage of the protocol is that any desired restriction site can be introduced by the primers, if it is not present in the remaining vector backbone.

The PfuUltra high-fidelity DNA polymerase from Stratagene is a genetically engineered optimized polymerase with an error rate of only 4.3×10^{-7} per base pair per duplication and an enhanced processivity compared to other available proofreading polymerases. In *E.coli* the genome is replicated prior to cell division mainly by the DNA polymerase III replisome. Thereby the rate of incorporation of mismatching bases is reported to range between 10^{-7}

and 10^{-9}. The DNA mismatch repair mechanism fixes about 99% of these errors resulting in an overall error rate of between 10^{-9} and 10^{-11} per base pair per replication.[153–155] In the first instant the lower error rate seems to speak for the amplification of the plasmid DNA in *E.coli* as it is conventionally done. But as bacterial cultures are inoculated with only a single colony it takes more than 30 duplications until a cell mass is reached, which is appropriate for the purification of the necessary amounts of plasmid DNA. With the PfuUltra high-fidelity DNA polymerase comparable amounts could be amplified during only 15-18 PCR cycles. With this it can be estimated that compared to the conventional amplification of plasmid DNA in bacteria the effective error rate, which has to be expected for the PCR based protocol, should be lower or equal.

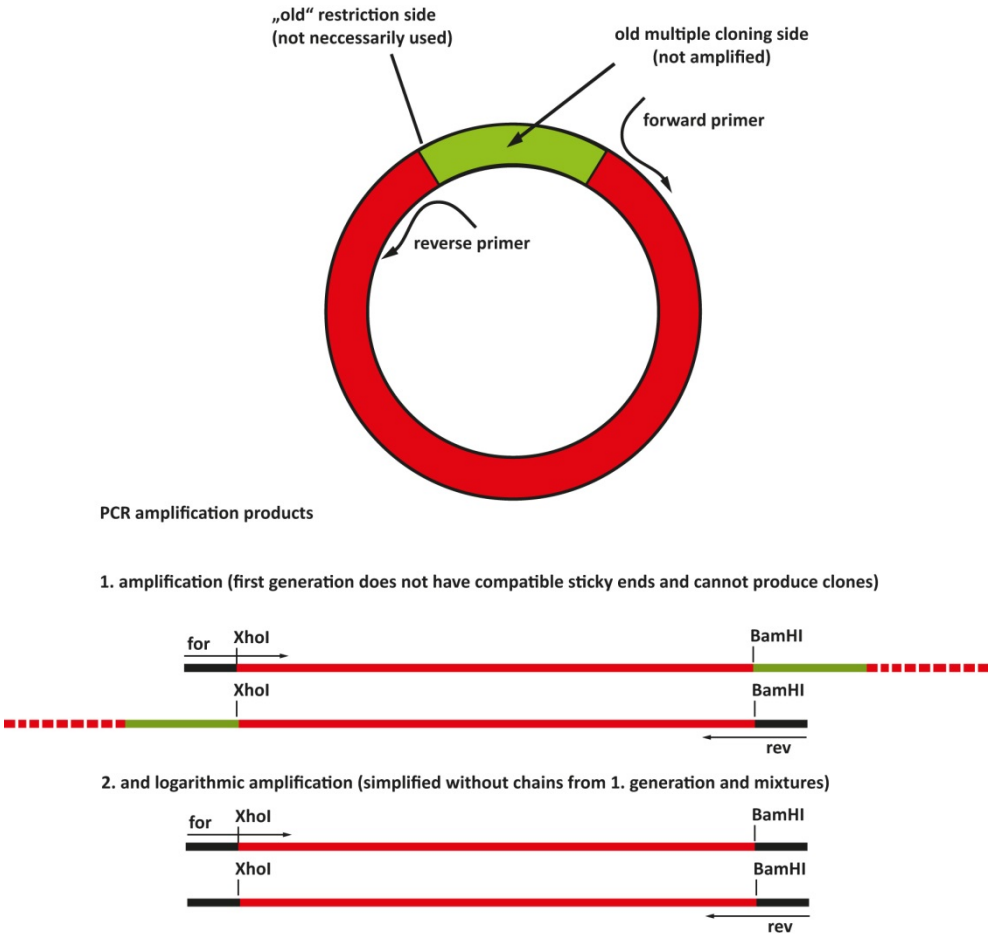


Figure 3-1 Scheme of the PCR based amplification of vector backbones. Forward and reverse primers are designed in such a way that the vector backbone can be amplified as a linearized construct with the desired restriction sides at the ends.

The template DNA used for the PCR based vector amplification protocol was purified from the methylating (crucial!) strain DH5α.

20 ng template DNA, 125 ng forward and reverse primer as well as 1 µl PfuUltra high fidelity DNA polymerase, 1 x polymerase buffer and 1 µl of 10 mM dNTP mix were used in a

total volume of 50 µl filled up with milli-Q H_2O. The dsDNA was initially denatured for 1 min at 95°C. Then 18 cycles followed with denaturation during 30 seconds at 95°C, 30 seconds at an appropriate annealing temperature (generally set about 3°C below the lower calculated primer melting temperature) and an extension period of 2 min per kbp of the target length at 68°C. The end extension was set to 10 min at 68°C.

Following PCR purification the template DNA was digested with Dpn I. The digestion with the desired pair of restriction enzymes could in many cases be performed in parallel. Otherwise an intermediate PCR purification step was necessary to exchange the buffer. After a final PCR purification step the vector DNA was usable for ligation.

When forward and reverse primer contain the same restriction site this protocol can as well easily be adapted to cut out sections from existing plasmids by simply religating the final vector preparation. This opportunity is especially useful to shorten expressed sequences at their N-terminus.

3.1.6. Transformation of bacteria

For transformation competent bacteria (DH5α or NEB T7 express) [156] were incubated with circular plasmid DNA or ligation mixtures for 10 min on ice, heat-shocked for 45 s at 42°C, recovered on ice for 5 min and incubated with 400 µl SOC medium at 37°C for another 30 min. Afterwards the cells were carefully pelleted at 4000 rpm for 1 min and most of the supernatant was discarded. The pelleted cells were resolubilized in the remaining volume (50-100 µl) and spread onto LB agar plates containing the appropriate antibiotics. The LB agar plates were incubated overnight at 37°C.

3.1.7. Plasmid DNA preparation from bacterial cultures

Plasmid DNA was purified from bacteria using either the Nucleo Spin Plasmid kit for mini preps (~ 5 ml culture volume) or Nucleo Bond PC 100 for midi preps (~ 100 ml culture volume) both from Macherey Nagel according to the manufacturer's instructions.

3.1.8. Determination of DNA concentration

DNA concentrations were determined spectrometrically by UV absorption at 260 nm using a NanoDrop 1000 spectrophotometer (Peqlab Biotechnologie GmbH). DNA purity was evaluated by the ratio of 260/280 nm; a ratio ≥ 1.8 was interpreted as a sign of pure DNA.

3.1.9. Site-directed mutagenesis

So called Quick Change PCR reactions for the site-directed introduction of specific mutations were carried out using either PfuTurbo or PfuUltra high fidelity DNA polymerase from Stratagene in a total volume of 50 µl according to the manufacturer's instructions. Primers were designed with the help of the web based program primerX and ordered from Biospring (Biospring, Frankfurt). After the PCR reaction the methylated template DNA was specifically digested by DpnI during incubation for at least 3 h at 37°C. Afterwards the reaction mixture was directly used to transform *E.coli* DH5α cells. Mutations were verified by DNA sequencing.

3.1.10. DNA sequencing

Sequencing was performed by Sequence Laboratories (Seqlab, Göttingen) with 300-600 ng plasmid DNA per reaction. Primer concentrations and total sample volume were adjusted as suggested by the company.

3.1.11. Agarose Gel Electrophoresis

To separate DNA fragments analytical as well as preparative agarose gel electrophoresis was performed using gels with 1-2% agarose in 1 x TAE buffer supplemented with EtBr (analytical) or crystal violet (preparative). Samples were prepared using the appropriate sample buffer and running conditions were 120 V for 20-60 min depending on the expected size of the DNA and as well the length and agarose concentration of the gel.

3.2. Protein biochemical methods

3.2.1. Determination of protein concentration

Routinely the protein concentrations of purified proteins were determined spectrophotometrically using a NanoDrop 1000 Spectrophotometer (Peqlab Biotechnologie GmbH, Germany). The absorption at 280 nm was determined and the concentration was then calculated using the extinction coefficient. The coefficients were determined using the web server operated program Protparam (Expasy). Two constructs corresponding to the TD of *C.int.* p53/p73-a with different truncations did not contain a tyrosine or tryptophan residue and phenylalanine does not significantly contribute to the extinction at 280 nm. Since the amino acid composition of the short peptides differed too much from available reference proteins such as bovine serum albumin (BSA), colorimetric assays like Bradford were as well not reliable. For the respective constructs highly concentrated stocks and from those NMR samples with a precise concentration of DSS were prepared. One dimensional proton spectra were recorded and the concentration of the stocks was calculated by comparison of the integrals from peaks corresponding to the DSS signal and those of well separated peaks from the protein in the amide and aliphatic regions.

3.2.2. Concentrating of protein samples

Protein samples were concentrated with either amicon centrifugation units (centriprep) or amicon stirring cells at 4°C. The pore sizes were chosen according to the size of the protein. For all proteins with a MW smaller than 20 kDa filter units and membranes with 3 kDa MWCO were used, and otherwise such with 10 kDa MWCO. The expected oligomeric states were not regarded.

3.2.3. Heterologous expression in *E. coli*

For recombinant protein production the expression vector encoding the desired protein was in general freshly transformed into one of the *E.coli* expression strains (2.9). The cells were spread onto LB agar plates containing the appropriate antibiotics for selection and incubated over night at 37°C. The next day colonies were scratched off the plate,

resuspended by gentle vortexing in 5 ml of fresh culture medium in a sterile culture tube and used to inoculate a 0.2-1.0 L LB preculture containing the appropriate antibiotics. Cells were grown at 37°C with constant shaking at 180 rpm until the preculture reached an OD_{600} of 1.0-1.2. For unlabelt expression the main cultures (1.2 L LB or 2XYT medium per 2 L flask) were then inoculated with the preculture to an OD_{600} of 0.1-0.2.

For the production of isotopically enriched proteins the amount of preculture necessary to inoculate the main culture (1.2 L M9 minimal medium in 2 L flask) to an OD_{600} of approximately 0.35 production was centrifuged at 4000 rpm (GS3 rotor) for 10 min at 4°C. The supernatant was removed and the cell pellet was resuspended in the main culture medium.

In either case the culturing temperature was reduced to 20-25°C when the OD_{600} had reached 0.5-0.6 and gene expression was induced by the addition of 1 mM IPTG at an OD_{600} of 0.7-0.8. If not mentioned differently expression was carried out overnight.

Afterwards the cells were harvested by centrifugation at 6000 rpm (GS3 rotor) at 4°C for 10 min. The pellets were resuspended in residual medium, transferred into 50 ml falcon tubes, pelleted again by centrifugation in an Eppendorf table top centrifuge and stored at -80°C until usage.

Since the genes for the *C.int.* p53/p73-a and -b proteins had not been codon optimized, cells carrying the pRARE plasmid were used for expression. This plasmid encodes tRNA genes for codons rarely used by *E.coli*. The selection pressure (34 µg/ml chloramphenicol) for this co-plasmid was only present on the LB agar plates. For all constructs encoding the ampicillin resistance procuring β-lactamase gene 100 µg/ml ampicillin was contained in the LB agar plates and the preculture but not in the main culture. Protein sequences, which needed selection of the encoding plasmid throughout the whole expression protocol, were cloned into a vector with a kanamycin resistance cassette. In such cases 30 µg/ml kanamycin were contained in the LB agar plates and all culture media.

3.2.4. Cell Lysis of *E. coli*

5 ml 10 x lysis buffer were mixed with 20 ml milli-Q H_20 and complemented with one aliquot of PIC as well as 0.5 ml and 1.0 ml, respectively, of the stock solutions of DNAse I and RNAse A. The mixture was precooled on ice before it was added to an *E.coli* cell pellet resulting from an expression in 1-3 L culture medium. The volume was filled up with precooled milli-Q H_2O to 45 ml and, if indicated by Cys residues in the primary sequence of the expressed protein, β-ME was added to give a final concentration of 10 mM in 50 ml. The pellet was resuspended by intensive vortexing, afterwards supplemented with a small amount of lysozyme and incubated on ice for 30-60 min before being subjected to sonication.

Sonication was performed with a Sonicator Labsonic U by 3 times 1 min sonication with an intensity of ~200 W and a sonication interval of 0.6 s. The lysate was constantly kept on ice and incubated for at least 3 min between the sonication circles.

After cell disruption the volume was filled up with milli-Q H_2O to 50 ml and the lysate was cleared from remaining insoluble parts by centrifugation at 17000 rpm (SS34 rotor) at 4°C for 45 min.

3.2.5. Protein purification by Immobilized Metal Ion Chromatography (IMAC)

For the purification of proteins harboring a polyhistidine-tag (e.g. the N-terminal 6xHis-tag of constructs based on pBH4) immobilized metal ion chromatography (IMAC) was applied by using an ÄKTAprime plus or an ÄKTApurifier FPLC system at 4°C. Using columns with 25 mm inner diameter flow rates were set to 5 ml/min for equilibration, stripping and ion-loading, and 3ml/min for washing and elution. When smaller columns with only 1.5 ml inner diameter were used the flow rates were set to 3 and 1.5 ml/min, respectively.

An OmniFit 25/100 column packed with 10 ml IMAC sepharose 6 fast flow resin or a prepacked 5 ml column with the same resin was charged with Ni(II) ions, equilibrated with IMAC buffer A1 mixed with 10% IMAC buffer B1 (50 mM imidazole final concentration) and equilibrated with a mixture of 95% IMAC buffer A1 and 5% IMAC buffer B1 (25 mM imidazole final concentration). If the protein of interest contained cysteine residues the reducing buffers IMAC buffer A2 and B2 were used instead for equilibration and further on.

Using an external pump the cell lysate was applied onto the column at a flow rate of 2 ml/min. The column was then washed with buffer containing 100 mM imidazole until the base line was reached again and eluted with buffer containing 500 mM (100% B1 or B2). 3 ml fractions were collected,inspected by SDS PAGE and those containing the protein of interest were pooled and further purified.

For regeneration the column was finally run with stripping buffer to remove the chelated Ni ions. Afterwards the column was recharged with Ni(II) sulfate solution and stored at 4°C.

3.2.6. Cleavage of fusion protein or purification tag with TEV protease

Tobacco etch virus (TEV) protease was expressed as a fusion protein with Maltose binding protein (MBP) at the N-terminus followed by a TEV cleavage site and a 6xHis-tag in front of the actual protease protein sequence. To increase the stability of the protease a 5xArg-tag was fused to the C-terminus. The TEV protease, which cleaved itself of the MPB already during expression, was itself purified in large quantities via the 6xHis-tag by Ni(II) IMAC.

If followed by an internal TEV cleavage site the N-terminal fusion protein (MBP or GST) or 6xHis-tag could be removed by proteolysis with TEV protease. TEV protease was added in a molar ratio of 1:25 and cleavage was performed overnight at 4°C while the reaction was dialyzed against a volume of IMAC buffer A2 sufficient to reduce the imidazole concentration to ≤ 20 mM.

3.2.7. Reversed Immobilized Metal Ion Chromatography (Re-IMAC)

To remove 6xHis-tag expression tags, TEV protease and uncleaved fusion proteins from TEV protease cleavage reactions gravity flow Ni(II) IMAC was used. 10 ml IMAC sepharose 6 fast flow resin in a gravity flow column loaded with Ni(II) ions were washed with 3 CV of IMAC buffer A1 mixed with 10% of IMAC buffer B1 and equilibrated with a mixture of 95% IMAC buffer A2 and 5% IMAC buffer B2 (25 mM imidazol final concentration). The dialyzed and TEV-cleaved protein solution was applied onto the column followed by washing with the IMAC buffer mixture, which was used for equilibration, until no more protein could be detected in the flow through using Bradford. All protein containing fractions were combined and the column was regenerated by stripping and recharging of Ni(II) ions as described in 3.2.5.

3.2.8. Size Exclusion Chromatography (SEC)

Size exclusion chromatography (SEC or gel filtration) was used as a polishing step during purification, for the characterization of the oligomeric state of proteins and for a qualitative reproducible exchange of the buffer.

Depending on the separation problem and the size of the protein or complex of interest different resins were used. Small proteins (3 kDa – 70 kDa) are best resolved with Superdex75™. For larger proteins or complexes columns with Superdex200™ (10 kDa – 600 kDa) were used. All gel filtrations were performed using an ÄKTAprime plus (only preparative) or ÄKTApurifier FPLC system at 4°C. Absorption was monitored at 280 as well as 260 nm (second wavelength only on ÄKTApurifier) during the run.

For analytical scale gel filtrations and the preparation of NMR samples 10/300 GL columns from GE Healthcare packed with the respective resin were used. Samples were loaded with a 0.5 ml sample loop onto the pre-equilibrated column. The flow rate was set to 0.5-0.7 ml/min and fractionation started after 0.2 CV. For preparative scale gel filtrations HiLoad™ 16/60 prep grade columns, which allowed for a maximal sample size of 5 ml, were used with flow rates of 1.0-1.4 ml/min.

3.2.9. Dialysis

Buffer exchange was performed by dialysis against 5 L of buffer. Proteins were transferred in SpectraPor dialysis membranes with a MWCO of 3.5 kDa and dialyzed at 4°C ON.

3.2.10. Tris-Tricine-SDS-PAGE (adapted from Schaegger)

Glycine-SDS-PAGE (often referred to as Laemmli-SDS-PAGE) and Tricine-SDS-PAGE, based on glycine-Tris and Tricine-Tris buffer systems, respectively, are the most commonly used SDS electrophoretic systems for the separation of proteins. Laemmli-SDS-PAGE is more frequently used, but offers only a limited resolution for proteins with a molecular weight smaller than 30 kDa compared to Tris-Tricine-PAGE.[157,158] Since many protein constructs studied in this thesis have a rather small (monomeric) molecular weight, an optimal resolution in the range of 5-15 kDa was crutial and for convenience all proteins were analyzed using Tris-Tricine-PAGE. To account for differences in molecular weight the polyacrylamide concentration in the separation gels were either 11% or 16%. The Mini-

Protean Tetra Cell System from BioRad was used with glass plates, spacers and combs for gels with a thickness of 1 mm. The volumes in the pipetting scheme (Table 3-1) are according to this system.

Table 3-1 Pipetting scheme for two 11 or 16% Tris-Tricine-SDS separating and two 4% stacking gels.

Compound	Stacking gel (4%)	Separating gel (11%)	Separating gel (16%)
30% (w/v) Polyacrylamid (37.5 : 1)	0.8 ml	4.4 ml	-
40% (w/v) Polyacrylamid (19 : 1)	-	-	4.8 ml
Gel buffer, 3-fold	1.5 ml	4 ml	4 ml
Glycerol 50% (w/v)	-	2.4	2.4
Mili-Q H_2O	3.7 ml	1.2 ml	0.8 ml
APS 10% (w/v)	45 µl	60	40
TEMED	4.5 µl	6	4
Total volume	~ 6 ml	~ 12 ml	~ 12 ml

Protein samples in 1 x reducing sample buffer were incubated at 37°C for 15-30 min or for up to 60 min for samples that were in pellet form. Tris-Tricine-SDS gels were run in 1 x cathode (inside the chamber) and anode buffer (surrounding the chamber). Proteins were separated at 80 V for 20 min followed by 150 V for 55 min (11% gels) or 75 min (16% gels).

3.2.10.1. Coomassie brilliant blue staining

When SDS-PAGE gel electrophoresis was finished, the gels were at first placed into fixing solution and incubated for 30 min at RT with shaking. Then the solution was replaced with Coomassie brilliant blue staining solution and further incubated at RT with shaking. After 1-2 hours the staining solution was removed, the gel over-laminated with water and cooked in the microwave. When cooled down to ≤ 60°C the water was supplemented with 10% (v/v) acetic acid and incubation under shaking was prolonged for 30 min or up to overnight until the background was as clear as desired.

3.3. Biophysical methods

3.3.1. Circular Dichroism spectroscopy

Circular dichroism (CD) spectra and melting curves were measured with a Jasco J-810 spectropolarimeter (Jasco Labortechnik, Gross-Umstadt, Germany) using a cuvette with 1 mm path length and a sample volume of 300 µl.

The respective purified proteins were transferred into the low salt circular dichroism buffer (50 mM NaPi pH 7.0) by gel filtration generally not more than 12 h before the measurement. Samples for CD spectra were diluted with 9/10 shares of degased filtered milli-Q H_2O resulting in an effective buffer concentration of 5 mM NaPi at pH 7.0 effectively. This low concentration was necessary to allow for wavelengths < 190 nm to be scanned. The protein concentration was adjusted to 10 µM, if not indicated differently. Spectra were recorded as an average of 3 to 6 measurements from 170-315 nm. The data interval was set to 0.1 nm, sensitivity to standard, D.I.T. to 1 s, bandwidth to 1 nm and the scanning speed to 100 nm/min. The control of the shutter and the photomultiplier voltage was set to auto. Baseline correction was done by subtracting a spectrum obtained with pure buffer under the same measurement settings. When indicated (SG25) the spectra were smoothed using the the Jasco spectra analysis software with the method of Savitzky and Golay using a convolution width of 25 data points. Melting curves were recorded with samples in 1 x CD buffer and a protein concentration of 100 µM, if not mentioned differently. The CD signal at 222 nm was recorded when the temperature was increased from 5°C to 110°C with a ramp speed of 1°C/min. All other settings were as described above. A blank measurement with buffer was subtracted as baseline correction.

3.3.1.1. Curve fitting of thermal denaturation curves determined for homo-tetramers characterized by reversible coupled disassembly and unfolding

The following equations and derivations are adapted from the “protein folding handbook” edited by J. Buchner and T. Kiefhaber, wikipedia and publications on protein science (Hagihara et al. 1994; Mateu et al. 1998).[159–161]

$K_U(T)$, the equilibrium constant of the unfolding reaction is related to the Gibbs energy change upon unfolding, ΔG_U , as:

$$K_U(T) = exp\left(-\frac{\Delta G_U(T)}{RT}\right)$$

The Gibbs energy of unfolding is defined by the following equation:

$$\Delta G_U(T) = \Delta H_U(T) - T \cdot \Delta S_U(T)$$

$\Delta H_U(T)$ is the change in enthalpy and $\Delta S_U(T)$ is the change in entropy upon protein unfolding. The heat capacity change upon unfolding, ΔCp, defines the temperature dependence of both parameters.

$$\Delta C_p = \frac{d\Delta H_U(T)}{dT} = T \cdot \frac{d\Delta S_U(T)}{dT}$$

When it is assumed that ΔCp does not change with temperature, the following terms for $\Delta H_U(T)$ and $\Delta S_U(T)$ can be written:

$$\Delta H_U(T) = \Delta H_U(T_0) + \Delta C_p \cdot (T - T_0)$$

$$\Delta S_U(T) = \Delta S_U(T_0) + \Delta C_p \cdot \ln\left(\frac{T}{T_0}\right)$$

where the terms $\Delta H_U(T_0)$ and $\Delta S_U(T_0)$ are the changes in enthalpy and entropy at a reference temperature T_0. When a monomeric protein unfolds in a two-state process there is one temperature at which the fractions of natively folded, F_N, and unfolded, F_U, protein are equally populated. This temperature is frequently referred to as the transition or melting temperature, T_m. It is convenient to set the reference temperature to this, since for a monomeric system the equilibrium constant at T_m is equal to 1 and thus the Gibbs energy is equal to zero.

$$\Delta G_U(T_m) = -RT \ln K_U(T_m) = \Delta H_U(T_m) - T_m \Delta S_U(T_m) = 0$$

This allows the entropy of unfolding at T_m to be rewritten:

$$\Delta S_U(T_m) = \frac{\Delta H_U(T_m)}{T_m}$$

And the entropy function can be translated into a term only depending on T_m, $\Delta H_U(T_m)$ and ΔC_p:

$$\Delta S_U(T) = \frac{\Delta H_U(T_m)}{T_m} + \Delta C_p \cdot \ln\left(\frac{T_m}{T}\right)$$

Thus the temperature dependence of the Gibbs energy upon unfolding can be expressed as a term only depending on the same three parameters:

$$\Delta G_U(T) = \Delta H_U(T_m) \cdot \left(1 - \frac{T}{T_m}\right) + \Delta C_p \cdot \left[T - T_m - T \ln\left(\frac{T}{T_m}\right)\right]$$

For multimeric systems where denaturation is coupled with dissociation of a complex this is more complicated. Here the case of a native homo-tetramer, N_4, is described, which dissociates into unfolded monomers, U, upon (thermal) denaturation.

$$N_4 \overset{K_d}{\leftrightarrow} 4 \cdot U$$

The equilibrium constant is defined as:

$$K_d = \frac{[U]^4}{[N_4]}$$

where $[N_4]$ and $[U]$ are the concentrations of native tetramers and unfolded monomers, respectively. The fractions of total protein, P_t, present in these two states are presented as:

$$[N_4] = \frac{1}{4} \cdot P_t \cdot F_{N_4} \qquad \text{and} \qquad [U] = P_t \cdot F_U$$

F_{N4} and F_U are the fractions of native tetramers and unfolded monomers, respectively, and the total protein concentration expressed in mole of monomers is:

$$P_t = 4 \cdot [N_4] + [U]$$

With that the equilibrium constant of coupled dissociation and unfolding can be expressed as a function of total protein concentration and the fraction of unfolded protein as:

$$K_d = \frac{4 \cdot [P_t]^3 \cdot {F_U}^4}{(1 - F_U)}$$

The fact that the total protein concentration does not cancel out makes this case much more complicated. ΔG_U is not zero at the transition midpoint of thermal unfolding, T_m, since the equilibrium constant is not equal to 1.

$$K_{d,F_U=0,5} = 4 \cdot ([P_t]/2)^3$$

The term $-RT \ln[K_d(T_m)]$, which accounts for the concentration dependency of ΔG_U, does not cancel out as in the case of a monomeric system, and hence, the modified Gibbs-Helmholtz equation for the extrapolation of thermodynamic parameters to different temperatures is:

$$\Delta G_U(T) = \Delta H_U(T_m) \cdot \left(1 - \frac{T}{T_m}\right) + \Delta C_p \cdot \left[T - T_m - T \ln\left(\frac{T}{T_m}\right)\right] - RT \ln[K_d(T_m)]$$

The thermodynamic parameters for the thermal unfolding of tetrameric proteins were determined by direct non-linear fitting to the experimental ellipticity values by a least squares approach using Micorsoft Excel with the add-in solver.

The calculated ellipticity was defined by the expression:

$$\theta = (1 - F_U)(\theta_N + m_N T) + F_U(\theta_U + m_U T)$$

with θ_N and θ_U being the intercept (ellipticity values at 0 K) to the native (N) and unfolded (U) state, and m_N and m_U being the slope with temperature, respectively.

Due to the very high thermal stability of the *C.int.* p53/p73-a TD only a limited temperature range above full denaturation could be measured. Since that especially the baseline for the unfolded monomer was not explicit, and hence, could not be separately determined by linear regression. Both baselines were assumed to be linear and their slope and intercepts were included in the fitting parameters. To facilitate a reasonable fitting the allowed range for the baseline of unfolded monomer was limited to values ≤ 0 at 0°C.

The fraction of unfolded monomer, F_U , had to be expressed as a term depending solely on the K_d at the respective temperature. Accordingly the root of the following fourth-order equation was solved:

$$\left(4 \cdot \frac{[P_t]^3}{K_d}\right) \cdot {F_U}^4 + F_U - 1 = 0$$

with the factor "a" being defined as

$$a = 4 \cdot \frac{[P_t]^3}{K_d} = 4 \cdot [P_t]^3 / exp\left(\frac{-\Delta G_U(T)}{RT}\right)$$

The four roots (x1, x2, x3, x4) of the general quartic equation

$$ax^4 + bx^3 + cx^2 + dx + e = 0$$

with a ≠ 0 are given by the following formulas:

$$x_{1,2} = -\frac{b}{4a} - S \pm \frac{1}{2}\sqrt{-4S^2 - 2p + \frac{q}{S}}$$

$$x_{3,4} = -\frac{b}{4a} + S \pm \frac{1}{2}\sqrt{-4S^2 - 2p - \frac{q}{S}}$$

where p and q are defined as

$$p = \frac{8ac - 3b^2}{8a^2}$$

and

$$q = \frac{b^3 - 4abc + 8a^2 d}{8a^3}$$

which simplifies in this special case with b=c=0 and d=1 to

$$p = 0$$

and

$$q = \frac{1}{a}$$

And where

$$S = \frac{1}{2}\sqrt{-\frac{2}{3}p + \frac{1}{3a}\left(Q + \frac{\Delta_0}{Q}\right)}$$

$$Q = \sqrt[3]{\frac{\Delta_1 + \sqrt{\Delta_1^2 - 4\Delta_0^3}}{2}}$$

with

$$\Delta_0 = c^2 - 3bd + 12ae$$

$$\Delta_1 = 2c^3 - 9bcd + 27b^2e + 27ad^2 - 72ace$$

simplifies with b=c=0, d=1 and e=-1 to

$$\Delta_0 = -12a$$

$$\Delta_1 = 27a$$

and

$$\Delta_1^2 - 4\Delta_0^3 = -27\Delta$$

$$\begin{aligned}\Delta = {} & 256a^3e^3 - 192a^2bde^2 - 128a^2c^2e^2 + 144a^2cd^2e - 27a^2d^4 + 144ab^2ce^2 \\ & - 6ab^2d^2e - 80abc^2de + 18abcd^3 + 16ac^4e - 4ac^3d^2 - 27b^4e^2 \\ & + 18b^3cde - 4b^3d^3 - 4b^2c^3e + b^2c^2d^2\end{aligned}$$

which simplifies in this special case with b=c=0, d=1 and e=-1 to

$$\Delta = -256a^3 - 27a^2$$

Hence, Q and S can be rewritten as:

$$Q = \sqrt[3]{\frac{27a + \sqrt{-27(-256a^3 - 27a^2)}}{2}} = \sqrt[3]{\frac{27a + \sqrt{729a^2 + 6912a^3}}{2}}$$

$$S = \frac{1}{2}\sqrt{\frac{1}{3a}\left(Q - \frac{12a}{Q}\right)} = \frac{1}{2}\sqrt{\frac{Q}{3a} - \frac{4}{Q}} = \sqrt{\frac{Q}{12a} - \frac{1}{Q}}$$

In my calculations with Microsoft Excel I determined the absolute value for $\left(\frac{Q}{12a} - \frac{1}{Q}\right)$ prior to extracting the root to prevent errors from incomplete cancelling out.

The following root was calculated

$$F_U = x_1 = -S + \frac{1}{2}\sqrt{-4S^2 + \frac{1}{a \cdot S}}$$

It is theoretically possible to include ΔC_p as a parameter in the fitting of thermal denaturation curves detected with CD, but there is generally not enough information to determine ΔC_p accurately. ΔC_p is the second derivative of ΔG with respect to temperature and because of that including the ΔC_p term can, in fact, actually worsen the agreement between T_m values determined by CD and DSC.[162] It is therefore common practice to set ΔC_p equal to zero. Johnson et al. (1995) had determined a value of ΔC_p=1.7 kcal/(K·mol of tetramer) for a slitely elongated construct of the TD of human p53.[163] This would correspond to 425 cal/(K·mol of monomer) or 1.78 kJ/(K·mol of monomer). Taking this into account the value for ΔC_p was fixed as equal to zero for the first rounds of iterative fitting. Afterwards ΔC_p was set to 2.5 kJ/(K·mol of monomer) as a starting point and iterative fitting was repeated with a limited allowed range for values of ΔC_p between 0-5 kJ/(K·mol of monomer). In most cases the value of ΔC_p was thereby changed back close or equal to zero.

3.3.2. Nuclear Magnetic Resonance (NMR) in solution

NMR experiments were performed on Bruker Avance spectrometers equipped with cryogenic 5 mm z-axis gradient triple resonance probes at proton frequencies of 950, 900, 800, 700 or 600 MHz, or on a Bruker Avance 500 MHz spectrometer with a room-temperature x,y,z-gradient, triple resonance probe.

3.3.2.1. Sample preparation for the C.int. p53/p73 project

NMR samples for the *C.int.* p53/p73-a TD project were prepared by gelfiltration in RE buffer (50 mM L-arginine & 50 mM L-glutamate at pH 7.0) and those for the *C.int.* p53/p73-b TD project in reducing RE buffer (RE buffer supplemented with 1 mM DTT). After concentrating using amicon stirring cells the samples were complemented with 5% (v/v) D_2O/DSS NMR standard and, if indicated, with protease inhibitor cocktail. The DSS (4,4-dimethyl-4-silapentane-1-sulphonate) was used as an internal reference for chemical shifts. Unless otherwise stated, all experiments for the *C.int.* p53/p73 project were measured at 298 K.

3.3.3. Spectra analysis

Spectra were processed with Bruker TopSpin™ 2.1 and analyzed with UCSF Sparky 3.114 [164].

3.3.4. Sequential assignments

The resonances of nuclei from the protein backbone were assigned with TROSY-based HNCACB, HNCOCA, HN(CA)CO and HNCO experiments.[165,166]

3.3.5. Hetero NOE experiments

HetNOEs were determined with a steady-state ^{15}N-[^{1}H]-hetero NOE-TROSY experiment.[167] At an interleaved manner two experiments were recorded, one with presaturation of the protons and one control experiment without presaturation (reference). The residue specific

hetNOE was calculated based on the intensities of the respective peaks in both spectra. The peak heights were determined with sparky and multiplied by 2^{NCproc}. The NC proc value is an intensity scaling factor. It is set by all processing commands when scaling the data down or up by a factor of 2. The actual hetNOE values were calculated by dividing the modified $I_{presaturated}/I_{reference}$. The backbone ^{15}N-[^{1}H]-hetero NOE provides information about the motion of individual N-H bond vectors. Those which undergo motion faster than the overall tumbling of the molecule show decreased NOE intensities relative to the average observed for the majority of the residues. Hetero NOE values around 0.8 were interpreted to represent rigid regions, and hetero NOE values < 0.4 to indicate unstructured, highly flexible regions.

3.3.6. TALOS (torsion angle likeliness obtained from shift & sequence similarity)

The program TALOS+ was routinely used for the identification of secondary structure elements.[168,169] The program makes use of the well-known observation that the differences between observed backbone chemical shifts and their corresponding random coil values (secondary chemical shifts) are highly correlated with aspects of local protein secondary structure. TALOS+ compares the protein sequence and chemical shifts of a target protein with a database of 200 NMR solution structures of different proteins and makes predictions for the backbone angles phi and psi. In addition an estimated backbone order parameter S2 is reported, which is derived from the chemical shifts by a method, called the random coil index (RCI).[170] This parameter can help to distinguish regions that are dynamically disordered from those which are highly structured. As the chemical shifts for HA and CO are not obtained when a sequential assignment is carried out based on HNCACB and HNCA spectra, only the backbone chemical shifts of HN, CA, CB and N were used as input for TALOS+, if not indicated differently.

3.3.7. Determination of secondary structure propensity based on chemical shifts

Not every secondary structure forms in an all-or-nothing fashion and many different experimental and computational methods have been developed aiming at characterizing the

conformational properties of proteins lacking distinct secondary structure, hence, unstructured states.

NMR chemical shifts provide ensemble-averaged information about all local conformations sampled by the protein up to the millisecond time scale.[171] The differences between measured and random coil chemical shifts, which are often referred to as chemical shift deviations (CSD) or secondary chemical shifts, can be used to define the presence of α-helices or β-strands in partially structured states.[172–175] And it was shown that CSDs are correlated with the percent helicity of a peptide sequence.[176] A given CA ensemble averaged CSD, taken as a fraction of the average CSD, expected for the case that the respective residue would locate within a fully formed helix, can be used as a quantitative measure of the percent helicity at that residue.[177]

Based on these findings recently three new methods have been developed, which aim at improving the quantitative characterization of secondary structure elements in disordered states. Forman-Kay and co-workers introduced the secondary structure propensity (SSP) algorithm, which reports the fractional populations of α-helices and β-strands.[178] The algorithm is run locally as a script, explicitly accounts for the differential sensitivity of different types of chemical shifts to secondary structure and includes a routine for the correction of the offset in the ^{13}C dimension. The algorithm ncSPC (neighbor-corrected Structural Propensity Calculator), is based on the same methodology, corrects in addition for effects from neighboring residues and is run as a webserver operated program.[179] The most recently developed method is the δ2D method.[180] It is able to translate a set of chemical shifts into probabilities of occupation of secondary structure elements and discriminates between four different local conformational states (random coil, α-helix, β-strand and polyproline II).

So far the methodology is still under debate and none of the methods is so widely accepted and used as for example TALOS+. For this reason I carried out the analysis using all three algorithms. The results obtained with the most sophisticated algorithm δ2D are presented under results and those obtained with SSP and ncSPC in the appendix, respectively. If not mentioned differently the backbone chemical shifts of HN, CA, CB and N were used as input.

3.3.8. Calculation of chemical shift perturbation

To compare chemical shifts of assigned backbone amide (N and HN) resonances of wild type sequences with those from truncated or mutated sequences the chemical shift perturbation vectors were calculated from the absolute values of the chemical shift differences. To facilitate a size corrected comparison the chemical shifts in the ^{15}N dimension were multiplied with a scaling factor. The backbone chemical shift vectors were calculated as follows:

$$\Delta^{2D} = [(\Delta^{HN})^2 + (\Delta^N R^N)^2]^{1/2}$$

where R^N=0.2, and Δ^{HN} and Δ^N are the chemical shift differences for a residue specific assigned backbone resonance in the two respective HSQC spectra to be compared.

3.3.9. Assignment of side chain resonances and pro-chiral methyl groups in *C.int.* p53/p73-b TD

For side chain resonance assignment was carried out based on 3D TOCSY H(C)CH, H(CCCO)NH and (H)C(CCO)NH [^{15}N, ^{1}H]-TROSY spectra.[165,166] As the respective protein sequence contains only three tyrosine and one phenylalanine residue the aromatic side chains were assigned directly in a 3D NOESY [$^{13}C_{aromatic}$, ^{1}H]-ct-TROSY spectrum and the corresponding aromatic 2D version. The prochiral methyl-groups of valine and leucine were assigned in a 2D [^{13}C, ^{1}H]-ct-HSQC recorded from a sample with 10% nonrandom ^{13}C labeling.[181]

3.3.10. Assignment of inter-monomer NOEs in the *C.int.* p53/p73-b TD

Two samples of *C.int.* p53/p73-b TD with labeling schemes of U-[^{13}C, ^{1}H] and U-[^{15}N, ^{1}H] were separately expressed and purified. After mixing of equimolar amounts the proteins were denatured with 6 M guanidinium hydrochloride and subsequently refolded by dilution with a 50-fold volume of reducing RE buffer. The protein solution was reconcentrated in an amicon stirring cell under constant pressure from nitrogen gas and applied onto preparative gel filtration using a Superdex75™ 16/60 column, equilibrated and run in reducing RE buffer. These samples with 1:1 mixed labeling of the individual monomers were used to assign those NOESY peaks, which result from inter-monomer contacts, from a 3D 15N-edited NOESY-[13C,1H]-HSQC, a 3D [13C,1H]-HMQC-NOESY-[15N,1H]-TROSY and a constant time J-resolved NOESY-[13C,1H]-HSQC experiment.[182] J-coupling was recorded as a pseudo 4th dimension. The types of interactions, which can be detected by the different experiments, are depicted in Figure 3-2.

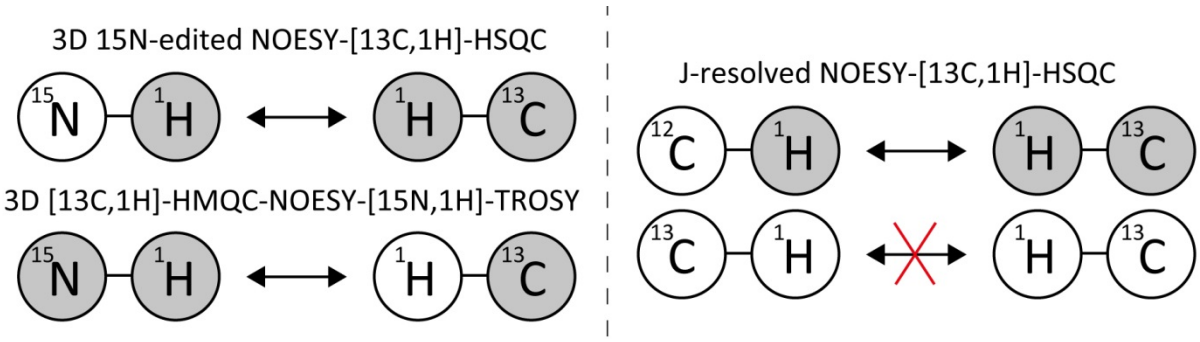


Figure 3-2 Scheme of the NMR experiments used to discriminate inter- from intra-monomer NOEs. The experiments were meant to specifically detect those NOE peaks, which result from contacts between individual monomers, using samples with a U-[^{13}C, ^{1}H] 1:1 U-[^{15}N, ^{1}H] labeling scheme. Nuclei in grey circles are resolved as individual dimensions.

3.3.11. Calculation of the solution structure of the *C.int.* p53/p73-b TD

The solution structure of the *Ciona int.* p53/p73-b TD was calculated by Henry Jonker from the group of Prof. Harald Schwalbe using the software package CYANA [183–185] in particular with modified protocols (CYANA 3.9 development version) and refinement was performed with ARIA/CNS.[186,187] The protein backbone and side chain chemical shifts were manually assigned and the NOESY peak lists were manually peak picked and assigned to a large extend using UCSF Sparky 3.114.[164]

Intersubunit distance constraints were obtained with 4D-CT-J-Resolved ^{13}C-separated NOESY (in D_2O, τ_M = 100 ms), with 3D ^{15}N-edited/^{13}C-separated NOESY (in H_2O, τ_M = 110 ms) and with 3D ^{15}N/^{13}C-separated NOESY (in H_2O, τ_M = 110 ms) experiments.

The resonances and 3D NOESY peak lists from 3D NOESY-[^{15}N, ^{1}H]-HSQC (in H2O, τ_M = 60 ms), 3D NOESY-[^{13}C, ^{1}H]-HSQC (aliphatic and aromatic, in H2O, τ_M = 60 ms and τ_M = 50 ms, respectively), as well as a the intermolecular 3D and 4D J-resolved NOESYs were used as input for automated NOESY cross peak assignment and calibration with CYANA.

The chemical shift tolerances were set to 0.035 ppm and 0.40 ppm for the protons and heavy atoms respectively. In addition to NOE data, hydrogen bond distances and amply defined dihedral angle restraints (based on TALOS+ predictions and confirmed with NOESY assignments and initial structure calculations) as well symmetry (distance difference) restraints were included in the standard structure calculation with CYANA 3.9 (100 structures per iteration, 20000 refinement steps). The final bundle of the 20 best structures was used as input for optimization with CNS 1.1 using adapted ARIA 1.2 setup and protocols for refinement in explicit water.[188] The restraints were converted to ARIA/CNS format and included in this refinement stage. The standard settings and allhdg5.3 force field were used with OPLS nonbonded parameters. The final structure bundle was analysed with Procheck 3.5.4.[169,189] The experiments used for assignment and structure calculation are listed in Table 3-2.

Table 3-2 NMR experiments used for the determination of the solution structure of *C.int.* p53/p73-b TD.

Experiment	Conc. (monomer)	Labeling	Mixing time	^{1}H frequency
2D [^{15}N,^{1}H]-HSQC	2.5 mM	U-[^{13}C, ^{15}N, ^{1}H]	-	900
HNCACB; 3D [^{15}N,^{1}H]-TROSY	2.5 mM	U-[^{13}C, ^{15}N, ^{1}H]	-	900
HNCO; 3D [^{15}N,^{1}H]-TROSY	2.5 mM	U-[^{13}C, ^{15}N, ^{1}H]	-	500
3D (H)C(CCO)NH-TOCSY	2.5 mM	U-[^{13}C, ^{15}N, ^{1}H]	21 ms	600
3D H(CCCO)NH-TOCSY	2.5 mM	U-[^{13}C, ^{15}N, ^{1}H]	21 ms	500
3D H(C)CH-TOCSY	2.5 mM	U-[^{13}C, ^{15}N, ^{1}H]	21 ms	500
2D [^{13}C,^{1}H]-HSQC	4 mM	10% ^{13}C, U-[^{15}N, ^{1}H]	-	800
2D [^{13}C,^{1}H]-ct-HSQC; τ_C=28.3 ms	4 mM	10% ^{13}C, U-[^{15}N, ^{1}H]	-	800
3D NOESY-[^{15}N,^{1}H]-TROSY	2.5 mM	U-[^{15}N, ^{1}H]	60 ms	800
3D NOESY-[^{13}C,^{1}H]-HSQC	2.5 mM	U-[^{13}C, ^{15}N, ^{1}H]	60 ms	800
3D NOESY [^{13}C$_{aromatic}$, ^{1}H]-ct-TROSY	4 mM	U-[^{13}C, ^{1}H] 1:1 U-[^{15}N, ^{1}H] *	50 ms	600
3D ^{15}N-edited NOESY-[^{13}C,1H]-HSQC; inter	4 mM	U-[^{13}C, ^{1}H] 1:1 U-[^{15}N, ^{1}H]	110 ms	800
3D [^{13}C,^{1}H]-HMQC-NOESY-[^{15}N,^{1}H]-TROSY; interC	4 mM	U-[^{13}C, ^{1}H] 1:1 U-[^{15}N, ^{1}H]	110 ms	900
CT-J-resolved NOESY-[13C,1H]-HSQC	4 mM	U-[^{13}C, ^{1}H] 1:1 U-[^{15}N, ^{1}H] in D_2O	100 ms	700

* the labeling scheme was not necessary for this experiment; at the time of recording the sample was simply available

4. Results

4.1. Alignment of *Ciona intestinalis* and *Ciona savignyi* p53/p73 proteins towards sequences from other species

Using clustalW2 the full length protein sequences of the p53/p73-like proteins from the two closely related tunicate species *Ciona intestinalis* (*C.int.*) and *Ciona savignyi* (*C.sav.*) were aligned (data not shown) against the p53, p63 and p73 proteins of five representative vertebrate species (*Homo sapiens, Mus musculus, Xenopus tropicalis, Danio rerio* and *Callorhynchus milii*) to localize those sections, which correspond to conserved domains. Protein sequences from three other species *Caenorhabditis elegans* (nematode), *Drosophila melanogaster* (arthropod) and *Branchiostoma floridae* (cephalochordate, lancelet), which are taxonomically more distant to vertebrates, were also included. The DBD is the only domain conserved in all included species. The NMR solution structure of the DBD of human p63 was used as a reference, and those sections, which aligned to the sequence spanned by the secondary structure elements (SSEs), were selected for pairwise alignments. The alignment scores in percent sequence identity (SI) for the alignment of the *Ciona* sequences versus those of the vertebrate and invertebrate example species are listed in Table 4-1. A complete table listing the percentages of all pairwise alignments between the sequences from the deuterostome species (*Ciona*, lancelet and vertebrates) is included in the appendix in section 7.1.8.1. Within the DBD the degree of sequence identity between the *Ciona* proteins and any of the vertebrate p53, p63 or p73 proteins ranges between 38 and 51% with no significant outlier. For the *Ciona* p53/p73-a and p53/p73-b proteins the sequence identity towards the two lancelet homologs is in the same range, but the percentages indicate a closer relationship towards lancelet p63/p73 (47-52% SI) compared to lancelet p53 (34-38% SI). *Drosophila melanogaster* p53 (Dmp53) has an even lower sequence identity within the DBD towards any of the *Ciona* proteins ranging between 29 and 32% and the sequence identity of the p53 family protein from *Caenorhabditis elegans* (Cep-1) is only 19%.

The sequence of the oligomerization (OD) or tetramerization domain (TD) is only conserved among deuterostome species. As revealed by the respective NMR solution structures the ODs of Dmp53 (pdb 2RP4) and Cep-1 (pdb 2RP5) have domain architectures,

which differ from those of vertebrate p53 family proteins and from one another.[113] For both proteins also gapped sequence alignments towards the TDs of the human proteins are very weak.[190] For this reason Dmp53 and Cep-1 were excluded. The NMR solution structure of the TD of human p73 was used as a reference, and those sections, which aligned to the sequence spanned by the SSEs, were selected for pairwise alignments. The alignment scores in percent amino acid identity for the alignment of the *Ciona* sequences versus those of the vertebrate and invertebrate example species are listed in Table 4-2. A complete table listing the percentages of all pairwise alignments between the sequences from the deuterostome species (*Ciona*, lancelet and vertebrates) is included in the appendix in section 7.1.8.2.

Table 4-1 Pairwise percentage amino acid identity within DNA binding domain of p53 family proteins from different species against p53/p73 proteins from *Ciona* species.

DBD alignment % amino acid identity	vertebrate															invertebrate			
	p53					p63					p73								
	human p53 aa 102-292	**mouse p53** aa 96-286	**frog p53** aa 76-267	**zebrafish p53** aa 70-260	**e. shark p53** aa 92-283	**human p63** aa 170-362	**mouse p63** aa 170-362	**frog p63** aa 76-273	**zebrafish p63** aa 74-266	**e. shark p63** aa 193-385	**human p73** aa 120-312	**mouse p73** aa 112-304	**frog p73** aa 120-312	**zebrafish p73** aa 127-319	**e. shark p73** aa 125-317	**lancelet p53** aa 99-284	**lancelet p63/p73** aa 169-361	**fruit fly p53** aa 86-278	**roundworm p53** aa 219-417
***C.int.* p53/p73-a** aa 184-375	**40**	**42**	**44**	**46**	**44**	**50**	**50**	**42**	**48**	**51**	**50**	**51**	**51**	**50**	**51**	**38**	**52**	**29**	**19**
***C.sav.* p53/p73-a** aa 180-371	**38**	**41**	**43**	**45**	**44**	**48**	**48**	**42**	**46**	**48**	**48**	**49**	**48**	**47**	**48**	**36**	**51**	**32**	**19**
***C.int.* p53/p73-b** aa 142-337	**41**	**43**	**43**	**41**	**46**	**45**	**45**	**39**	**46**	**46**	**45**	**47**	**47**	**47**	**47**	**34**	**48**	**31**	**19**
***C.sav.* p53/p73-b** aa 135-326	**40**	**42**	**43**	**41**	**45**	**47**	**47**	**40**	**46**	**46**	**45**	**45**	**46**	**46**	**44**	**37**	**47**	**32**	**19**

Protein sequences belonging to the p53 family of transcription factors from the tunicate species *Ciona intestinalis* (*C.int.*) and *Ciona savignyi* (*C.sav.*) were aligned against those of the vertebrate species *Homo sapiens* (human), *Mus musculus* (mouse), *Xenopus tropicalis* (frog), *Danio rerio* (zebrafish) and *Callorhynchus milii* (e. shark), and the invertebrate species *Branchiostoma floridae* (lancelet), *Drosophila melanogaster* (fruit fly) and *Caenorhabditis elegans* (roundworm). The full length protein sequences were aligned using clustalW2. The sections corresponding to the DNA binding domain (DBD) were identified as those spanned by the secondary structure elements in the NMR solution structure of the DBD of human p63 (pdb 2RMN). Percent amino acid identity was determined for pairwise alignments using Jalview and rounded. The accession numbers are listed in the appendix in section 7.1 and the amino acid (aa) numberings correspond to the respective canonical isoforms.

The sequences of the *Ciona* p53/p73-a proteins, which correspond to the TD, show between 34 and 43% sequence identity towards the vertebrate proteins, irrespective of being it p53, p63 or p73. The values for the same comparison are throughout lower for the *Ciona* p53/p73-b proteins and range between only 22 to 30%. As all *Ciona* and vertebrate

p53 family proteins derived from the same common ancestor this indicates that after the dissociation the evolution of the *Ciona* p53/p73-b proteins might have been characterized by a higher degree of degeneration with respect to the TD. By trend the degree of sequence identity towards the *Ciona* p53/p73-a and p53/p73-b proteins is lower for the two lancelet p53 family proteins compared to the vertebrate proteins. For this reason the *Ciona* p53/p73 proteins were further only compared to the corresponding vertebrate sequences. The alignment of protein sequences corresponding to the tetramerization domain of p53/p63/p73-like proteins from five vertebrate species versus those from *C.int.* and *C.sav.* is shown in Figure 4-2**.** Using the PDBePISA server the different interfaces, which are formed between the monomers, were determined.

Table 4-2 Pairwise percentage amino acid identity within tetramerization domain of p53 family proteins from different deuterostome species against p53/p73 proteins from *Ciona* species.

TD alignment % amino acid identity	vertebrate															invertebrate			
	p53					p63					p73								
	human p53 aa 324-367	**mouse p53** aa 318-361	**frog p53** aa 302-338	**zebrafish p53** aa 300-343	**e. shark p53** aa 313-356	**human p63** aa 397-440	**mouse p63** aa 397-440	**frog p63** aa 304-347	**zebrafish p63** aa 298-341	**e. shark p63** aa 421-464	**human p73** aa 351-394	**mouse p73** aa 343-386	**frog p73** aa 351-394	**zebrafish p73** aa 358-401	**e. shark p73** aa 356-399	**lancelet p53** aa 321-364	**lancelet p63/p73** aa 399-442	**fruit fly p53** aa	**roundworm p53** aa
***C.int.* p53/p73-a** aa 415-458	**41**	**43**	**38**	**41**	**36**	**41**	**41**	**39**	**43**	**39**	**39**	**39**	**43**	**36**	**43**	**34**	**39**	**X**	**X**
***C.sav.* p53/p73-a** aa 409-452	**39**	**41**	**38**	**39**	**34**	**43**	**43**	**39**	**43**	**39**	**36**	**36**	**41**	**34**	**43**	**34**	**32**	**X**	**X**
***C.int.* p53/p73-b** aa 375-418	**36**	**32**	**24**	**25**	**30**	**27**	**27**	**25**	**25**	**23**	**30**	**25**	**27**	**23**	**25**	**25**	**27**	**X**	**X**
***C.sav.* p53/p73-b** aa 351-394	**32**	**30**	**22**	**23**	**27**	**23**	**23**	**28**	**30**	**25**	**27**	**23**	**25**	**25**	**25**	**20**	**25**	**X**	**X**

Protein sequences belonging to the p53 family of transcription factors from the tunicate species *Ciona intestinalis* (*C.int.*) and *Ciona savignyi* (*C. sav.*) were aligned against those of the vertebrate species *Homo sapiens* (human), *Mus musculus* (mouse), *Xenopus tropicalis* (frog), *Danio rerio* (zebrafish) and *Callorhynchus milii* (e. shark), and the invertebrate species *Branchiostoma floridae* (lancelet). The full length protein sequences were aligned using clustalW2. The sections corresponding to the tetramerization domain (TD) were identified as those spanned by the secondary structure elements in the NMR solution structure of the TD of human p73 (pdb 2KBY). Percent amino acid identity was determined for pairwise alignments using Jalview and rounded. The accession numbers are listed in the appendix in section 7.1 and the amino acid (aa) numberings correspond to the respective canonical isoforms.

All isoforms of the *Ciona* p53/p73 proteins, which possess a full length N-terminus, do have a conserved binding motif for MDM2. These isoforms likewise correspond to those isoforms of the vertebrate proteins, which have a full length N-terminus and a transcription activation domain (TA), Figure 4-1.

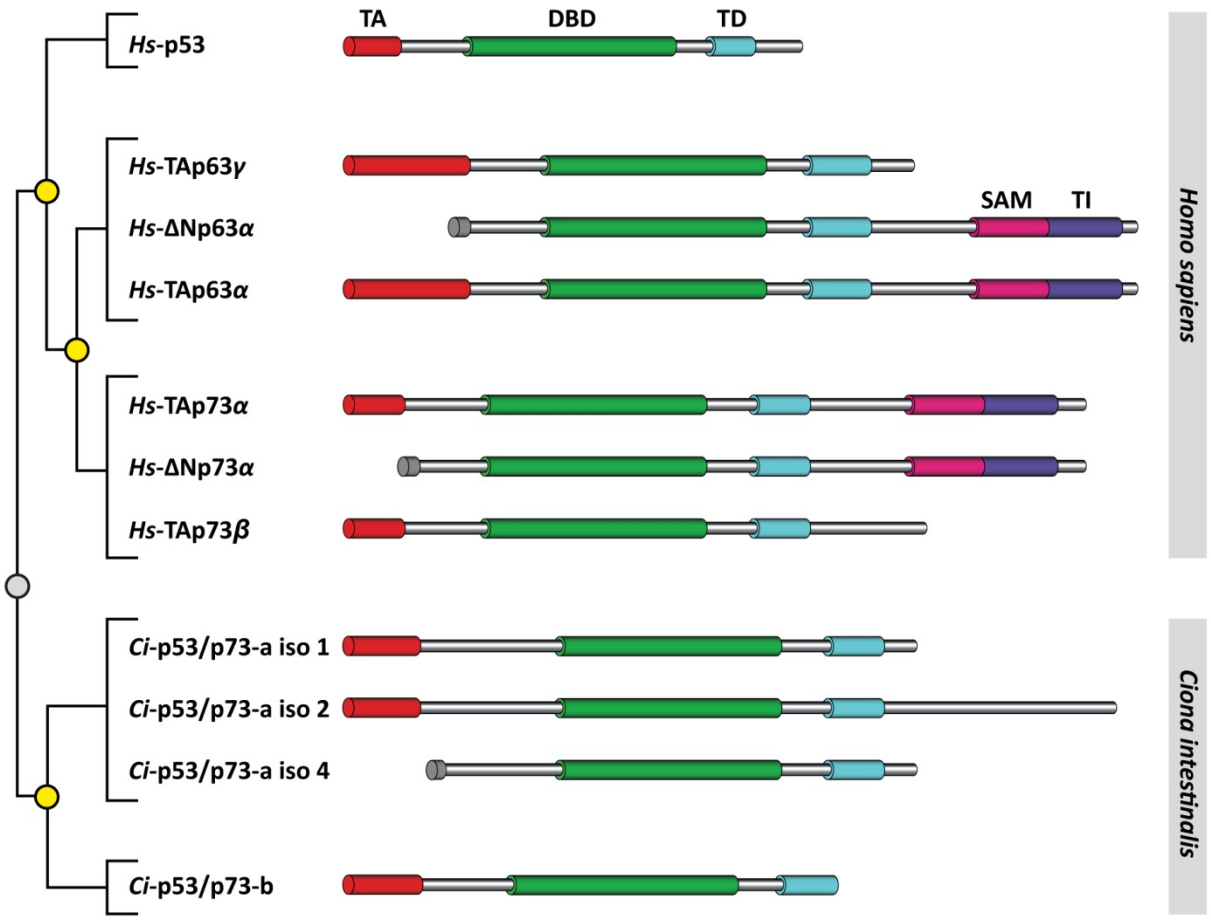


Figure 4-1 Phylogenetic tree and domain overview of the p53 family of transcription factors from *Homo sapiens* and the orthologous proteins from *Ciona intestinalis*. In vertebrates the p53 family of transcription factors consists of three paralogous proteins: p53, p63 and p73. These are represented here by the respective proteins from *Homo sapiens* (*Hs*). *Hs*-p53 is well-known for its importance as a tumor suppressor and consists of three domains: transactivation (TA) domain, DNA binding domain (DBD) and oligomerization domain, which is also known as tetramerization domain (TD). This canonical domain architecture is extended in *Hs*-p63 and *Hs*-p73, which belong to the same family of transcription factors. Due to different promoters both proteins exist in isoforms with differing N-termini. The TA isoforms possess a full transactivation domain, while the ΔN isoforms feature a truncated N-terminus. Alternative splicing results in variations in the length of the C-terminus. The alpha isoforms possess two additional domains: a sterile alpha motif (SAM) and a transactivation inhibitory (TI) domain. Two different proteins, which are encoded by the genome of the ascidian *Ciona intestinalis* (*Ci*), are evident on transcript level and belong to the p53 family.[140] The phylogenetic tree on the left depicts the homology relationship. All proteins derived from a common ancestor. Following speciation into two distinct evolutionary lines (marked with a gray circle) independent duplication events (marked with yellow circles) resulted in two (*Ci*) and three (vertebrates, *Hs*) genomic copies, respectively.[1,191]
The relative lengths of the depicted proteins match the differences both of total sequences and single domains. Domain annotations for the TA, SAM and TI domains of the human proteins were taken from the uniprot database. Those for the DBD and TD domains were done according to the corresponding protein structures and represent the sections spanning all secondary structure elements of the respective domains. The DBD and TD domains of the *Ciona* proteins were annotated based on the alignment to p53, p63 and p73 proteins from five representative vertebrate species (human, mouse, frog, zebrafish and elephant shark). The accession numbers of the respective proteins are listed in the appendix in section 7.1. The TA domains of the *Ciona* proteins were annotated based on the conserved binding motif for MDM2, but their indicated lengths could only be imprecisely predicted.

The isolated TA does not possess a defined tertiary structure or SSEs (personal communication) and with the exception of the MDM2 binding motif the alignment within the N-terminus is very low even across the vertebrate sequences. Data on the function of the N-terminal section of the *Ciona* p53/p73 proteins with respect to different truncations

are not available. Consequently the TA domains of the *Ciona* proteins had to be annotated solely based on the conserved binding motif for MDM2, and their lengths could only be imprecisely predicted.

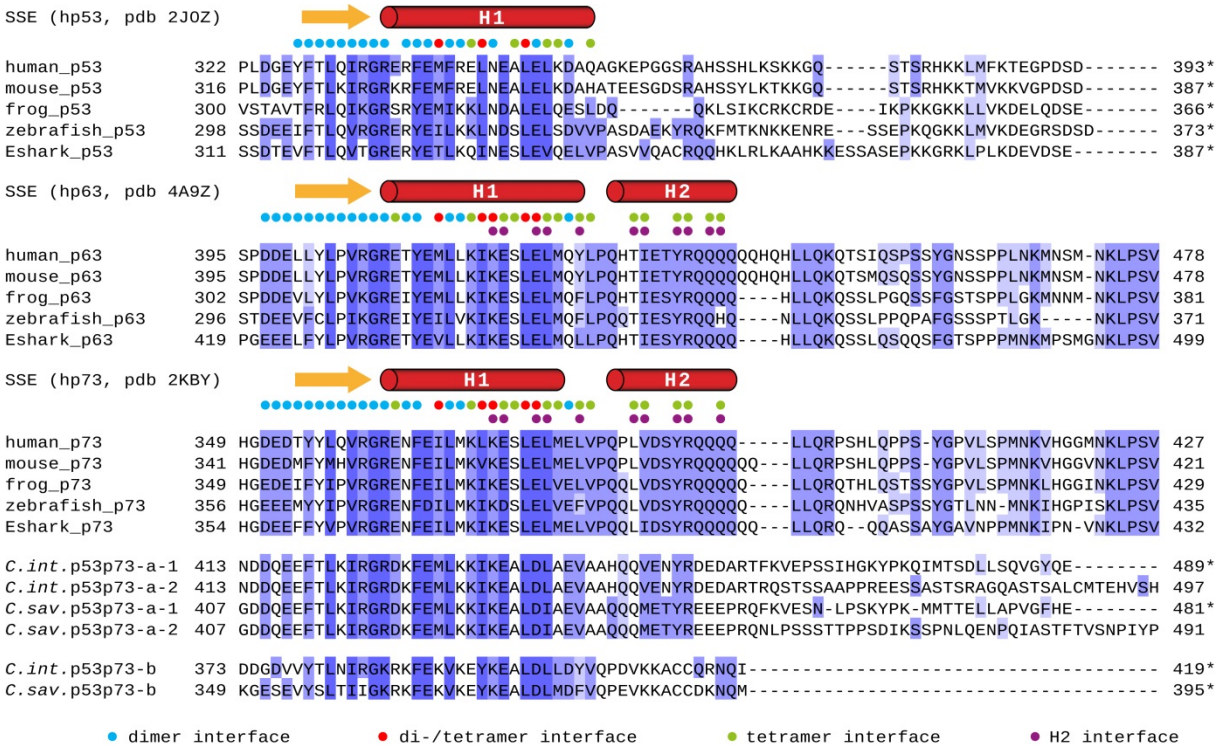


Figure 4-2 Alignment and functional annotation of protein sequences corresponding to the tetramerization domain of p53/p63/p73-like proteins from vertebrate species versus those from *Ciona intestinalis* (*C.int.*) and *Ciona savignyi* (*C.sav.*). The blue coloration of the alignment is according to percent of similar residues. Protein sequences are grouped by homology. An asterisk indicates the C-terminal end of the natural protein sequence. As a prefix to the vertebrate sequences the secondary structure elements (SSEs) of a crystal (hp63, pdb entry 4A9Z) or NMR solution structure (hp53, pdb entry 2J0Z; hp73, pdb entry 2KBY) of the respective human protein is depicted. The PDBePISA server was used to determine which residues make up the different interfaces. The accession numbers of the proteins are listed in the appendix in section 7.1.

With respect to human p53 the alpha isoforms of human p63 and p73 possess two additional domains. A sterile alpha motif (SAM) and a transcription inhibitory domain (TI) are located within the C-terminal section following the TD. The SAM domain is the third most conserved domain with amino acid identities ranging between 42 and 54% for the comparison of vertebrate p63 and p73 proteins. The pairwise identity scores for the corresponding sequences of the five examined vertebrate species are listed in 7.1.8.3. The in group identity scores are even higher with up to 93% between p73 and up to 100% between p63 proteins. So far no structural information is available for the TI (databases and personal communication). The hydrophobic sequence of this domain makes it prone to the formation of aggregates hampering studies of the isolated domain. As a consequence the boundaries

for this domain differ a lot in the literature. For this reason no isolated alignments were calculated.

4.2. Structural analysis of the TD and C-terminal sections of *Ciona intestinalis* p53/p73-a

4.2.1. Detailed sequence comparisons of the TD and C-terminal sections of the *Ciona intestinalis* and *Ciona savignyi* p53/p73-a isoforms

As indicated in the annotation of the gene locus of *C.int.* p53/p73-a, in Figure 1-9, several isoforms with differing C-termini result from a splice site located right after the TD. The position of the splice site is conserved in the respective gene in *Ciona savignyi*. Figure 4-3 shows the alignment of the sections from the respective first isoforms of the p53/p73-a proteins from *C.int.* and *C.sav.*, which correspond to the TD and the C-terminus. Throughout the full length protein sequences the two proteins share 75,6% amino acid identity with maxima within the sections corresponding to the DBD (88% identity) and the TD (84% identity). Within the isoform specific C-terminal sequences the degree of amino acid identity is only 55,2%, but some additional residues are functionally conserved.

As indicated in the annotation of the gene locus of *C.int.* p53/p73-a, the sequence of isoform 3 ends at the splice site after the TD and isoform 4 shares the same C-terminus with isoform 1; Figure 1-9. This made separate alignments of these isoforms obsolete.

Figure 4-3 Alignment of the tetramerization domain and C-terminal sequences of p53/p73-a isoform 1 from *Ciona intestinalis* (*C.int.*) and *Ciona savigny* (*C.sav.*) The blue bar above the sequences indicates the section spanned by the secondary structure elements in the tetramerization domain of human p73. For both proteins the sequences of isoform 1 match those of the corresponding isoform 2 from the N-terminus to the splice site located right after the TD. The position, which corresponds to the splice site, is marker with a red arrow. An asterisk indicates the C-terminal end of the natural protein sequence.

As depicted in Figure 4-1 the C-terminus of *C.int.* p53/p73-a isoform 2 is elongated with respect to isoform 1 as a result of different splicing at the before mentioned site. For *C.int.* both isoforms are evident on transcript level,[140] and transcripts for both isoforms were as well predicted based on the genomic sequence of the closely related species *C.sav.*

The sections of the *Ciona* p53/p73-a isoform 2 proteins, which correspond to their elongated C-termini, do not align to any of the p53 family proteins from vertebrate species or lancelet. A blast search against the ncbi databases of non-redundant protein sequences or reference proteins returned no aligning sequences beside the one from the other *Ciona* species, respectively. An according search with the respective nucleotide sequences as queries did as well not return any hit. The protein sequences, which correspond to the elongated C-termini of *C.int.* and *C.sav.* p53/p73-a isoform 2, seem to be without analogy. Consequently the sections of the two proteins, which correspond to the TDs and the elongated C-termini, were aligned separately against another as shown in Figure 4-4. The sequences C-terminal to the splice site share 41% identity.

```
C.sav.p53p73-a-2  407 GDDQEEFTLKIRGRDKFEMLKKIKEALDIAEVAAQQQMETYREEEPRQNLPSSSTTPPSDIKSSPNLQENPQIASTFTVSNPITP  491
C.int.p53p73-a-2  413 NDDQEEFTLKIRGRDKFEMLKKIKEALDLAEVAAHQQVENYRDEDARTRQSTSSAAPPREESSASTSRAGQASTSALCMTEHVSH  497

C.sav.p53p73-a-2  492 DSQGFSLTPTHSSHPLTATPNLDLP--------TGNS---------PYWSVDADMGCDVPDTNQPSTSA-GLTQWAEQPNTDCTP  558
C.int.p53p73-a-2  498 GSQGFSLTPTHSSHPLTVTPTLEPPIVEGKTPYTHNTSAYPGMVDQSYWSTESDLGYDIPDTKQPTTTGNGLPNWPNQESLDSNP  582

C.sav.p53p73-a-2  559 RLVRDLSGLSTTSSLSSSQMNS-SGRFVTRVTLRQRVTLN-QEPTTDFWKPNASEATNGVKVERKD-DVESPPITGEESFEFL   638*
C.int.p53p73-a-2  583 KLLRNISGLSSVSSLTSSQLSSQTGRFVTRVTLRQTVALNPDMPNSQRWDHDTTSDNE--DVTRNDVKREHDNVTNTDSFDFL   663*
```

conserved in homologous proteins from deuterostome species (degree of conservation > >)

conserved only in C.int. and C.sav.

Figure 4-4 Alignment of the tetramerization domain and C-terminal sequences of p53/p73-a isoform 2 from *Ciona intestinalis* (*C.int.*) and *Ciona savigny* (*C.sav.*) The blue bar above the sequences indicates the section spanned by the secondary structure elements in the tetramerization domain of human p73. For both proteins the sequences of isoform 2 match those of the corresponding isoform 1 from the N-terminus to the splice site located right after the TD. The position, which corresponds to the splice site, is marker with a red arrow. An asterisk indicates the C-terminal end of the natural protein sequence.

4.2.2. C-terminus specific to *Ciona intestinalis* p53/p73-a isoform 2 is unstructured

The C-terminus of *C.int.* p53/p73-a isoform 2 is considerably extended with respect to isoform 1, 3 and 4. A corresponding transcript is as well predicted from the sequence of the respective *Ciona savignyi* gene, but the sequences are without analogy in the ncbi database. To investigate whether the isoform specific C-terminus contains secondary or tertiary structure at first the relative hydrophobicity along the primary sequence was determined using ProtScale (expasy) with the hydrophobicity indices at pH 7.5 determined by HPLC.[192] This was used to identify those regions, which could contain potential domains and guided the design of the expression constructs. The sections, which are depicted in Figure 4-5, were cloned into the expression vector pBH4. Expression was tried out for all constructs in *E.coli* NEB T7 express cells. As the DNA sequences had not been optimized with respect to the different codon usage in *E.coli* the cells were co-transformed with the pRARE co-plasmid. Many longer sequences were not soluble expressed even when the incubation temperature during induction of expression was only 18°C, and some longer constructs, which did not comprise the region corresponding to the TD, completely failed to express. The co-expression with chaperones in *E.coli* Arctic express (Stratagene) or the recloning into an expression vector, from which the respective sequences were expressed as a fusion protein with (maltose binding protein) MBP-linker-[8xHistag]-[TEV cleavage site] at the N-terminus, did not improve the problem. The majority of the shorter constructs, which were contained in the supernatant after cell lysis, showed a high degree of fragmented products being the result of either incomplete expression or, more likely, proteolytic digestion.

Finally only three constructs could be evaluated for structural investigations. The corresponding sequences are shown in Figure 4-6 as an alignment towards the section of *C.int.* p53/p73-a isoform 2, which spans the TD and the isoform specific elongated C-terminus. Expression, cell lysis as well as the purification by IMAC, TEV cleavage and reverse IMAC followed by gel filtration were performed according to the standard protocols for this project as described in the method section. The proteins were purified to ≥ 95% as judged from Coomassie stained Tricine-SDS-PAGE gels.

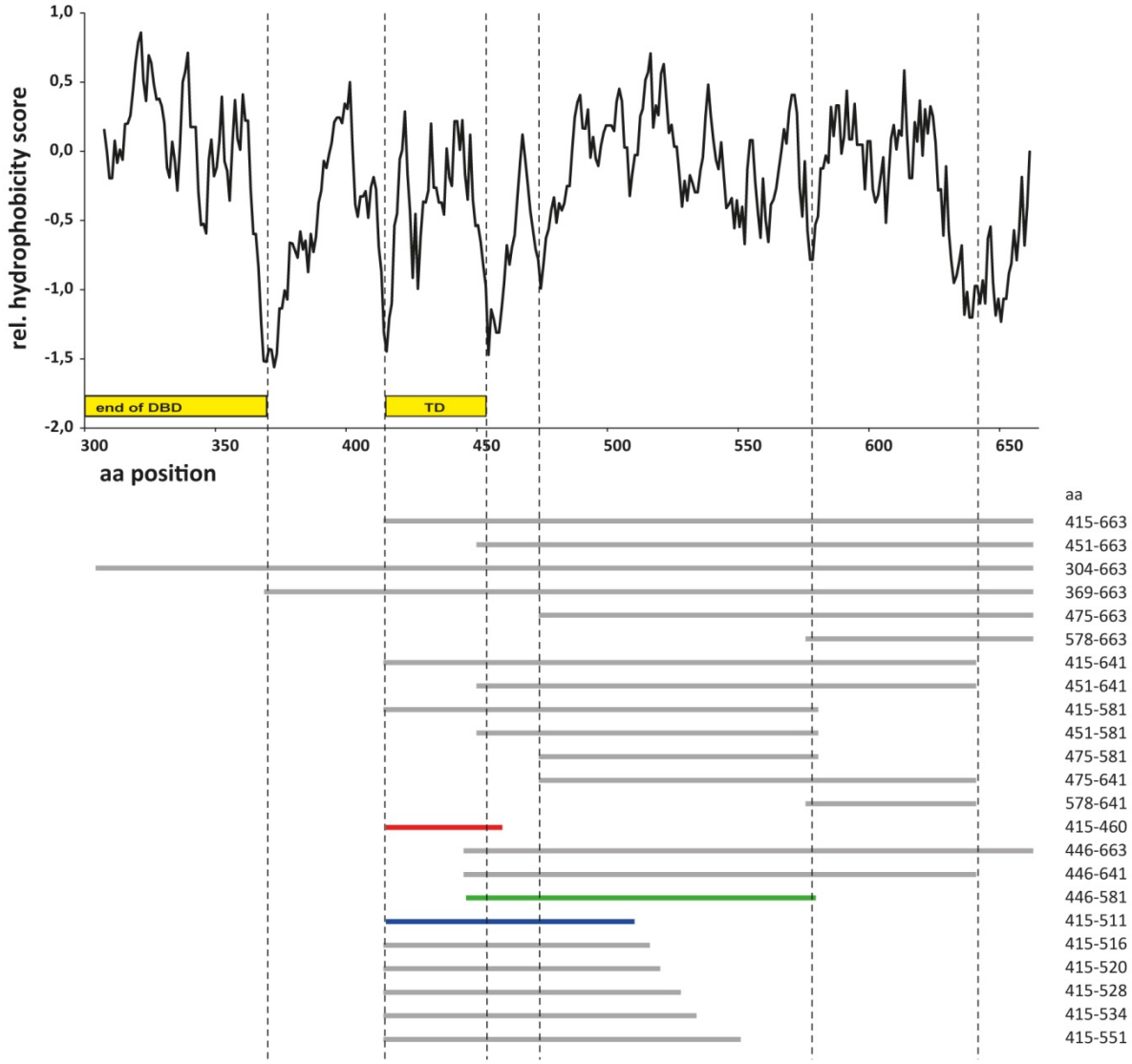


Figure 4-5 Relative hydrophobicity along the primary sequence of *Ciona intestinalis* p53/p73-a isoform 2 (C-terminal half) compared to the sections screened for expression. The relative hydrophobicity along the primary sequence of the C-terminal half of isoform 2 was calculated using ProtScale (expasy) with the hydrophobicity indices at pH 7.5 determined by HPLC and standard settings for window size and weighting.[192] Depicted below are the sections, which were cloned into the expression vector pHB4. Only the three constructs which are colored in red, green and blue could be validated for structural investigations.

To investigate whether the C-terminus of isoform 2 alone contains any secondary structure the construct spanning aa 446-581, which can be stably expressed without any degradation products, was analyzed by circular dichroism (CD). The spectrum is shown in Figure 4-7.

```
C.int.p53p73-a-2  413 NDDQEEFTLKIRGRDKFEMLKKIKEALDLAEVAAHQQVENYRDEDARTRQSTSSAAPPREESSASTSRAGQASTSALCMTEHVSH 497
aa 446-581        446                                GSAHQQVENYRDEDARTRQSTSSAAPPREESSASTSRAGQASTSALCMTEHVSH 497
aa 415-511        415 GSDQEEFTLKIRGRDKFEMLKKIKEALDLAEVAAHQQVENYRDEDARTRQSTSSAAPPREESSASTSRAGQASTSALCMTEHVSH 497
aa 415-460        415 GSDQEEFTLKIRGRDKFEMLKKIKEALDLAEVAAHQQVENYRDEDART                                        460*

C.int.p53p73-a-2  498 GSQGFSLTPTHSSHPLTVTPTLEPPIVEGKTPYTHNTSAYPGMVDQSYWSTESDLGYDIPDTKQPTTTGNGLPNWPNQESLDSNP 582
aa 446-581        498 GSQGFSLTPTHSSHPLTVTPTLEPPIVEGKTPYTHNTSAYPGMVDQSYWSTESDLGYDIPDTKQPTTTGNGLPNWPNQESLDSN  581*
aa 415-511        498 GSQGFSLTPTHSSH                                                                          511*

C.int.p53p73-a-2  583 KLLRNISGLSSVSSLTSSQLSSQTGRFVTRVTLRQTVALNPDMPNSQRWDHDTTSDNEDVTRNDVKREHDNVTNTDSFDFL     663*
```

Figure 4-6 Alignment of the corresponding construct sequences towards the section of *Ciona intestinalis* (*C.int.*) p53/p73-a isoform 2 encompassing the tetramerization domain (TD) and the elongated C-terminus. The blue bar above the sequences indicates the section spanned by the secondary structure elements in the tetramerization domain of human p73. The position of the isoform specific splice site right after the TD is marked with a red arrow. An asterisk indicates the C-terminal end of the natural protein sequence or the respective construct. The lower three lines of the alignment show the investigated constructs with the coloration used in the following spectra overlays. For technical reasons all constructs have two additional residues at the N-terminus, which are marked with grey background. This GS dipeptide is a remnant from the N-terminal TEV cleavage site and does not align to the protein sequence under investigation. For this reason it is left out in the numbering and description of the constructs.

The construct corresponding to aa 446-581 showed a CD spectrum typical for random coil with a broad minimum at 195-200 nm. In the presence of α-helices the spectrum would be dominated by this component showing strong minima at 222 and 208-209 nm. β-strands would result in a maximum at 190-195 nm, but with a much smaller molar ellipticity compared to α-helices. Hence, the presence of α-helices could be excluded based on the CD spectrum. To ensure, that the construct does not contain any β-strand conformation a [^{15}N, ^{1}H]-TROSY spectrum was recorded, which is presented in Figure 4-8.

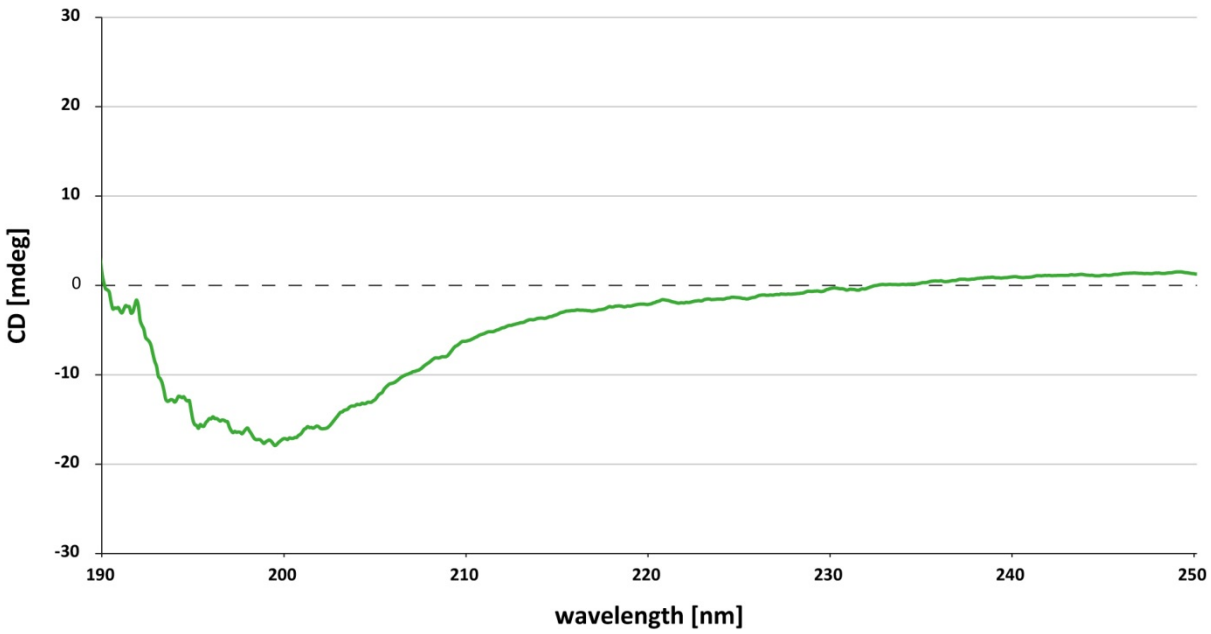


Figure 4-7 Circular dichroism spectrum of *Ciona intestinalis* p53/p73-a isoform 2 residues 446-581. The spectrum was measured at 298 K with a protein concentration of 5 µM buffered in 50 mM NaH_2PO_4 pH 7.0.

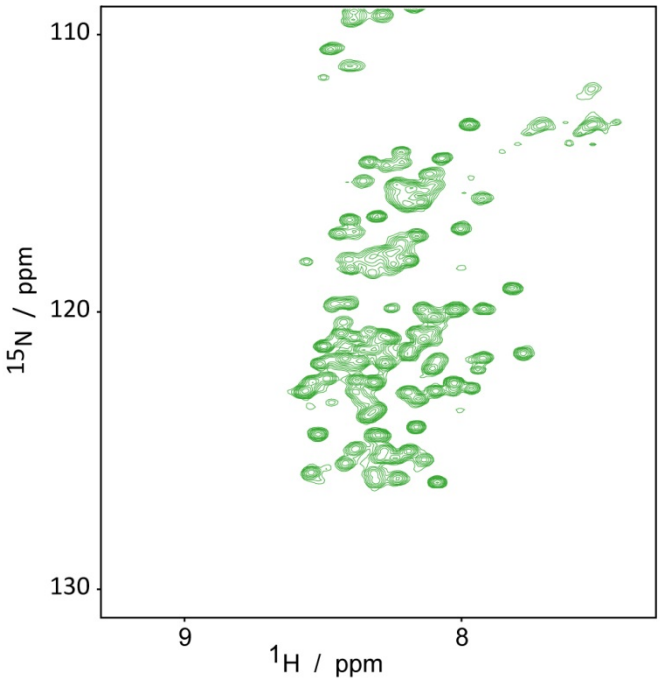


Figure 4-8 [^{15}N, ^{1}H]-TROSY spectrum of the construct encompassing residues 446-581 of *Ciona intestinalis* p53/p73-a isoform 2. The spectrum was recorded at 25°C.

The TROSY spectrum clearly confirmed the absence of β-strands. All peaks clustered in the range of 7.7 - 8.6 ppm (^{1}H frequency), but ^{15}N, ^{1}H cross peaks from the backbone amide of residues with β-strand conformation typically have ^{1}H frequency > 8.6 ppm and are generally more dispersed as well.

Next the two constructs were evaluated, which comprise the sequence aligning to the tetramerization domain of vertebrate p53-like proteins. [^{15}N, ^{1}H]-TROSY spectra were recorded and overlaid with one another as shown in Figure 4-9. Both spectra contain peaks, which have ^{1}H frequencies > 8.6 ppm indicating the presence of at least one β-sheet. This would be expected as the very N-terminus, common to both constructs, contains the sequence corresponding to the β-sheet in the TD of human p53.

The splice site, which results in the massive C-terminal elongation of isoform 2 with respect to isoform 1, 3 and 4, resides right after the sequence spanned by SSEs in the TD of human p73 or human p63. Since that the objective in the structural investigation of isoform 2 was to determine if the isoform specific C-terminus interacts with the TD. The construct encompassing aa 415-460, which is depicted in red, ends with the last residue common to all isoforms. All peaks in the corresponding [^{15}N, ^{1}H]-TROSY spectrum (red), with ^{1}H frequencies > 8.6 ppm, are perfectly matched by peaks from the spectrum of the construct elongated up to aa 511 (blue). From this it can already be excluded that the section corresponding to aa 461-511 interacts with the β-sheet of the TD. In total 26 peaks perfectly overlay with their

respective counterparts and for many of the remaining 20 peaks this could only not be determined due to overlap. As a result of the elongation at the C-terminus 5 to 7 peaks would anyhow have been expected to be effected by massive shifts. This leaves some 13 to 15 peaks, for which the corresponding residues are expected to mainly locate within aa 440-455. In the alignment towards human p73 this section corresponds to the last residues of helix one and the first half of the second helix. Taking the structure of human p73 TD into account massive shifts as well on the β-sheet would be expected if this section would adopt different conformations depending on the C-terminus. But this is obviously not the case. Hence, the conformation of the TD is unaffected by the C-terminal following sequence of aa 461-511, which is specific for isoform 2.

This left the question open if the section corresponding to aa 461-511 folds any independent SSEs. With the exception of only a single one all peaks from the spectrum of the C-terminal elongated construct, which lack an overlaying peak, locate with the range of 7.7 - 8.6 ppm (^{1}H frequency). This excludes β-strands and is indicative for random coil, although it does not principally exclude the presence of rather short α-helical segments.

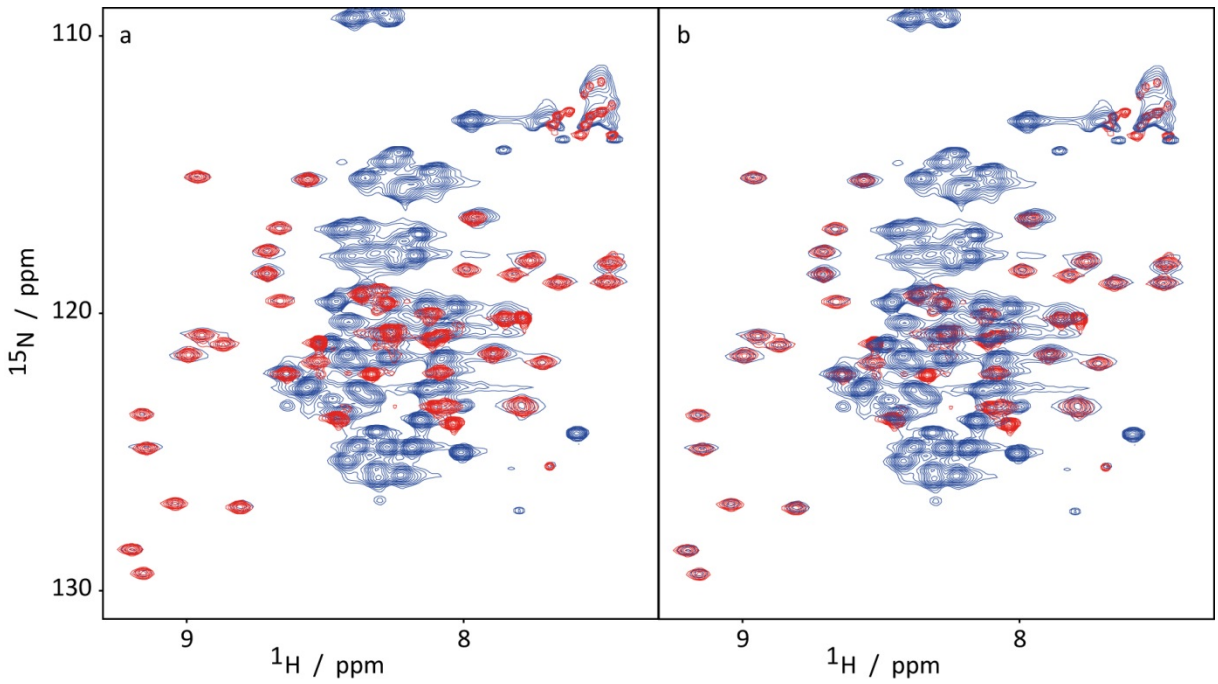


Figure 4-9 Overlay of [^{15}N, ^{1}H]-TROSY spectra of constructs encompassing either aa 415-460 or aa 415-511 of *Ciona intestinalis* p53/p73-a isoform 2. (a) Both constructs start with residue 415 and encompass the tetramerization domain. The shorter construct ends with aa 460, which is the last aa before the isoform specific splice site of the gene. The spectrum of the longer construct (aa 415-489) is depicted in blue and overlaid with the one of the shorter construct depicted in red. (b) The overlay is reversed. Both spectra were recorded at 25°C.

Circular dichroism spectroscopy was used to exclude as well the presence of additional α-helices within the C-terminal segment. Wavelength scans from both constructs were recorded from samples with identical molar protein concentrations and overlaid for comparison as presented in Figure 4-10. The spectrum of the construct encompassing the C-terminal elongation appears in total flatter when compared to the spectrum of the shorter construct. The maximum at 190-195 nm is lower and the minima at 208 and 222 nm are not only less pronounced but their ratio as well changed at the costs of the minimum at 222 nm. Denotative is also that the global minimum shifted from 208 to 204-205 nm. All this can best be explained by the addition of pure random coil spectral component as a result of the C-terminal elongation. Taken together NMR and CD clearly show that the isoform 2 specific sequence, which follows after the TD, is pure random coil and does not interact with the TD.

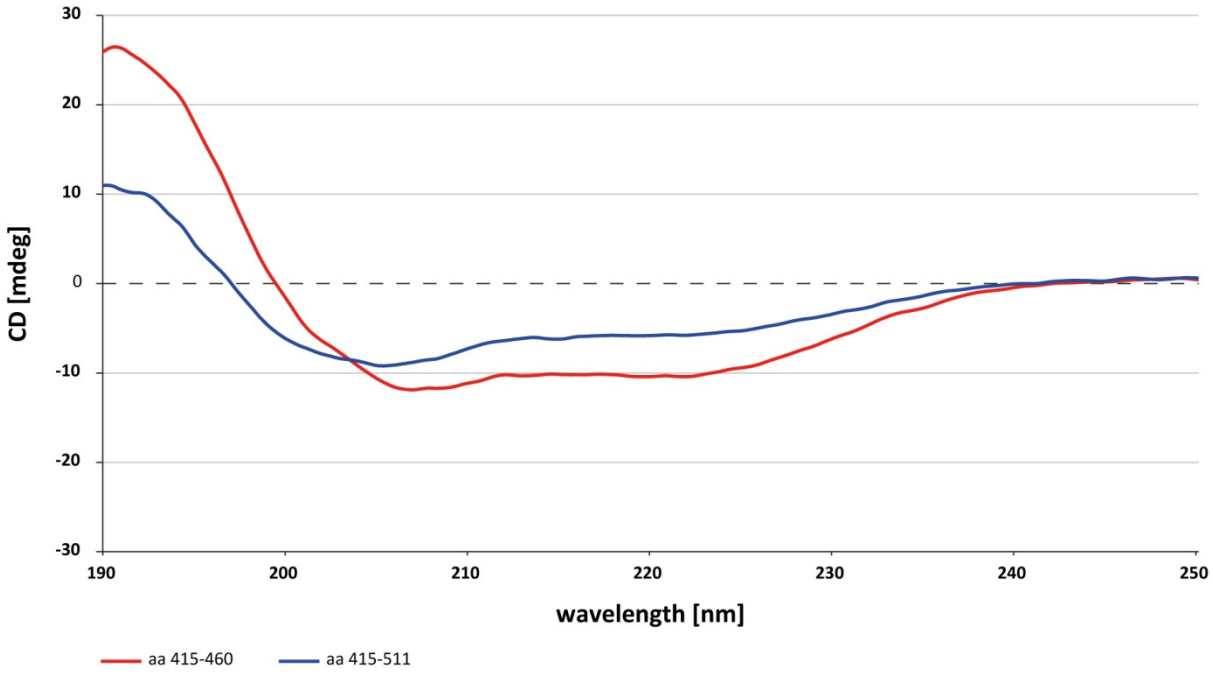


Figure 4-10 Circular dichroism spectra of constructs encompassing either aa 415-460 or aa 415-511 of *Ciona intestinalis* p53/p73-a isoform 2. The spectra were measured at 298 K with protein concentration of 10 µM buffered in 50 mM NaH_2PO_4 pH 7.0.

4.2.3. Determination of section with secondary structure in C-termini of *Ciona int.* p53/p73-a isoform 1, 3 and 4

Already from the length of the sequence following the TD it is obvious that isoform 1, 3 and 4 of *C.int.* p53/p73-a cannot contain a larger additional domain C-terminal to the TD such as the SAM domain in the alpha isoforms of vertebrate p63 or p73. To determine which section of the C-terminus of *C.int.* p53/p73-a isoform 1, 3 or 4 possesses secondary structure several differently truncated constructs were cloned into pBH4. The C-termini of isoform 1 and 4 are identical. Isoform 3 is truncated and ends at position 450. As isoform 1 and 3 differ from one another only with respect of this truncation the numbering of the two isoforms is the same. The following aa numbers always refer to isoform 1. In Figure 4-11 the constructs are aligned against the C-terminus of isoform 1. The first two aa, GS, at the N-terminus of the constructs result from the C-terminal portion of the TEV cleavage site and, hence, do not align. For this reason they are left out in the description and the comparison of the different truncations. In the recorded NMR spectra their backbone NH cross peaks were anyhow not detectable. The construct corresponding to aa 415-511 was cloned first using the natural cDNA as template, hence, was not codon optimized. This resulted in a low expression yield and the expression of partially incomplete proteins due to interrupted translation breakdown. The side directed mutagenesis of the codons for Leu 481 and Leu 482 to “ctg” solved both problems at once and the construct was used as the template for all other cloning related to isoform 1.

```
C.int.p53p73-a-1  400 QTAPAKRKCAMNGNDDQEEFTLKIRGRDKFEMLKKIKEALDLAEVAAHQQVENYRDEDARTFKVEPSSIHGKYPKQIMTSDLLSQVGYQE 489*
aa 446-581        405    GSKRKCAMNGNDDQEEFTLKIRGRDKFEMLKKIKEALDLAEVAAHQQVENYRDEDARTFKVEPSSIHGKYPKQIMTSDLLSQVGYQE 489*
aa 415-511        415              GSDQEEFTLKIRGRDKFEMLKKIKEALDLAEVAAHQQVENYRDEDARTFKVEPSSIHGKYPKQIMTSDLLSQVGYQE 489*
aa 415-464        415              GSDQEEFTLKIRGRDKFEMLKKIKEALDLAEVAAHQQVENYRDEDARTFKVE                           464
aa 415-460        415              GSDQEEFTLKIRGRDKFEMLKKIKEALDLAEVAAHQQVENYRDEDART                               460
aa 415-450        415              GSDQEEFTLKIRGRDKFEMLKKIKEALDLAEVAAHQQV                                         450
aa 415-445        415              GSDQEEFTLKIRGRDKFEMLKKIKEALDLAEVA                                              445
```

Figure 4-11 Alignment of C-termini of *Ciona intestinalis* p53/p73-a isoform 1 versus corresponding constructs. The blue bar above the sequences indicates the section spanned by the secondary structure elements in the tetramerization domain of human p73. The position of the isoform specific splice site right after the TD is marked with a red arrow. An asterisk indicates the C-terminal end of the natural protein sequence. The upper line shows the C-terminal aa sequences of the isoforms 1. The lower six lines show the investigated constructs with the coloration used in the following spectra overlays and analysis of chemical shift perturbations. For technical reasons all constructs have two additional residues at the N-terminus, which are marked with grey background. This GS dipeptide is a remnant from the N-terminal TEV cleavage site and does not align to the protein sequence under investigation. For this reason it is left out in the numbering and description of the constructs.

The expression in NEB T7 express cells with pRARE co-plasmid, the cell lysis as well as the purification by IMAC, TEV cleavage and reverse IMAC were performed following the

standard protocols for this project as described in the method section. All proteins were purified to ≥ 98% as judged from Coomassie stained Tricine-SDS-PAGE gels. Gel filtration using a preparative scale Superdex75™ 16/60 column with settings as described was included as a last purification set (polishing) and to verify homogeneity. An analytical scale Superdex75™ 10/30 column was used to determine the oligomeric state and to prepare samples in different buffers for NMR and CD. All NMR spectra for this subproject were recorded from samples with ≥ 500 µM monomer concentration and at a temperature of 25°C, if not indicated differently.

The spectrum of the construct encompassing aa 415-489 is shown in Figure 4-12. Under the conditions applied certain NH cross peaks had much higher intensities compared to the remaining. These intensive peaks clustered mainly within 7.7-8.6 ppm (^{1}H frequency) indicating that the construct encompasses some flexible unstructured sections.

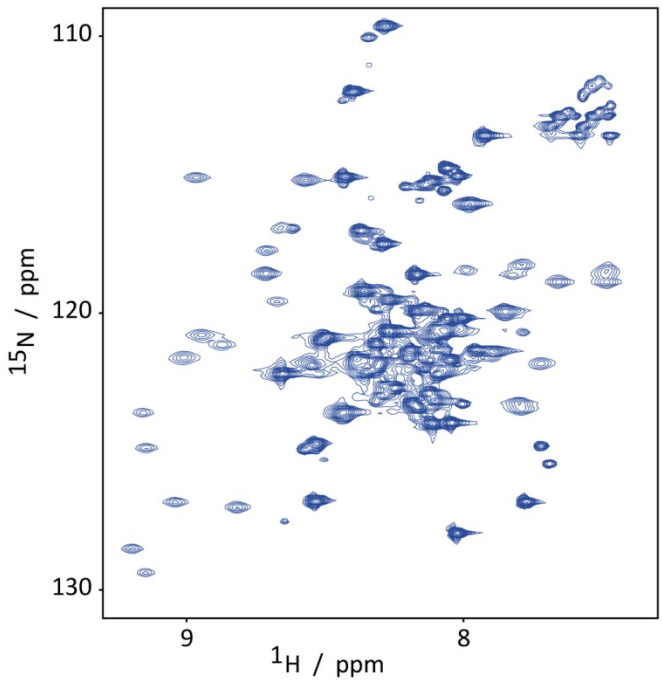


Figure 4-12 [^{15}N, ^{1}H]-TROSY spectrum of aa 415-489 of *Ciona intestinalis* p53/p73-a isoform 1. The spectrum was recorded at 25°C.

To facilitate a valid comparison of the other differently truncated constructs to this one, additional 3D TROSY-based HNCACB and HNCA spectra were recorded, but the very different peak intensities prevented a complete assignment of the protein backbone; data not shown. For this reason HNCACB spectra of the C-terminal shorter constructs (aa 415-464), (aa 415-

460) and (aa 415-450) were recorded as well. This enabled a complete assignment of all four constructs and advanced the following analysis of the effect of the different truncations.

The overlay of the spectra of the two constructs, which encompass the complete C-terminus, is shown in Figure 4-13. The spectrum corresponding to aa 405-489 revealed that the sequence stretch upstream of Asp 415 is most likely unstructured. With respect to aa 415-489 additional peaks clustered within 7.7 - 8.6 ppm (^{1}H frequency). This is typical for residues with no distinctive secondary structure. The peaks had as high intensities as those peaks, which could be assigned to the unstructured C-terminal end. Although many peaks in the spectrum of the N-terminal longer construct (aa 405-489) are much less resolved, it was partially possible to translate the assignment.

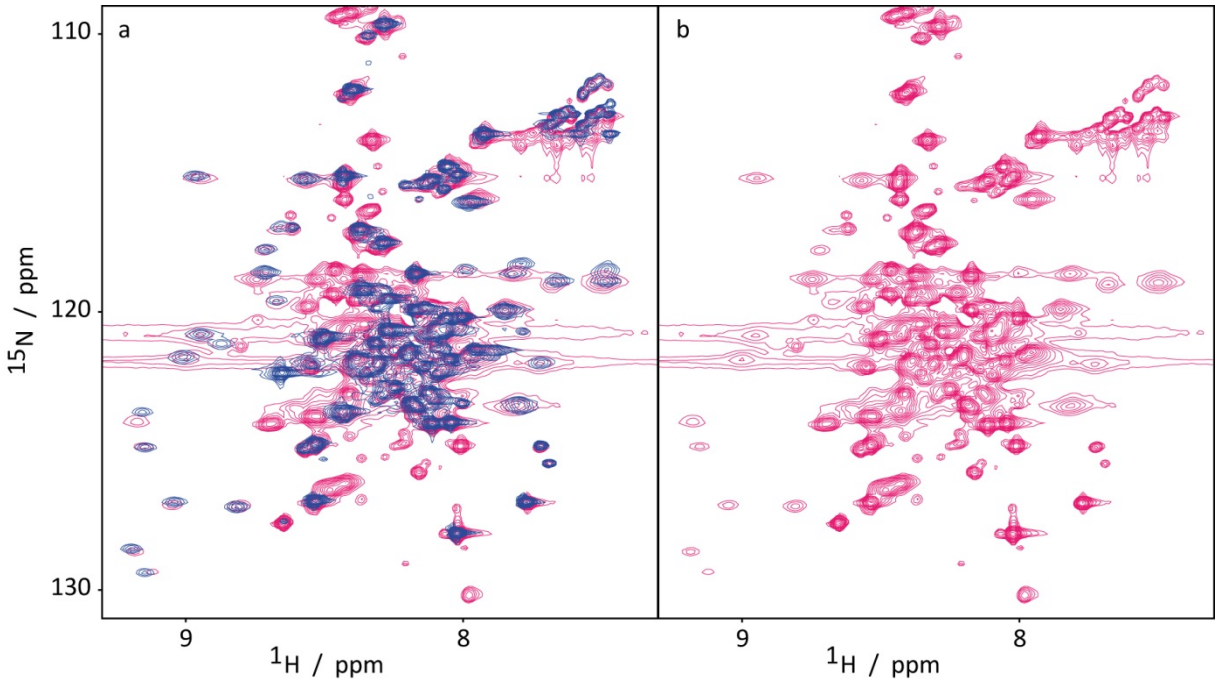


Figure 4-13 Overlay of [15N, 1H]-TROSY spectra of two different constructs each encompassing the TD and the remaining C-terminus of *Ciona intestinalis* p53/p73-a isoform 1 and 4. (a) The spectrum of the N-terminal longer construct (aa 405 – 489) is shown in pink. The spectrum of the N-terminal shorter construct (aa 415-489) is shown in blue. (b) Only the spectrum of the construct with the longer N-terminus is shown. Both spectra were recorded at 25°C.

The chemical shift perturbations (CSPs) of aa 405-489 with respect to aa 415-489 are depicted in Figure 4-14. The CSPs are relatively constant along the protein sequence, albeit, frequently about 0.1. This was attributed to the lower resolution of the spectrum corresponding to aa 405-489 and higher overlap. As higher CSPs cluster nowhere along the sequence, it can be excluded that the N-terminal elongation of aa 405-489 with respect to

aa 415-489 encompasses residues, which make contacts to any residue following later in the sequence.

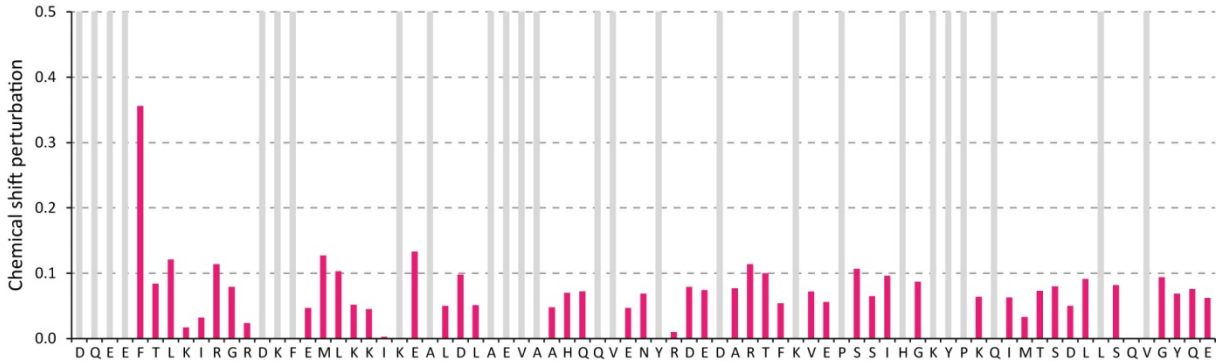


Figure 4-14 Chemical shift perturbation of construct aa 405-489 with respect to construct aa 415-489. The construct corresponding to aa 415-489 of *Ciona intestinalis* p53/p73-a isoform 1 was elongated at the N-terminus by ten residues. Wherever possible the assignment was transmitted and the NH chemical shift perturbation was calculated as described under 3.3.8. Residues, which could not be assigned, are marked with grey bars.

Figure 4-15 shows the overlay of the [^{15}N, ^{1}H]-TROSY spectra corresponding to aa 415-489 and the C-terminal truncated construct corresponding to aa 415-464 of *C.int.* p53/p73-a isoform 1.

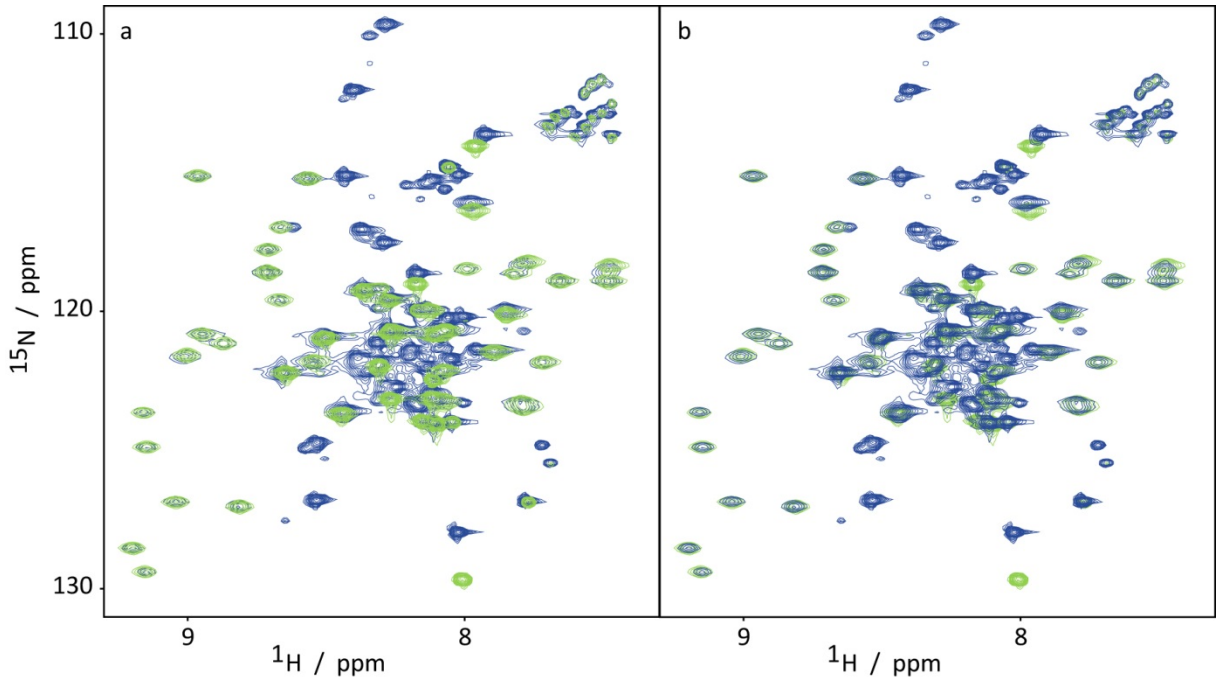


Figure 4-15 Overlay of [^{15}N, ^{1}H]-TROSY spectra of constructs encompassing either aa 415-464 or aa 415-489 of *Ciona intestinalis* p53/p73-a isoform 1. (a) Both constructs start with residue 415. The spectrum of the construct encompassing the remaining C-terminus is shown in blue (aa 415 – 489). The spectrum of the C-terminal truncated construct (aa 415-464) is overlaid and depicted in green. (b) The overlay is reversed. Both spectra were recorded at 25°C.

Almost every peak within the spectrum of the longer construct, which has a lower intensity and locates outside the range of 7.7 - 8.6 ppm (^{1}H frequency), is perfectly matched by one peak from the spectrum of the C-terminally truncated (aa 415-464) construct. As shown in Figure 4-16, the truncation results in only minor CSPs. Only the last 3 to 6 residues show CSPs higher than 0.1, but this has to be expected as those residues are close to the truncation.

Hereby it was confirmed that residues 465-489 do not interact with residues 415-464.

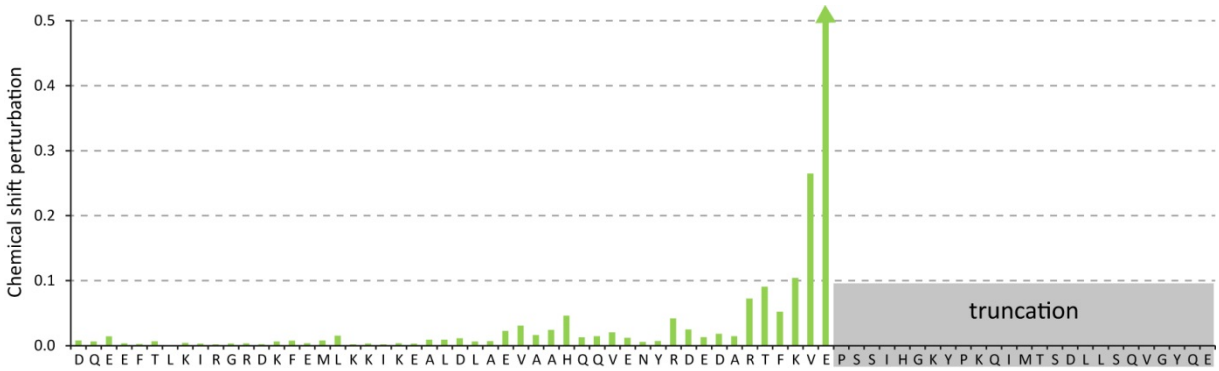


Figure 4-16 Chemical shift perturbation upon C-terminal truncation after aa 464 of construct encompassing aa 415-489 of *Ciona intestinalis* p53/p73-a isoform 1. Peaks in both spectra shown in Figure 4-15 were assigned to their corresponding residues. The NH chemical shift perturbation was calculated as described under 3.3.8. Chemical shift perturbations exceeding 0.5 are marked with arrow heads.

The CSP analyses of the three shorter truncations are depicted in Figure 4-17. Since aa 460 is the last residue before the splice site causing the different C-termini, the construct encompassing aa 415-460 corresponds to the sequence common to all isoforms. Compared to the construct with the full length C-terminus the truncation after aa 460 resulted in significant CSPs not only of residues close to the truncation site. The section of aa 443-456 showed as well CSPs of up to 0.1. This was especially interesting as this section partially aligns to the residues right after the first helix and up to the middle of the second helix in the TD of human p73, as shown in Figure 4-2. But the construct still encompasses the complete section, which is spanned by the SSEs in the TD of human p73. Therefore CSPs like the ones observed would only be expected for the case that in this shorter truncation some residues would be missing, which would otherwise at least with some respect influence the fold in the section corresponding to a potential second helix. Hence, when comparing this analysis to the one for the truncation ending with aa 464, Figure 4-16, it is obvious that the first four

residues from the specific C-terminal sequence of isoform 1 and 4 should be comprised in constructs, which are used for structural investigations of a truncated sequence.

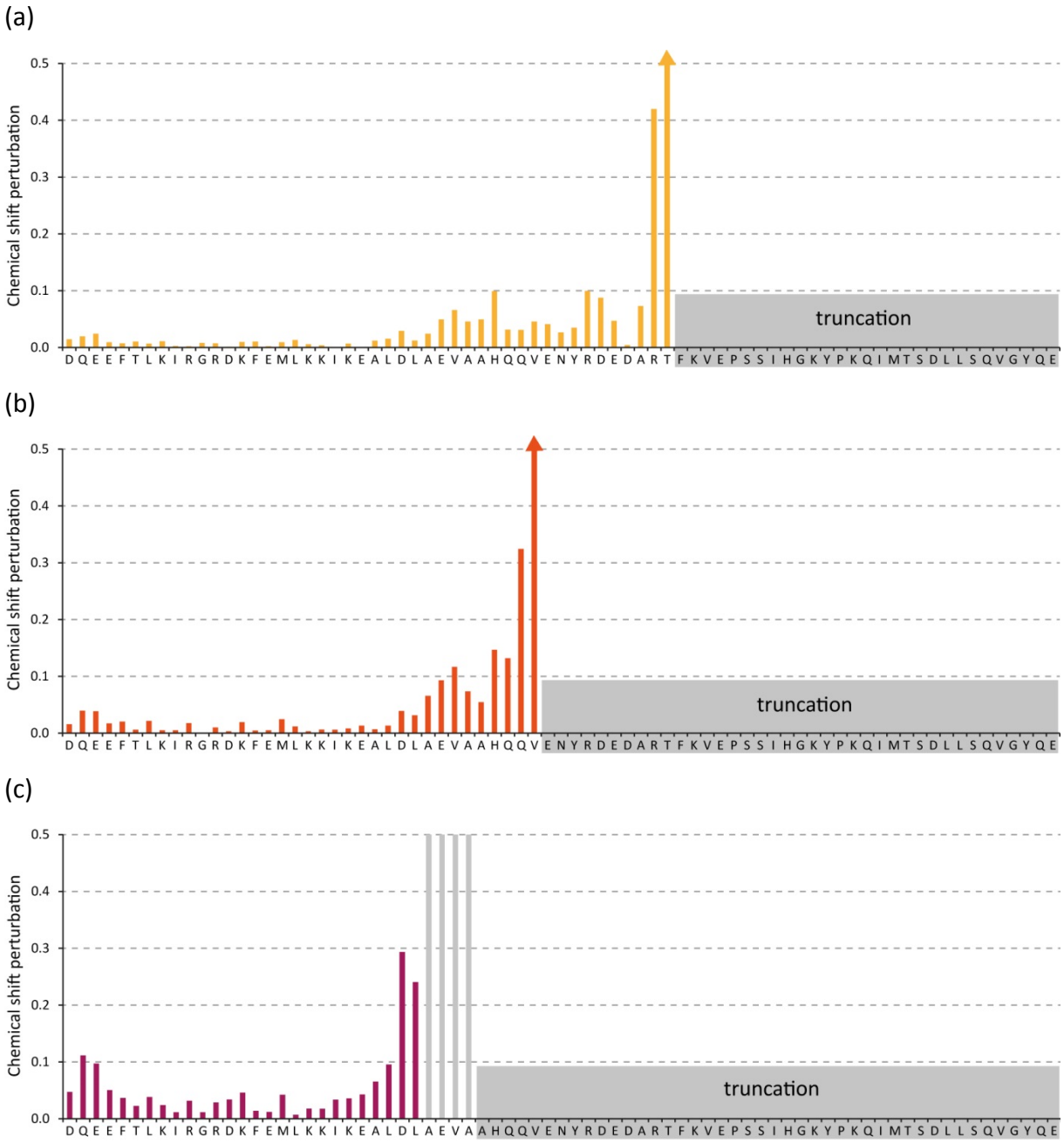


Figure 4-17 Chemical shift perturbation upon different C-terminal truncations with respect to the construct encompassing the TD and the remaining C-terminus (aa 415-489) of *Ciona intestinalis* p53/p73-a isoform 1. The NH chemical shift perturbations were calculated as described under 3.3.8. Chemical shift perturbations exceeding 0.5 are marked with arrow heads. Each bar diagram shows the chemical shift perturbations resulting from the corresponding truncation with respect to the construct encompassing the TD and the complete C-terminus (aa 415-489). The coloration of the bars is according to the color of the examined truncated construct in the alignment shown in Figure 4-11. (a) Chemical shift perturbations resulting from the truncation after aa 460. This truncation corresponds to the C-terminus of isoform 3 and ends with the last residue common to all isoforms, which contain the TD; isoform 1, 3, 4 and as well isoform 2. (b) Chemical shift perturbations resulting from the truncation after aa 450. (c) Chemical shift perturbations resulting from the truncation after aa 445. The last four residues could not be assigned and are marked with grey bars.

As expected the truncation after aa 450 resulted in even larger CSPs from residues, aligning to positions after the end of the first helix in the TD of human p73. But interestingly as before no significant CSPs were detected for the residues, locating N-terminal to the highly conserved glycine residue in position 425.

These would correspond to the β-sheet at the beginning of the TD, and hence, the environment of residues within the β-sheet is obviously not affected by this truncation. In the TD of human p73 the second helices wrap around the adjacent dimers in a clamp like fashion, and this results in interactions between the second helices and the β-sheets. In this truncation the section, which corresponds to the second helix, is largely missing, and hence, some significant CSPs from peaks, corresponding to the β-sheet, would have been expected for the case that the TD of *C.int.* p53/p73-a would have a stable folded tertiary structure closely resembling those of the TD of human p73. The section, which aligns to the first helices, was as well not affected by significant CSPs. Therefore it can be stated that the construct encompassing aa 415-450 represents the core TD, which should have a tertiary structure closely resembling those of the TD of human p53.

The shortest construct encompassed aa 415-445. This deletion of five additional residues at the C-terminus resulted in significant CSPs also from residues residing in the β-strand and in the first helix. This indicates that the conformation of the core TD composed of the β-sheets and the first helices was changed as a result of the truncation. Therefore, the construct was not any further examined.

4.2.4. The truncated construct corresponding to aa 415-450 of *Ciona intestinalis* p53/p73-a forms tetramers

The construct encompassing aa 415-450 was found to represent the core TD with a presumably p53-like domain architecture. To prove that this minimal domain is capable to form tetramers its sedimentation coefficient was determined by ultracentrifugation.

After purification, removal of the 6xHis-tag and concentration, a sample of the protein was applied onto a Superdex 75 16/60 column equilibrated and run in Nickel buffer A1. Opposed to NMR the detection in analytical ultracentrifugation is not hampered by high salt concentrations and thus this buffer having a higher ionic strength was selected to ensure that no aggregates interfere with a valid determination of the MW. The sample for the ultracentrifugation experiment was taken from the middle of the single elution peak.

The tetramer has a calculated molecular weight (MW) of 17492 Da. As shown in Figure 4-18 the peak maximum corresponds to a MW being approximately 2 kDa lower. Hence, the truncated construct is indeed a tetramer in solution and the more C-terminal sections are not essential to keep the dimeric subunits together. But although the samples had a monomer concentration of about 150 µM the protein might not have been exclusively tetrameric. As the construct does only possess two Phe residues but no Tyr or Trp its extinction coefficient is extremely low. To accurately determine the ratios between dimers and tetramers at concentrations ≤ the approximate K_d of tetramersization (~ 100 nM – 1 µM for human p53)[50] ultracentrifugation with fluorescence detection would have been the method of choice but such a device was not available.

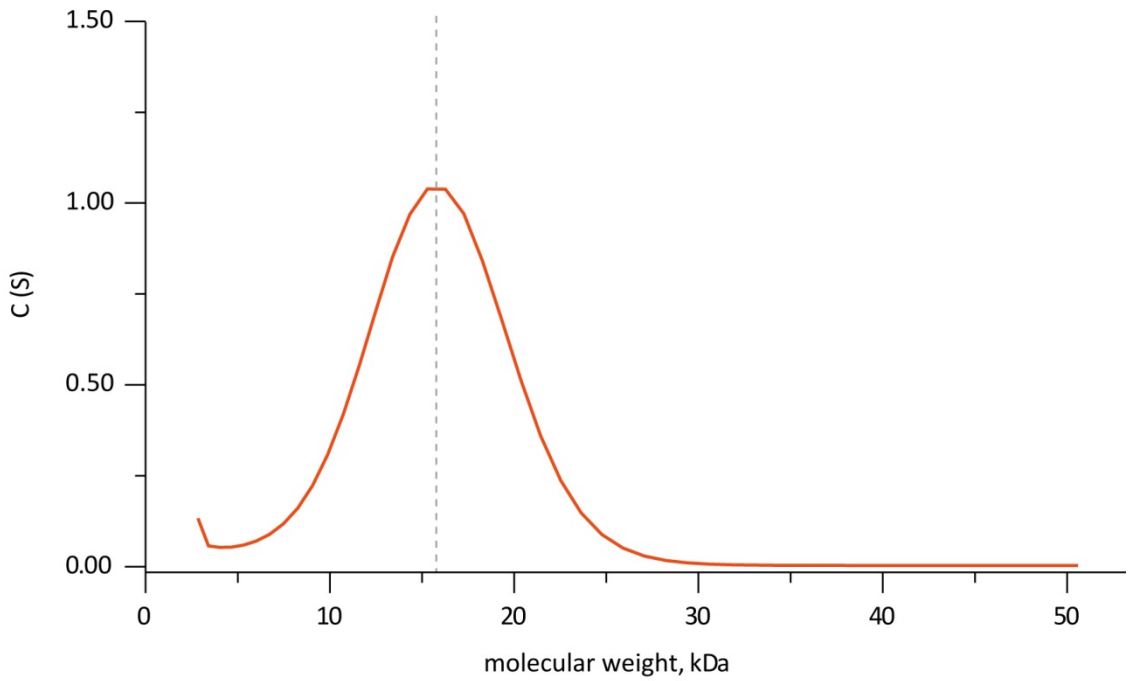


Figure 4-18 Determination of the oligomeric state of the truncated construct aa 415-450 by ultracentrifugation. The sedimentation coefficient (C(S)) of a sample of *Ciona intestinalis* p53/p73-a aa 415-450 (plus preceding dipeptide GS) was determined by ultracentrifugation. The protein monomer concentration was 150 µM buffered in 400 mM NaCl, 25 mM Tris pH 7.8. The calculated molecular weight of the tetramer is 17492 Da.

4.2.5. *Ciona int.* p53/p73-a TD of isoform 1 and 2 does not form a second helix

Truncation after aa 464 had been shown to cause only minor CSPs on residues in the section, which aligns to the TDs of vertebrate p53, p63 and p73. The backbone chemical shifts of the corresponding construct, encompassing aa 415-464, were completely assigned, and the chemical shifts of CA, CB, HN and N were used as input for TALOS+. The assignment of the respective section within the construct, encompassing the complete C-terminus of isoform 1, was not possible but the C-terminal section could be assigned without gaps. The

available CA, CB, HN and N chemical shifts were as well as used as input for TALOS+. The respective predictions of secondary structure and RCI S2 values are presented in Figure 4-20 (aa 415-489) and Figure 4-19 (aa 415-464). As expected the secondary structure prediction clearly indicates the presence of an N-terminal β-strand immediately followed by an α-helix. In all TDs from vertebrate p53, p63 and p73 like proteins the highly conserved glycine locates at the resulting sharp kink between the two SSEs. The secondary structure prediction by TALOS+ for *C.int.* p53/p73-a is in full agreement with such a motif. Surprisingly only one helix is predicted. As shown in Figure 4-2 the sequences of *C.int.* and also *C. sav.* p53/p73-a show conservation of some residues within the second helix of the TDs of human p63 and human p73, which interact with the core (β-sheets and first α-helices).[88,89,106] Especially the dipeptide Tyr-Arg was identified to be essential for the stability of the second helix in the TD of human p73.[88] The sequence alignment is not gapped between these residues and the highly conserved glycine residue at the kink between β-strand and first helix. Therefore it would be expected that the interaction motifs of core and second helix should as well be proper positioned to one another if helices with identical lengths and boundaries would fold. But this precondition seems not to be fulfilled as the helix is predicted to be elongated by about 3-4 residues with respect to the C-terminal end of the first helix in the TD of human p73. This might at least to some degree impede the formation of a second helix. It is interesting that the RCI S2 order value is not simply decreasing after the end of the helix. Especially from the prediction for aa 415-464 it is evident that the value reaches an additional local maximum within the section, which aligns to the second helix within the TD of human p73. This might be an indication for residual structure, which could serve as a nucleus for folding of an additional SSE.

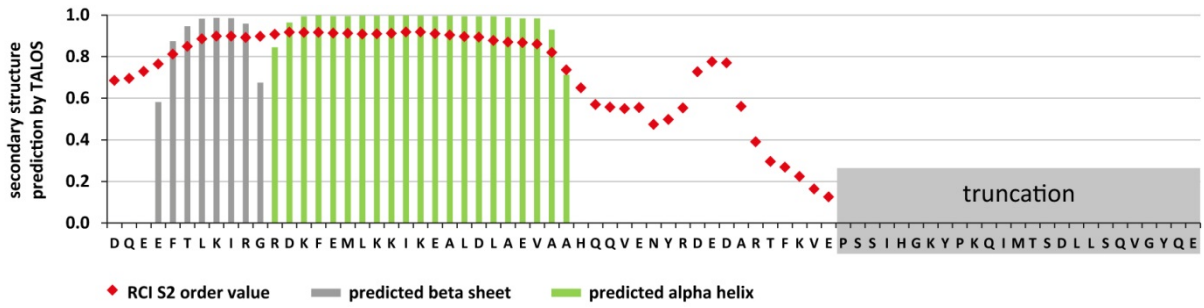


Figure 4-19 Prediction of secondary structure and RCI S2 order value by TALOS+ for *C.int.* p53/p73-a isoform 1 aa 415-464. The secondary structure and the RCI S2 order value were predicted using TALOS+ as described in 3.3.6. The chemical shifts of CA, CB, HN and N were used as inputs. The bars colored in grey represent predictions for β-strand conformation, those colored in green α-helical conformation. The RCI S2 order value is presented as red tilted squares.

With a comparatively high confidence level α-helical conformation was predicted for two additional neighboring residues within the C-terminal section. This was consistent with predictions, which were carried out using several different web server operated programs, using just the primary sequence as input (data not shown). Taking these findings into account it seems possible that for example as a result of post-translational modifications the C-terminus could as well adopt a different conformation with a higher content of secondary structure.

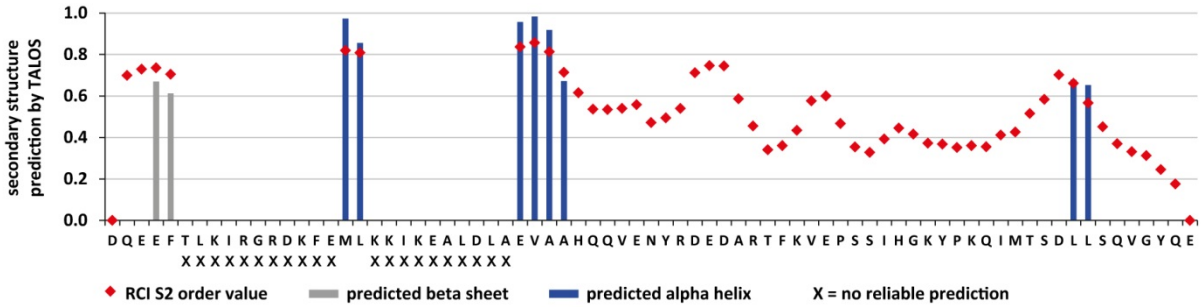


Figure 4-20 Prediction of secondary structure and RCI S2 order value by TALOS+ for *C.int.* p53/p73-a isoform 1 aa 415-489. The secondary structure and the RCI S2 order value were predicted using TALOS+ as described in 3.3.6. The chemical shifts of CA, CB, HN and N were used as inputs. As the assignment of the protein sequence was incomplete reliable predictions were not possible for several residues. The respective residues are labeled with an X. The bars colored in grey represent predictions for β-strand conformation, those colored in blue α-helical conformation. The RCI S2 order value is presented as red tilted squares.

The buffer, used for the NMR experiments, contains 50 mM L-arginine and 50 mM L-glutamate and is advantage due to low conductivity.[193] But as it is not commonly used for other technics one might argue that the buffer could have interfered with complete folding of the protein. Therefore the construct encompassing the full length C-terminus was screened in a variety of different buffers (listed in 2.11.11) and for each condition an [^{15}N, ^{1}H]-TROSY spectrum was recorded. By overlaying the spectra it was obvious that the conformation was unchanged, and hence, not affected by the buffer used (data not shown).

4.2.6. The core tetramerization domain of *C.int.* p53/p73-a isoform 1 and 2 is capable to support a second helix

It was shown that independent of the respective isoform the TD of *C.int.* p53/p73-a is consisting of an N-terminal β-strand followed by only a single α-helix. This is surprising as from the alignment towards vertebrate p63 or p73 it seems likely that the protein would resemble a comparable tertiary structure encompassing a second helix. Those residues

within the first helix, which interact with the second helix, are all at least functionally conserved (aa with similar side chain). And the essential Tyr-Arg dipeptide within the second helix is as well conserved in *C.int.* and in *C. sav.* p53/p73-a, as well.

Hence it had to be investigated if the section corresponding to β-strand and first α-helix would indeed be capable to interact with the second helix motifs. For this purpose the sequence following the first helix was from different positions onwards exchanged against the corresponding sequence from human p73. [^{15}N, ^{1}H]-TROSY spectra were recorded and overlaid with the construct corresponding to *C.int.* p53/p73-a isoform 1 aa 415-464 as reference for wild type conformation.

A chimeric protein was generated by fusing aa 415-444 from *C.int.* p53/p73-a isoform 1 (N-terminal) to aa 381-404 from human TAp73α (C-terminal). The corresponding [^{15}N, ^{1}H]-TROSY spectrum is presented in Figure 4-21 as an overlay with the spectrum of the wild type reference and showed massive CSPs on rather all residues, including those from the β-sheet. This revealed that the section from aa 415-444, which in the wild type likely adopts a conformation closely resembling the tertiary structure of the TD of human p53, in this construct interacts with the human p73 second helix sequence. Therefore the backbone chemical shifts (CA, CB, HN and N) of the chimeric construct were assigned and used as input for a secondary structure prediction with TALOS+, which is shown in Figure 4-22.

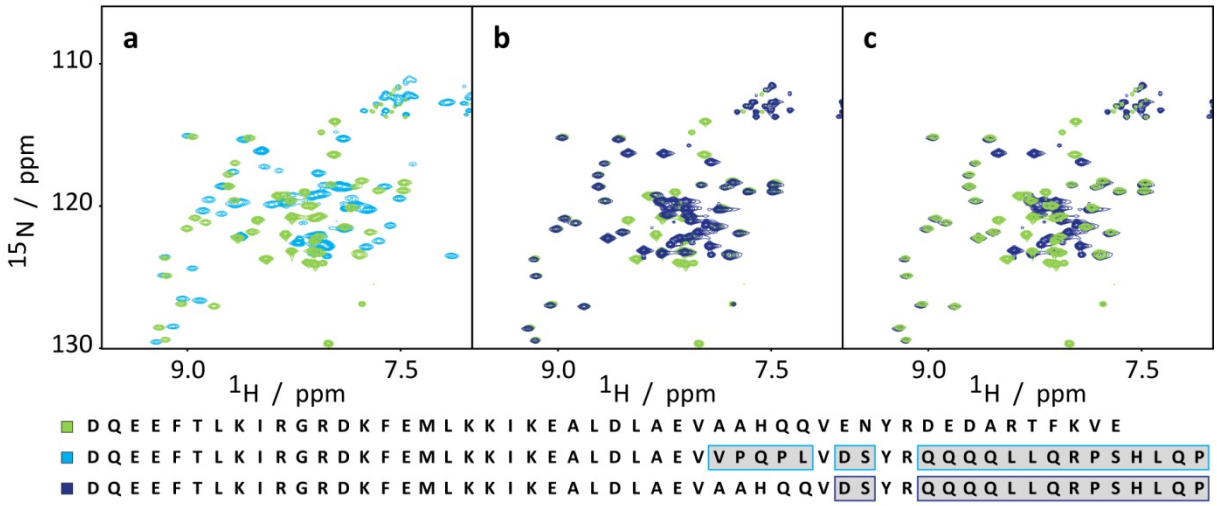


Figure 4-21 Overlay of [^{15}N, ^{1}H]-TROSY spectra of *Ciona intestinalis* (*C.int.*) p53/p73-a isoform 1 aa 415-489 and in comparison chimeric constructs with sequences being partially exchanged towards the corresponding sequence from human p73 either starting from aa 445 or from aa 451 onwards. All constructs start with residue 415. The spectrum of the *C.int.* wild type construct (aa 415-464) is always depicted in green. All spectra were recorded at 25°C. Below the spectra the sequences of the respective constructs are aligned. Residues which differ from the *C.int.* wild type construct are underlined in grey and boxed in the color corresponding to the respective mutant. (a) The spectrum of the chimeric construct with sequence exchange from aa 445 onwards is depicted in cyan and overlaid with the one of the wild type construct. (b) The spectrum of the chimeric construct with sequence exchange from aa 451 onwards is depicted in dark blue on top of the spectrum of the wild type construct. (c) The overlay from (b) is reversed.

The section, which corresponds to the second helix in the TD of human p73, was indeed predicted with very high confidence to adopt α-helical conformation. Compared to the wild type the first helix was predicted to be 3 residues shorter at its C-terminus. This supported the assumption that the first helix has to adopt circa the same length as in human p73 to allow for a proper positioning of the interacting residues within the core and the second helix.

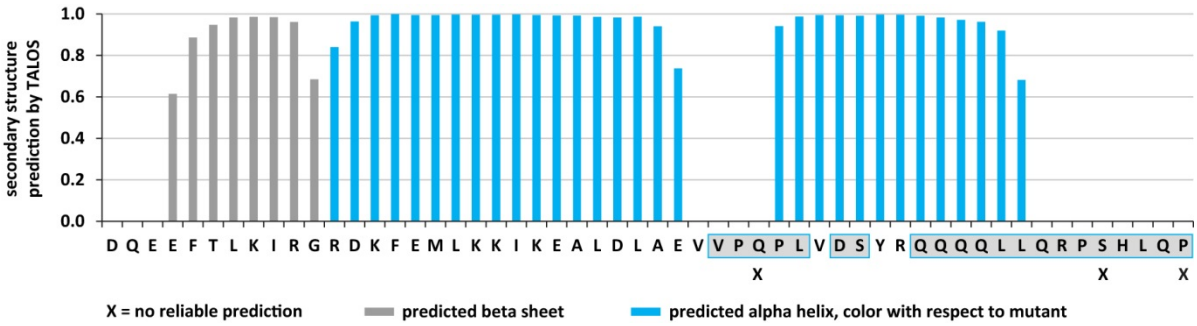


Figure 4-22 Prediction of secondary structure by TALOS+ for the chimeric protein encompassing aa 415-444 from *C.int.* p53/p73-a isoform 1 fused to aa 381-404 from human TAp73α. The secondary structure was predicted using TALOS+ as described in 3.3.6 with the chemical shifts of CA, CB, HN and N as input. Residues for which no reliable prediction was possible are labeled in an X. The bars colored in grey represent predictions for β-strand conformation, those colored in light blue α-helical conformation. Residues, which differ from the *C.int.* wild type sequence, are underlined in grey and boxed in light blue.

Thermal melting curves from the wild type and as well the chimeric protein were recorded with detection of circular dichroism at a wavelength of 222 nm to determine whether the second helix stabilizes the overall stability of the TD. The raw data were analyzed as described in 3.3.1.1. The plots of folded fractions versus temperature are presented in Figure 4-23.

Although the TD of the wild type protein already showed a very high T_m the midpoint of thermal denaturation was further increased by the formation of the second helix in the chimeric protein. To investigate whether the sequence of the second helix from human p73 would also alone be sufficient another chimeric protein was generated in whose sequence the transition from the *C.int.* towards the human sequence was shifted six residues to the C-terminus. The corresponding [^{15}N, ^{1}H]-TROSY spectrum is overlaid with the spectrum of the wild type in Figure 4-21. The spectra match one another to a very high degree and all peaks, which represent residues within the SSEs of the wild type, perfectly overlay with their respective counterparts. Hence, the sequence of the second helix alone is not sufficient and

if the second helix actually folds depends at least partially as well on the sequence between the two helices.

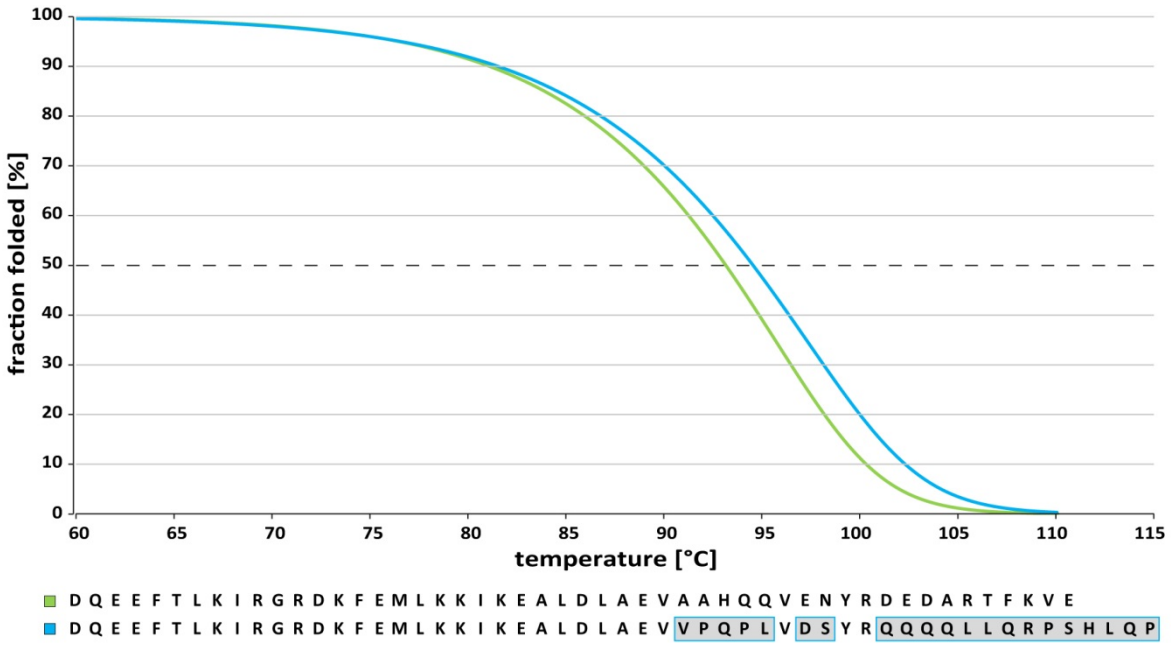


Figure 4-23 Fraction folded versus temperature plots calculated from thermal denaturation curves with recorded circular dichroism at 222 nm from *Ciona intestinalis* (*C.int.*) p53/p73-a isoform 1 aa 415-489 and a chimeric protein encompassing aa 415-444 from *C.int.* p53/p73-a isoform 1 fused to aa 381-404 from human TAp73α. Samples with 100 µM protein concentration (monomer) were freshly prepared in 1 x CD buffer. Experimental settings are described in section 3.3.1 and the applied methodologies of curve fitting and the transformation into fraction folded are described in section 3.3.1.1.

It had to be validated if the second helix in the chimeric construct packs onto the core in a way comparable to the situation in human p73. Therefore a double mutation changing YR to AA was introduced into the chimeric construct for which the second helix had been validated. The corresponding [^{15}N, ^{1}H]-TROSY spectrum is presented in Figure 4-24.

The comparison between the respective overlays with either the wild type or the folding chimeric protein shows that the mutation of Tyr-Arg to Ala-Ala leads to a break up. All peaks, which represent residues within the β-sheet, and many other peaks, which represent residues from the section common to all three constructs, overlay with their counterparts in the wild type spectrum. This proves that in the folding chimeric protein the second helix adopts a position towards the core closely resembling the packing of the SSEs in the TD of human p73.

To narrow down which part of the sequence between first and second helix is being most decisive, an additional chimeric protein was generated in whose sequence the transition from the *C.int.* towards the human sequence was shifted two residues to the C-terminus with respect to the position in the folding chimeric protein.

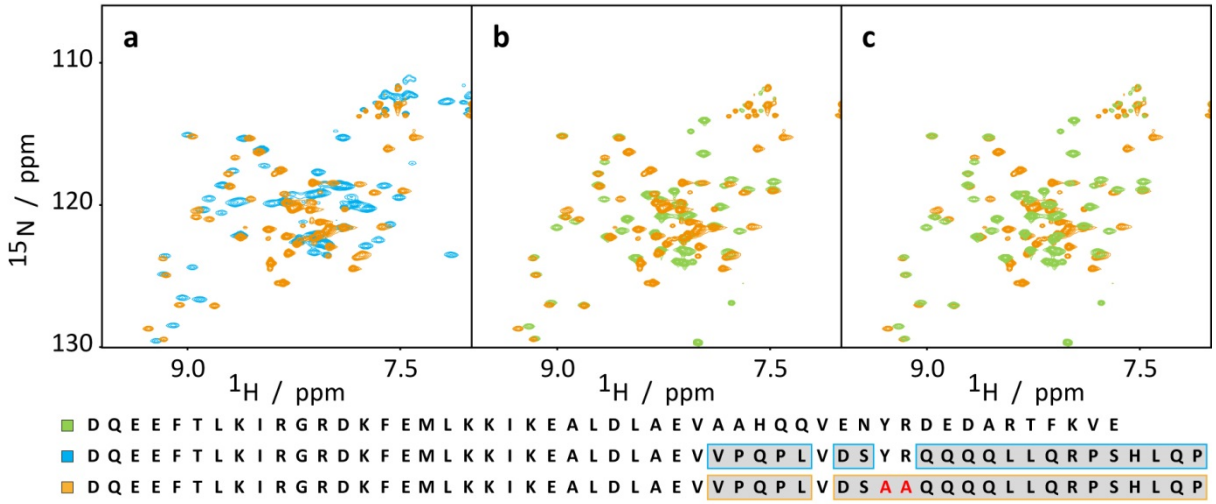


Figure 4-24 Overlays of [^{15}N, ^{1}H]-TROSY spectra of *Ciona intestinalis* (*C.int.*) p53/p73-a isoform 1 aa 415-489 wild type, the chimeric construct with a sequence being partially exchanged towards the corresponding sequence from human p73 starting from aa 445 onwards as well as the chimeric construct with an additional YR→AA mutation. All constructs start with residue 415. The spectrum of the chimeric construct with sequence exchange starting from aa 445 onwards and the additional mutation of YR to AA is always depicted in orange. All spectra were recorded at 25°C. Below the spectra the sequences of the respective constructs are aligned. Residues, which differ from the *C.int.* wild type construct, are underlined in grey and boxed in the color corresponding to the respective mutant. (a) The spectrum of the chimeric construct with sequence exchange from aa 445 onwards is depicted in cyan and overlaid with the spectrum of the construct, which has the same residues exchanged but in addition the mutation of YR to AA. (b) The spectrum of the wild type construct is depicted in green and overlaid with the spectrum of the chimeric construct with the additional YR to AA mutation. (c) The overlay from (b) is reversed.

In Figure 4-25 the corresponding [^{15}N, ^{1}H]-TROSY spectrum is presented. The comparison between the respective overlays with either the folding chimeric protein or the wild type shows that the presence of the Ala-Ala dipeptide, within the otherwise identical chimeric sequence, results in a wild type like conformation. Most peaks, which represent residues within the SSEs of the wild type, overlay better with their respective counterparts in the wild type spectrum than to those in the spectrum of the folding chimeric protein.

Hence, the formation of a second helix is strictly depending on a rather short section between the end of the first helix and the beginning of the second helix.

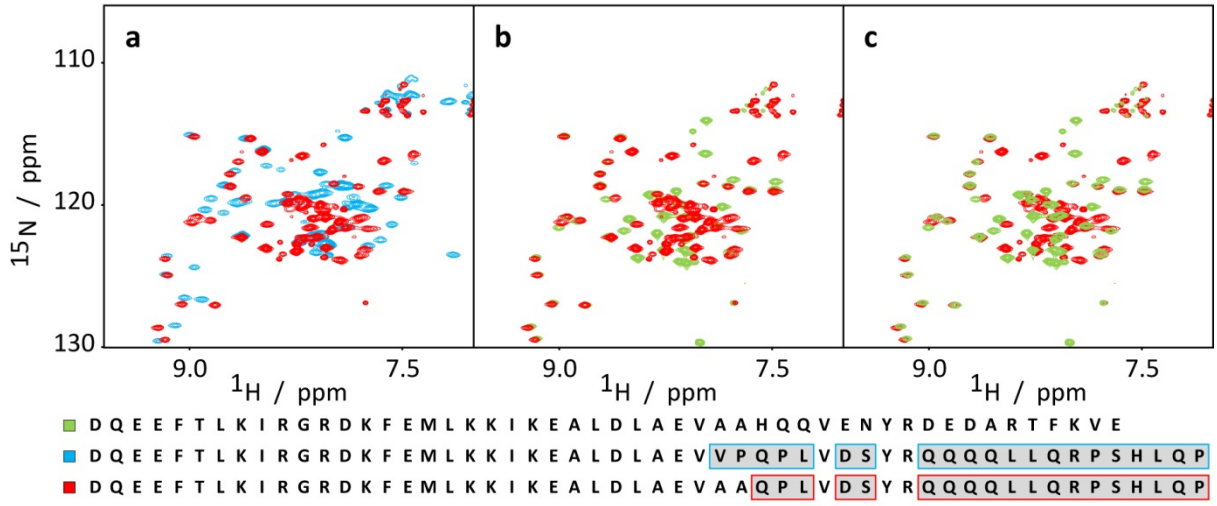


Figure 4-25 Overlays of [^{15}N, ^{1}H]-TROSY spectra of *Ciona intestinalis* (*C.int.*) p53/p73-a isoform 1 aa 415-489 wild type and the chimeric constructs with sequences being partially exchanged towards the corresponding sequence from human p73 either starting from aa 445 or from aa 448 onwards. All constructs start with residue 415. The spectrum of the chimeric construct with sequence exchange starting from aa 448 onwards is always depicted in red. All spectra were recorded at 25°C. Below the spectra the sequences of the respective constructs are aligned. Residues, which differ from the *C.int.* wild type construct, are underlined in grey and boxed in the color corresponding to the respective mutant. (a) The spectrum of the chimeric construct with sequence exchange from aa 445 onwards is depicted in cyan and overlaid with the spectrum of the one with sequence exchange starting from aa 448 onwards. (b) The spectrum of the wild type construct is depicted in green and overlaid with the spectrum of the chimeric construct with sequence exchange from aa 448 onwards. (c) The overlay from (b) is reversed.

4.2.7. Determination of minimal number of mutations causing a second helix

The hinge region, spanning from the end of the first up to the beginning of the second helix, had been identified to be decisive for the stabilization of the second helix, and specially the dipeptide Val-Pro was found to be important. To investigate which additional residues contribute to stabilizing interactions the respective section within the structure of the TD of human p73 (pdb code: 2KBY) was analyzed using PyMOL and PDBePISA. As reported in Table 4-3 the formation of the interface results in a gain in buried surface area on five residues, which locate within a seven residues long section. Figure 4-26 shows the structure in carton and surface representation with the seven residues depicted as sticks.

Table 4-3 Determination of buried surface area upon formation of the hinge region in the TD of human p73 using PDBePISA with the pdb entry 2KBY as input. The corresponding residues from *C.int.* p53/p73-a are compared with respect to functional conservation.

hinge region in TD of human p73 (pdb 2KBY)				*C.int.* p53/p73-a TD in comparison		
	res	ASA, Å²	BSA, Å²	conservation		res
Glu	379	150.06	0.00	conserved	Glu	443
Leu	380	115.75	77.21	smaller (hydrophobic)	Val	444
Val	381	42.49	35.83	smaller (hydrophobic)	Ala	445
Pro	382	77.44	12.88	-	Ala	446
Gln	383	152.87	0.00	-	His	447
Pro	384	102.28	0.00	-	Gln	448
Leu	385	106.27	37.10	-	Gln	449
Val	386	19.76	10.47	conserved	Val	450
Asp	387	88.50	0.00	similar	Glu	451

Accessible surface area (ASA) within monomer
Buried surface area (BSA) due to formation of the interface formed up between the monomers A+C or B+D

Out of the five residues, which contribute to the interface, the last one is being conserved in *C.int.* p53/p73-a. As two side chains orient to the surface of the protein and as a result do not contribute to the interface it was assumed that the TD of *C.int.* p53/p73-a isoform 1 does not form a second helix due to deviations from the motif L-V-P-X-X-L within the hinge region.

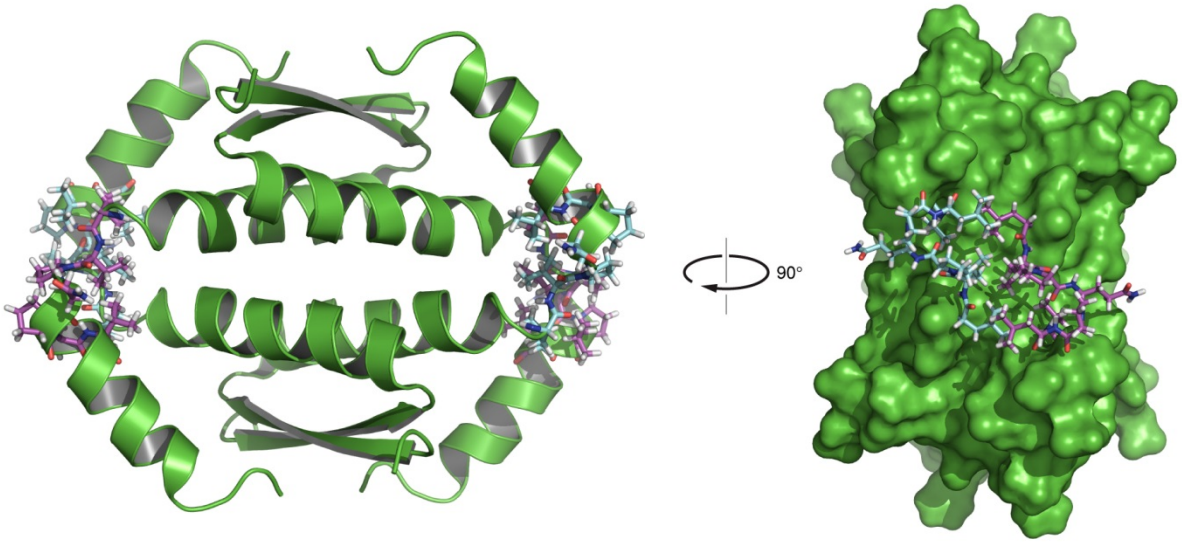


Figure 4-26 Cartoon and surface representation of the TD of human p73 (pdb entry 2KBY) with the seven residue long hinge region interface separately depicted as sticks. The sections of the interfaces formed between the monomers A+C and B+D are depicted as sticks with coloration in purple (primary dimer A+B) or in cyan (primary dimer C+D). The remaining structure is shown in carton representation or as surface representation after being rotated by 90 degree.

It was now time to determine how many out of these four residues would have to be simultaneously introduced into the *Ciona* sequence to cause the expected second helix to fold. The prediction of secondary structure using TALOS+ had indicated α-helical conformation for two neighboring residues residing in the very C-terminus of *C.int.* p53/p73 isoform 1; Figure 4-20. Hence it could not be excluded that the formation of the proposed second helix would in case be coupled with the formation of an additional helix within the C-terminus. The tertiary structure of the TD of Cep-1, the protein being the p53 homolog from *Caenorhabditis elegans*, can be described as a fusion of the canonical TD structure as found in human p53 and a SAM domain, which is folded by the sequence immediately following this core TD.[113] This made it even more likely that the formation of several SSEs could be coupled. Therefore it was decided to investigate the effect of mutations within the hinge region in the context of the construct with the full length C-terminus. All four residues were permutated by side directed mutagenesis resulting in a total of sixteen constructs including the wild type. The respective constructs are depicted in Figure 4-27 in comparison to the alignment of the wild type sequences of *C.int.* p53/p73-a isoform 1 and the corresponding sequence in human p73.

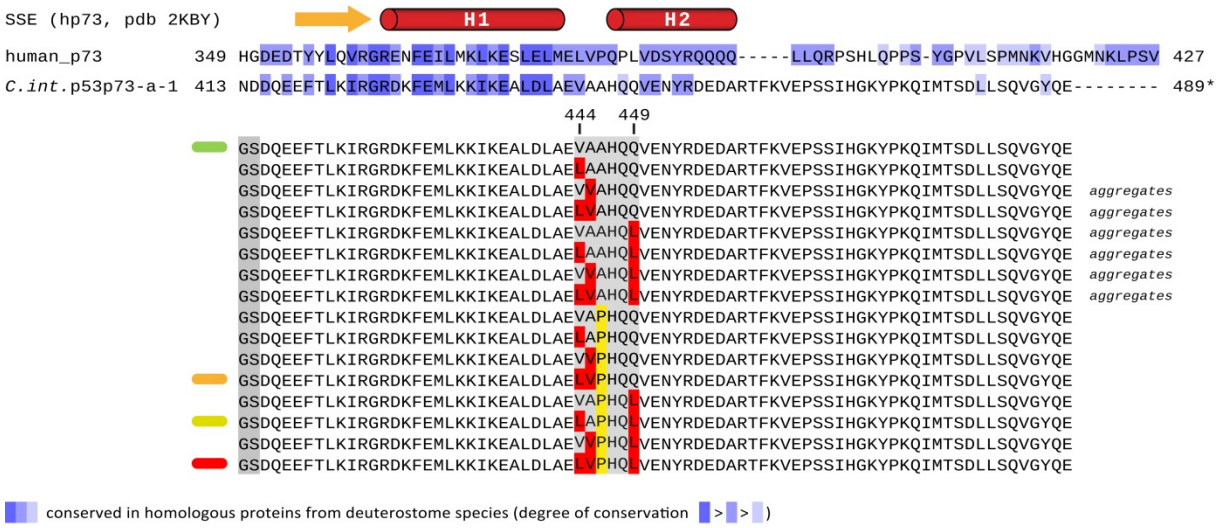


Figure 4-27 Alignment of the protein sequences corresponding to the tetramerization domain of human p73 versus those from *Ciona intestinalis* (*C.int.*) and in comparison as well 16 constructs with different hinge region mutations in the context of the *Ciona* sequence. The blue coloration of the alignment of the wild type protein sequences is according to percent of similar residues within p53, p63 and p73 like proteins from deuterostome species. An asterisk indicates the C-terminal end of the natural protein sequence. As a prefix to the sequence alignment the SSEs of the NMR solution structure of human p73 (pdb 2KBY) is depicted. The lower sixteen lines of the alignment show the investigated constructs, which contain different combinations of mutations within the hinge region. Mutations towards larger hydrophobic side chains are underlined in red, those which change the respective residue into proline are underlined in yellow. For technical reasons all constructs have two additional residues at the N-terminus, which are marked with dark grey background. This GS dipeptide is a remnant from the N-terminal TEV cleavage site and does not align to the wild type protein sequence. For this reason it is left out in the description of the constructs. Those constructs, which were based on the analysis of CSPs selected for further investigations, are marked with colored bars at their N-termini. The respective colors were used in the following analysis.

From the group of mutants without the proline residue only the one with the Val 444 to Leu mutation did not form aggregates. This reduced the set of mutants to 10 including the wild type construct. From these [^{15}N, ^{1}H]-TROSY spectra were recorded and the position of those peaks, which in the wild type had been assigned to residues from within the β-strand, the first helix and as well the C-terminus, were compared between the spectra. The shifts of peaks corresponding to residues within six positions upstream or downstream of the hinge region were not taken into account as they were expected to be affected by the change of the primary sequence.

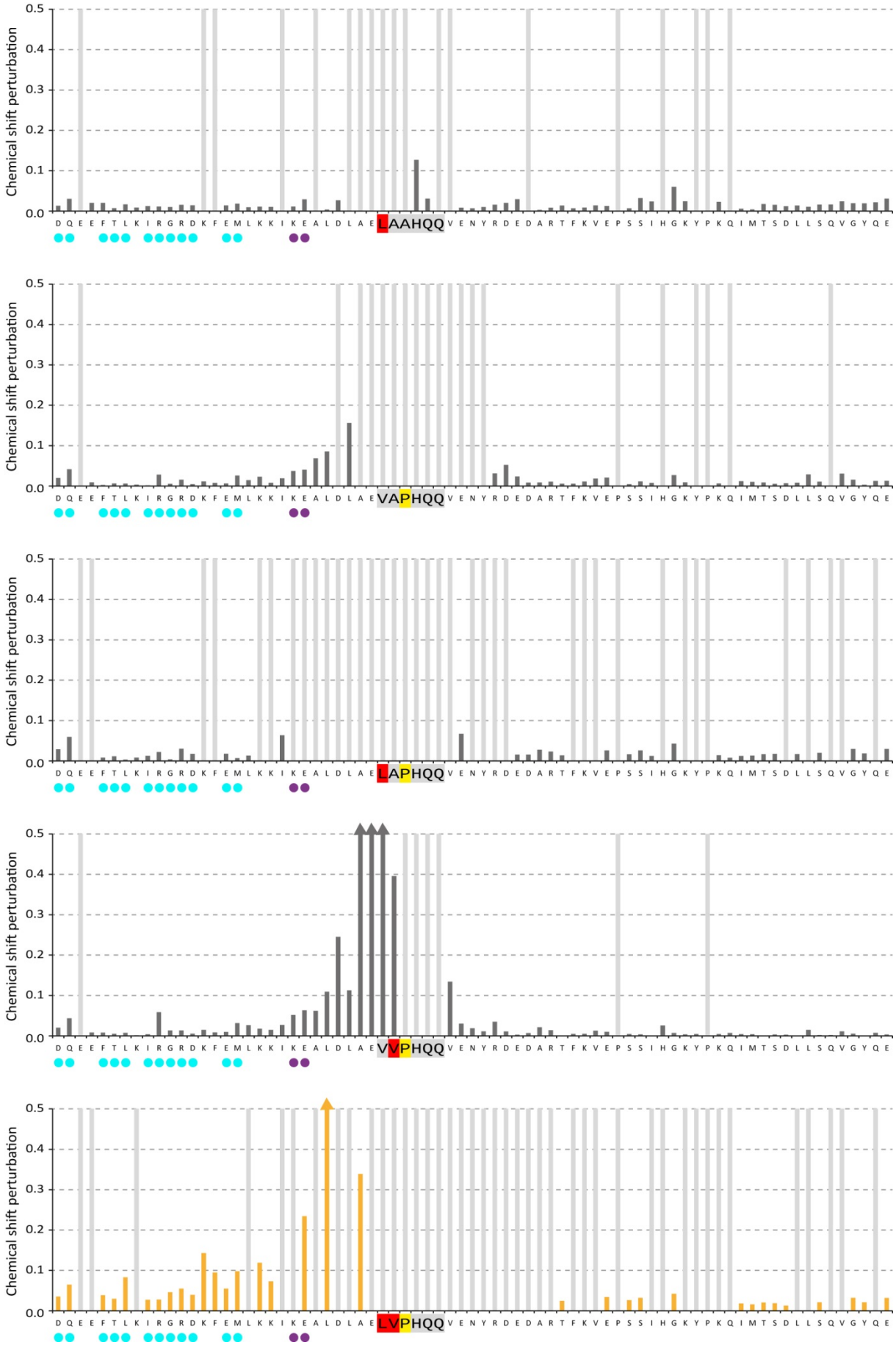
The section C-terminal to residue 464 was unaffected in all mutant constructs, and for most of the corresponding peaks it was possible to translate the assignment from the peaks of the wild type. This showed that the C-terminus did not form SSEs but stayed random coil.

Generously the shifts on peaks from residues within the β-sheet were rather low and by comparing the 2D spectra simply by eye it was not possible to group the mutants into those, which caused large changes of the tertiary structure, and those, which did not. Therefore in

a first attempt the CSPs relative to the wild type spectrum were calculated for some of the few highly dispersed peaks. But when two mutants were compared with respect to the ratios of the CSPs on three to six peaks these ratios were frequently inconsistent, and hence, based solely on the 2D spectra it could not be determined which combination of mutations did show the greatest effect. This made it necessary to at least partially assign those peaks, which correspond to the β-strand and the first helix, within all mutants. But as it had already been seen for the wild type construct the long random coil section at the C-terminus drastically increases the rotational correlation time of the rigidly folded parts. And this made it impossible to assign the N-terminal folded section from HNCACB or HNCA spectra of the construct with the full length C-terminus. Therefore the truncation after aa 464 was introduced into some of the nine mutant constructs and several HNCACB spectra were recorded. Within the context of the wild type sequence the truncation after aa 464 had been validated not to cause CSPs on residues from the β-strand and the first α-helix. For the mutants this was as expected also the case. Since that it was possible to translate the assignment of the respective peaks onto the spectrum of the corresponding longer construct. For some mutants no 3D spectra were recorded and the assignments were instead wherever possible concluded from the spectra of constructs with similar mutations in the hinge region. By this it was possible to assign to a large extent as well the section corresponding to the β-strand and the first α-helix. For the nine mutated constructs the CSPs towards the wild type were calculated as described in 3.3.8, and the respective values along the primary sequences are shown in Figure 4-28 for all nine mutants.

The predictive power of the CSPs is dependent on the position of the corresponding residue within the primary sequence and the tertiary structure, as well. As mentioned before CSPs on residues, which are close to the hinge region (+/- six residues), were not taken into account. To estimate in which position the assigned residues would locate with respect to the proposed second helix a similarity model was calculated using swiss-model. The NMR solution structure of the TD of human 73 was used as template and the query sequence contained all four mutations. The model is shown in Figure 4-29. Within the β-strand and the first α-helix those residues are colored in cyan, which could be assigned for all nine mutant constructs and which have at least the minimal distance towards the hinge region.

The constructs with single mutations of position 444 (to Leu) or position 446 (to Pro) and as well the construct, which contained the combination of both, did show only minor CSPs.


Chemical shift perturbation
0.5
0.4
0.3
0.2
0.1
0.0
D Q E E F T L K I R G R D K F E M L K K I K E A L D L A E LAAHQQ V E N Y R D E D A R T F K V E P S S I H G K Y P K Q I M T S D L L S Q V G Y Q E
D Q E E F T L K I R G R D K F E M L K K I K E A L D L A E VAPHQQ V E N Y R D E D A R T F K V E P S S I H G K Y P K Q I M T S D L L S Q V G Y Q E
D Q E E F T L K I R G R D K F E M L K K I K E A L D L A E LAPHQQ V E N Y R D E D A R T F K V E P S S I H G K Y P K Q I M T S D L L S Q V G Y Q E
D Q E E F T L K I R G R D K F E M L K K I K E A L D L A E VVPHQQ V E N Y R D E D A R T F K V E P S S I H G K Y P K Q I M T S D L L S Q V G Y Q E
D Q E E F T L K I R G R D K F E M L K K I K E A L D L A E LVPHQQ V E N Y R D E D A R T F K V E P S S I H G K Y P K Q I M T S D L L S Q V G Y Q E

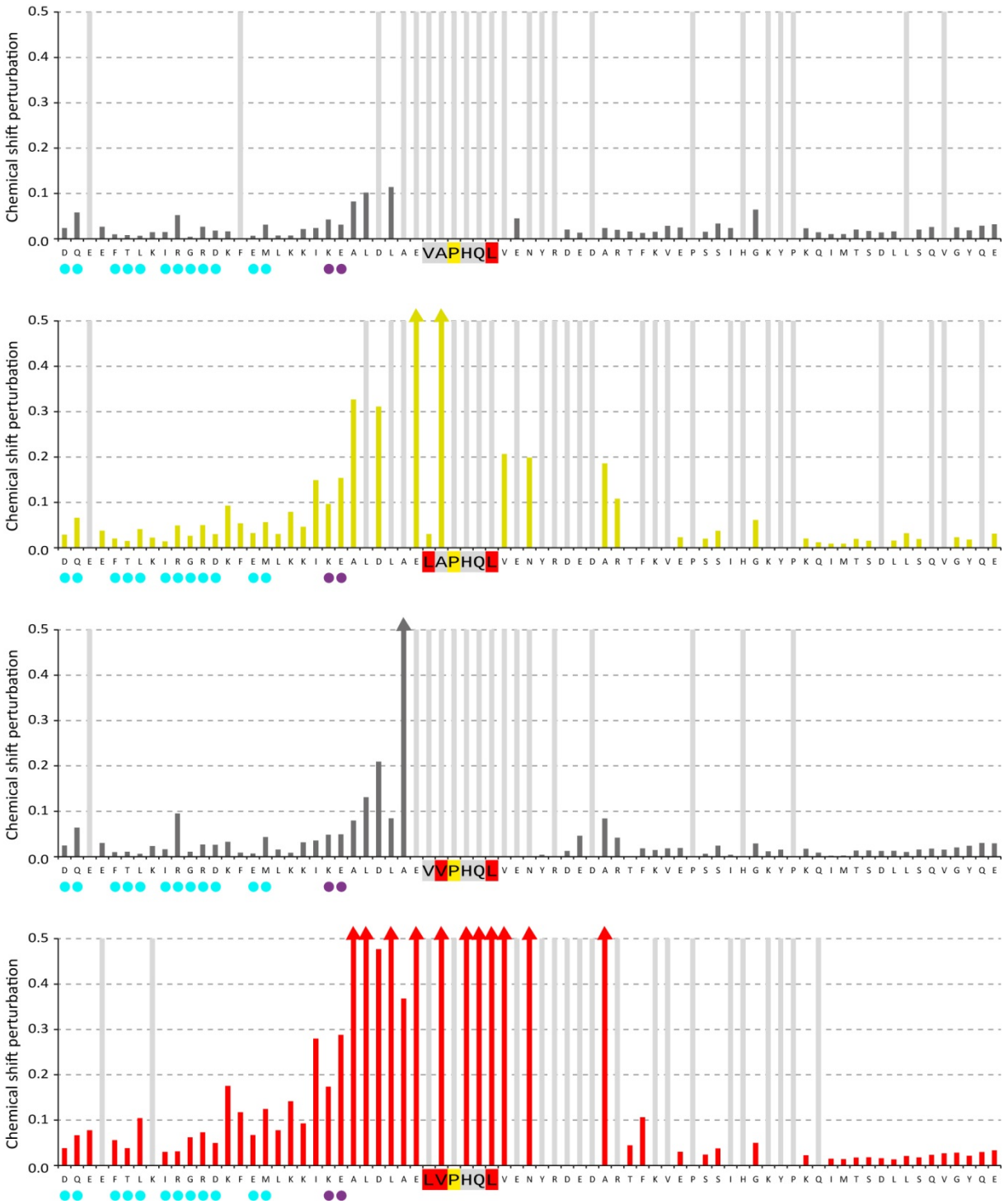


Figure 4-28 Analysis of the chemical shift perturbations (CSPs) observed for nine different combinations of mutations within the hinge. The chemical shifts observed for nine constructs with different mutations within the hinge region were compared to those observed for the wild type. All constructs comprised aa 415-489 of *C.int.* p53/p73-a isoform 1. Mutations towards larger hydrophobic side chains are underlined in red and those which change the respective residue into proline are underlined in yellow. The NH chemical shift perturbations were calculated as described under 3.3.8. Chemical shift perturbations exceeding 0.5 are marked with arrow heads. Each bar diagram shows the chemical shift perturbations resulting from the respective mutations. The coloration of the bars is either dark grey or colored according to the color used in the following analysis. Based on the analysis of the CSPs the respective constructs were selected for further investigations. Unassigned residues are marked with bars in light grey.

And the combinations of Pro in position 446 with either Val in position 445 or with Leu in position 449 did also show only minor effects on most of the chemical shifts, included in the analysis. The only exception was the Arg residue at the end of the β-strand. Compared to the before mentioned mutants the simultaneous mutation of all positions, detailed [VVPHQL], caused a little more intense CSPs but those determined for the dipeptide Lys-Glu in position 436-437 stayed low. The two residues are strictly conserved in vertebrate p63 and p73 as well as in all *Ciona* p53/p73 proteins (a and b from *C.int.* and as well *C.sav.*), and they are part of the interface formed by the second helix and the core TD (β-sheets and first helices). In the primary sequence they have the necessary distance towards the mutated residues, and hence, the respective CSPs on these to residues had a high predictive power.

The other two triple mutations, detailed [LAPHQL] and [LVPHQQ], and as well the combination of all four, detailed [LVPHQL], showed more intense CSPs on residues from within the β-strand and the first helix, and especially those determined for the Lys-Glu dipeptide were significantly more intense. Therefore the three mutants were selected for further investigations.

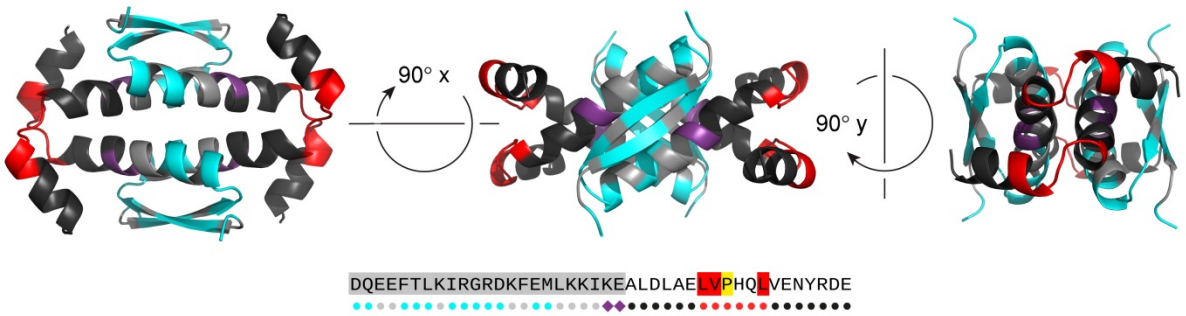


Figure 4-29 Similarity model of *C.int.* p53/p73-a isoform 1 aa 415-456 with four mutations within the hinge region between first and second helix. Based on the NMR solution structure of the TD of human p73 (pdb entry 2KBY) a similarity model was calculated using swiss-model. The *Ciona* query sequence contained the four mutations within the hinge region between the end of the first and the beginning of the second helix. Compared to the wild type sequence position 444 onwards was mutated to Leu-Val-Pro-His-Gln-Leu. The structure is shown in cartoon representation and the coloration discriminates between those residues which were included in the analysis of the CSPs, and those which reside too close to the mutated hinge region. The primary sequence is presented below the model with the mutated residues being underlined in red and yellow. As indicated by the dots below the sequence the corresponding section has red coloration in the model. The flanking six residue long sections are shown in black. CSPs from residues locating in these sections were not included in the analysis as they were assumed to be always affected simply by the mutation of the primary sequence, and hence, not necessarily reporting changes in the tertiary structure. Residues, having the necessary distance towards the permutated hinge region, are underlined in grey. From these the ones with assignment for all nine mutants do have cyan color in the model. The dipeptide Lys-Glu is shown in purple and marked with tilted squares below the protein sequence. The two residues are strictly conserved in *C.int.* p53/p73-a and vertebrate p63 or p73 proteins. In the TD of human p73 they are part of the interface between the second helix and the core TD. Therefore the CSPs on these residues would have been of high interest. However it was not possible to assign them in all nine mutants.

Figure 4-30 shows the overlay of the [^{15}N, ^{1}H]-TROSY spectra of these three mutants and the wild type.

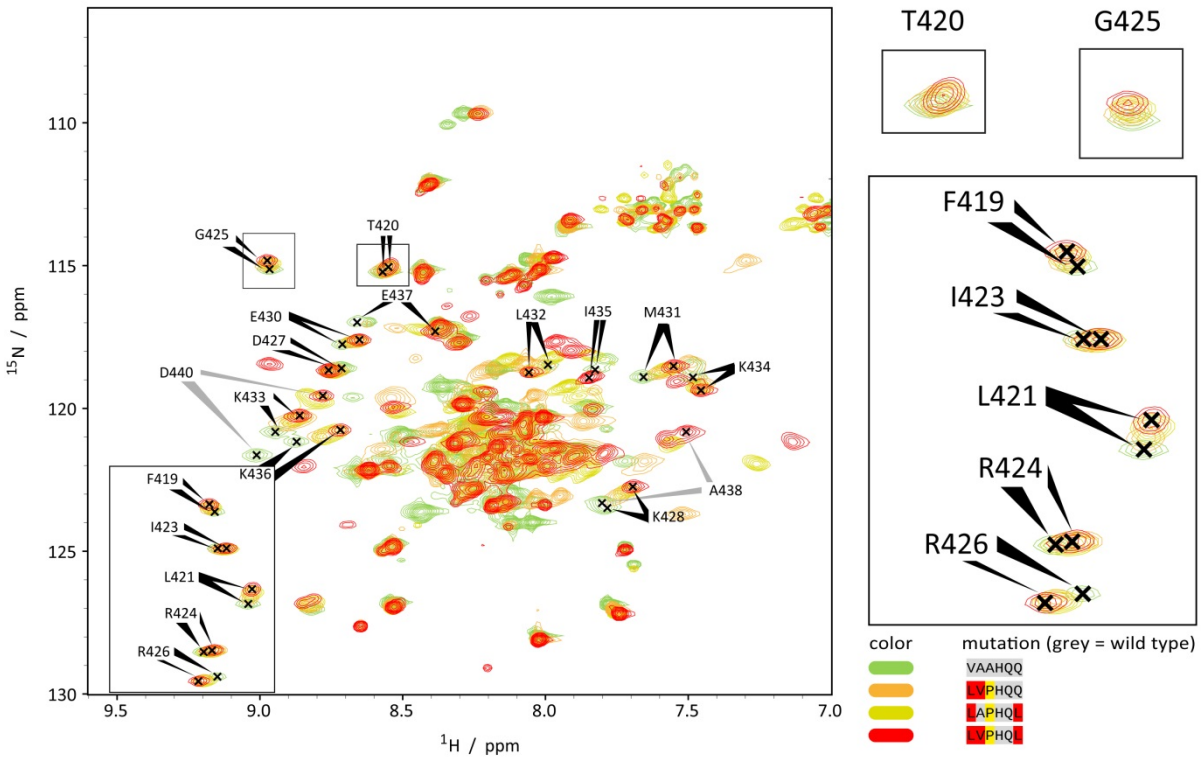


Figure 4-30 Overlay of [^{15}N, ^{1}H]-TROSY spectra of wild type and mutant constructs of *C.int.* p53/p73-a isoform 1 aa 415-489. All spectra were recorded at 25°C. The caption in the lower right relates the colors of the spectra to the respective sequence of aa 444-449. Residues corresponding to the wild type sequence are underlined in grey. Mutations towards larger hydrophobic side chains are underlined in red, and those which change the respective residue into proline are underlined in yellow. The assignments always refer to the respective peaks in the spectrum of the wild type (green) and the spectrum of the construct having all four mutations (red).

4.2.8. Formation of second helix upon mutations in hinge region

It should be investigated to which extent the mutations in the hinge region change the secondary structure content. For this purpose and to enable a complete backbone assignment the truncation after aa 464 was introduced into the respective constructs possessing the three combinations of mutations for which the most intense CSPs had been determined. The backbone chemical shifts for CA, CB, HN and N were assigned from HNCACB spectra using Sparky 3.114 and used as input for TALOS+. The respective secondary structure predictions are presented in Figure 4-31 in comparison to the prediction for the wild type and those for the chimeric construct, for which the formation of a second helix had been shown (4.2.6).

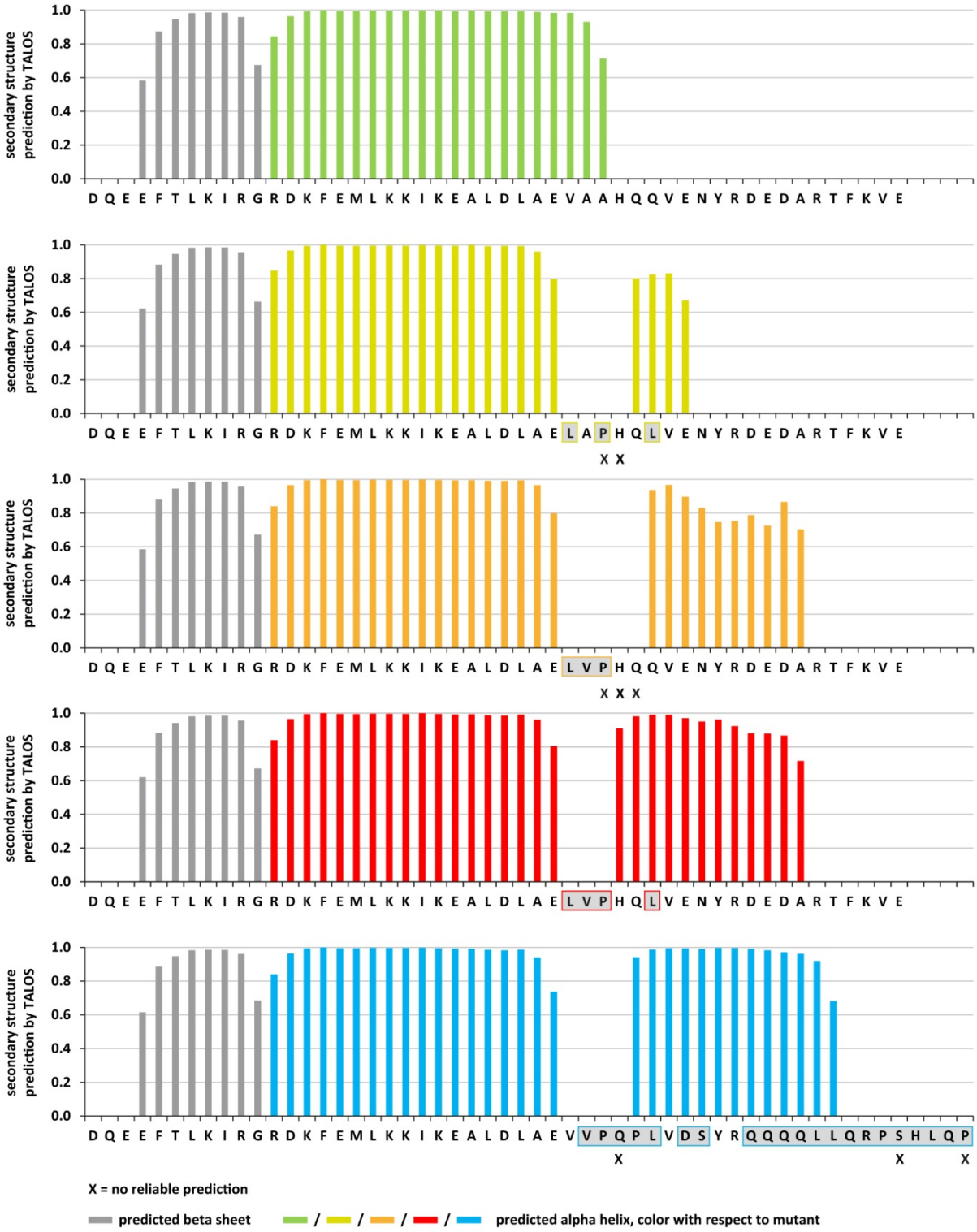


Figure 4-31 Predictions of secondary structure by TALOS+ for several partially mutated constructs encompassing aa 415-464 of *C.int.* p53/p73-a isoform 1 and in comparison those predicted for the chimeric protein in which aa 415-444 from *C.int.* p53/p73-a were fused to aa 381-404 from human TAp73α. The secondary structure elements (SSEs) were predicted using TALOS+ as described in 3.3.6 with the chemical shifts of CA, CB, HN and N as input. The chemical shift values were obtained from spectra recorded at 25°C. Unassigned residues and those for which no reliable prediction was possible are labeled with an X. Each bar diagram shows the SSEs predicted for the respective construct, whose protein sequence is written below. Bars depicted in dark grey represent predictions for β-strand conformation. The bars representing predictions for α-helical conformation have the color, which was assigned to the respective construct. Residues, which differ from the *C.int.* wild type, are underlined in light grey and boxed in the color assigned to the respective mutant.

TALOS+ predicts the formation of a second helix for all three inspected mutants. With respect to its length in the wild type the first helix is in all mutants predicted to be shortened at its C-terminal end by about three residues, and hence, to end with Glu 443, the last residue before the hinge region. The two mutants with either LVPHQL or LVPHQQ had been shown to cause the most and the second most intense CSPs compared to the wild type, respectively. For both mutants TALOS+ predicts a second helix, which extends to Ala 458, and for the mutant LVPHQL His 447 was predicted to be the first residue with helical conformation. Hence, proline is in the N-cap position just like in human p63 and human p73. Due to gaps in the assignments the beginning of the second helix could not be predicted for the mutants LVPHQQ and LAPHQL but because of the preference for proline in the N-cap position it is assumed that the second helix as well starts with His 447, respectively. For the mutant LAPHQL the second helix is predicted to only reach up to Glu 415. Hence, the dipeptide Tyr-Arg, which is known to intensively interact with the core and thereby to stabilize the second helix, is not predicted to adopt α-helical conformation. It is conspicuous that the level of confidence for the prediction of α-helical conformation for residues within the second helix varies between the three hinge region mutants whereas the confidence levels of the prediction for residues within the β-strand and the first helix do not. For the construct with all four mutations the second helix was predicted with the highest average confidence. However for the corresponding residues in the chimeric construct the predictions had even higher confidence levels, and in addition the second helix was predicted to be C-terminally longer in the chimeric construct. This was attributed to the very different sequences C-terminal to the Tyr-Arg motif. Like that for instance *C.int.* p53/p73-a does not have a Gln residue at three positions downstream of the Tyr-Arg motif but this Gln is conserved in vertebrate p63 as well as p73 and this residue is a part of the second helix to core interface. Hence, this missing potential interface residue could be one reason for the second helix to be shorter. Another interesting observation, worth mentioning, is that the second helix in the mutants LVPHQL and LVPHQQ ends with circa those residues for which in the wild type no secondary structure but a surprisingly high RCI S2 order value had been predicted; Figure 4-19.

4.2.9. The second helix resulting from mutations in the hinge region adopts a position relative to the core TD like in human p73

The length of the second helix was predicted by TALOS+ to differ considerably as a function of the combination of mutations within the hinge region and for the mutant LAPHQL the helix was even predicted not to span the Tyr-Arg motif, which is known to be essential for the stability of the second helix.

Hence, it had to be validated if the second helix, which folds as a result of the mutations in the hinge region, indeed packs onto the core in a way comparable to the situation in human p73. Therefore a double mutation changing Tyr-Arg to Ala-Ala was introduced into the construct possessing all four mutations. The [^{15}N, ^{1}H]-TROSY spectrum of the new construct was compared to the respective spectra of the wild type and the construct with all four hinge region mutations.

On most residues within the β-strand and those within the first helix, which can be evaluated, the CSPs between the wild type and the mutant LVPHQL are ≤ 0.15; Figure 4-28. As expected the position of the corresponding peaks in the spectrum of the new construct with the Tyr-Arg to Ala-Ala mutation is always in between the positions, which were determined for the respective residues in the spectra of the wild type and of the mutant LVPHQL. This leads necessarily to even lower CSPs, and hence, an analysis by simply overlaying the spectra was not sufficient. Therefore, to the possible extent the assignment was inferred from the TROSY spectra of the wild type and the different hinge region mutants. The CSPs were then calculated with respect to the wild type and as well in respect to the mutant LVPHQL to elucidate which conformation is being more closely resembled.

As shown in Figure 4-32 the chemical shifts determined for the hinge region mutant, which additionally possesses the Tyr-Arg to Ala-Ala mutation, consistently deviate less from those of the wild type. This reveals that the tertiary structure resembles more closely those of the wild type, which does not fold a second helix. And thus it was shown, that for the formation of a second helix the Tyr-Arg motif is equally essential as in p73.[88] Those residues from within the first helix, which in human p73 make up the second helix to core interface, are conserved in *C.int.* p53/p73-a. Thus supposedly the second helix, which is caused by the mutations within hinge region, adopts a position relative to the core TD which very closely resembles the tertiary structure of the TD of human p73.

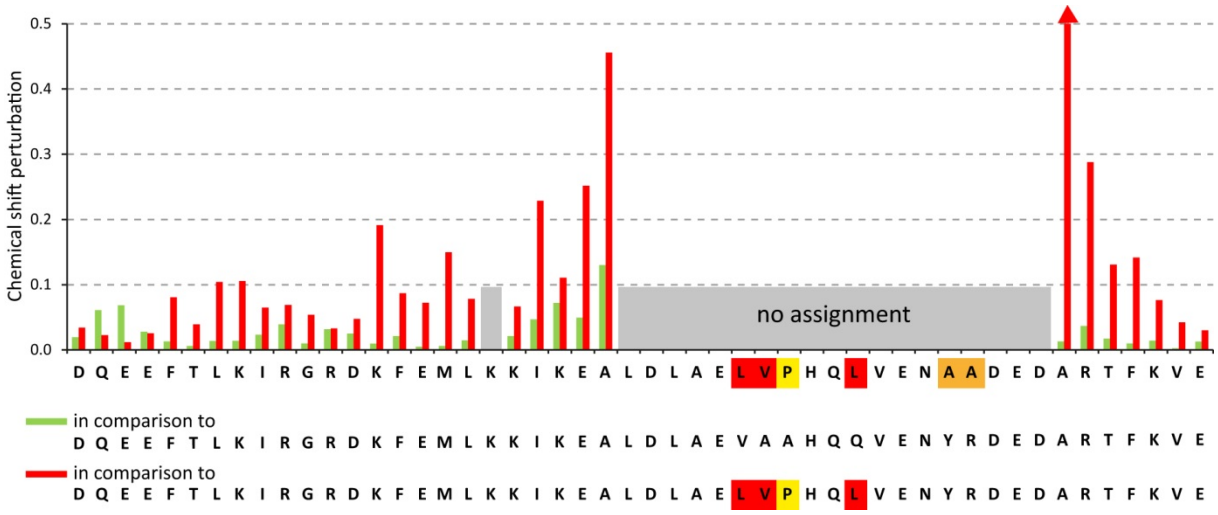


Figure 4-32 Analysis of the chemical shift perturbations (CSPs) observed upon the mutation of all crucial residues within the hinge region and the simultaneous mutation Tyr-Arg to Ala-Ala. The chemical shifts observed for the construct with the protein sequence, indicated under the bar diagram, were compared to those observed for constructs with the other two protein sequences, listed below. All constructs comprised aa 415-464 of *C.int.* p53/p73-a isoform 1 and deviate from each other in the underlined residues. The protein sequence in the middle corresponds to the wild type. Mutations towards larger hydrophobic side chains are underlined in red, those which change the respective residue into proline are underlined in yellow and the mutation of Tyr-Arg to Ala-Ala is indicated in orange. The NH chemical shift perturbations were calculated as described under 3.3.8. Chemical shift perturbations exceeding 0.5 are marked with arrow heads. Unassigned residues are marked with bars in light grey.

4.2.10. Change of secondary structure propensity upon mutation of hinge region

The TALOS+ predictions for α-helical conformation within the section corresponding to the second helix had differing confidence levels, and even for the construct, possessing all four mutations, the α-helical conformation was for each residue consistently predicted with a lower confidence as compared to the respective predictions for the chimeric construct. Furthermore the second helix in the mutant LAPHQL was predicted not to encompass the Tyr-Arg motif, which forms essential stabilizing interactions with the core. When referenced against the wild type the CSPs for residues within the β-strand and the first helix were consistently lower for the hinge region mutant compared to the chimeric construct (CSP data not shown; see TROSY spectra: Figure 4-21a and Figure 4-30).

Taken together these data suggested a more detailed analysis of the chemical shift data with respect to the fractions of the underlying conformers which actually really adopt α-helical conformation. Based on the CA, CB, HN and N chemical shifts the propensities for secondary structures were calculated using the program δ2D, which is described under 3.3.7. The propensities calculated with δ2D are presented in Figure 4-33.

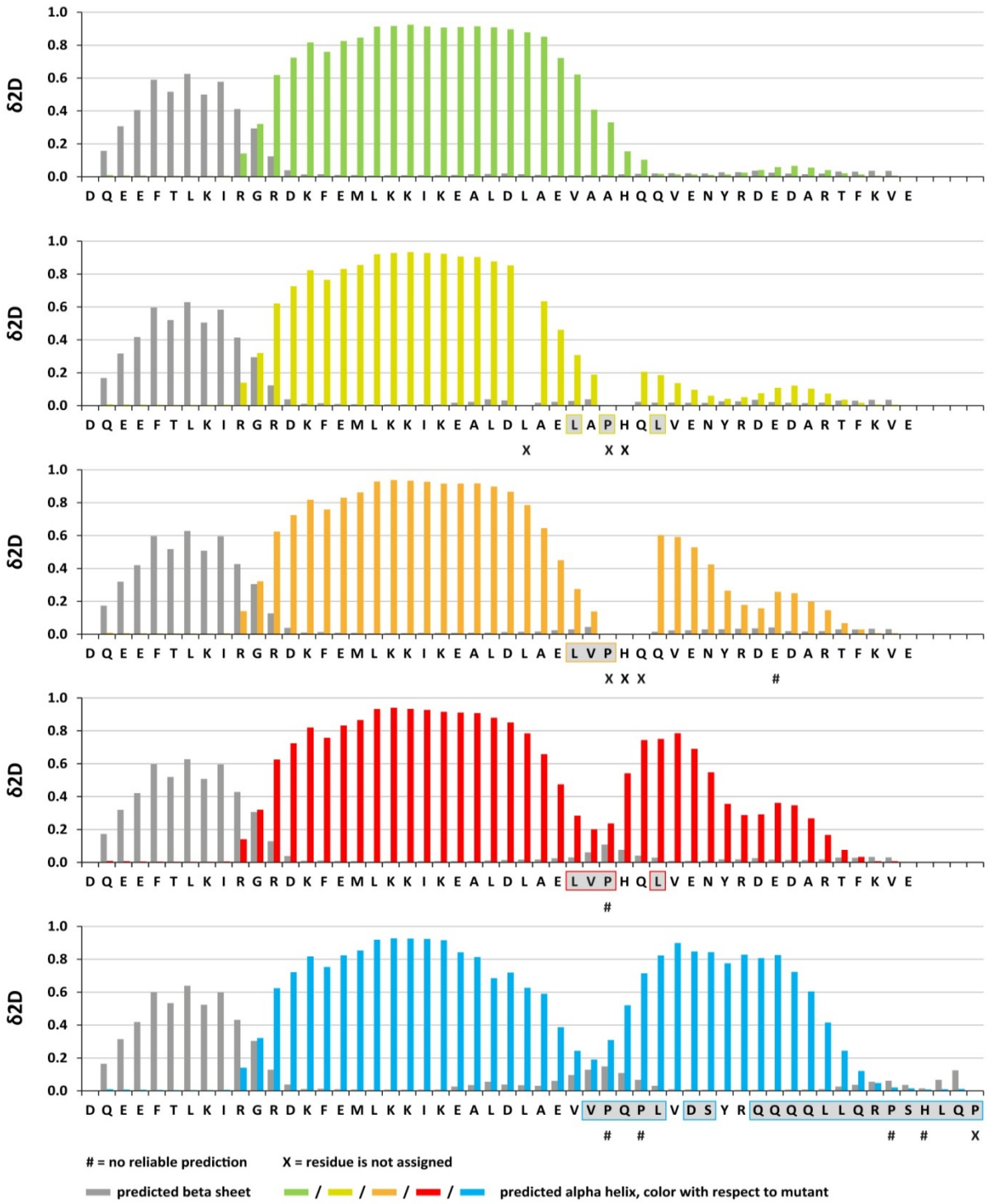


Figure 4-33 Propensities for β-strand and α-helical conformation as calculated with the δ2D method for several partially mutated constructs encompassing aa 415-464 of *C.int.* p53/p73-a isoform 1 and in comparison those for the chimeric protein in which aa 415-444 from *C.int.* p53/p73-a were fused to aa 381-404 from human TAp73α. The probability of occupation of secondary structure (secondary structure propensity) was calculated using the δ2D method as described in 3.3.7 with the chemical shifts of CA, CB, HN and N as input. The chemical shift values were obtained from spectra recorded at 25°C. Residues for which no reliable prediction was possible are labeled with # and unassigned residues are labeled with an X. For clarity only the propensities of β-strand and α-helical conformation are depicted; those for random coil and polyproline II are not reported. Each bar diagram shows the propensities for β-strand or α-helical conformation determined for the respective protein sequence, written below. Bars depicted in dark grey represent the propensity of the respective residue to adopt β-strand conformation. The bars representing the propensity towards α-helical conformation have the color, which was assigned to the respective construct. Residues, which differ from the *C.int.* wild type sequence, are underlined in light grey and boxed in the color assigned to the respective mutant.

The δ2D method is the latest and most advanced algorithm for the calculation of secondary structure propensities from chemical shifts.[171] However, the optimal methodology is still a subject of debate and to account for this I repeated the calculations with the two most popular earlier developed algorithms. The propensities calculated with the secondary structure propensity algorithm (SSP) and those calculated with the neighbor corrected structural propensity calculator (ncSPC) are reported in the appendix in section 7.2. The propensities calculated with the three programs were generously in good agreement, and therefore only the results obtained with δ2D are described here.

In short, the prediction of the protein backbone to adopt β-strand or α-helical conformation is based on the deviation of chemical shifts from their expected random coil values (chemical shift deviation = CSD). If a SSE is not rigidly folded but more an ensemble of different states, which only partially adopt β-strand or α-helical conformation, this might not be resolved when the different states exchange fast on the NMR time scale. In this case the determined CSDs are an average of the CSDs, which correspond to the different underlying conformational states, and hence, can be used to calculate the relative population size of conformers forming the respective SSE. Thus the calculated secondary structure propensities can be interpreted as the fraction of conformers in which the backbone of the respective residue adopts β-strand or α-helical conformation.

For residues, residing within the β-strand and the N-terminal half of the first helix, the respective calculated propensities to adopt β-strand or α-helical conformation were identical in all four constructs. Hence, the rigidity of the SSE was not modified by the mutations in the hinge region or the exchange of the C-terminal half of the protein sequence towards the respective sequence from human p73. Interestingly, the residues in the C-terminal half of the first helix were predicted to adopt α-helical conformation with a lower propensity when the sequence, corresponding to the hinge region and second helix, was exchanged towards the sequence from human p73. Consistent with the prediction by TALOS+ the first helix was predicted to be longer in the wild type than in the mutants and in the chimeric protein.

For residues within the section, corresponding to the second helix, the propensity to adopt α-helical conformation was extremely dependent on the sequence in the hinge region. In the wild type no residue within this section was predicted to adopt α-helical conformation in more than 0.6% of all conformers. But it shall be noted that the local maximum again aligned with those residues for which although being unstructured a surprisingly high RCI S2

order value had been predicted by TALOS+; Figure 4-19. For the construct with the hinge region and second helix protein sequence from human p73 residues in the middle of the second helix were predicted to adopt α-helical backbone conformation in more than 82% of all conformers. Hence, the second helix in this construct can be stated to be rigidly folded. Among the three investigated hinge region mutants the relative propensities resembled the TALOS+ prediction. The mutation LVPHQL resulted in the strongest gain in conformers possessing a second helix, and the mutation towards LAPHQL had by far the lowest effect. For the latter mutant residues within the section corresponding to the second helix were predicted to possess α-helical backbone conformation in not more than 20% of all conformers. Hence, averaged over time this mutant does not form a second helix and the TALOS+ prediction of a shortened second helix seems to be wrong although the confidence levels were rather high. For the mutants LVPHQL and LVPHQQ the second helix always showed two local maxima with very different propensities for α-helical backbone conformation. The higher maximum was between the hinge region and the Tyr-Arg motif and reaches up to 60% and 78% for the mutants LVPHQQ and LVPHQL, respectively. The position of the lower local maximum in α-helical propensity aligns with the position of the local maximum of the RCI S2 order value predicted by TALOS+ for the wild type construct. But surprisingly for the mutants LVPHQQ and LVPHQL the residues in this section were predicted to adopt α-helical conformation in only 25% and 36% of all conformers. Hence, even in the mutant, which possesses all four stabilizing mutations within the hinge region, averaged over time residues within the C-terminal half of the second helix and also the residues Tyr-Arg do not adopt α-helical backbone conformation.

Since that it had to be determined if in comparison to the wild type the change in tertiary structure caused by the mutations within the hinge results in a gain or in a loss of stability.

4.2.11. Determination of stabilizing effect of hinge region mutations by CD

It was shown that mutations in three to four crutial positions within the hinge region are sufficient to cause the formation of a second helix also in the TD of *C.int.* p53/p73-a. In the structure of the TD of human p73 the respective residues within the hinge regions of two adjacent subunits form a cluster of stabilizing hydrophobic interactions. But the mutations cause the formation of the second helix at the costs of the length of the first helix, which is shortened at its C-terminal end by about 3-4 residues with respect to the wild type. In

addition the C-terminal half of the second helix is not rigidly folded but adopts only partially α-helical backbone conformation. Therefore it was investigated by circular dichroism to which extent the respective mutations change the overall thermal stability of the TD.

Thermal melting curves were recorded by detection of the change in circular dichroism at 222 nm upon thermal denaturation of the proteins. In Figure 4-34 the raw data and the fitted melting curves are presented for the wild type of the *C.int.* p53/p73-a isoform 1 aa 415-464, the three hinge region mutants and as well the chimeric protein.

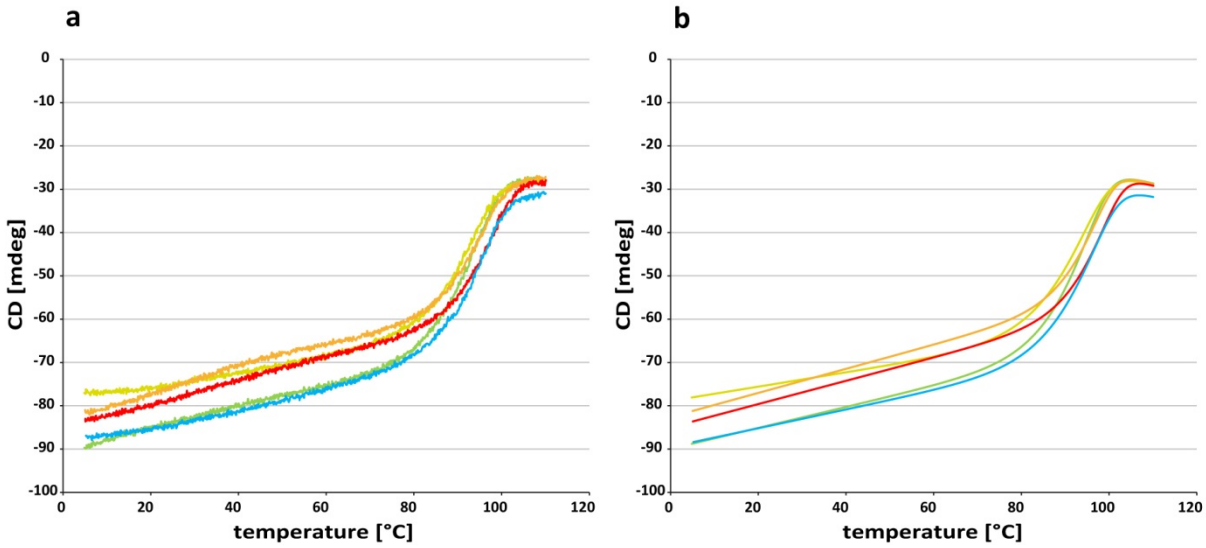


Figure 4-34 Thermal denaturation curves with recorded circular dichroism at 222 nm from wild type and several hinge region mutants of *C.int.* p53/p73-a isoform 1 aa 415-464 and as well the chimeric protein encompassing aa 415-444 from *C.int.* p53/p73-a fused to aa 381-404 from human TAp73α. Samples with 100 µM protein concentration (monomer) were freshly prepared in 1 x CD buffer. Experimental settings are described in section 3.3.1. Non-linear curve fitting by a least squares approach was performed as described in section 3.3.1.1. (a) raw data (b) fitted melting curves. The curves have the colors, which were as well assigned to the respective proteins in Figure 4-35.

To smooth the melting curves and to calculate from these data the fractions of folded protein as a function of temperature it was necessary to fit the curves to an appropriate model. In a quantitative analysis of the thermodynamic stability of the isolated TD from human p53 it had been shown that its thermal denaturation can be described as a reversible, two-state transition in which the folded tetramers are directly converted into unfolded monomers.[163] Hence, for this model to be appropriate intermediate states of denaturation are not allowed to be populated and the denaturation has to be reversible. And indeed for all investigated constructs of *C.int.* p53/p73-a TD the observed thermal unfolding was fully reversible. Upon lowering the temperature the signal corresponding to the folded protein was completely recovered (data not shown). The second precondition had

to be assumed from the homology of the proteins. In all known structures of tetramerization domains from proteins of the p53-family the single monomers are highly intertwined with one another and the assemblies are dominated by inter-molecular forces. The amino acid sequence within the core of the tetramerization domain (β-sheets and first α-helices) of *C.int.* p53/p73-a is highly conserved with respect to the TD of human p53 and the wild type predictively adopts a comparable tertiary structure. For the mutants it was shown that the second helix adopts a position with respect to the core that is comparable to the topology in the TD of human p73. The second helix unfolds when essential residues within the interface between core and second helix are mutated. Hence, the second helices wrap around the neighboring dimer in a clamp like fashion and can only fold in the tetramer. And the hydrophobic cluster in the hinge region, which is formed by the mutated residues, is as well only formed in the tetramer as the participating monomers do not originate from the same primary dimer.

Therefore, it was reasonable to assume that the thermal unfolding of *C.int.* p53/p73-a follows also a reversible, two-state transition in which disassembly of the tetramers and unfolding into monomers were coupled without intermediate states being populated. As these essential preconditions were fulfilled it was possible to determine the thermodynamic parameters of the thermal unfolding process by applying non-linear least squares curve fitting according to the characteristics of the unfolding process. The math is described in section 3.3.1.1.

The calculated fractions of folded tetramer as a function of temperature are presented in Figure 4-35 for the chimeric protein, which possesses hinge region and second helix sequence from human p73, as well as for the wild type of *C.int.* p53/p73-a isoform 1 aa 415-464 and the three mutants with the hinge region sequences LAPHQL, LVPHQQ and LVPHQL. Thermal denaturation was as well investigated for three of the hinge region mutants, which had shown rather low CSPs and which therefore were not investigated by NMR in more detail. The respective calculated fractions of folded tetramer versus temperature are presented in Figure 4-36 in comparison to the wild type (raw data and smoothened melting curve are not shown). The melting temperature (T_m) defined as the midpoint of thermal denaturation is throughout very high and thus the literature was searched for the respective value for wild type p53. In several studies employing CD the midpoint of thermal denaturation of human p53 TD was estimated to be between 68 and

73.5°C.[161,68,194] But as reported by Johnson et al. the range can be much bigger.[163] They determined the T_m by differential scanning calorimetry (DSC) with monomer concentrations ≥ 30 µM and found that it varies between 34.3 and 85.3°C with a strong dependency on the pH (3.0 – 7.0). With this information the high T_m values, which were determined for the wild type and the mutant forms of the TD of *C.int.* p53/p73-a, lie within the expectation range for a protein with a partially different sequence and therefore potentially additional stabilizing interactions. The high pH dependency of the thermal stability of human p53 can easily be explained as the primary dimers are stabilized by inter-chain salt bridges.[64,65] As the respective residues are conserved in *C.int.* p53/p73-a it was expected that the stability of the core would as well be highly dependent on pH, and therefore extreme care was taken to exclude pH variations between the different samples.

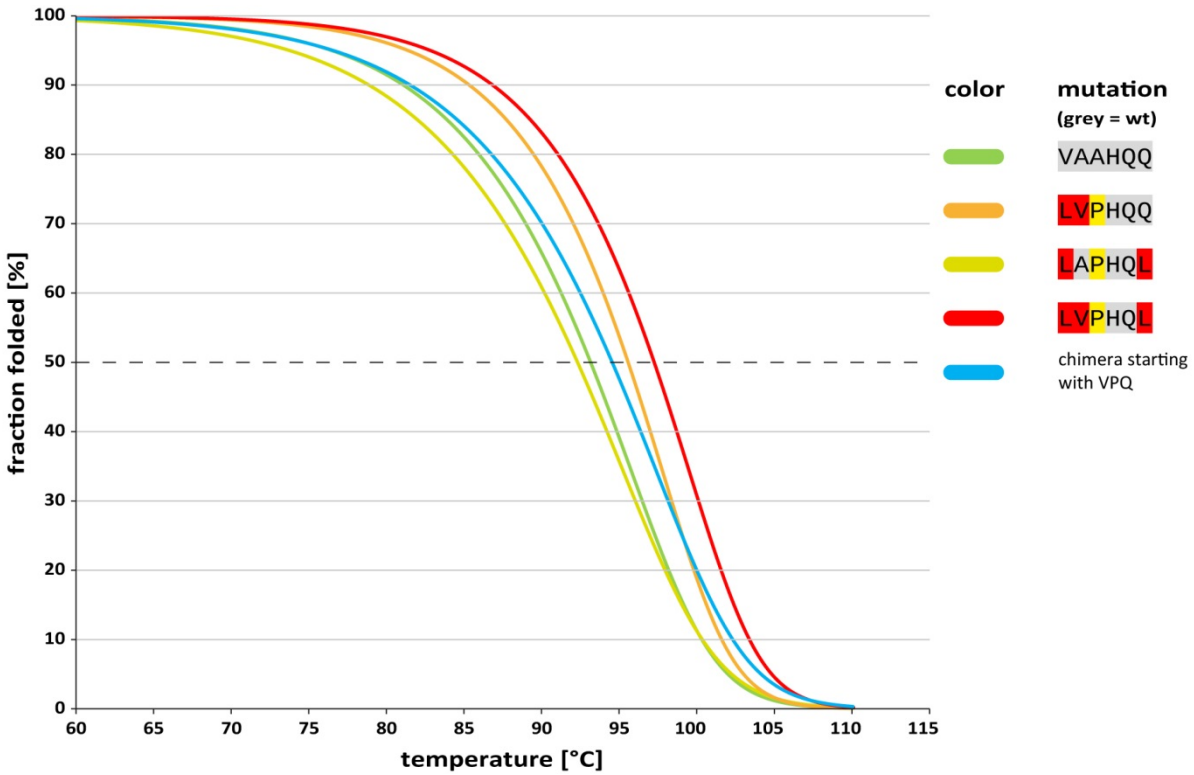


Figure 4-35 Fraction folded versus temperature plots calculated from thermal denaturation curves with recorded circular dichroism at 222 nm from wild type and several hinge region mutants of *C.int*. p53/p73-a isoform 1 aa 415-464 and also the chimeric protein encompassing aa 415-444 from *C.int.* p53/p73-a fused to aa 381-404 from human TAp73α. Samples with 100 µM protein concentration (monomer) were freshly prepared in 1 x CD buffer. Experimental settings are described in section 3.3.1. A two-state transition with coupled disassembly of the tetramers and unfolding of the monomers was presupposed and non-linear curve fitting was performed according to a least squares approach as described in section 3.3.1.1.

The thermodynamic parameters for the wild type and the mutant forms are reported in Table 4-4. For the wild type a T_m of 93.2°C was determined and only the chimeric protein as

well as the two mutants with the hinge region sequences LVPHQQ and LVPHQL were capable to increase this high value any further. All other combinations of mutations result in a reduction of thermal stability.

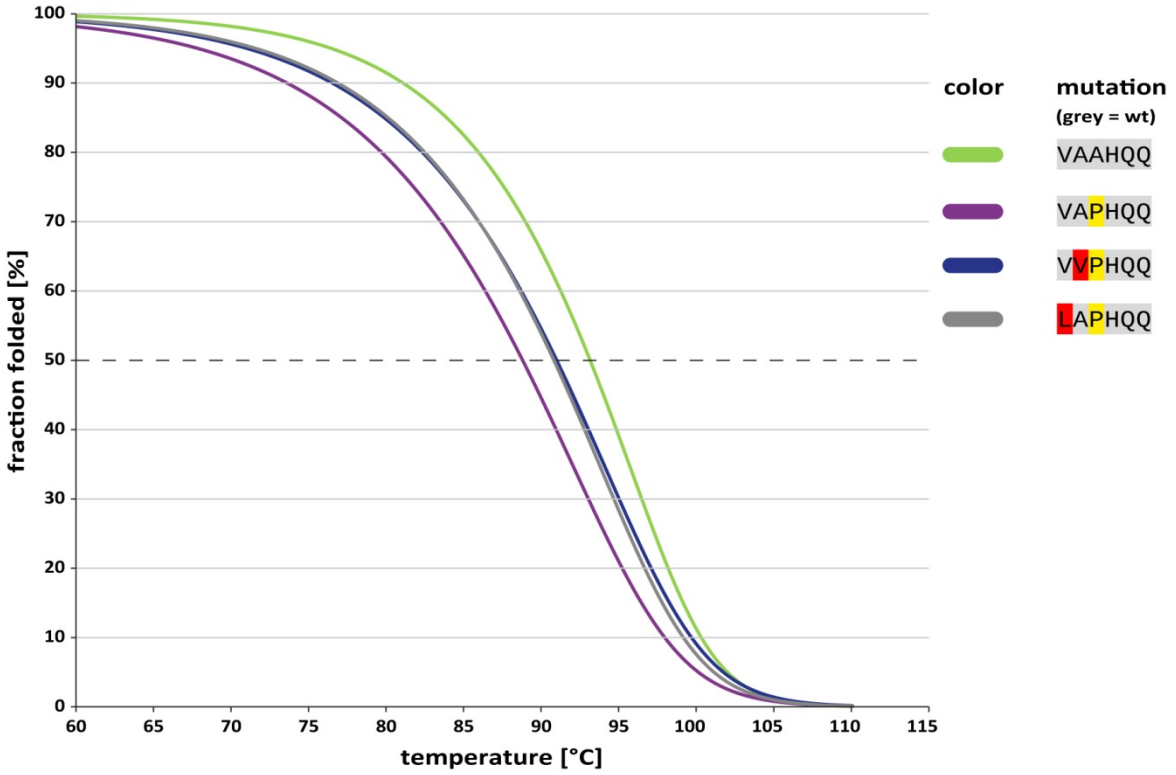


Figure 4-36 Fraction folded versus temperature plots calculated from thermal denaturation curves with recorded circular dichroism at 222 nm from wild type of *C.int*. p53/p73-a isoform 1 aa 415-464 and three mutants with only up to two mutated residues within the hinge region. Samples with 100 µM protein concentration (monomer) were freshly prepared in 1 x CD buffer. Experimental settings are described in section 3.3.1. A two-state transition with coupled disassembly of the tetramers and unfolding of the monomers was presupposed and non-linear curve fitting was performed according to a least squares approach as described in section 3.3.1.1.

Table 4-4 Thermodynamic parameters for wild type and mutant *C.int.* p53/p73-a isoform 1 aa 415-464

hinge region sequence	T_m [°C]	ΔH_U^{Tm} [kcal/mol]	ΔG_U^{Tm} [kcal/mol]	$\Delta\Delta G_U^{Tm}$ [kcal/mol] referenced to wild type
VAAHQQ (wild type)	**93.2**	**150.1**	**20.91**	**=**
VAPHQQ	**88.8**	**115.3**	**20.18**	**-0.73**
VVPHQQ	**91.0**	**122.8**	**20.30**	**-0.61**
LAPHQQ	**90.8**	**127.6**	**20.45**	**-0.46**
LAPHQL	**92.3**	**133.5**	**20.77**	**-0.15**
LVPHQQ	**95.6**	**180.2**	**20.78**	**-0.13**
LVPHQL	**97.3**	**178.7**	**20.84**	**-0.08**
chimeric protein*	**94.6**	**139.8**	**20.68**	**-0.23**

* aa 415-444 from *C.int.* p53/p73-a fused to aa 381-404 from human TAp73α

For human p53 TD Mateu et al. calculated values of ΔH_U^{Tm} between 155 and 165 kcal/mol, which differed with the length of the investigated constructs.[161] The respective value for wild type *C.int.* p53/p73-a lies in a comparable range and was determined to be 150.1 kcal/mol. This indicates that the thermal unfolding happens with an equally high degree of cooperativeness. The two mutants with the hinge region sequences LVPHQQ and LVPHQL were found not only to increase the thermal stability but also to increase the cooperativity of unfolding as indicated by values of ΔH_U^{Tm} of 180.2 and 178.7 kcal/mol, respectively. The calculated values of ΔH_U^{Tm} for all other combinations of mutations were lower than the value of the wild type. Interestingly, the chimeric protein was found to be characterized by a lower degree of cooperativeness, as well, although the T_m of this construct lies in between the wild type and the stabilizing hinge region mutants.

4.2.12. Determination of the gain in stability of the second helix by measurements of heteronuclear nuclear Overhauser effects

Using the δ2D method the secondary structure propensity along the sequences of wild type and mutants were calculated based on the chemical shifts. This revealed that residues residing within the section corresponding to the second helix do not constantly exist with α-helical backbone conformation. Furthermore, the differences in thermal denaturation revealed that the ability of a combination of mutations within the hinge region to stabilize the TD correlates with the induced increase in time averaged helical content within this second helical section. This opened up the question how flexible the corresponding residues move with respect to the core TD and how a decreased temperature might influence the structure of the wild type. At first ^{15}N-[^{1}H]-heteronuclear NOE-TROSY experiments were recorded with the aim to investigate how the mutations within the hinge region influence the molecular motion of residues within the second helical section. In Figure 4-37 the determined heteronuclear nuclear Overhauser effects (hetNOEs) along the protein sequences are reported for the wild type and the three hinge region mutants LAPHQL, LVPHQQ and LVPHQL.

For the section, spanning the β-strand and the first helix, hetNOEs between 0.6 and 0.8 were determined for all constructs denoting that, as shown before, the core of the TD is stably folded. The hetNOEs determined for residues residing within the section corresponding of the second helix were on the contrary highly affected by substitutions

within the hinge region and revealed that with respect to the wild type all three mutants resulted in a gain in rigidity. The hetNOEs, which were determined for the respective residues in the wild type, form a plateau around zero. The absolute values clearly show the absence of stable structure but the plateau is particularly noticeable, since actually the hetNOE values would be expected to more or less constantly further decrease with increasing distance from the folded core. The most suitable explanation for this observation seems to be a local stabilization limited to approximately four to six residues in the rear range. And indeed on basis of the chemical shift data a local maximum of the RCI S2 order value was calculated, Figure 4-19, whose position would fit to this assumption. The plateau formation was partially also observed for the mutants and here particularly for those with the weaker stabilizing effect. The degree to which the three mutants increased the rigidity of the second helical section reflected their order of rank from the previous analysis by TALOS+ and δ2D.

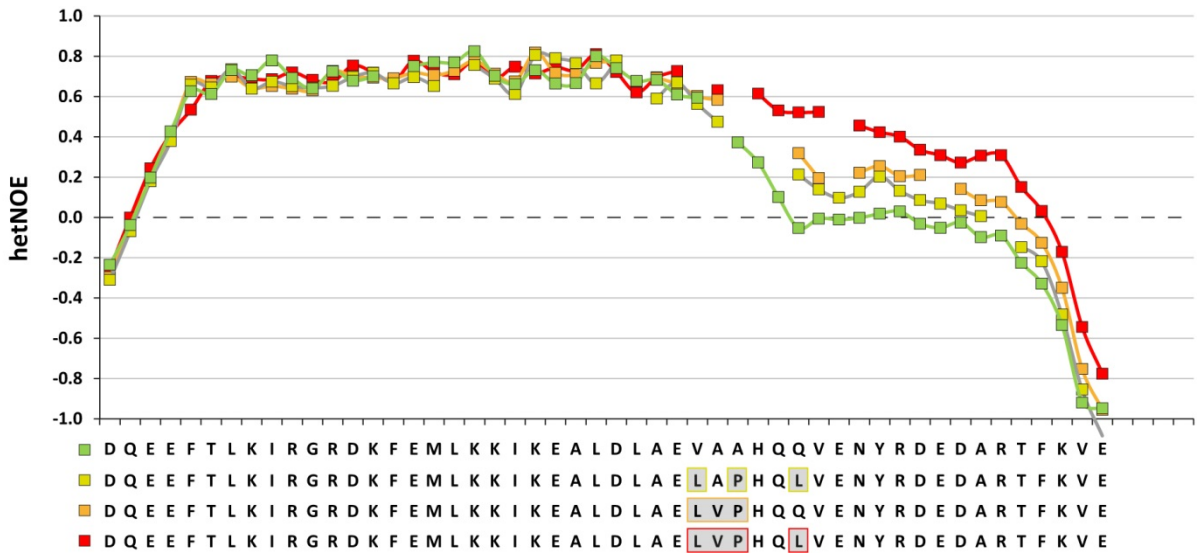


Figure 4-37 ^{15}N-[^{1}H] heteronuclear nuclear Overhauser effects (hetNOEs) determined at 25°C for wild type and several hinge region mutants of *C.int*. p53/p73-a isoform 1 aa 415-464. ^{15}N-[^{1}H]-heteronuclear NOE-TROSY spectra were recorded and evaluated as described in section 3.3.5. The samples contained ≥ 1.5 mM protein monomer in RE buffer. Spectra were recorded at 25°C. Residues, which differ from the *C.int.* wild type sequence, are underlined in light grey and boxed in the color assigned to the respective mutant.

Only the mutant with all four amino acid substitutions increased the hetNOE to values above 0.4 and this as well only for residues in the upper half of the second helix. The threshold of 0.4 is important as values below are commonly interpreted as (mainly) unstructured. In the wild type the hetNOE values drastically drop after the end of the first helix. This indicates highly independent movement of the core TD and C-terminal section. In

the stabilized mutant with the hinge region sequence LVPHQL the hetNOE values do not form a local minimum between the core TD and the second helix but constantly decrease in small steps between the end of the first helix and the projected end of the second helix. This proves that the second helix is folding onto the core TD. But at the same time it also reveals that this change in the overall tertiary structure is particularly stabilized by the hinge region and only subordinately by the formation of the second helix. The constantly decreasing hetNOE values indicate that the C-terminal half of the second helix is affected by helical fraying and caused by a lack of stabilizing interactions present in a variety of different, only partially helical conformations. The lower degree of stabilization within the C-terminus of the second helix makes it likely that no residue located C-terminal to the conserved Tyr-Arg motif is engaged in stabilizing interactions with the core TD.

4.2.13. Lower temperatures result in stabilization of the second helix

With the aim to investigate whether the structure of the wild type protein might be affected by a reduction in temperature an ^{15}N-[^{1}H]-heteronuclear NOE-TROSY experiment was recorded at 10°C, and to facilitate the translation of the assignment as well several [^{15}N, ^{1}H]-TROSY experiments were recorded at temperature intervals of 3°C between 10-25°C. In Figure 4-38 the respective determined hetNOE values along the protein sequence are reported for 10°C and 25°C measuring temperature.

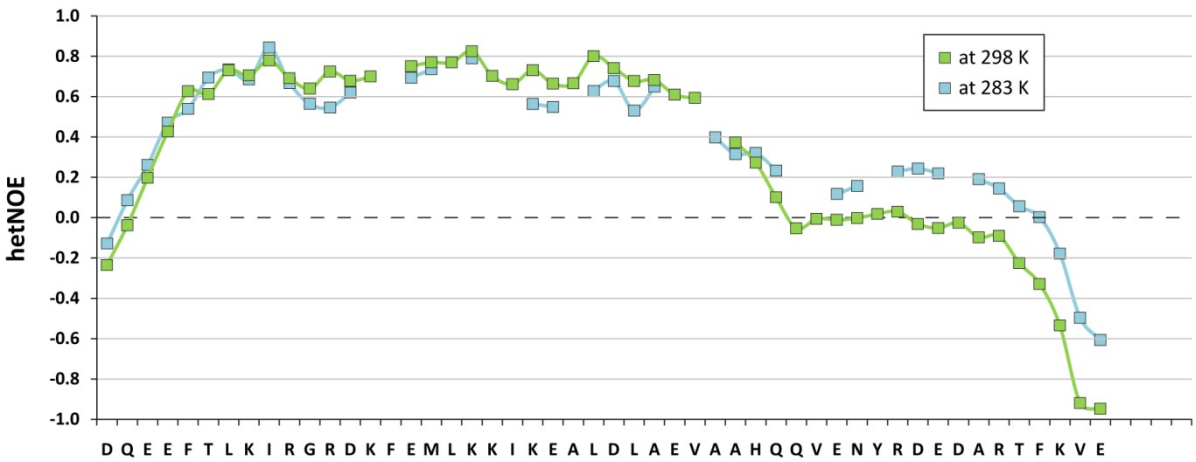


Figure 4-38 ^{15}N-[^{1}H] heteronuclear nuclear Overhauser effects (hetNOEs) determined for wild type *C.int*. p53/p73-a isoform 1 aa 415-464 at 10°C and 25°C. ^{15}N-[^{1}H]-heteronuclear NOE-TROSY spectra were recorded and evaluated as described in section 3.3.5. The sample used for both measurements contained 2.5 mM protein monomer in RE buffer. Spectra were recorded at 10°C and 25°C.

Upon a reduction of the temperature the section, corresponding to the second helix, shows a gain in rigidity. The hetNOEs are still in the range, generally interpreted as indicative for rather unstructured and flexible regions. But the gain of up to 0.2 is remarkably. With respect to the stabilizing effect of the mutations within the hinge region the mechanism underlying this temperature driven stabilization seems to be different. Although the assignment has some gaps it is obvious that the hetNOE values for residues within the second helical section follow an independent curve. This indicates that the second helix is preferentially stabilized by internal interactions.

The observed local minimum between the core TD and the stabilized C-terminal section aligns with the positions, which would correspond to the first turn of the second helix in human p63 or p73. This is remarkably, since residues with local maxima in backbone flexibility do generally not reside within folded SSEs but in loops and at hinges. At first sight it therefore seems to be excluded that a second helix, whose formation could potentially be triggered by a reduction in temperature, would start at the same position as in human p73. Furthermore the first helix appears to maintain about the same length as at 25°C, therefore presumably still being circa one turn longer than the corresponding helices in human p63 and p73, and spanning at least the first half of the hinge region. And this again would be expected to prevent the protein from adopting a comparable tertiary structure.

But the picture changes when it is taken into account that the relative sizes of the underlying conformational populations, with and without a second helix, deviate strongly from each other. The determined hetNOE values for residues within the center of the second helix are about 0.2, and for residues within rigidly folded SSEs with no conformational exchange generally values around 0.8 can be expected. For the purpose of an estimation, it was assumed that if the conformers possessing a second helix could be separately assayed the respective hetNOE values for this section would as well be 0.8 or less. It was further assumed that the exchange between the different conformers is as fast on the NMR time scale that the determined hetNOE values report on the population weighted average of the conformational rigidity of the underlying residues. According to these assumptions the second helix is folded to ≤ 25% in the temporal means, and hence unfolded to ≥ 75%. The conformers, which are lacking the second helix, dominate the spectrum and it had been shown that at 25°C the first helix is partially spanning the hinge region; Figure 4-31. For the first three residues (VAA) the propensities for α-helical backbone conformation as calculated

with the δ2D method lie in between 0.328 and 0.619; Figure 4-33. Although the fraction of conformers, lacking the second helix, is reduced to 75% at 10°C, therefore, the average helicity of these residues presumably still exceeds the helicity of the residues from the center of the second helix. And this in turn accounts for the observed shift of the local maximum of local backbone flexibility from the hinge region towards the residues residing within the first turn of the second helix. Thus it could be accepted that by temperature decrease gradually the formation of a second helix could be induced, which possesses similarity to those in human p73.

This interpretation was further supported by the observation that at 10°C several residues within the β-strand and the first helix show reductions of the hetNOE with respect to the values determined at 25°C. Hence, the local rigidity of the backbone must be reduced as a result of increased conformational exchange. The most affected residues within the core TD are Arg, at one position after the highly conserved Gly residue between the β-strand and the first α-helix, and the motif Lys-Glu within the first helix. The Arg residue was always among those residues, which showed the most intense CSPs upon aa exchanges in the hinge region. And this can presumably be attributed to a change in the angle between the β-strand and the first helix, which is caused by the second helices packing onto the core.

The gain in conformational exchange on the motif Lys-Glu was particularly interesting, since the two residues are highly conserved in vertebrate p63 and p73, and in the respective structures of the TDs they are identified as directly interacting with the second helix.

In summary it can be stated that the folding of a second helix can be gradually induced by decreasing the temperature. The resulting overall tertiary structure likely resembles those of the TDs of human p63 and p73.

4.3. Structural analysis of the TD of *Ciona intestinalis* p53/p73-b

4.3.1. Detailed sequence comparison of the TD sequences of *Ciona intestinalis* and *Ciona savignyi* p53/p73-b

The protein p53/p73-b is the result of a relatively recent gene duplication event, which happened independently of the gene duplications during the evolution of p53-like proteins in vertebrates; Figure 4-1. This duplication is specific to the *Ciona* lineage and occurred after the divergence from the last common ancestor shared by *Ciona* species and vertebrates. The corresponding genomic sequence does not contain any introns, and therefore, retro-transposition was suggested as the mechanism most likely associated with this duplication.[139] As a result of the lack of introns or internal splice sites only a single isoform of *Ciona* p53/p73-b is expressed. Figure 4-39 shows the alignment of the very C-terminal sections from *C.int.* and *C.sav.* p53/p73-b, which correspond to the TD. Throughout the full length protein sequences the two proteins share 50% amino acid identity with maxima within the sections corresponding to the DBD (65% identity) and the TD (75% identity). Compared to the respective degrees of conservation determined for the comparison of the p53/p73-a proteins of the two species (total sequence of isoform 1: 75.6%, DBD: 88%, TD 84%) this is much lower. This indicates a generally higher degree of degeneracy on the new genomic copy p53/p73-b. Remarkably, the sequence of the TD is least affected.

```
C.sav.p53p73-b   324 KKGVLENKKSSKR-----------PLLPKKDNL--NIDKGESEVYSLTIIGKRKFEKVKEYKEALDLMDFVQPEVKKACCDKNQM 395*
C.int.p53p73-b   335 KKNVTSQNKSRKRLCKSATNSKSIPVVQKNEENKVDNDDDGDVVYTLNIRGKRKFEKVKEYKEALDLLDYVQPDVKKACCQRNQI 419*
```

conserved in homologous proteins from deuterostome species (degree of conservation > >)

conserved only in C.int. and C.sav.

Figure 4-39 Alignment of the tetramerization domain sequences of p53/p73-b from *Ciona intestinalis* (*C.int.*) and *Ciona savigny* (*C.sav.*) The blue bar above the sequences indicates the section spanned by the secondary structure elements in the tetramerization domain of human p73. An asterisk indicates the C-terminal end of the natural protein sequence.

4.3.2. Determination of section spanned by secondary structure elements within the C-terminus of *Ciona intestinalis* p53/p73-b

The sequence which aligns to the TDs of vertebrate p53-like proteins resides at the very C-terminal end of *C.int.* p53/p73-b. Hence, within the C-terminal section SSEs, which would not be part of the TD, could be excluded. To determine a minimal sequence, which

comprises all SSEs, several differently truncated constructs were cloned into pBH4. In Figure 4-40 the constructs are presented as an alignment with the natural protein sequence. The first two aa, GS, at the N-terminus of the constructs result from the C-terminal portion of the TEV cleavage site, and hence, do not align. For this reason they are left out in the description and the comparison of the different truncations. In the recorded NMR spectra their backbone NH cross peaks were anyhow not detectable.

```
C.int.p53p73-b    335 KKNVTSQNKSRKRLCKSATNSKSIPVVQKNEENKVDNDDDGDVVYTLNIRGKRKFEKVKEYKEALDLLDYVQPDVKKACCQRNQI 419*
aa 357-419        357                       GSIPVVQKNEENKVDNDDDGDVVYTLNIRGKRKFEKVKEYKEALDLLDYVQPDVKKACCQRNQI 419*
aa 374-419        374                                   GSDGDVVYTLNIRGKRKFEKVKEYKEALDLLDYVQPDVKKACCQRNQI 419*
aa 374-410        374                                   GSDGDVVYTLNIRGKRKFEKVKEYKEALDLLDYVQPDVK          410
```

Figure 4-40 Alignment of the C-terminus of *Ciona intestinalis* p53/p73-b and the corresponding constructs. The blue bar above the sequences indicates the section spanned by the secondary structure elements in the tetramerization domain of human p73. An asterix indicates the C-terminal end of the natural protein sequence. The upper line shows the aa sequence of the C-terminal end of the natural protein. The lower three lines show the investigated constructs with the coloration used in the following spectra overlays and analysis of chemical shift perturbations. For technical reasons all constructs have one or two additional residues at the N-terminus, which are marked with grey background. This GS dipeptide is a remnant from the N-terminal TEV cleavage site and does not align to the protein sequence under investigation. For this reason it is left out in the numbering and description of the constructs.

The expression in NEB T7 express cells, the cell lysis as well as the purification by IMAC, TEV cleavage and reverse IMAC were performed following the standard protocols for this project as described in the method section. Since the C-terminus of the protein contains two cysteine residues the respective reducing buffers were used throughout all purification steps. All proteins were purified to ≥ 98% as judged from Coomassie stained Tricine-SDS-PAGE gels. Gel filtration using a preparative scale Superdex75™ 16/60 column with settings as described was included as a last purification set (polishing) and to verify homogeneity. An analytical scale Superdex75™ 10/30 column was used to determine the oligomeric state and to prepare samples for NMR. All NMR spectra for this subproject were recorded from samples with ≥ 1 mM monomer concentration and at a temperature of 25°C; if not indicated differently.

An overlay of the [^{15}N, ^{1}H]-TROSY spectra of the two constructs, which encompass the complete C-terminus, is shown in Figure 4-41. The spectrum corresponding to aa 357-419 revealed that the sequence stretch upstream of Asp 374 is most likely unstructured. With respect to aa 374-419 additional peaks clustered exclusively within the range of 7.7 – 8.6 ppm (^{1}H frequency) which indicates that the residues do not have a distinct backbone conformation. To ensure that the N-terminal truncation at Asp 374 would not affect the structure of the TD, the CSPs were calculated, Figure 4-42. At first the construct

encompassing aa 374-419 was sequentially assigned and then the assignment was translated to the corresponding peaks within the spectrum of the N-terminally longer construct.

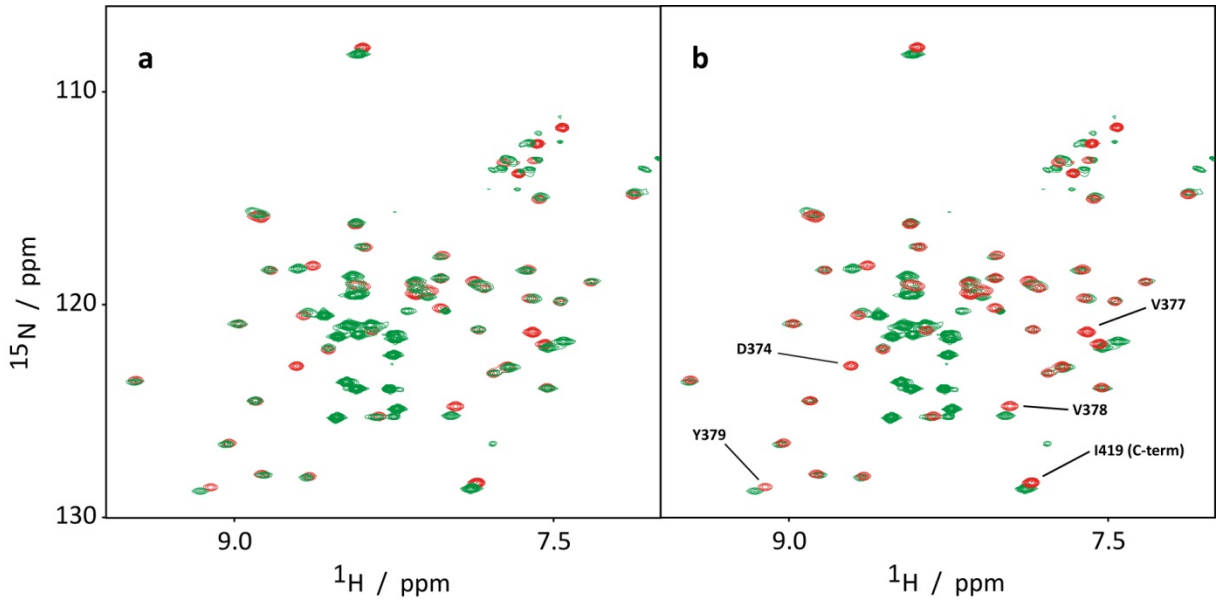


Figure 4-41 Overlay of [^{15}N, ^{1}H]-TROSY spectra of two different constructs each encompassing the TD located in the very C-terminal section of *Ciona intestinalis* p53/p73-b but with different N-terminal lengths. (a) The spectrum of the N-terminal longer construct (aa 357-419) is depicted in green and overlaid onto the spectrum of the N-terminal shorter construct (aa 374-419) depicted in red. (b) The overlay is reversed. Both spectra were recorded at 25°C.

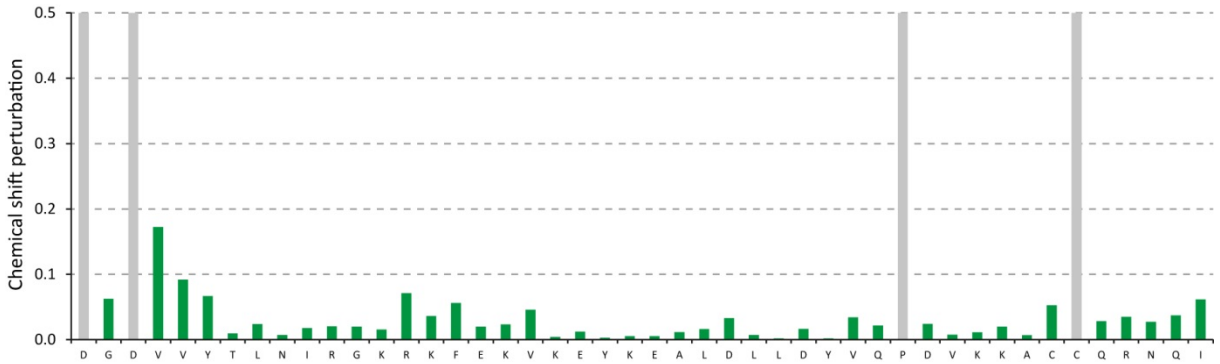


Figure 4-42 Chemical shift perturbation upon N-terminal elongation of the construct encompassing aa 374-419 of *Ciona intestinalis* p53/p73-b by 17 residues. The sequential assignment was done for the construct encompassing aa 374-419. Wherever possible the assignment was transmitted onto the corresponding peaks in the spectrum of the N-terminal longer construct, encompassing aa 357-419. The NH chemical shift perturbation was calculated as described under 3.3.8. Residues, which could not be assigned, are marked with grey bars.

The majority anyhow overlaid pretty good and only the first five to six residues at the N-terminus showed noteworthy CSPs. But these could be explained by the proximity of the truncation site and were not differently expected. The small CSP which was observed for the very C-terminus was attributed to possible minor errors in the adjustment of the pH of the samples. Taken together the truncation at Asp 374 was shown not to affect the TD.

Next the truncation of the last nine residues at the C-terminus was investigated. Figure 4-43 shows the overlay of the [^{15}N, ^{1}H]-TROSY spectra corresponding to aa 374-419 and the C-terminal truncated construct corresponding to aa 374-410 of *C.int.* p53/p73-b.

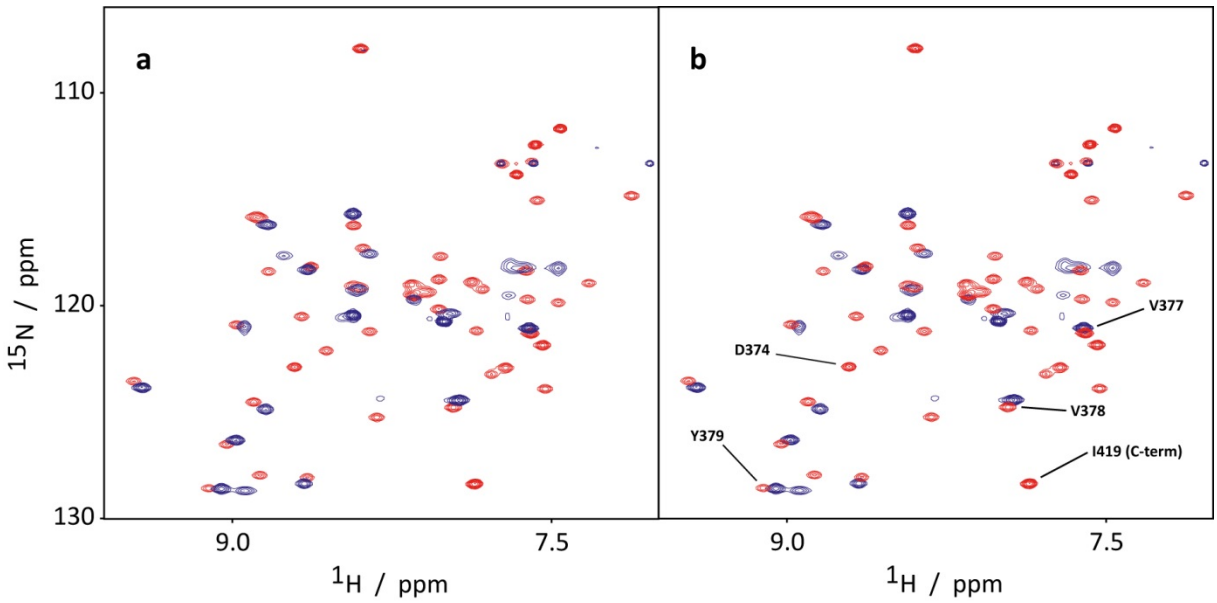


Figure 4-43 Overlay of [^{15}N, ^{1}H]-TROSY spectra of constructs encompassing either aa 374-419 or aa 374-410 of *Ciona intestinalis* p53/p73-b. (a) The spectrum of the C-terminal truncated construct (aa 374-410) is depicted in blue and overlaid onto the spectrum of the construct encompassing the full C-terminus (aa 374-419) depicted in red. (b) The overlay is reversed. Both spectra were recorded at 25°C.

Almost every peak within the spectrum of the C-terminally truncated construct is shifted with respect to the respective peak in the spectrum of the construct encompassing the full C-terminus. Therefore, as reported in Figure 4-44, the CSPs for many residues were respectable. With that it was shown that the truncation of the last nine residues massively affects the conformation of the TD. Therefore the construct corresponding to aa 374-419 was selected for further structural investigation.

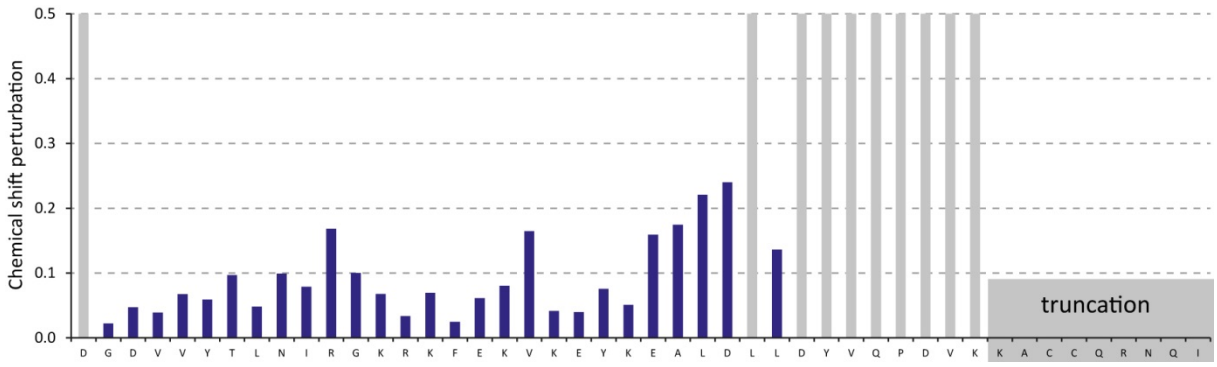


Figure 4-44 Chemical shift perturbation upon C-terminal truncation after aa 410 with respect to the construct encompassing aa 374-419 of *Ciona intestinalis* p53/p73-b. The sequential assignment was done for the construct encompassing aa 374-419. Wherever possible the assignment was transmitted onto the corresponding peaks in the spectrum of the C-terminal truncated construct, encompassing aa 357-410. The NH chemical shift perturbation was calculated as described under 3.3.8. Residues, which could not be assigned, are marked with grey bars.

4.3.3. *Ciona intestinalis* p53/p73-b TD possesses a stably folded second helix

To investigate how many SSEs compose the TD of *C.int.* p53/p73-b, the backbone chemical shifts of the construct, encompassing aa 374-419, were completely assigned as described in method sections 3.3.4 and 3.3.9. The chemical shifts of CA, CB, CO, HA, HN and N were then used as input for TALOS+. The prediction of the most probable backbone conformations, which is shown in Figure 4-45, revealed that the TD possesses a second helix. And this helix was predicted with maximal values of confidence for the residues within its central positions.

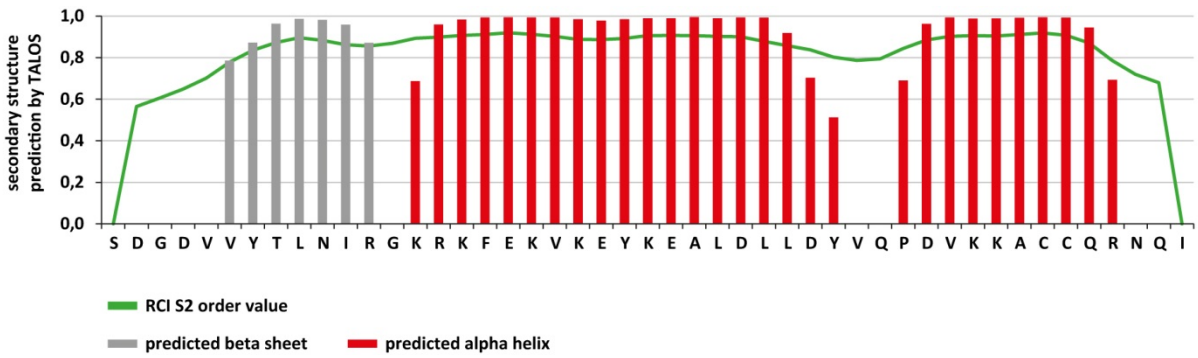


Figure 4-45 Prediction of secondary structure and RCI S2 order value by TALOS+ for *C.int.* p53/p73-b aa 374-419. The secondary structure and the RCI S2 order value were predicted using TALOS+ as described in 3.3.6. The chemical shifts of CA, CB, CO, HA, HN and N were used as inputs. The bars colored in grey represent predictions for β-strand conformation, those colored in red α-helical conformation. The RCI S2 order value is presented as a green line.

In agreement with that also the RCI S2 order value was predicted to have the same high level as for residues, residing within the first helix. The second helix was therefore predicted

to have the same conformational stability as the core of the domain. This was as well confirmed by the calculation of secondary structure propensities using the δ2D method; data not shown. The identification of this stably folded second helix was important in particular, since the respective sequence does not align to any extent with human p63 and p73. Most noticeably the Tyr-Arg motif is not conserved. In addition the second helix starts with a proline, which is therefore not in the N-cap position. With respect to the proline residue in the N-cap position of the second helix in the TD of human p63 and p73 this proline residue is shifted by one position to the C-terminus. Therefore this proline is not conserved but the second helix starts at the same position.

4.3.4. The second helix in the tetramerization domain of *Ciona intestinalis* p53/p73-b is not essential for the oligomeric state

To investigate if the oligomeric state of the tetramerization domain of *C.int.* p53/p73-b depends on the second helix the sedimentation coefficient of a truncated construct (aa 374-410) was determined by ultracentrifugation.

After purification, removal of the 6xHis-tag and concentration a sample of the protein was applied onto a Superdex 75 16/60 column equilibrated and run in Nickel buffer A1. Opposed to NMR the detection in analytical ultracentrifugation is not hampered by high salt concentrations and thus this buffer having a higher ionic strength was selected to ensure that no aggregates interfere with a valid determination of the MW. The sample for the ultracentrifugation experiment was taken from the middle of the single elution peak.

The tetramer has a calculated molecular weight (MW) of 18120.4 Da. As shown in Figure 4-46 the peak maximum appeared about 2 to 2.5 kDa lower. Hence, the truncated construct indeed forms predominantly tetramers in solution and the second helix is not essential to keep the dimeric subunits together. But the protein might not have been exclusively tetrameric although the concentration of the sample was 30 µM, and thus, fairly above the expected range for the K_d of tetramerization (~ 100 nM – 1 µM for human p53)[50]. However, higher concentrations did not affect the position of the peak maximum with respect to MW and measurements with lower concentrated samples were hampered by limited detection.

The truncated construct encompassing aa 374-410 is still spanning the residues corresponding to the first turn of the second helix; Figure 4-45. Therefore a further truncated construct (aa 374-406) was cloned, which is ending with the residue preceding the

proline residue at the beginning of the second helix. But this construct expressed in the form of soluble aggregates and precipitated after the removal of the 6xHis-tag.

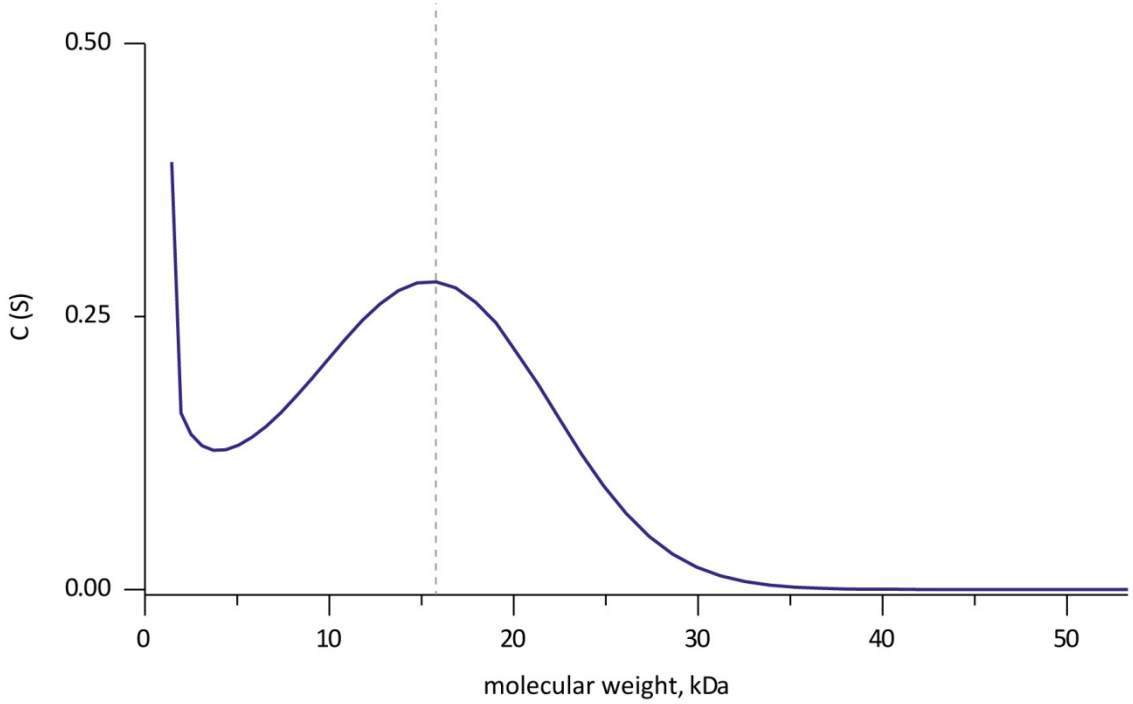


Figure 4-46 Determination of the oligomeric state of the second helix depleted construct aa 374-410 by ultracentrifugation. The sedimentation coefficient (C(S)) of a sample of *Ciona intestinalis* p53/p73-b aa 374-410 (plus preceding dipeptide GS) was determined by ultracentrifugation. The construct corresponds to the tetramerization domain being truncated within the second helix. The protein monomer concentration was 30 µM buffered in 400 mM NaCl, 25 mM Tris pH 7.8. The calculated molecular weight of the tetramer is 18120.4 Da.

4.3.5. Determination of the NMR solution structure of *C.int.* p53/p73-b TD

When being compared to vertebrate p53-like proteins the sequence of the TD of *C.int.* p53/p73-b shows a high degree of conservation for residues, residing within the β-strand and the first α-helix. In contrast to that the sequence of the second helix, which is located at the very C-terminus of the protein, does not possess residues with conservation in vertebrate homologs. This qualified the TD of *C.int.* p53/p73-b as an interesting target for detailed structural investigation.

In order to enable the calculation of the NMR solution structure of *C.int.* p53/p73-b aa 374-419 the extensive assortment of experiments, which is listed in Table 3-2, had to be recorded. The respective samples contained (monomer) protein concentrations between 2.5 and 4 mM buffered in reducing RE buffer, and all experiments were recorded at 25°C. The assignment of the side chain resonances and the pro-chiral methyl groups of Leu and Val was

carried out as described in the method section 3.3.9. The homo-tetrameric structure of the protein is a special challenge for NMR since each residue is present four times in an chemically in each case perfectly identical environment. This makes it impossible to assign resonances to specific monomers if the respective samples are uniformly labeled. But it is essential for the calculation of a structure on the basis of NOE derived distance constraints to specifically assign within which monomers the involved nuclei reside with respect to each other. To address this problem a set of three different NOESY experiments was recorded. The experiments all rely on a non-uniform labelling scheme of the respective samples, and the tetramers needed to consist predominantly of monomers with different labeling. As the K_D of tetramer formation is (presumably) in about the same low range as determined for other TDs from p53-like proteins it was necessary to ensure efficient mixing by denaturation and refolding. The NOESY experiments used for the discrimination of inter- from intra-chain proximity and the preparation of the respective samples are described in detail in method section 3.3.10. The assignment of the respective NOESY peak lists was done exclusively manually and the identification of the positions with close proximity towards at least one other monomer facilitated the assignment of the spectra recorded with the not monomer-selective, regular NOESY experiments.

The calculation of the structure was performed by Henry Jonker using the software CYANA and ARIA/CNS as described in detail in method section 3.3.11. The statistics of the final structure after water refinement are reported in Table 4-5.

The structure, which is presented in Figure 4-47, has great similarity to the TDs of human p63 and p73. It can equally well be described as a dimer of dimers. Two monomers form a symmetrical primary dimer via intermolecular antiparallel β-sheets and antiparallel helix packing. Within the individual chains the β-strand is followed by a sharp turn at whose central position a glycine residue is highly conserved. As shown in Figure 4-48 these primary dimers are stabilized by salt bridges and hydrogen bonds, which are formed between residues at the adjacent ends of the antiparallel oriented first helices, and hence always occur two times within each primary dimer. Asp 400 and Asp 403 form salt bridges with Lys 388 from the adjacent monomer and this cluster of polar interactions is further extended by an intra-chain salt bridge between Asp 400 and Lys 396 as well as a hydrogen bond between the side chain amino group of Lys 396 and the main chain carbonyl of Asn 382 from the other monomer.

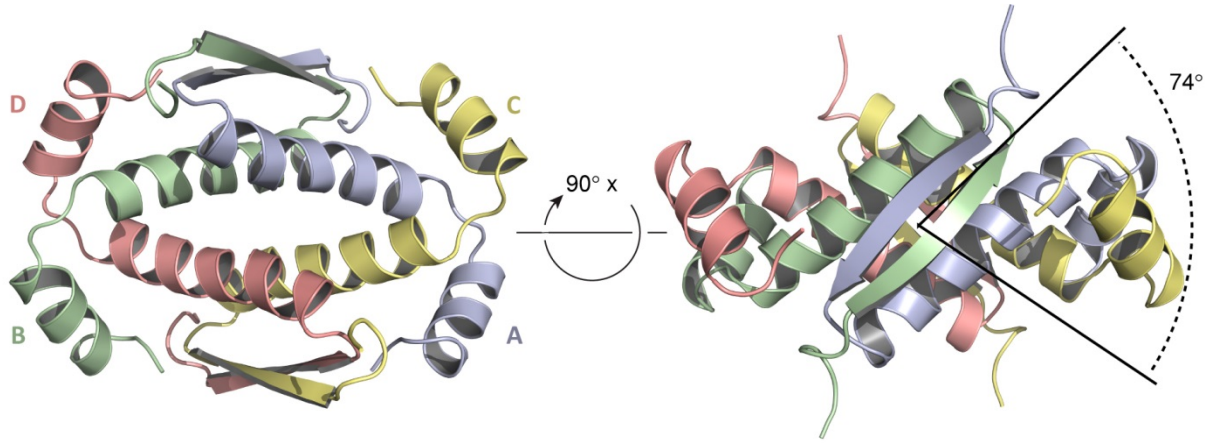


Figure 4-47 NMR solution structure of *Ciona intestinalis* p53/p73-b TD. The tetramerization domain (TD) of *C.int.* p53/p73-b was solved as an NMR solution structure from a construct starting with aa 374 and ending with the very C-terminus of the natural protein sequence. All spectra used for determination of the structure were recorded at 25°C.

The TD of human p53 is stabilized by a comparable cluster of salt bridges, which was shown to be essential for its high thermal stability.[49] In the crystal structure (pdb entry 1C26) only a single salt bridge between the guanidinium moiety of Arg 337 and the carboxylate group of Asp 352 from the adjacent monomer is resolved.[64] But molecular dynamics simulations revealed that this single salt bridge is joined by Arg 333 and Glu 349 resulting in the formation of a fluid salt-bridging cluster.[65] Arg 337 (in human p53) is not conserved in vertebrate p63 or p73 proteins, but p53/p73-a as well as p53/p73-b from *C.int.* and *C.sav.* all possess a Lys residue at the corresponding position; Figure 4-2.

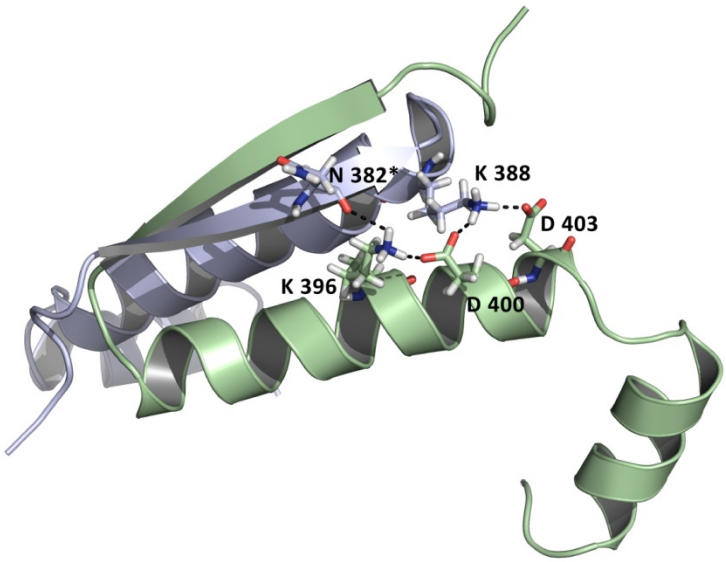


Figure 4-48 Hydrogen bonds and salt bridges spanning across the dimer interface within the TD of *C.int.* p53/p73-b. The involved residues are depicted as sticks, and if only the peptide backbone is engaged this is indicated by a star after the residue number. A hydrogen bond and a network of salt bridges are stabilizing the primary dimers. The side chain of Lys 396 is donating a hydrogen bond to the main chain carbonyl of Asn 382 located in the β-strand of the adjacent monomer. Asp 400 and Asp 403 form salt bridges to Lys 388. These interactions each occur four times in the tetramer. For clarity the second dimer is not shown and the polar interactions are only shown on one side of the primary dimer.

Table 4-5 Statistics derived from structure calculation with CYANA and ARIA/CNS for the *C.int.* p53/p73-b TD structure

NOE assignment		
Total number of peaks	2778	
Assigned peaks (total number of peaks)	2487 (2778)	90 %
3D ^{15}N-resolved NOESY-HSQC	440 (511)	86 %
3D ^{13}C-resolved NOESY-HSQC (aliphatic)	1640 (1768)	93 %
3D ^{13}C-resolved NOESY-HSQC (aromatic)	109 (112)	97 %
3D ^{15}N/^{13}C-separated NOESY	48 (87)	55 %
3D ^{15}N-edited/^{13}C-separated NOESY	67 (101)	66 %
4D-CT-J-Resolved ^{13}C-separated NOESY	183 (199)	92 %
Unassigned peaks	291	10 %
NOE derived distance constraints		
Total number of NOE derived upper limit distance restraints	5669	
Ambiguous NOEs	0	
Unambiguous NOEs	5669	
Inter-subunit	1785	
A-B	296	
A-C	108	
A-D	39	
Intra-subunit	3884	
Intra-residue	1235	
Inter-residue	2649	
Short range ($\|I - j\| \leq 1$)	2548	
Medium range ($1<\|I - j\| <5$)	1184	
Long range ($\|I - j\| \geq 5$)	1937	
Hydrogen bonds		
Total number of hydrogen bonds	80	
Inter-subunit hydrogen bonds per dimer	6	
Intra-subunit hydrogen bonds per monomer	17	
Dihedral restraints derived by Talos+		
Dihedral angle restraints (ϕ/ψ) per monomer	30 / 30	
Restraint violations		
Max. distance restraint violation (Å)		0.361
Number of violated distance restraints > 0.3 Å per monomer		2
Number of violated dihedral angle constraints		0
Ramachandran plot		
Residues in most favored regions		81.5 %
Residues in additionally allowed regions		18.5 %
Residues in generously allowed regions		0.0 %
Residues in disallowed regions		0.0 %
RMSD		
Backbone average RMSD to mean (Å)		0.33
Heavy atom average RMSD to mean (Å)		0.53
All atom average RMSD to mean (Å)		0.68

The tetramer is composed of two primary dimers which are related by 2-fold symmetry and pack at an angle of 74° as depicted in Figure 4-47. This causes the respective first helices to orient in such a way that they cross each other whereby they form the (main) tetramerization interface.

The polar interactions across this core tetramerization interface are shown in Figure 4-49. The residue pairs of Lys 391 and Glu 394 from two adjacent monomers, respectively, form in each case in both directions a total of four salt bridges. As shown in Figure 4-49 (B) the aromatic side chains from the Tyr residues in position 395 from all four monomers cluster together in the very center of the TD and thereby arrange in such a way that a remarkably big cavity is formed.

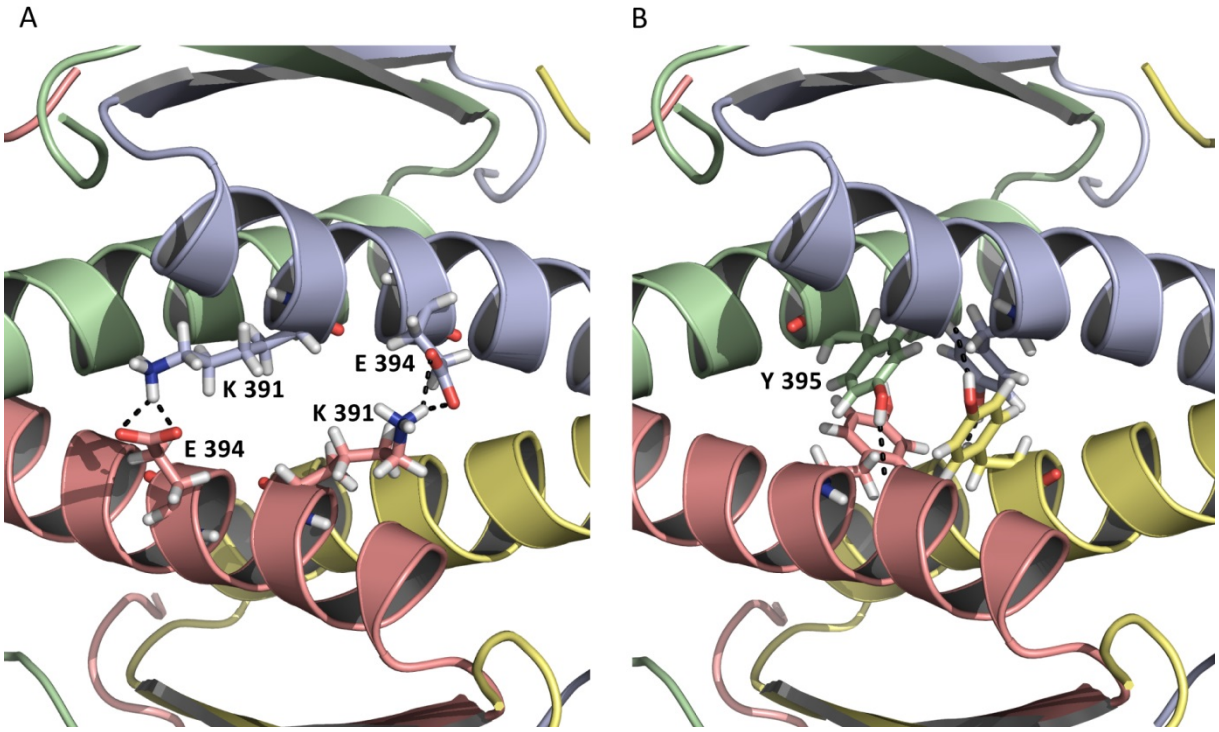


Figure 4-49 Hydrogen bonds and salt bridges spanning across the tetramerization interface formed by the respective first helices within the core TD of *C.int.* p53/p73-b. (A) In reverse in each case the residues Lys 391 and Glu 394 from two adjacent monomers form salt bridges across the tetramerization interface. For clarity these interactions are only shown on one pair of monomers. (B) The residues Tyr 395 from all four chains reside in the very middle of the TD structure and form the center of a large hydrophobic core. In addition the hydroxyl groups of the Tyr side chains presumably donate hydrogen bonds to the main chain carbonyls of Lys 391, respectively.

Each Tyr side chain is engaged in the network of hydrogen bonds and polar interactions which involve the main chain carbonyl of Lys 391 from an opposing monomer. Since the carbonyl should already be saturated due to the regular hydrogen bonding network within the helix, it is questionable if the hydroxyl group of the Tyr side chain indeed donates a hydrogen bond. But on the other hand the first helix was fond to be somewhat kinked at the

position of Lys 391 which reduces the distance between the Tyr hydroxyl group and the carbonyl group of Lys 391. And this in turn would suggest that a rather strong interaction is formed. With respect to the salt bridges these presumed hydrogen bonds are respectively formed back and forth between the other possible inter-dimer monomer pairs.

Vertebrate p53, p63 and p73 do not have a Tyr residue in the respective position but instead Val, Ile or Leu, and hence the presumed stabilizing inter-dimer hydrogen bonds are missing. The position of the salt bridge between Lys 391 and Glu 394 is also unique among deuterostome p53-like proteins. In all other sequences the position corresponding to Lys 391 is occupied with a hydrophobic residue, most often Met or Ile.

Only human and mouse p53 possess a Glu residue (Glu 343, human p53) at the position corresponding to Glu 394. In *Ciona* p53/p73-a as well as vertebrate p63 and p73 instead Lys is highly conserved at the respective position. In p53 the Glu residue forms a salt bridge across the tetramerization interface with a different Lys residue (Lys 351, human p53), which belongs to the other monomer from the opposing primary dimer. This salt bridge is the only inter-dimer one in the TD of human p53 and is important for its thermal stability and the oligomeric state.[68] It is assumed that in the TD of *C.int.* p53/p73-b the salt bridge between Lys 391 and Glu 394 is of similar importance.

The interface between the first helices is further stabilized by extensive hydrophobic interactions. As shown in Figure 4-50 the (respective part of the) hydrophobic core is formed by the side chains of the already mentioned Tyr 395 and is further extended by Ala 398 and Val 405 as well as the three Leu residues in position 399, 401 and 402. The side chains of Lys 391 and Glu 394 cover one side of the hydrophobic core and their side chain methylene groups are as well involved in hydrophobic interactions.

The tetramer is further stabilized by the C-terminal second helices, which clasp the opposite primary dimer and pack against the respectively most adjacent first helix. As shown in Figure 4-51 the resulting interface is characterized by various polar as wells as hydrophobic interactions.

In the TD of *C.int* p53/p73-b the primary dimers pack against each other at an angle of 74°; Figure 4-47. For the crystal structure of the TD of human p73 the respective angle has been reported to be about 65°,[106] and in the NMR solution structure the same angle measures 69°. Hence, with respect to the situation in human p73 the primary dimers in the TD of *C.int.* p53/p73-b are rotated against each other by additional 5-9° and orient closer to

orthogonal. For many residues in the core tetramerization interface, which is formed between the first helices, this twist does not have a greater impact and consequently it is tolerated that many residues are conserved. But the C-terminal ends of the first helices and also the immediately following sections get much more separated. As a consequence the hydrophobic interface, which in human p73 is formed by the hinge regions between the respective first and second helices from two adjacent monomers, is missing in *C.int.* p53/p73-b. For *C.int.* p53/p73-a exactly this interface had been identified to be decisive for the formation of the second helix (chapter 4.2) and the most important residues are highly conserved in vertebrate p73. Thus, in *C.int.* p53/p73-b the second helix instead needs to be stabilized internally and by direct interactions with the adjacent primary dimer.

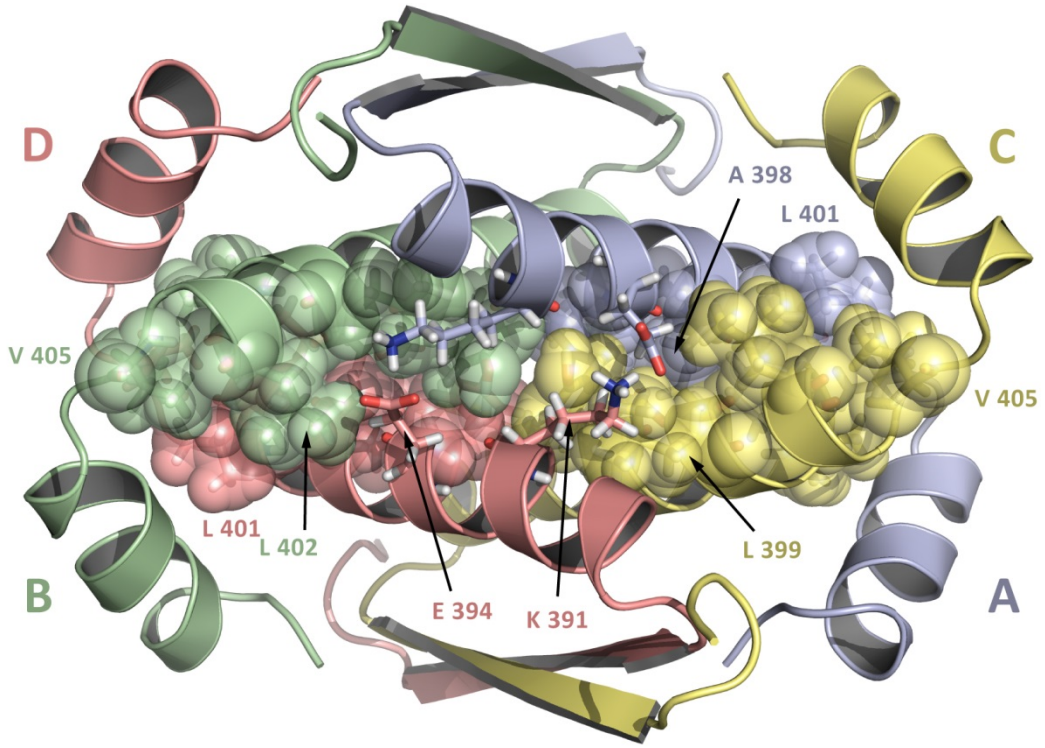


Figure 4-50 Hydrophobic core within the tetramerization interface of the core TD of *C.int.* p53/p73-b. The side chains of Tyr 395, Ala 398 and Val 405 as well as the three Leu residues in position 399, 401 and 402 form numerous hydrophobic interactions. Their side chains are depicted as spheres. The methylene groups from the side chains of Lys 391 and Glu 394 are as well involved in hydrophobic interactions but this side chains are depicted as a stick model to allow a better view into the core.

The first turn of the second helix is locked by an internal salt bridge between Asp 408 and Lys 411, and this should at least partially compensate for the lack of an N-capping motif. Because the Pro residue is not in the N-cap position but the first residue with α-helical backbone conformation and its carbonyl is engaged in respective hydrogen bonding. Hence,

the beginning of the second helix is not shifted with respect to human p73. And the Pro side chain is not engaged in hydrophobic interactions but oriented towards the surface of the protein like the Gln residue in the corresponding position in human p73.

The last three C-terminal residues of *C.int.* p53/p73-b do not have α-helical backbone conformation but contribute to various stabilizing interactions with residues from the other primary dimer. Therefore in the following they were assigned as part of the interface residues of the second helix.

As shown in Figure 4-51 (A) the side chains of Lys 410 and Asn 417 as well as the C-terminal carboxyl group form salt bridges to three residues (Glu 397, Asp 403 and Lys 396, respectively), which reside within the first helix of the most adjacent monomer from the other primary dimer. In addition the main chain carbonyl groups of Arg 416 and Asn 417 accept hydrogen bonds donated by the guanidinium moiety of Arg 384 of the other monomer from the opposing primary dimer.

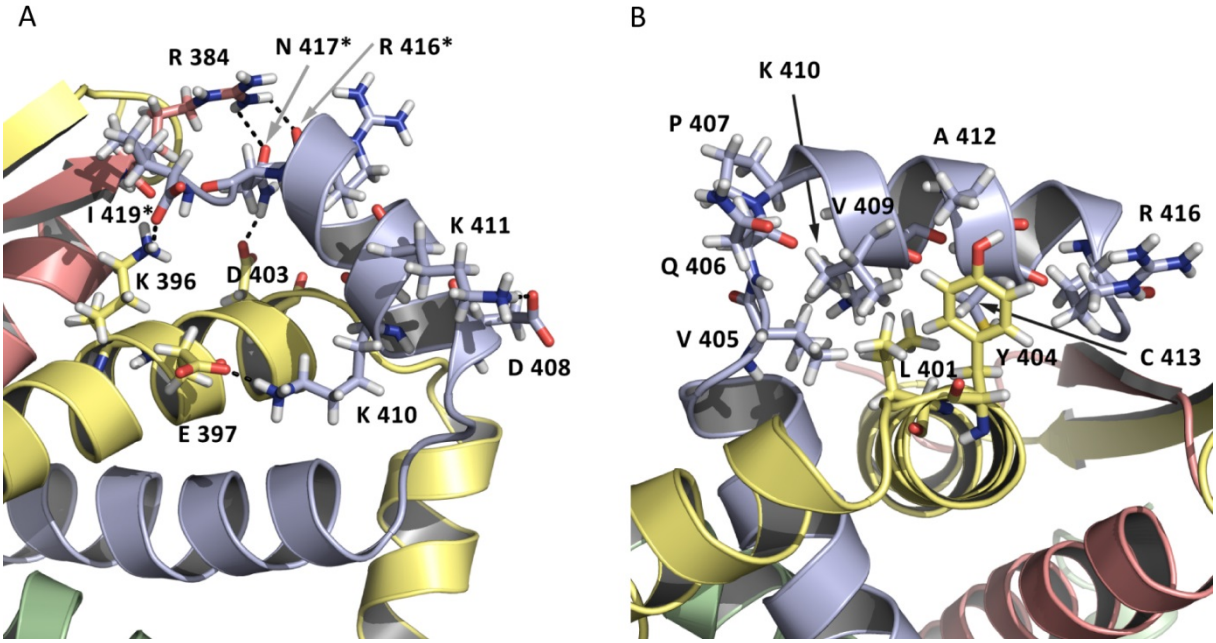


Figure 4-51 Polar and hydrophobic interactions across the interface between the core TD (β-sheet and first helix) and the second helix as well as a stabilizing intra-chain salt bridge within the second helix of the TD of *Ciona intestinalis* p53/p73-b. The second helices from one primary dimer are wrapping around the adjacent primary dimer in a clamp like fashion. Thereby the second helix is stabilized by various polar as well as hydrophobic interactions. The involved residues are depicted as sticks. (A) Polar interactions stabilizing the second helix. Asp 408 and Lys 411 form an intra-chain salt bridge and thereby lock the first turn of the second helix. Lys 410 is forming a salt bridge with Glu 397, which is located in the first helix of the most adjacent monomer from the other primary dimer. Lys 396 and Asp 403 reside in the same helix and form salt bridges to the C-terminal carboxyl and the side chain of Asn 417, respectively. The main chain carbonyls of Arg 416 and Asn 417 are accepting hydrogen bonds donated by the side chain of Arg 384 from a third monomer. If in at least one interaction only the peptide backbone is engaged this is indicated by a star after the residue number. (B) The second helix is stabilized by a hydrophobic cluster involving Leu 401 and Tyr 404, residing at the C-terminal end of the first helix of the most adjacent monomer from the other primary dimer.

As shown in Figure 4-51 (B) the second helix is furthermore stabilized by a hydrophobic cluster build up by various hydrophobic side chains which all point towards Leu 401 and Tyr 404 from the first helix of the other monomer it is wrapping around. In comparison to human p73 the most obvious difference is that the aromatic ring (Tyr 404), which is making big contributions to this hydrophobic interface, does not belong to a residue from the second helix (Tyr 389 in human p73) but to the last residue of the first helix. Interestingly, some vertebrate p63 proteins also possess a Tyr (e.g. human p63) or a Phe residue (e.g. zebrafish p63) in the respective position at the end of the first helix.

An overview on which residues are making up the different interfaces is provided in Figure 4-52. In comparison to human p63 or p73 in total many more residues are involved in the interfaces connecting the second helix onto the core TD.

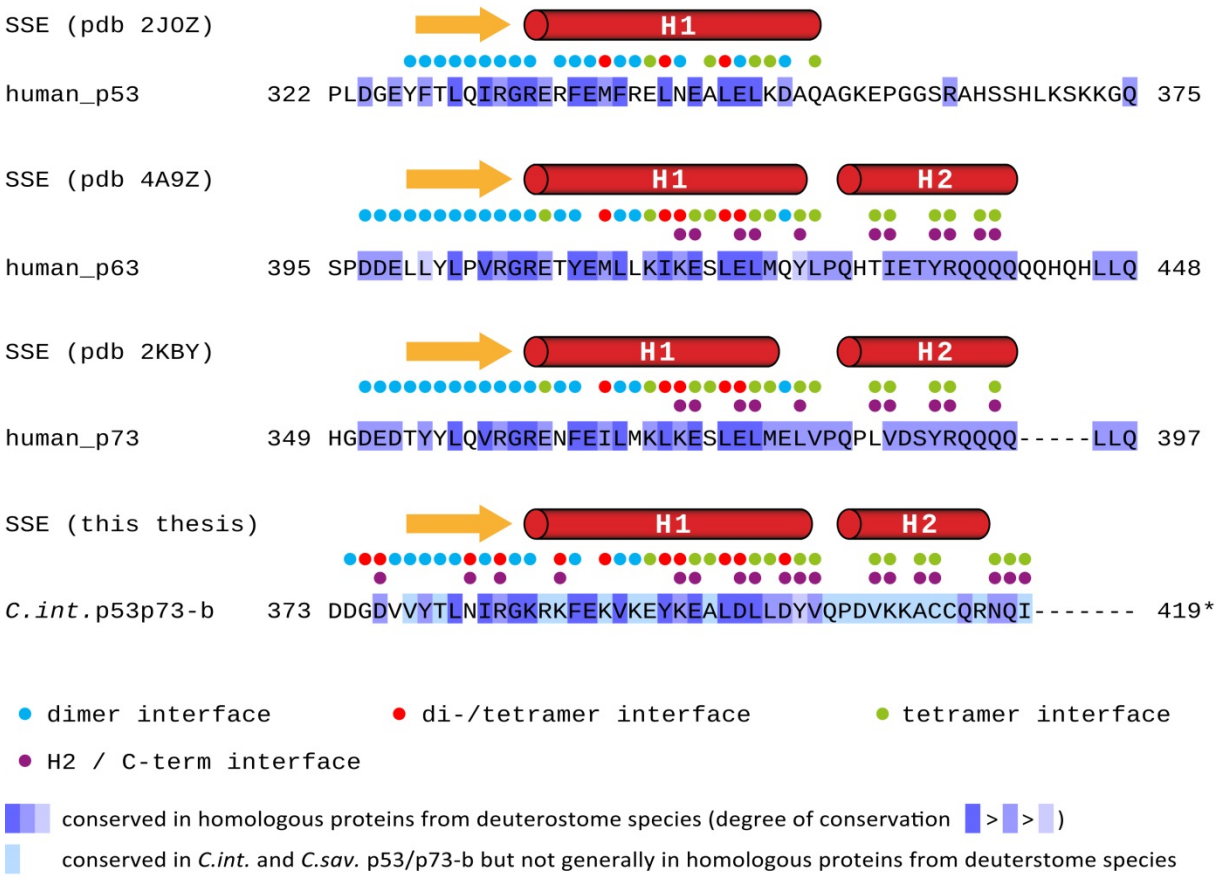


Figure 4-52 Alignment and functional annotation of the tetramerization domain of *Ciona intestinalis* p53/p73-b in comparison to human p53, p63 and p73. The blue coloration of the alignment is according to percent of similar residues in homologous proteins from deuterostome species, and cyan coloration indicates residues which are only conserved in the protein sequences of p53/p73-b from *Ciona intestinalis* (*C.int.*) and *Ciona savignyi*. An asterisk indicates the C-terminal end of the natural protein sequence. As a prefix to the sequences the secondary structure elements (SSEs) are depicted. The PDBePISA server was used to determine which residues make up the different interfaces.

4.4. Structural comparison of the TD of *C.int.* p53/p73-b and TDs of p53 and p73-like proteins from vertebrate species

The TD of human p53 is generally seen as the canonical motif because its monomers do not possess a second helix. And accordingly truncated constructs from other members of the protein family are frequently referred to as the minimal motif or the core TD. In the TDs of human p63 and p73 this canonical motif is extended by a second helix which has been proven to be essential for stability as well as the overall architecture of the domain.[88,89,106] For human p63 it was shown that the truncation of the second helix drastically affects the angle at which the primary dimers pack against each other, while the structure of the primary dimers is largely maintained.[89] The change in the packing angle in turn has a great influence on the main tetramerization interface, which forms between the first helices. And this is the reason why the respective angle is often referred to when the TDs of different proteins are compared, especially when the focus lies on the effect, which a potential second helix has on the overall structure. However, the published values are often only roughly estimated or if actually measured they cannot be reproduced, partly because of missing method description.

Hence, to facilitate a reasonable comparison of the TD of *C.int.* p53/p73-b towards already published TDs of vertebrate p53 as well as p63 or p73 proteins the most characteristic angles, at which the different SSEs orient with respect to each other, had to be determined. As a set of reference structures the TDs of human p53 and p73 were selected since for both proteins, respectively, a crystal structure as well as a NMR solution structure is available. The just recently published crystal structure of the TD of zebrafish p53, which interestingly possesses a second helix, was as well included for comparison. The structures were analyzed using PyMOL with the python script “angle_between_helices” and the “loop_orientation” method. The selection of the residues to be assigned to individual SSEs was done with care to ensure reasonable orientation vectors, and therefore in most cases the residues at the very ends of the SSEs were not included. The packing and orientation angles are reported in Table 4-6.

The angle at which the two primary dimers pack against each other has to be indirectly calculated. In a first step the orientation vectors along the respectively first helices in the monomers A and D are determined. Thereby the numbering of the monomers has to be

according to the scheme shown in Figure 4-53. The first helix is the longest SSE, and hence is best suitable. In a second step the angle between the two helices is determined and its value subtracted from 180 degree. For the TD of *C.int.* p53/p73-b the dimer-dimer packing angle was calculated to measure 74.1°. In human p53 the angle differs drastically with respect on the method used for structure determination and was calculated to be 72.4° in the solution and 82° in crystal structure. For the TD of human p73 the angle is also different in the two structures and measures 69° in the solution and 63.9° in the crystal structure. In zebrafish p53 the determination is complicated by the fact that the symmetry within the primary dimers is imperfect. To account for this the angles between the respective first helices of the monomer pairs A and D as well as B and C were calculated. The average was then subtracted from 180° and the packing angle calculated to measure 74°. The comparison revealed that the angle at which the primary dimers pack against each other is the same in the TDs of *C.int.* p53/p73-b and zebrafish p53. In both proteins the dimers are more orthogonal orientated with respect to their relative positions in human p73. This can as well the visually estimated in the top view perspectives presented in the middle row of Figure 4-53.

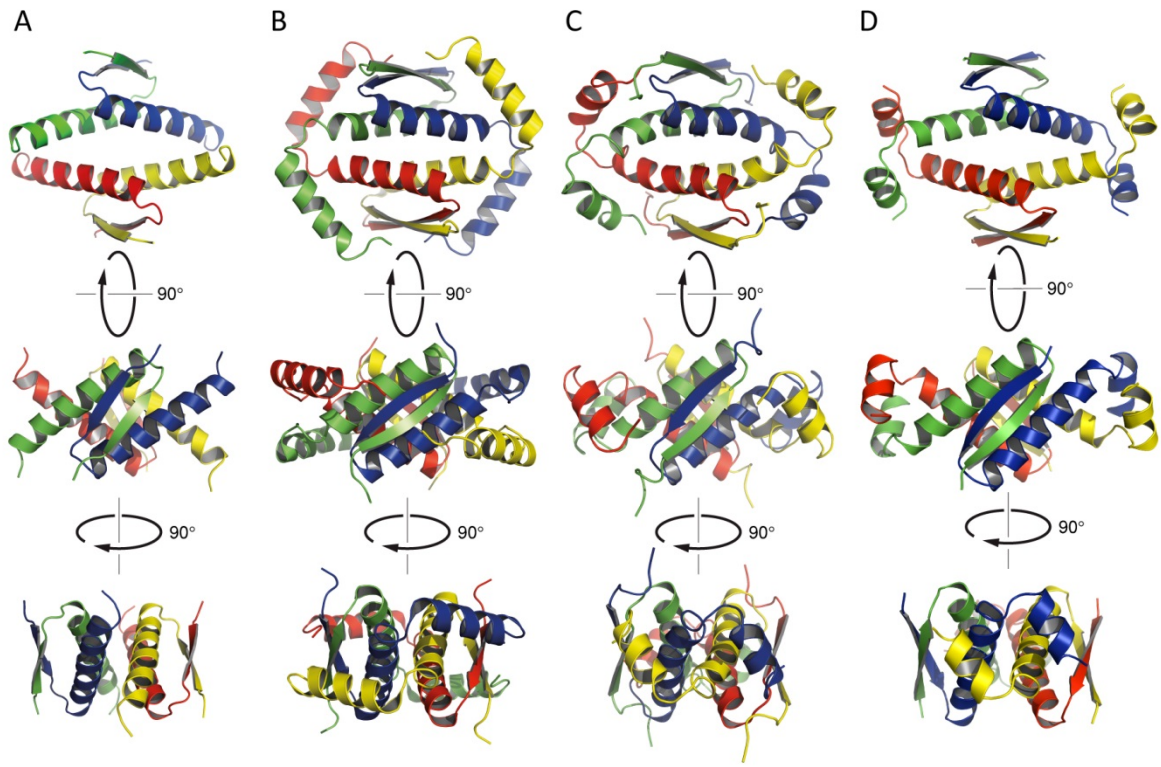


Figure 4-53 Comparison of tetramerization domain architectures. The coloration of the monomers follows the scheme: A = blue; B = green; C = yellow; D = red. (A) Solution structure of human p53 tetramerization domain (TD) (PDB entry 2J0Z). (B) Solution structure of human p73 TD (PDB entry 2KBY). (C) Solution structure of *Ciona intestinalis* p53/p73-b TD (PDB entry 2MW4). (D) Crystal structure of zebrafish TD (PDB entry 4D1M, chains IJKL).

Table 4-6 Most characteristic packing and orientation angles in the TDs of *C.int.* p53/p73-b, human p53, human p73 and zebrafish p53. The respective monomers were named according to Figure 4-53.

***C.int.* p53/p73-b TD**, solution structure, pdb entry 2MW4 monomer A = chain A; monomer B = chain B; monomer C = chain C, monomer D = chain D selected as SSE for calculation of orientation vectors: β-strand = V378-I383; helix 1 = K388-D403; helix 2 = D408-Q415								
	H1 (A)	H1(B)	H1 (C)	H1 (D)	strand (B)	strand (C)	strand (D)	H2 (C)
strand (A)	155.8				155.0	95.4	89.87	
H1 (A)		160.7	78.5	105.9				
H2 (A)	126.9		85.7				81.3	138.2
***Human* p73 TD**, solution structure, pdb entry 2KBY monomer A = chain A; monomer B = chain B; monomer C = chain C, monomer D = chain D selected as SSE for calculation of orientation vectors: β-strand = T354-V359; helix 1 = E363-E376; helix 2 = Q383-Q393								
	H1 (A)	H1(B)	H1 (C)	H1 (D)	strand (B)	strand (C)	strand (D)	H2 (C)
strand (A)	163.2				159.6	78.7	104.9	
H1 (A)		167.6	70.5	111.0				
H2 (A)	115.7		121.4				123.2	108.5
***Human* p73 TD**, crystal structure, pdb entry 2WQI monomer A = chain A; monomer B = chain B; monomer C = chain C, monomer D = chain D selected as SSE for calculation of orientation vectors: β-strand = T354-V359; helix 1 = E363-E376; helix 2 = Q383-Q393 *instead the angle between H2(D) and B1(A) was determined								
	H1 (A)	H1(B)	H1 (C)	H1 (D)	strand (B)	strand (C)	strand (D)	H2 (C)
strand (A)	160.8				162.1	80.4	100.8	
H1 (A)		163.1	67.5	116.1				
H2 (A)	111.8		130.6				128.2*	105.6
***Human* p53 TD**, solution structure, pdb entry 2J0Z monomer A = chain A; monomer B = chain B; monomer C = chain D, monomer D = chain C selected as SSE for calculation of orientation vectors: β-strand = F328-I332; helix 1 = E336-D352								
	H1 (A)	H1(B)	H1 (C)	H1 (D)	strand (B)	strand (C)	strand (D)	
strand (A)	151.0				152.5	108.5	77.1	
H1 (A)		157.1	77.0	107.6				
***Human* p53 TD**, crystal structure, pdb entry 1C26 Crystal structure contains single chain, monomers were generated as symmetry mates and named respectively. selected as SSE for calculation of orientation vectors: β-strand = F328-I332; helix 1 = E336-D352								
	H1 (A)	H1(B)	H1 (C)	H1 (D)	strand (B)	strand (C)	strand (D)	
strand (A)	152.3				149.8	110.4	77.7	
H1 (A)		160.1	85.4	98.0				
***zebrafish* p53 TD**, crystal structure, pdb entry 4D1M monomer A = chain I; monomer B = chain J; monomer C = chain L, monomer D = chain K selected as SSE for calculation of orientation vectors: β-strand = I304-V309; helix 1 = E313-L327; helix 2 = A333-Y339 due to an imperfect symmetry the angles deviate when being determined for a different set of monomers (Δ ≤ 7 degree are not reported separately) *the respective angle between H1(B) and H1(C) is 113.4 degree								
	H1 (A)	H1(B)	H1 (C)	H1 (D)	strand (B)	strand (C)	strand (D)	H2 (C)
strand (A)	149.9				156.9	98.1	88.01	
H1 (A)		152.3	81.0	98.6*				
H2 (A)	115.0		73.0				63.3	177.9

The most striking difference in the overall structural architectures of the TDs of human p73, zebrafish p53 and *C.int.* p53/p73-b is the difference in the orientation of the respective second helices with respect to the core. To quantify this the three structures were aligned and the angle between the second helices were determined as described in the subtext of Figure 4-54. Interestingly the second helix of *C.int.* p53/p73-b orients towards the aligned core TDs in a position right in the middle of the second helices from human p73 and zebrafish p53. With respect to the TD of human p73 the second helix in *C.int.* p53/p73-b is skewed by 41°. This in turn results in a more orthogonal orientation towards the first helix it is wrapping around (85.7° versus 121.4° in solution structure of human p73).

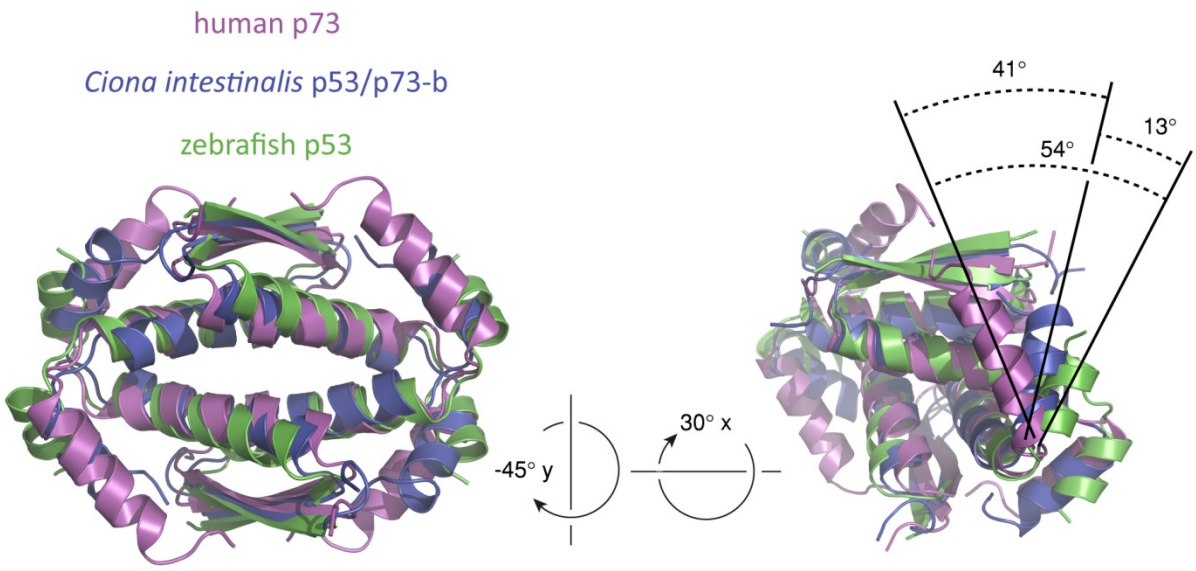


Figure 4-54 Overlay of the tetramerization domains of human 73, zebrafish 53 and *Ciona intestinalis* p53/p73-b. Using pymol the structures were aligned accordingly to give the best fit with respect to the β-sheets and the first helices. Within the overlay the angles between the respective second helices were determined using the python script "angle_between_helices" with the "loop_orientation" method. The residues, which were included in the calculation as part of the second helices, are listet in Table 4-6. Structures used for the overlay are the NMR solution structures of human p73 (pdb entry 2KBY) and *Ciona intestinalis* (*C.int.*) p53/p73-b (pdb entry 2MW4) as well as the crystal structure of zebrafish p53 (pdb entry 4D1M with chains I, J, K and L). The primary dimer of zebrafish p53 is not perfectly symmetrical, and this causes the second helices to adopt slightly different angles with respect to the other dimer.[195] Consequently the angles, which are shown in the figure only apply on one side of the primary dimers. On the other monomer the second helix of zebrafish p53 is skewed with respect to human p73 and *C.int.* p53/p73-b by 59° and 20°, respectively.

4.5. The TDs of *Ciona intestinalis* p53/p73-a and p53/p73-b do not form hetero oligomers

For human p63 and p73 it was shown that the TDs of the two proteins *in vitro* preferentially form hetero-tetramers from homo-dimers,[88] and thereby the dissociation of the kinetically more stable p73 homo-tetramer was identified as the rate-limiting step.[89] The formation of such mixed tetramers suggests functional cross talk, and therefore it was an important finding that human p53 does not form hetero-tetramers with its family members.[88]

C.int. p53/p73-b is the result of a relatively recent gene duplication event caused by retro-transposition and most likely the ancient genomic copy of p53/p73-a served as the template. Therefore it was a pressing question whether the two paralogous p53/p73 proteins from *C.int.* are (still) capable to form hetero-tetramers or if such TD driven interactions are inhibited by accumulated aa substitutions on both proteins.

A ^{15}N labeled sample of the TD of *C.int.* p53/p73-b (aa 374-419, construct the NMR solution structure has been solved of) was mixed with a 10-fold molar excess of unlabeled *C.int.* p53/p73-a aa 415-464. The latter construct had been shown to dominantly exist in an overall structure lacking a second helix (section 4.2.5 and 4.2.9) and to only weakly fold a second helix even at lower temperatures (section 4.2.13). After incubation at RT for 2 days a [^{15}N, ^{1}H]-TROSY experiment was recorded but the spectrum showed no shifts with respect to pure samples of *C.int.* p53/p73-b TD (data not shown). Therefore the proteins were denatured with 6 M guanidium hydrochloride and subsequently refolded by dilution with a 50-fold volume of reducing RE buffer. The protein solution was reconcentrated in an amicon sterring cell under constant pressure from nitrogen gas and the guanidium hydrochloride was quantitatively removed by several times filling up with reducing RE buffer (effective dilution effect ≥ 10000). To ensure that the final molar ratio after refolding would still be the same it was separately assayed for both constructs that no loss of protein occurs during the described denaturation and refolding process, and it was in addition shown that the [^{15}N, ^{1}H]-TROSY spectrum of a pure sample of *C.int.* p53/p73-a aa 415-464 was as well as unchanged after the process.

Even after denaturation and refolding of the 1 to 10 mixed sample the [^{15}N, ^{1}H]-TROSY spectrum showed no shifts with respect to the reference sample of pure *C.int.* p53/p73-b TD

(data not shown). This revealed that the TDs of *C.int.* p53/p73-a and –b do not form hetero-tetramers.

As shown in Figure 4-55 many residues within the sections, corresponding to the β-strand and the first helix, are identical or at least functionally conserved between *C.int.* p53/p73-a and –b. In contrast to that no conserved residues are cotained within the section corresponding to the second helix in *C.int.* p53/p73-b. This argued for the possibility that this second helix might prevent the formation of hetero-tetramers and that the core TDs could be capable to interact with each other. In order to examine this possibility, the before described experiment was repeated with truncated constructs.

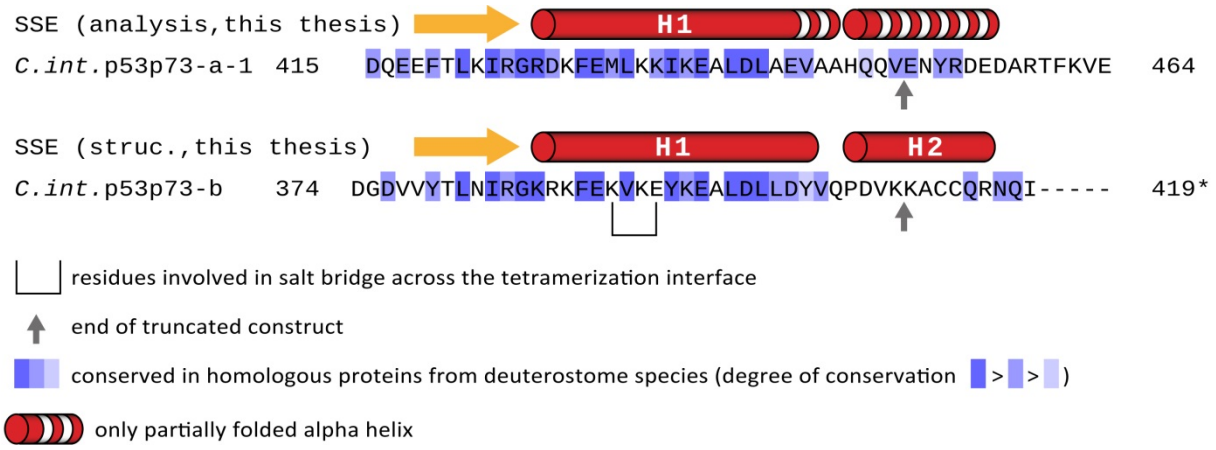


Figure 4-55 Alignment of the TDs of *C.int.* p53/p73-a and –b with annotation of the determined secondary structure elements (SSEs).

The construct of *C.int.* p53/p73-b aa 374-410 corresponds to the TD lacking the last nine residues, and thus possessing only the first four residues from the section which would otherwise form the second helix. Hence, it can be assumed that this construct does not fold a second helix. For *C.int.* p53/p73-a the construct encompassing aa 415-450 was selected because it is the shortest truncated construct which still entirely spans the sequence of the first helix.

A ^{15}N labeled sample of the truncated p53/p73-b protein was mixed with a 10-fold molar excess of unlabeled truncated p53/p73-a and the mixture was subjected to the same protocol of denaturation and refolding as described before. The [^{15}N, ^{1}H]-TROSY spectrum again exactly resembled that of the reference sample of pure *C.int.* p53/p73-b aa 374-410.

Hence, it can be stated that also the core TDs of *C.int.* p53/p73-a and –b do not form hetero-tetramers, and thus functional cross talk via the TDs can be excluded.

As shown in Figure 4-49 the side chains of Lys 391 and Glu 394 form a salt bridge across the tetramerization interface in the TD of *C.int.* p53/p73-b. In *C.int.* p53/p73-a the corresponding residues are instead Met and Lys. Therefore, if hetero-tetramers would consist of homo-dimers four such salt bridges would get lost and in addition the two directly opposing Lys side chains would result in charge repulsion. There is only one possible combination with two monomers of each type where at least two of the inter-dimer salt bridges could form. But in this tetramer structure the second helix of *C.int.* p53/p73-b would miss the Tyr residue on the other first helix it is wrapping around. As a result the hydrophobic cluster, which is shown in Figure 4-51 (B) would be disturbed and this in turn would presumably have a strong destabilizing effect. Beside other probably not well suitable aa substitutions within the two *C.int.* p53/p73 proteins surely these incompatibilities are among the most crucial ones accounting for the observed strictly separated oligomerization of the TD peptides.

4.6. Prediction of a single genomic copy of a p53/p73 like protein in the genome of *Botryllus schlosserie*

For the evaluation of the determined structural organization of the TDs of the *C.int.* p53/p73 proteins it would have been beneficial to compare the respective sequences to those of the corresponding proteins from a variety of other tunicate species, being less closely related than *C.sav.*. However, for nearly all other species from this subphylum sufficient genomic information is not available. The only exception is *Botryllus schlosserie* (*B.schl.*) whose genome has just recently been sequenced with 11-fold sequence coverage.[133] But until today the genome has only partially been annotated and gene models as well as ESTs are missing. As therefore protein sequences were not available, a blast search within the *B.schl.* genome was performed on the botryllus genome browser using blastn against the botznik-ctg database. The protein sequences of the DBD of *C.int.* and *C.sav.* p53/p73-a were used as queries, and a single genomic copy of a p53 family protein was detected (botctg012291). Blastn searches with the full length sequences or the sequences of the p53/p73-b proteins as queries resulted in the same single hit. For consistency the *B.schl.* p53 family protein was named *B.schl.* p53/p73. In a next step the web server Spaln2 was

used to predict splice sides and transcripts for the respective genomic section of *B.schl.* chromosome 5 by calculating spliced alignments towards the evaluated isoforms of *C.int.* and *C.sav.* p53/p73-a. *C.int.* and *C.sav.* p53/p73-b, which resulted from a recent lineage-specific duplication event, were not used as queries since only a single gene had been identified in the genome of *B.schl.*. The alignment of the protein sequences, which were predicted using the isoforms 1, 2 and 4 of *C.int.* as well as *C.sav.* p53/p73-a as references, revealed that the predictions were all consistent with respect to the last 306 amino acids. Hence, the sequence comprising the TD of *B.schl.* p53/p73 was always identical. The DBD was also well covered by all six predictions and DBD as well as TD were detected as such by the conserved domain blast search of NCBI. An isoform with an elongated C-terminus, as it would correspond to *C.int.* and *C.sav.* p53/p73-a isoform 2, was not predicted.

The protein sequence, which was forecasted by spliced alignment against *C.int.* p53/p73-a isoform 1, is listed in the appendix, 7.1.5. All residue numbers refer to this preliminary sequence. In Figure 4-56 the part of the predicted protein sequence, which corresponds to the TD, is presented as an alignment towards the TDs of human p63 and p73 as well as the TD of *C.int.* p53/p73-b.

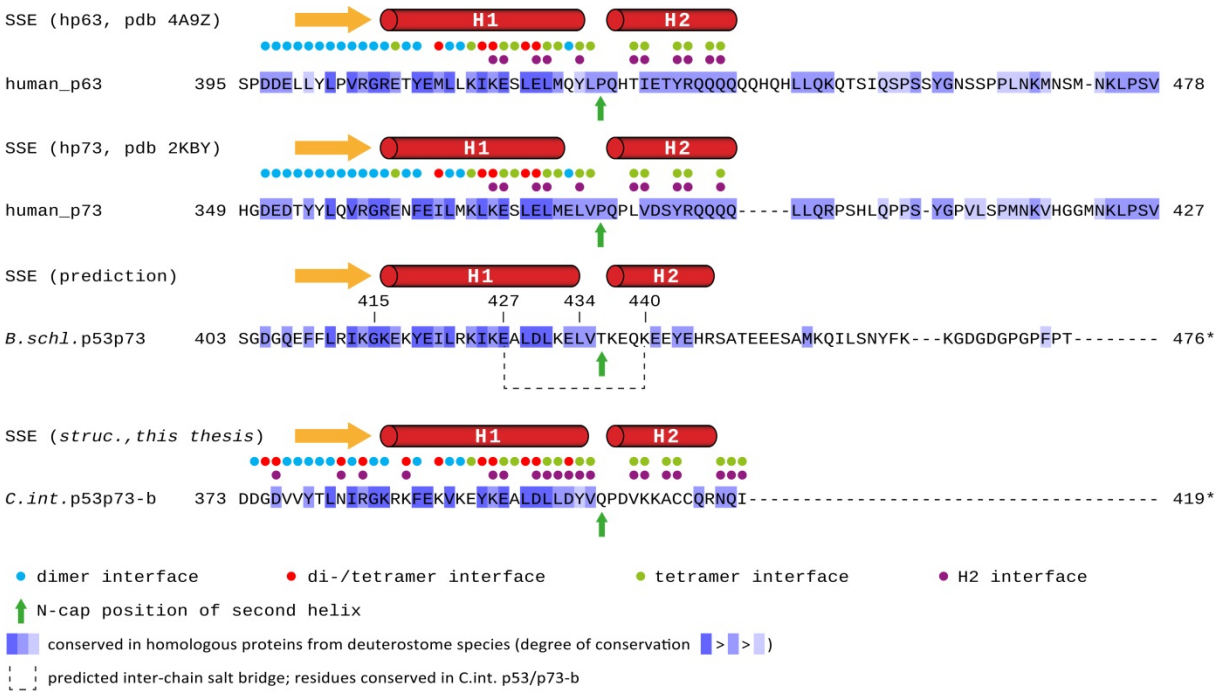


Figure 4-56 Predicted TD sequence of *Botryllus schlosserie* (*B.schl.*) p53/p73 aligned towards the TDs of human p63 and p73 as well as the TD of *Ciona intestinalis* (*C.int.*) p53/p73-b. The blue coloration of the alignment is according to percent of similar residues in homologous proteins from deuterostome species. An asterisk indicates the C-terminal end of the natural protein sequence. As a prefix to the sequences the secondary structure elements (SSEs) are depicted. The PDBePISA server was used to determine which residues make up the different interfaces.

As expected the sections, which correspond to the β-strand and the first helix, show a high content of conserved residues.

The degree of conservation is much lower in the section, which presumably forms the second helix. But several important motifs are conserved, and on basis of these it is to assume that the second helix is (rigidly) folded. First of all the characteristic Tyr residue (aa 443) is conserved, which has been shown to be essential for the second helix in human p63 and p73 to fold. In addition all residues within the first helix, which in human p63 or p73 directly contact the second helix, are functionally conserved. In the mutational analysis of *C.int.* p53/p73-a (results section 4.2.8) the first two residues within the hinge region between the first and the second helix had been identified as among the three most decisive positions. In *B.schl.* p53/p73 the residues Leu and Val (aa 434 and 435) are conserved with respect to human p73, and this makes it likely that also in the TD of *B.schl.* p53/p73 two oppositely positioned hinge regions will form a hydrophobic cluster which stabilizes the second helices, respectively. The Pro residue in the N-cap position of the second helix in human p63 and p73 is instead Thr, but Thr is as well capable to stabilize the N-terminus of an α-helix. In the so called capping box Thr in the N-cap position forms reciprocal hydrogen bonds with the residue in position N3 (third residue with α-helical backbone conformation). Thereby its side chain hydroxyl group accepts a hydrogen bond from the otherwise unsaturated main chain amine of residue N3 and in turn the main chain amine of Thr donates a hydrogen bond if the side chain of the residue in position N3 possesses a prober functional group. This makes Glu the most frequently observed residue in position N3, but Gln as in the sequence of *B.schl.* p53/p73 would as well be possible.

As shown in Figure 4-51 in the TD of *C.int.* p53/p73-b the second helix is beside others stabilized by an inter-chain salt bridge between a Lys residue in position N4 of the second helix and a highly conserved Glu residue in the first helix of the other monomer, which the second helix is wrapping around (Lys 410 and Glu 397). The corresponding residues are conserved in *B.schl.* p53/p73 (Lys 440 and Glu 427) making this stabilizing interaction possible, as well.

Taken together it appears likely that the TD of *B.schl.* p53/p73 possesses a stably folded second helix and an overall topology closely resembling the TDs of human p63 and p73.

5. Discussion

5.1. Phylogenetic relationship between p53-like transcription factors from vertebrate and tunicate species

The alpha isoforms of vertebrate p63 and p73 possess a highly conserved sterile alpha motif (SAM) domain, which is located C-terminal to the tetramerization domain (TD) and missing in p53. Therefore this feature is widely used in comparative phylogenetic studies to assign proteins, which belong to the family of p53-like transcription factors, as being of the p53 or the p63/p73 type. Until today no discrete function could be assigned to the SAM domain, but proteins of the p63/p73 type occur throughout the phylogenetic tree of this protein family.

The genomes of all vertebrate species encode one protein of the p53 and at least one but mostly two proteins of the p63/p73 type. Cephalochordata have been shown to be more distinct from vertebrates than tunicata, but the genome of the representative species *Branchiostoma floridae* (lancelet) codes as well for one protein of the p53 and one of the p63/p73 type; appendix 7.1.6.1. The purple sea urchin (*Strongylocentrotus purpuratus*) belongs to the phylum echinodermata, which is a sister phylum of the phylum chordata, and is hence even more distinct. Interestingly, its genome contains as well one gene for a p53 type (UniprotKB/Swiss-Prot: H9N2D2) and one for a p63/p73 type protein (UniprotKB/Swiss-Prot: H9N2D3).

From a functional perspective it therefore suggested itself that tunicate species should most likely as well possess at least one copy of both the p53 and the p63/p73 type, respectively. The genome of the tunicate species *Ciona intestinalis* (*C.int.*) is completely sequenced, well annotated and since *C.int.* is a popular model organism, comprehensive databases of ESTs are available as well. Hence, it can be assumed that all genes, which are encoded by the genome of this species, are known. And indeed the genome of *C.int.* as well as the genome of the very closely related species *Ciona savignyi* (*C.sav.*) contain two genomic copies for p53-like transcription factors. But the protein named p53/p73-b is the product of a recent lineage-specific gene duplication,[139] and therefore it can be excluded that a p53 type and a p63/p73 type protein were maintained continuously throughout

evolution. Just recently the genome of the colonial tunicate *Botryllus schlosserie* (*B.schl.*), which belongs to the same class of tunicates, was also sequenced, although not yet completely annotated.[133] A blast search against the protein sequence of the highly conserved DNA binding domain (DBD) identified only a single genomic copy for a p53 family protein, which for consistency was named *B.schl.* p53/p73. And by spliced alignment using the *Ciona* p53/p73-a proteins as references it was possible to forecast a preliminary protein sequence; results section 4.6.

The observation that the genome of *B.schl.* encodes only a single p53 family protein puts the conclusion close that already the last common ancestor of *B.schl.* and *C.int.* did possess only a single p53 family protein. Hence the lineage-specific duplication, which resulted in *C.int.* p53/p73-b, was not a necessary precondition for the loss of another before existing gene. As the p53/p73-b gene is completely free of introns retro-transposition has already been proposed as the most likely mechanism underlying this gene duplication event.[139] The obvious assumption that predating the duplication only a single genomic copy had existed now strongly suggests that the gene for *Ciona* p53/p73-a must have been the template, which establishes a paralogous relationship between the two *Ciona* p53/p73 proteins. Generally, the mechanisms underlying the retention of duplicated gene copies are believed to be neofunctionalization and subfunctionalization,[196] the latter of which especially applying when only single genes are duplicated, as in the case of *Ciona* p53/p73-b.[197,198] This suggests that the functions, which are fulfilled by *B.schl.* p53/p73, might at least partially be split between the *Ciona* proteins p53/p73-a and p53/p73-b.

The fact that the cephalochordate model organism possesses a p53 type and a p63/p73 type protein indicates that the last common ancestor of vertebrates and tunicates already possessed a diversified p53 family. But phylogenetic studies have revealed that all vertebrate p53 family proteins form a monophyletic group, which evolved from a single cephalochordate-like p63/p73 type protein after the divergence of cephalochordates and chordates.[139] Hence, during vertebrate evolution the cephalochordate-like p53 type protein was deleted. Within the DBD the *Ciona* proteins also show a higher degree of sequence identity to the cephalochordate p63/p73 than to its p53 type protein; Table 4-1. This indicates that *C.int.* p53/p73-a as well evolved from the same cephalochordate-like p63/p73 type protein as the vertebrate p53 family. But in this work it was shown that neither of the expressed *C.int.* p53/p73-a isoforms nore *C.int.* p53/p73-b possesses a sterile alpha motif

(SAM) domain. This permits two possible phylogenetic relations of the vertebrate and tunicate p53 family proteins. Either the gene duplication, which gave rise to vertebrate p53 and p63/p73, happened before the divergence of vertebrates and tunicates followed by silencing of the p63/p73 gene in the tunicate lineage or the SAM domain was deleted independently in vertebrate p53 and in *Ciona* p53/p73-a.

5.2. Tunicate p53/p73 proteins do not possess a SAM domain

Among all isoforms of the *Ciona* p53/p73 proteins only isoform 2 of p53/p73-a has a sequence C-terminal to the TD, which is long enough to comprise an additional domain. But the protein sequence of the isoform-specific extended C-terminus of *C.int.* p53/p73-a isoform 2 has no equivalent in the ncbi database besides the respective isoform of *C.sav.* p53/p73-a, and by the structural investigations presented in section 4.2.2 it was shown that the 203 aa long isoform-specific sequence does not fold any secondary structure elements (SSEs). A truncation at the isoform specific splice site directly behind the TD leaves the TD unaffected, which reveals that the C-terminal sections do have no direct influence on the oligomeric state as well. Taken together it therefore appears questionable if the C-terminal sequence of isoform 2 has any distinct function.

So far *C.int.* p53/p73-a and -b have been specific subjects in only one publication. Both proteins were reported to regulate the zygotic ZicL expression and the differentiation of the embryo notochord, thereby being expressed from maternal RNA.[140] In eggs the mRNA of *C.int.* p53/p73-a isoform 2 accounts for about 30% of the total *C.int.* p53/p73-a mRNA, and isoform 1 mRNA accounts for the remaining 70%. By the 64-cell stage the mRNA for isoform 1 is completely degraded but the mRNA for isoform 2 persists much longer with remaining amounts being detectable until the early tailbud stage. Therefore one could argue if the prolonged C-terminus might just have resulted from the need for an mRNA, which is protected against degradation by its prolonged 3’ end or potential complex RNA structures, and that the actual encoded protein sequence could be rather unimportant. An observation supporting this theory is that the sections from *C.int.* and *C.sav.* p53/p73-a isoform 2, which correspond to the isoform-specific C-terminus, do have a significantly lower degree of sequence identity compared to the rest to the protein; results section 4.2.1..

Tunicates have a simplified life style with respect to the last common ancestor of tunicates and vertebrates, and it is presumably owed to this circumstance that during the following separate evolution ancestral genes and even complete gene families were much more likely lost in *C.int.* than in vertebrates.[132] This might as well be the reason why in *C.int.* there had no longer been the need for a p53-like transcription factor with a SAM domain, i.e. one of the p63/p73 type.

Most frequently SAM domains function as protein-protein interaction (PPI) motifs, and if this should apply as well for the SAM domain of p63 and p73, it is likely that the respective interaction partner is conserved in cephalochordates and vertebrates but lost in *C.int.*. Hence, the ancestral genes, which were specifically lost in *C.int.*, are the pool in which to search to identify potential interaction partners, which bind human p63 or p73 specifically via their SAM domains.

Until today only one group of regulatory proteins has been ascribed to bind the SAM domains of p63 and p73, but it is questionable if the SAM domain indeed contributes to the interaction. The alpha isoforms of human p63 and p73 have been reported to interact via their SAM domains with E3 ubiquitin ligases such as Itch; an interaction, which results in their proteolytic degradation. Thereby the PPPY motif, which is strictly conserved in p63 and p73, is bound by the WW2 domains of Itch,[82,199] but the PPPY motif is located some residues upstream of the first SSE of the SAM domain (pdb entries: 1RG6 and 2Y9T), has not been shown to interact with residues from the SAM domain, and hence is not part of the domain. It has not yet been investigated whether the PPPY motif alone, in the absence of the SAM domain, would be sufficient for the interaction with Itch, but the observation that the mutation to PPPF can abolish Itch binding argues strongly for the PPPY motif being the only essential interaction side.[83] Only a single publication has reported that also residues from the SAM domain itself were part of the minimal interaction motif for Itch binding.[199] But in the respective binding studies the p63 SAM domain was only represented by 18mer peptides encompassing the PPPY motif, the short linker connecting it to the SAM domain and the first four residues from the SAM domain, which possess discrete secondary structure. Therefore it is impossible that these peptides resemble the native fold, which the respective sequence adopts within the protein. Furthermore, from the two residues, which were identified by this publication as additionally belonging to the interface for Itch binding, one is not surface exposed but completely buried. To enable binding of Itch to the proposed motif the SAM

domain would have to break up, and this seems very unlikely since the SAM domain forms a large hydrophobic core. In summary, it is highly questionable whether the SAM domains of p63 and p73 indeed participate in the interaction with Itch or other E3 ubiquitin ligases. Since however the SAM domain is highly conserved it is in turn very likely that at least one real interaction partner exists, and there is some probability that it is not encoded by the genome of *C.int.* or *B.schl.*.

5.3. TD of *C.int.* p53/p73-a does not possess a stably folded second helix

In the TDs of human p63 and p73 each monomer comprises an additional second helix at its C-terminus, which is not present in the TD of human p53. Several residues within the respective protein sequences from vertebrate p63 and p73 proteins are equally highly conserved as residues within the core TD (β-sheet and first α-helices), which represents the minimal TD as seen in human p53. Strictly conserved are the residues Tyr-Arg, which were found to make various contacts with the core TD, therefore being essential for the stability of the second helix. Exactly this motif is also conserved in *C.int.* and *C.sav.* p53/p73-a and therefore it had been expected that the TD of *C.int.* p53/p73-a would also possess a stably folded second helix, but as shown in section 4.2.5 this is not the case.

The backbone chemical shifts were determined from HNCACB spectra recorded at 25°C and were analyzed using the programs TALOS+ and δ2D; 4.2.5 and 4.2.10 (Figure 4-33; wild type). This revealed that with respect to human p73 the first helix of the TD is elongated at its C-terminal end by about 3-4 residues and that in ≤ 7% of all conformers the section corresponding to the predicted second helix adopts α-helical backbone conformation.

Those residues, which in the TD of human p53 form salt bridges spanning across the tetramerization interface, are not conserved in *C.int.* p53/p73-a, and the stability of the tetrameric state of the TDs of human p63 and p73, which as well lack such salt bridges within the core TD, highly depends on the second helix, respectively. Therefore, it suggested itself that the tetrameric state of the TD of *C.int.* p53/p73-a could be unstable, as well. But surprisingly, ultracentrifugation experiments with a construct representing the core TD of *C.int.* p53/p73-a revealed that the domain is primarily tetrameric at 150 µM.

Since many residues within the β-sheet and the first helices are conserved between *C.int.* p53/p73-a and vertebrate p63 and p73 it was tested if this core TD would be capable to

support a second helix. Utilizing chimeric constructs, in which starting at different positions the *Ciona* protein sequence was exchanged for the corresponding protein sequence from human p73, it was shown that the core TD is indeed able to support a stably folded second helix and furthermore that the hinge region between the first and the second helix is decisive for the formation of the second helix; results section 4.2.6. Based on the structure of the TD of human p73 those residues within the hinge region were identified, which presumably contribute to stability the most. Out of the identified residues those four, which are not conserved in both proteins, were introduced into *C.int.* p53/p73-a in different combinations. The motif Leu-Val-Pro was found to induce the fold of the second helix at least to such an extent that the overall thermal stability of the TD was increased; results section 4.2.11. However, the determination of heteronuclear nuclear Overhauser effects (hetNOEs) as well as the calculation of propensities for secondary structure revealed that, even when all four critical residues within the hinge region were substituted, the section corresponding to the second helix did still not gain full helicity; result sections 4.2.8 and 4.2.10. The portion of conformers, in which the residues from the section, corresponding to the second helix, possessed α-helical backbone conformation, was still significantly lower than expected for a stably folded helix. And in agreement with that hetNOEs from the respective section were indicating a higher degree of flexibility than expected for a rigidly folded SSE. For comparison the same experiments were also performed with the chimeric protein in which the section, encompassing the hinge region and the second helix, had been completely exchanged for the corresponding sequence from human p73. These experiments revealed that in contrast to the mutant proteins, with substitutions exclusively within the hinge region, this chimeric protein indeed folds a stable second helix. It had therefore been compellingly necessary to verify that the partially folded helix, induced by mutations within the hinge region, indeed packs onto the core TD in a similar way as in human p73. For this purpose the additional substitution of the Tyr-Arg motif with Ala-Ala was introduced into the construct with all four critical residues within the hinge region. The comparison of the respective HSQC spectra showed that with respect to the core TD this substitution almost completely restored the structure of the wild type; result section 4.2.9. And this in turn confirmed that, if folded, the second helix in any case packs onto the core TD via its Tyr-Arg motif, thereby closely resembling the structure of the TD of human p73.

In summary, the experiments revealed that the TD of *C.int.* p53/p73-a does not fold a second helix according to an all-or-nothing scheme, but the degree of helicity of the disputed sequence can rather be fine-tuned by mutations in the hinge region.

The residue-specific hetNOEs at lower temperatures revealed that temperature decrease can influence the fold of the second helix in a similar manner as the introduction of stabilizing residues within the hinge region; results section 4.2.13. At 10°C an estimated fraction of 25% of all conformers possesses a second helix. Since in all experiments only a single NH cross peak per residue was detected, it is to be assumed that the exchange between the two states with and without a folded second helix is fast on the NMR time scale, and hence cannot be resolved. A gain in the portion of one state at the cost of the other as a result of an amino acid substitution or lower temperature only resulted in a shift of the NH cross peaks of the affected amino acid, but neither in line broadening nor the appearance of additional peaks. But all detected NH cross peaks corresponding to residues from the core TD could be correlated to conformers with stably folded β-sheet and first helices, and this implies that the second helices are folding independently of the core. Notably, the data merely imply that under the conditions applied the second helix is at the brink to formation and do not imply that the folding of the second helix is non-cooperative. On the contrary, as judged from the higher enthalpy change upon thermal denaturation those hinge region mutants, which increase the overall thermal stability of the TD, also increase the folding cooperativity in comparison to the wild type TD. Interestingly, the chimeric protein, which possesses a stably folded second helix, shows a lower folding cooperativity compared to the wild type TD.

5.4. The only partially folded second helix in *C.int.* p53/p73-a could function in regulating interactions or in fine tuning the K_d of tetramerization

The second helical motif of *C.int.* p53/p73-a is not stably folded, but folding can be induced by temperature decrease. Interestingly the Tyr-Arg motif is conserved and if folded the second helix packs onto the core TD in a position closely resembling the TD of human p73. It does not seem likely that exactly these two characteristic essential residues have been retained in the p53/p73-a proteins of both *Ciona* species coincidentally, whereas all neighboring residues are not conserved. In turn this suggests that the weakly folded second

helix and the temperature induced change in the overall structural topology of the TD of *C.int.* p53/p73-a could fulfill a specific function. Several different functions would be conceivable.

The stability of the second helix is obviously a function of temperature, and if folded the second helix wraps around the neighboring dimer in the same clamp like fashion as in the TD of human p73. Therefore this SSE will influence the change in the K_d of tetramerization, which results from a change in temperature, and at lower temperatures the K_d of tetramerization will presumably be lower than it would be without the instable second helix. However, it is questionable if this stabilizing effect, which occurs without any doubt, is functionally relevant. Arguing against this theory is the observation that the thermal stability of the TD of *C.int.* p53/p73-a, as determined by circular dichroism (CD), is even higher than those of the TD of human p53. On the other hand analytical ultracentrifugation suggested that even at concentrations fairly exceeding the assumed K_d of tetramerization the construct lacking the second helix might not have been exclusively tetrameric, and if this applied it would indeed strongly argue for a function of the second helix in regulating the oligomeric state as a function of temperature.

Another possibility would be that the second helix serves as a regulator of inter- or intra-molecular interactions. Theoretically, proteins which interact with the TD could need the second helix as part of their binding interface. If this is the case the second helix would be stabilized by conformational selection upon binding and this would in turn stabilize the interaction as well as the tetrameric state. Comparable binding mechanisms have been reported in particular for intrinsically disordered proteins whose interaction sites possess transiently populated secondary structures. However, most likely the second helix is packing onto the core TD as in the TD of human p73, and in this conformation the only conserved residues, Tyr-Arg, are not surface exposed but instead involved in interactions with the core TD.

This speaks for the theory that the second helix might not itself be part of an interaction site but instead blocking one on the core TD. For human p63 it has been shown that the transactivation domain (TA) of the protein can as well bind to the TD.[108] Interestingly, the TA thereby binds onto the core TD and the TA-TD interaction is stronger when the second helix is truncated. Hence, the interface the TA is binding to obviously overlaps with the residues the second helices are packing onto. On the side of the p63 TA the MDM2 binding motif was

shown to be essential for binding onto the core TD. Since the MDM2 binding motif is conserved in the TA domain of *C.int.* p53/p73-a as well, it seems possible that the TA-core TD interaction could also occur in *C.int.* p53/p73-a. And if this should apply, the second helix, whose stability is tunable, would compete with this interaction in a temperature dependent manner.

Another possibility would be that the second helix regulates an intermolecular interaction. In human the E3 ubiquitin ligase Pirh2 negatively regulates p53,[200] and its interaction pattern is thereby fundamentally different from that observed for MDM2 and MDMX. Pirh2 interacts via its C-terminal zinc-binding motif with the TD of human p53, and the interaction strictly requires the TD to be stably tetrameric. This in turn results in the preferential ubiquitination of tetrameric, thus transcriptionally active p53.[201] Pirh2 was reported to bind p53 with a K_d on the order of 600 µM, thus the interaction is rather transient. Human p73 was as well found to be targeted by Pirh2 for proteasome-dependent degradation.[202] And although the direct physical interaction between Pirh2 and p73 was only generally confirmed, it seems likely that Pirh2 might as well specifically bind the TD of p73. But with respect to the TD of human p53 each monomer in the TD of human p73 is extended by a stably folded second helix, and those patches, which would be exposed on the surface of the TD of human p53, will be hidded to a large extent by the second helices once the TD of human p73 has fully folded. It can therefore be reasonably assumed that in any respect the formation of the second helix should have an effect on the Pirh2 mediated control of the protein level of human p73. The genome of *C.int.* codes for a Pirh2 ortholog (UniProtKB/Swiss-Prot: F6RD37-1), whose full length protein sequence aligns to the human protein with 51% sequence identity. This puts the possibility close that the temperature-dependently stable second helix of *C.int.* p53/p73-a could be a regulator for Pirh2 mediated proteasomal degradation.

5.5. The TD of *C.int.* p53/p73-b possesses a rigid second helix with an alternative protein sequence

The gene encoding *C.int.* p53/p73-b resulted from a relatively recent lineage-specific gene duplication caused by retrotransposition.[139] According to phylogenetic data most likely an mRNA transcript for *C.int.* p53/p73-a, or more precisely the ancestral version of *C.int.*

p53/p73-a existing at that time, was spontaneously reverse transcribed back into DNA and inserted into chromosomal DNA. Therefore a high degree of sequence identity between the paralogous proteins would be expected, but interestingly, within the TD residues with conservation in both proteins are restricted to the section spanning the β-strand and the first helix. Hence, *C.int.* p53/p73-b lacks the characteristic Tyr-Arg motif, which is decisive for the formation of the second helix in human p73 and p63 as well as in *C.int.* p53/p73-a, and it had from there been predicted that *C.int.* p53/p73-b possessed only a minimal / core TD.

However, the structural investigation revealed that the TD of *C.int.* p53/p73-b possesses a rigidly folded second helix. In the NMR solution structure this second helix adopts a position with respect to the core TD, closely resembling the overall architecture of the TD of human p73. The second helix is also wrapping around the neighboring primary dimer in a clamp like fashion and hence its formation depends on the tetrameric state. Due to technical limitations it could not be determined to what extent the second helix increases the stability of the tetramer, but the gain is presumably large since various residues from the second helix are engaged in an extensive network of inter-dimer salt bridges and hydrogen bonds, as well as hydrophobic contacts.

Interestingly, analytical ultracentrifugation experiments revealed that the second helix is albeit not required for tetramerization. This can presumably be attributed to the fact that the core TD of *C.int.* p53/p73-b is stabilized by salt bridges involving residues from both primary dimers (Lys 391 to Glu 394). It has been shown for the TD of human p53 that its tetrameric state as well depends on such inter-dimer salt-bridges, but interestingly, in *C.int.* p53/p73-b these salt bridges form between different pairs of monomers.

An observation, which is interesting with respect to the evolution of the motifs potentially stabilizing a second helix, is that in *C.int.* p53/p73-b the hydrophobic interactions between the core TD and residues from the second helix to a large extent depend on a Tyr residue at the end of the first helix. In *C.sav.* p53/p73-b the respective position is occupied with a Phe residue, and hence, functionally conserved. But neither *C.int.* p53/p73-a or *C.sav.* p53/p73-a nor *B.schl.* p53/p73 does possess an aromatic residue in the respective position. However, in the respective position Tyr (e.g. human and mouse p63) or Phe (e.g. zebrafish p63 and p73) are frequently found in vertebrate proteins which suggests that the ancestral protein, which had been the template for the retrotransposition, quite probably had already possessed a

Tyr or Phe residue at this important position. If this did apply no adaptive mutations at this position would have been required to allow for the stabilization of the second helix.

5.6. The second helix in the course of deuterostome evolution

In the TDs of human p63 and p73 each monomer comprises an additional second α-helix at its C-terminal end, which is not present in the TD of human p53. But recently Joerger et al. reported that in contrast to human p53 the TD of zebrafish p53 possesses this second helix.[195] Furthermore, they suggested based on sequence alignments that the single ancestral p63/p73 type protein, from which the vertebrate p53 family evolved through two sequential gene duplication events, possessed the second helix, and that it got lost on the p53 gene at different stages of vertebrate evolution. Like human also zebrafish possesses three p53 family genes, p53, p63 and p73, and therefore the second helix on p53 obviously got lost not before the second gene duplication during vertebrate evolution, which resulted in p63 and p73. *In vitro* and *in vivo* studies have shown that human p63 and p73 are able to form hetero-tetramers via their TDs, whereas none of the proteins forms mixed tetramers with p53.[88,106,109] Hence, the TD of p53 lost its second helix at a stage in evolution where two additional family members already existed whose functional cross-talk, mediated by the formation of mixed tetramers, is obviously needed or at least not problematic. From a functional perspective this puts the possibility close that prior to the second duplication event the TDs of p53 and the p63/p73 precursor had been able to form hetero-tetramers, as well. Afterwards the duplicated genes were most likely retained as a result of subfunctionalization and to facilitate this the TD of p53 had to get incompatible towards the TDs of p63 and p73. During evolution p53 accumulated substitutions within the tetramerization interface, became stabilized by the innovation of salt bridges spanning between the primary dimers and in turn did no longer require its second helix to stabilize the conformation of the tetramer. The low degree of conservation between the respective sequences from closely related vertebrate species also suggests that the degeneracy of the second helix in p53 must have occurred relatively recent. As the TD of p53 became incompatible the TDs of p63 and p73 could keep an more ancient domain architecture and retained their second helices.

Based on conserved motifs being suitable to stabilize the second helix it was predicted with high confidence that the TD of *B.schl.* p53/p73 also possesses a stably folded second helix.

After the lineage-specific gene duplication the degeneration into two no more interacting proteins developed differently in *Ciona* than in vertebrates. *C.int.* p53/p73-b was newly generated most probably by retrotransposition with the ancestral version of *C.int.* p53/p73-a serving as the template. The TD of *C.int.* p53/p73-b possesses a rigidly folded second helix, and this helix interestingly folds from an amino acid sequence, which lacks those residues being decisive for the stability of the second helices in the TDs of human p63 and p73. In contrast to this the Tyr-Arg motif is conserved in *C.int.* p53/p73-a, but its TD does not possess a rigidly folded second helix. However, the consensus of both protein sequences with respect to residues with conservation in vertebrate p63 and p73 combined with the observation of a proper N-cap for the second helix in *B.schl.* p53/p73 puts the suggestion close that prior to the gene duplication event the ancestral version of *C.int.* p53/p73-a had as well possessed a stable helix. After this gene duplication the TDs then accumulated amino acid substitutions, which prevent functional crosstalk via the formation of hetero-oligomers and in the consequence facilitating subfunctionalization. The TD of *C.int.* p53/p73-b is stabilized by an inter-dimer salt bridge, which specifically stabilizes the homo-tetramer, and its second helix presumably depends on hydrophobic inter-dimer interactions with a Tyr residue, which would not be in place if the respective monomer would be exchanged against *C.int.* p53/p73-a. In combination this surely contributes a large portion to the fact that the formation of hetero-oligomers does not occur. The separated oligomerization therefore does not depend on a rigidly folded second helix within the TD of *C.int.* p53/p73-a. And this in turn allowed for degenerating amino acid substitutions, which result in the drastically reduced and temperature dependent stability of the second helix of *C.int.* p53/p73-a.

As a summary fFigure 5-1 illustrates the phylogenetic relationship of the p53 family proteins of species from the two subphyla vertebrata and tunicata, and compares the architectures of their TDs. Vertebrata and tunicata are the closest sibling clades to one another and in this thesis it was shown that each tunicate species has at least one p53 family protein with a TD closely resembling the topology of the TDs of human p63 and p73. This strongly suggests that the last common ancestor of vertebrate and tunicate p53 family proteins did already possess a TD stabilized by a rigidly folded second helix.

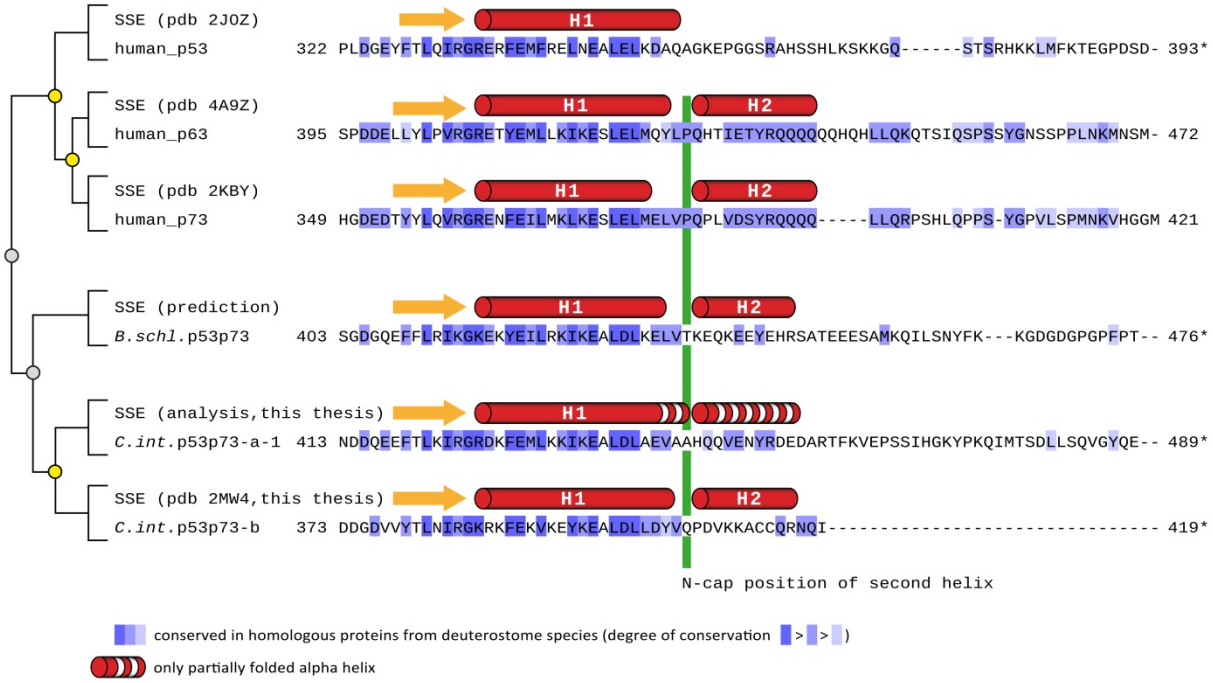


Figure 5-1 Phylogenetic tree and alignment of the tetramerization domains of human p53, p63 and p73 as well as the p53/p73 proteins from the tunicate species *Botryllus schlosserie* und *Ciona intestinalis*. The phylogenetic tree on the left depicts the homology relationship. All proteins derived from a common ancestor. A speciation event (marked with a gray circle) gives to orthologous proteins, whereas after a gene duplication event (marked with yellow circles) the duplicates will have a paralogous relationship. An asteris indicates the C-terminal end of the natural protein sequence. As a prefix to the sequences the secondary structure elements (SSEs) are depicted.

5.7. Outlook: the colonial ascidian *Botryllus schlosseri*

The genome of the colonial ascidian *Botryllus schlosseri* (*B.schl.*) was recently sequenced,[133] and several features connected with its colonial life style could make *B.schl.* a promising next target for further investigations of ancestral functions and domain structures of p53/p63/p73 like proteins conserved in ascidians.

After sexual reproduction *B.schl.* follows the standard plan of chordate development. But after the adult has settled it in addition invokes a stem cell-mediated budding program with several rounds of asexual reproduction.[203] Colonies from the same genetic origin are able to fuse their blood vessels to form a single organism, whereas incompatible colonies reject one another. Once the blood vessels have been fused, circulating stem cells from one part of the colony can compete and replace the germline or somatic cells in another part of the colony.[204–208] With respect to the function of human p63 as the guardian of the female germline and in the regulation of stem cells it might also be worth investigating the structure of the TD of *B.schl.* p53/p73. However, so far it had only been possible to forecast a preliminary protein sequence. As a compelling precondition to structural investigations this would first have to be finally approved.

6. References

1. Gabaldón, T. Large-scale assignment of orthology: back to phylogenetics? *Genome Biol.* **9,** 235 (2008).

2. Lane, D. P. & Crawford, L. V. T antigen is bound to a host protein in SV40-transformed cells. *Nature* **278,** 261–3 (1979).

3. Linzer, D. I. & Levine, A. J. Characterization of a 54K dalton cellular SV40 tumor antigen present in SV40-transformed cells and uninfected embryonal carcinoma cells. *Cell* **17,** 43–52 (1979).

4. Eliyahu, D., Raz, A., Gruss, P., Givol, D. & Oren, M. Participation of p53 cellular tumour antigen in transformation of normal embryonic cells. *Nature* **312,** 646–9

5. Parada, L. F., Land, H., Weinberg, R. A., Wolf, D. & Rotter, V. Cooperation between gene encoding p53 tumour antigen and ras in cellular transformation. *Nature* **312,** 649–51

6. Baker, S. J. *et al.* Chromosome 17 deletions and p53 gene mutations in colorectal carcinomas. *Science* **244,** 217–21 (1989).

7. Finlay, C. A., Hinds, P. W. & Levine, A. J. The p53 proto-oncogene can act as a suppressor of transformation. *Cell* **57,** 1083–93 (1989).

8. Lane, D. P. Cancer. p53, guardian of the genome. *Nature* **358,** 15–6 (1992).

9. Vogelstein, B., Lane, D. & Levine, A. J. Surfing the p53 network. *Nature* **408,** 307–10 (2000).

10. Donehower, L. A. *et al.* Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* **356,** 215–21 (1992).

11. Hollstein, M., Sidransky, D., Vogelstein, B. & Harris, C. C. p53 mutations in human cancers. *Science* **253,** 49–53 (1991).

12. Eliyahu, D., Michalovitz, D., Eliyahu, S., Pinhasi-Kimhi, O. & Oren, M. Wild-type p53 can inhibit oncogene-mediated focus formation. *Proc. Natl. Acad. Sci. U. S. A.* **86,** 8763–7 (1989).

13. Bargonetti, J., Reynisdóttir, I., Friedman, P. N. & Prives, C. Site-specific binding of wild-type p53 to cellular DNA is inhibited by SV40 T antigen and mutant p53. *Genes Dev.* **6,** 1886–98 (1992).

14. Kern, S. E. *et al.* Oncogenic forms of p53 inhibit p53-regulated gene expression. *Science* **256,** 827–30 (1992).

15. Martinez, J., Georgoff, I. & Levine, A. J. Cellular localization and cell cycle regulation by a temperature-sensitive p53 protein. *Genes Dev.* **5,** 151–9 (1991).

16. Milner, J. & Medcalf, E. A. Cotranslation of activated mutant p53 with wild type drives the wild-type p53 protein into the mutant conformation. *Cell* **65,** 765–74 (1991).

17. Conseiller, E. *et al.* CTS1: a p53-derived chimeric tumor suppressor gene with enhanced in vitro apoptotic properties. *J. Clin. Invest.* **101,** 120–7 (1998).

18. Waterman, M. J., Waterman, J. L. & Halazonetis, T. D. An engineered four-stranded coiled coil substitutes for the tetramerization domain of wild-type p53 and alleviates transdominant inhibition by tumor-derived p53 mutants. *Cancer Res.* **56,** 158–63 (1996).

19. Toledo, F. & Wahl, G. M. Regulating the p53 pathway: in vitro hypotheses, in vivo veritas. *Nat. Rev. Cancer* **6,** 909–23 (2006).

20. Zhou, B. B. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408,** 433–9 (2000).

21. Freiman, R. N. & Tjian, R. Regulating the regulators: lysine modifications make their mark. *Cell* **112,** 11–7 (2003).

22. Haupt, Y., Maya, R., Kazaz, A. & Oren, M. Mdm2 promotes the rapid degradation of p53. *Nature* **387,** 296–9 (1997).

23. Perry, M. E., Piette, J., Zawadzki, J. A., Harveyt, D. & Levinet, A. J. The mdm-2 gene is induced in response to UV light in a p53-dependent manner. **90,** 11623–11627 (1993).

24. Wu, X., Bayle, J. H., Olson, D. & Levine, A. J. The p53-mdm-2 autoregulatory feedback loop. *Genes Dev.* **7,** 1126–32 (1993).

25. Lev Bar-Or, R. *et al.* Generation of oscillations by the p53-Mdm2 feedback loop: a theoretical and experimental study. *Proc. Natl. Acad. Sci. U. S. A.* **97,** 11250–5 (2000).

26. Appella, E. Modulation of p53 function in cellular regulation. *Eur. J. Biochem.* **268,** 2763–2763 (2001).

27. Khosravi, R. *et al.* Rapid ATM-dependent phosphorylation of MDM2 precedes p53 accumulation in response to DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* **96,** 14973–7 (1999).

28. Maya, R. *et al.* ATM-dependent phosphorylation of Mdm2 on serine 395: role in p53 activation by DNA damage. *Genes Dev.* **15,** 1067–77 (2001).

29. Mayo, L. D., Turchi, J. J. & Berberich, S. J. Mdm-2 phosphorylation by DNA-dependent protein kinase prevents interaction with p53. *Cancer Res.* **57,** 5013–6 (1997).

30. Pomerantz, J. *et al.* The Ink4a tumor suppressor gene product, p19Arf, interacts with MDM2 and neutralizes MDM2's inhibition of p53. *Cell* **92,** 713–23 (1998).

31. Honda, R. & Yasuda, H. Association of p19(ARF) with Mdm2 inhibits ubiquitin ligase activity of Mdm2 for tumor suppressor p53. *EMBO J.* **18,** 22–7 (1999).

32. Honda, R. & Yasuda, H. Activity of MDM2, a ubiquitin ligase, toward p53 or itself is dependent on the RING finger domain of the ligase. *Oncogene* **19,** 1473–6 (2000).

33. Shvarts, A. *et al.* MDMX: a novel p53-binding protein with some functional properties of MDM2. *EMBO J.* **15,** 5349–57 (1996).

34. Marine, J.-C. & Jochemsen, A. G. Mdmx as an essential regulator of p53 activity. *Biochem. Biophys. Res. Commun.* **331,** 750–60 (2005).

35. Stad, R. *et al.* Mdmx stabilizes p53 and Mdm2 via two distinct mechanisms. *EMBO Rep.* **2,** 1029–34 (2001).

36. De Graaf, P. *et al.* Hdmx protein stability is regulated by the ubiquitin ligase activity of Mdm2. *J. Biol. Chem.* **278,** 38315–24 (2003).

37. Kawai, H. *et al.* DNA damage-induced MDMX degradation is mediated by MDM2. *J. Biol. Chem.* **278,** 45946–53 (2003).

38. Pan, Y. & Chen, J. MDM2 promotes ubiquitination and degradation of MDMX. *Mol. Cell. Biol.* **23,** 5113–21 (2003).

39. Gilkes, D. M. *Multiple Modes of Mdmx Regulation Affect P53 Activation*. Book ISBN 0549774025 (ProQuest, 2008). at <http://books.google.com/books?id=Fy4xCF1rxO4C&pgis=1>

40. Lambert, P. F., Kashanchi, F., Radonovich, M. F., Shiekhattar, R. & Brady, J. N. Phosphorylation of p53 serine 15 increases interaction with CBP. *J. Biol. Chem.* **273,** 33048–53 (1998).

41. Ito, A. *et al.* p300/CBP-mediated p53 acetylation is commonly induced by p53-activating agents and inhibited by MDM2. *EMBO J.* **20,** 1331–40 (2001).

42. Gu, W. & Roeder, R. G. Activation of p53 sequence-specific DNA binding by acetylation of the p53 C-terminal domain. *Cell* **90,** 595–606 (1997).

43. Liu, L. *et al.* p53 sites acetylated in vitro by PCAF and p300 are acetylated in vivo in response to DNA damage. *Mol. Cell. Biol.* **19,** 1202–9 (1999).

44. Kitayner, M. *et al.* Structural basis of DNA recognition by p53 tetramers. *Mol. Cell* **22,** 741–53 (2006).

45. Halazonetis, T. D. & Kandil, A. N. Conformational shifts propagate from the oligomerization domain of p53 to its tetrameric DNA binding domain and restore DNA binding to select p53 mutants. *EMBO J.* **12,** 5057–64 (1993).

46. Imagawa, T., Terai, T., Yamada, Y., Kamada, R. & Sakaguchi, K. Evaluation of transcriptional activity of p53 in individual living mammalian cells. *Anal. Biochem.* **387,** 249–56 (2009).

47. Kawaguchi, T. *et al.* The relationship among p53 oligomer formation, structure and transcriptional activity using a comprehensive missense mutation library. *Oncogene* **24,** 6976–81 (2005).

48. Davison, T. S., Yin, P., Nie, E., Kay, C. & Arrowsmith, C. H. Characterization of the oligomerization defects of two p53 mutants found in families with Li-Fraumeni and Li-Fraumeni-like syndrome. *Oncogene* **17,** 651–6 (1998).

49. DiGiammarino, E. L. *et al.* A novel mechanism of tumorigenesis involving pH-dependent destabilization of a mutant p53 tetramer. *Nat. Struct. Biol.* **9,** 12–6 (2002).

50. Rajagopalan, S., Huang, F. & Fersht, A. R. Single-Molecule characterization of oligomerization kinetics and equilibria of the tumor suppressor p53. *Nucleic Acids Res.* **39,** 2294–303 (2011).

51. Gaglia, G., Guan, Y., Shah, J. V & Lahav, G. Activation and control of p53 tetramerization in individual living cells. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 15497–501 (2013).

52. Rajagopalan, S., Jaulent, A. M., Wells, M., Veprintsev, D. B. & Fersht, A. R. 14-3-3 activation of DNA binding of p53 by enhancing its association into tetramers. *Nucleic Acids Res.* **36,** 5983–91 (2008).

53. Sakaguchi, K. *et al.* Phosphorylation of serine 392 stabilizes the tetramer formation of tumor suppressor protein p53. *Biochemistry* **36,** 10117–24 (1997).

54. Weinberg, R. L., Veprintsev, D. B. & Fersht, A. R. Cooperative binding of tetrameric p53 to DNA. *J. Mol. Biol.* **341,** 1145–59 (2004).

55. Schumacher, B., Mondry, J., Thiel, P., Weyand, M. & Ottmann, C. Structure of the p53 C-terminus bound to 14-3-3: implications for stabilization of the p53 tetramer. *FEBS Lett.* **584,** 1443–8 (2010).

56. Ma, L. *et al.* A plausible model for the digital response of p53 to DNA damage. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 14266–71 (2005).

57. Bode, A. M. & Dong, Z. Post-translational modification of p53 in tumorigenesis. *Nat. Rev. Cancer* **4,** 793–805 (2004).

58. Brooks, C. L. & Gu, W. p53 regulation by ubiquitin. *FEBS Lett.* **585,** 2803–9 (2011).

59. Fernandez-Fernandez, M. R., Veprintsev, D. B. & Fersht, A. R. Proteins of the S100 family regulate the oligomerization of p53 tumor suppressor. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 4735–40 (2005).

60. Friedman, P. N., Chen, X., Bargonetti, J. & Prives, C. The p53 protein is an unusually shaped tetramer that binds directly to DNA. *Proc. Natl. Acad. Sci. U. S. A.* **90,** 3319–23 (1993).

61. Lee, W. *et al.* Solution structure of the tetrameric minimum transforming domain of p53. *Nat. Struct. Biol.* **1,** 877–890 (1994).

62. Stenger, J. E., Mayr, G. A., Mann, K. & Tegtmeyer, P. Formation of stable p53 homotetramers and multiples of tetramers. *Mol. Carcinog.* **5,** 102–6 (1992).

63. Davison, T. S. *et al.* Structure and functionality of a designed p53 dimer. *J. Mol. Biol.* **307,** 605–17 (2001).

64. Jeffrey, P. D., Gorina, S. & Pavletich, N. P. Crystal structure of the tetramerization domain of the p53 tumor suppressor at 1.7 angstroms. *Science* **267,** 1498–502 (1995).

65. ThuZar, L., Durant, J. J. & Bashford, D. A Fluid Salt-Bridging Cluster and the Stabilization of p53. *J. Mol. Biol.* **373,** 1334–1347 (2007).

66. Brokx, R. D., Bolewska-Pedyczak, E. & Gariépy, J. A stable human p53 heterotetramer based on constructive charge interactions within the tetramerization domain. *J. Biol. Chem.* **278,** 2327–32 (2003).

67. Ishioka, C. *et al.* Oligomerization is not essential for growth suppression by p53 in p53-deficient osteosarcoma Saos-2 cells. *Biochem. Biophys. Res. Commun.* **232,** 54–60 (1997).

68. Muscolini, M. *et al.* Characterization of a new cancer-associated mutant of p53 with a missense mutation (K351N) in the tetramerization domain. *Cell Cycle* **8,** 3396–405 (2009).

69. Kaghad, M. *et al.* Monoallelically expressed gene related to p53 at 1p36, a region frequently deleted in neuroblastoma and other human cancers. *Cell* **90,** 809–19 (1997).

70. Yang, A. *et al.* p63, a p53 homolog at 3q27-29, encodes multiple products with transactivating, death-inducing, and dominant-negative activities. *Mol. Cell* **2,** 305–16 (1998).

71. Mills, A. A. *et al.* p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature* **398,** 708–13 (1999).

72. Yang, A. *et al.* p63 is essential for regenerative proliferation in limb, craniofacial and epithelial development. *Nature* **398,** 714–8 (1999).

73. Yang, A. *et al.* p73-deficient mice have neurological, pheromonal and inflammatory defects but lack spontaneous tumours. *Nature* **404,** 99–103 (2000).

74. Melino, G., Lu, X., Gasco, M., Crook, T. & Knight, R. A. Functional regulation of p73 and p63: development and cancer. *Trends Biochem. Sci.* **28,** 663–70 (2003).

75. Melino, G., De Laurenzi, V. & Vousden, K. H. p73: Friend or foe in tumorigenesis. *Nat. Rev. Cancer* **2,** 605–15 (2002).

76. Rufini, A. *et al.* p73 in Cancer. *Genes Cancer* **2,** 491–502 (2011).

77. MacPartlin, M. *et al.* p300 regulates p63 transcriptional activity. *J. Biol. Chem.* **280,** 30604–10 (2005).

78. Zeng, X. *et al.* The N-terminal domain of p73 interacts with the CH1 domain of p300/CREB binding protein and mediates transcriptional activation and apoptosis. *Mol. Cell. Biol.* **20,** 1299–310 (2000).

79. Zdzalik, M. *et al.* Interaction of regulators Mdm2 and Mdmx with transcription factors p53, p63 and p73. *Cell Cycle* **9,** 4584–4591 (2010).

80. Little, N. A. & Jochemsen, A. G. Hdmx and Mdm2 can repress transcription activation by p53 but not by p63. *Oncogene* **20,** 4576–80 (2001).

81. Rossi, M. *et al.* The ubiquitin-protein ligase Itch regulates p73 stability. *EMBO J.* **24,** 836–48 (2005).

82. Rossi, M. *et al.* Itch/AIP4 associates with and promotes p63 protein degradation. *Cell Cycle* **5,** 1816–22 (2006).

83. Rossi, M. *et al.* The E3 ubiquitin ligase Itch controls the protein stability of p63. *Proc. Natl. Acad. Sci. U. S. A.* **103,** 12753–8 (2006).

84. Jost, C. A., Marin, M. C. & Kaelin, W. G. p73 is a simian [correction of human] p53-related protein that can induce apoptosis. *Nature* **389,** 191–4 (1997).

85. De Laurenzi, V. *et al.* Two new p73 splice variants, gamma and delta, with different transcriptional activity. *J. Exp. Med.* **188,** 1763–8 (1998).

86. Melino, G. *et al.* p73 Induces apoptosis via PUMA transactivation and Bax mitochondrial translocation. *J. Biol. Chem.* **279,** 8076–83 (2004).

87. Keyes, W. M. *et al.* p63 deficiency activates a program of cellular senescence and leads to accelerated aging. *Genes Dev.* **19,** 1986–99 (2005).

88. Coutandin, D. *et al.* Conformational stability and activity of p73 require a second helix in the tetramerization domain. *Cell Death Differ.* **16,** 1582–9 (2009).

89. Natan, E. & Joerger, A. C. Structure and kinetic stability of the p63 tetramerization domain. *J. Mol. Biol.* **415,** 503–13 (2012).

90. Grob, T. J. *et al.* Human delta Np73 regulates a dominant negative feedback loop for TAp73 and p53. *Cell Death Differ.* **8,** 1213–23 (2001).

91. Stiewe, T., Theseling, C. C. & Pützer, B. M. Transactivation-deficient Delta TA-p73 inhibits p53 by direct competition for DNA binding: implications for tumorigenesis. *J. Biol. Chem.* **277,** 14177–85 (2002).

92. Nakagawa, T. *et al.* Autoinhibitory regulation of p73 by Delta Np73 to modulate cell survival and death through a p73-specific target element within the Delta Np73 promoter. *Mol. Cell. Biol.* **22,** 2575–85 (2002).

93. Kartasheva, N. N., Contente, A., Lenz-Stöppler, C., Roth, J. & Dobbelstein, M. p53 induces the expression of its antagonist p73 Delta N, establishing an autoregulatory feedback loop. *Oncogene* **21,** 4715–27 (2002).

94. Tomasini, R. *et al.* TAp73 knockout shows genomic instability with infertility and tumor suppressor functions. *Genes Dev.* **22,** 2677–91 (2008).

95. Concin, N. *et al.* Transdominant DeltaTAp73 isoforms are frequently up-regulated in ovarian cancer. Evidence for their role as epigenetic p53 inhibitors in vivo. *Cancer Res.* **64,** 2449–60 (2004).

96. Casciano, I. *et al.* Expression of DeltaNp73 is a molecular marker for adverse outcome in neuroblastoma patients. *Cell Death Differ.* **9,** 246–51 (2002).

97. Puig, P. *et al.* p73 Expression in Human Normal and Tumor Tissues: Loss of p73{alpha} Expression Is Associated with Tumor Progression in Bladder Cancer. *Clin. Cancer Res.* **9,** 5642–5651 (2003).

98. Ponting, C. P. SAM: a novel motif in yeast sterile and Drosophila polyhomeotic proteins. *Protein Sci.* **4,** 1928–30 (1995).

99. Kantaputra, P. N. *et al.* Mutation in SAM domain of TP63 is associated with nonsyndromic cleft lip and palate and cleft palate. *Am. J. Med. Genet. A* **155A,** 1432–6 (2011).

100. Kantaputra, P. N., Hamada, T., Kumchai, T. & McGrath, J. A. Heterozygous mutation in the SAM domain of p63 underlies Rapp-Hodgkin ectodermal dysplasia. *J. Dent. Res.* **82,** 433–7 (2003).

101. Tsutsui, K. *et al.* A novel p63 sterile alpha motif (SAM) domain mutation in a Japanese patient with ankyloblepharon, ectodermal defects and cleft lip and palate (AEC) syndrome without ankyloblepharon. *Br. J. Dermatol.* **149,** 395–9 (2003).

102. Bakkers, J. & Camacho-Carvajal, M. Destabilization of ΔNp63α by Nedd4-Mediated Ubiquitination Ubc9-Mediated Sumoylation, and Its Implications on Dorsoventral Patterning of the Zebrafish Embryo. *Cell Cycle* **4,** 790–800 (2005).

103. Li, Y., Zhou, Z. & Chen, C. WW domain-containing E3 ubiquitin protein ligase 1 targets p63 transcription factor for ubiquitin-mediated proteasomal degradation and regulates apoptosis. *Cell Death Differ.* **15,** 1941–51 (2008).

104. Serber, Z. *et al.* A C-terminal inhibitory domain controls the activity of p63 by an intramolecular mechanism. *Mol. Cell. Biol.* **22,** 8601–11 (2002).

105. Straub, W. E. *et al.* The C-terminus of p63 contains multiple regulatory elements with different functions. *Cell Death Dis.* **1,** e5 (2010).

106. Joerger, A. C. *et al.* Structural evolution of p53, p63, and p73: implication for heterotetramer formation. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 17705–10 (2009).

107. Suh, E.-K. *et al.* p63 protects the female germ line during meiotic arrest. *Nature* **444,** 624–8 (2006).

108. Deutsch, G. B. *et al.* DNA damage in oocytes induces a switch of the quality control factor TAp63α from dimer to tetramer. *Cell* **144,** 566–76 (2011).

109. Luh, L. M. *et al.* Analysis of the oligomeric state and transactivation potential of TAp73α. *Cell Death Differ.* **20,** 1008–16 (2013).

110. Brodsky, M. H. *et al.* Drosophila melanogaster MNK/Chk2 and p53 regulate multiple DNA repair and apoptotic pathways following DNA damage. *Mol. Cell. Biol.* **24,** 1219–31 (2004).

111. Derry, W. B. *et al.* Regulation of developmental rate and germ cell proliferation in Caenorhabditis elegans by the p53 gene network. *Cell Death Differ.* **14,** 662–70 (2007).

112. Schumacher, B. *et al.* C. elegans ced-13 can promote apoptosis and is induced in response to DNA damage. *Cell Death Differ.* **12,** 153–61 (2005).

113. Ou, H. Der, Löhr, F., Vogel, V., Mäntele, W. & Dötsch, V. Structural evolution of C-terminal domains in the p53 family. *EMBO J.* **26,** 3463–73 (2007).

114. Gee, H. *Before the Backbone: Views on the Origin of the Vertebrates*. Book ISBN 0412483009 (Chapman & Hall, 1996).

115. Hall, B. K. *Evolutionary Developmental Biology*. Book ISBN 0412785900 (Chapman & Hall, 1999).

116. Adoutte, A. *et al.* The new animal phylogeny: Reliability and implications. *Proc. Natl. Acad. Sci.* **97,** 4453–4456 (2000).

117. Darwin, C. *The Descent of Man*. (Murray, 1871).

118. Delsuc, F., Brinkmann, H., Chourrout, D. & Philippe, H. Tunicates and not cephalochordates are the closest living relatives of vertebrates. *Nature* **439,** 965–8 (2006).

119. Satoh, N. *Developmental Genomics of Ascidians*. Book ISBN 1118656040 (2013).

120. Nakatani, Y., Moody, R. & Smith, W. C. Mutations affecting tail and notochord development in the ascidian Ciona savignyi. *Development* **126,** 3293–301 (1999).

121. Dehal, P. *et al.* The draft genome of Ciona intestinalis: insights into chordate and vertebrate origins. *Science* **298,** 2157–67 (2002).

122. Simmen, M. W., Leitgeb, S., Clark, V. H., Jones, S. J. M. & Bird, A. Gene number in an invertebrate chordate, Ciona intestinalis. *Proc. Natl. Acad. Sci.* **95,** 4437–4440 (1998).

123. Friedman, R. & Hughes, A. L. The temporal distribution of gene duplication events in a set of highly conserved human gene families. *Mol. Biol. Evol.* **20,** 154–61 (2003).

124. Gu, X., Wang, Y. & Gu, J. Age distribution of human gene families shows significant roles of both large- and small-scale duplications in vertebrate evolution. *Nat. Genet.* **31,** 205–9 (2002).

125. McLysaght, A., Hokamp, K. & Wolfe, K. H. Extensive genomic duplication during early chordate evolution. *Nat. Genet.* **31,** 200–4 (2002).

126. Friedman, R. & Hughes, A. L. Pattern and timing of gene duplication in animal genomes. *Genome Res.* **11,** 1842–7 (2001).

127. Ohno, S. *Evolution by gene duplication*. Book ISBN 0045750157 (Springer, 1970).

128. Abi-Rached, L., Gilles, A., Shiina, T., Pontarotti, P. & Inoko, H. Evidence of en bloc duplication in vertebrate genomes. *Nat. Genet.* **31,** 100–5 (2002).

129. Ambreen, S., Khalil, F. & Abbasi, A. A. Integrating large-scale phylogenetic datasets to dissect the ancient evolutionary history of vertebrate genome. *Mol. Phylogenet. Evol.* **78C,** 1–13 (2014).

130. Rutkowski, R., Hofmann, K. & Gartner, A. Phylogeny and function of the invertebrate p53 superfamily. *Cold Spring Harb. Perspect. Biol.* **2,** 1–13 (2010).

131. Miwata, K. *et al.* Systematic analysis of embryonic expression profiles of zinc finger genes in Ciona intestinalis. *Dev. Biol.* **292,** 546–54 (2006).

132. Hughes, A. L. & Friedman, R. Loss of ancestral genes in the genomic evolution of Ciona intestinalis. *Evol. Dev.* **7,** 196–200 (2005).

133. Voskoboynik, A. *et al.* The genome sequence of the colonial chordate, Botryllus schlosseri. *Elife* **2,** e00569 (2013).

134. Terajima, D. *et al.* Identification of candidate genes encoding the core components of the cell death machinery in the Ciona intestinalis genome. *Cell Death Differ.* **10,** 749–53 (2003).

135. Takada, N. *et al.* The cell death machinery controlled by Bax and Bcl-XL is evolutionarily conserved in Ciona intestinalis. *Apoptosis* **10,** 1211–20 (2005).

136. Satou, Y. *et al.* A cDNA resource from the basal chordate Ciona intestinalis. *Genesis* **33,** 153–4 (2002).

137. Tassy, O. *et al.* The ANISEED database : Digital representation , formalization , and elucidation of a chordate developmental program. 1459–1468 (2010). doi:10.1101/gr.108175.110.

138. Endo, T. *et al.* CIPRO 2.5: Ciona intestinalis protein database, a unique integrated repository of large-scale omics data, bioinformatic analyses and curated annotation, with user rating and reviewing functionality. *Nucleic Acids Res.* **39,** D807–14 (2011).

139. Nedelcu, A. M. & Tan, C. Early diversification and complex evolutionary history of the p53 tumor suppressor gene family. *Dev. Genes Evol.* **217,** 801–6 (2007).

140. Noda, T. The maternal genes Ci-p53/p73-a and Ci-p53/p73-b regulate zygotic ZicL expression and notochord differentiation in Ciona intestinalis embryos. *Dev. Biol.* **360,** 216–29 (2011).

141. Yasuo, H. & Satoh, N. Conservation of the Developmental Role ofBrachyuryin Notochord Formation in a Urochordate, the AscidianHalocynthia roretzi. *Dev. Biol.* **200,** 158–170 (1998).

142. Yagi, K., Satou, Y. & Satoh, N. A zinc finger transcription factor, ZicL, is a direct activator of Brachyury in the notochord specification of Ciona intestinalis. *Development* **131,** 1279–88 (2004).

143. Danilova, N., Sakamoto, K. M. & Lin, S. P53 Family in Development. *Mech. Dev.* **125,** 919–31 (2008).

144. Wallingford, J. B., Seufert, D. W., Virta, V. C. & Vize, P. D. P53 activity is essential for normal development in Xenopus. *Curr. Biol.* **7,** 747–757 (1997).

145. Sah, V. P. *et al.* A subset of p53-deficient embryos exhibit exencephaly. *Nat. Genet.* **10,** 175–80 (1995).

146. Armstrong, J. F., Kaufman, M. H., Harrison, D. J. & Clarke, A. R. High-frequency developmental abnormalities in p53-deficient mice. *Curr. Biol.* **5,** 931–6 (1995).

147. Satou, Y. *et al.* Improved genome assembly and evidence-based global gene model set for the chordate Ciona intestinalis: new insight into intron and operon populations. *Genome Biol.* **9,** R152 (2008).

148. Ferrè, F. & Clote, P. DiANNA 1.1: an extension of the DiANNA web server for ternary cysteine classification. *Nucleic Acids Res.* **34,** W182–5 (2006).

149. Ferrè, F. & Clote, P. DiANNA: a web server for disulfide connectivity prediction. *Nucleic Acids Res.* **33,** W230–2 (2005).

150. Ferrè, F. & Clote, P. Disulfide connectivity prediction using secondary structure information and diresidue frequencies. *Bioinformatics* **21,** 2336–46 (2005).

151. Saiki, R. K. *et al.* Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* **239,** 487–91 (1988).

152. Mullis, K. B. & Faloona, F. A. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol.* **155,** 335–50 (1987).

153. Fijalkowska, I. J., Schaaper, R. M. & Jonczyk, P. DNA replication fidelity in Escherichia coli: a multi-DNA polymerase affair. *FEMS Microbiol. Rev.* **36,** 1105–21 (2012).

154. Schaaper, R. M. Base selection, proofreading, and mismatch repair during DNA replication in Escherichia coli. *J. Biol. Chem.* **268,** 23762–5 (1993).

155. Kuban, W. *et al.* Role of Escherichia coli DNA polymerase IV in in vivo replication fidelity. *J. Bacteriol.* **186,** 4802–7 (2004).

156. Inoue, H., Nojima, H. & Okayama, H. High efficiency transformation of Escherichia coli with plasmids. *Gene* **96,** 23–8 (1990).

157. Schägger, H. Tricine-SDS-PAGE. *Nat. Protoc.* **1,** 16–22 (2006).

158. Schägger, H. & von Jagow, G. Tricine-sodium dodecyl sulfate-polyacrylamide gel electrophoresis for the separation of proteins in the range from 1 to 100 kDa. *Anal. Biochem.* **166,** 368–79 (1987).

159. Hagihara, Y., Oobatake, M. & Goto, Y. Thermal unfolding of tetrameric melittin: comparison with the molten globule state of cytochrome c. *Protein Sci.* **3,** 1418–29 (1994).

160. Buchner, J. & Kiefhaber, T. *Protein Folding Handbook 5-volume set*. Book ISBN 3527307842 (Wiley-VCH, 2005).

161. Mateu, M. G. & Fersht, a R. Nine hydrophobic side chains are key determinants of the thermodynamic stability and oligomerization status of tumour suppressor p53 tetramerization domain. *EMBO J.* **17,** 2748–58 (1998).

162. Greenfield, N. J. Analysis of circular dichroism data. *Methods Enzymol.* **383,** 282–317 (2004).

163. Johnson, C. R., Morin, P. E., Arrowsmith, C. H. & Freire, E. Thermodynamic analysis of the structural stability of the tetrameric oligomerization domain of p53 tumor suppressor. *Biochemistry* **34,** 5309–16 (1995).

164. Goddard, T. D. & Kneller, D. G. SPARKY 3. *Univ. California, San Fr.*

165. Salzmann, M., Pervushin, K., Wider, G., Senn, H. & Wüthrich, K. TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc. Natl. Acad. Sci. U. S. A.* **95,** 13585–90 (1998).

166. Salzmann, M., Wider, G., Pervushin, K., Senn, H. & Wüthrich, K. TROSY-type Triple-Resonance Experiments for Sequential NMR Assignments of Large Proteins. *J. Am. Chem. Soc.* **121,** 844–848 (1999).

167. Ferrage, F., Piserchio, A., Cowburn, D. & Ghose, R. On the measurement of 15N-{1H} nuclear Overhauser effects. *J. Magn. Reson.* **192,** 302–13 (2008).

168. Shen, Y., Delaglio, F., Cornilescu, G. & Bax, A. TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts. *J. Biomol. NMR* **44,** 213–23 (2009).

169. Cornilescu, G., Delaglio, F. & Bax, A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13,** 289–302 (1999).

170. Berjanskii, M. V & Wishart, D. S. The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts. *Nucleic Acids Res.* **35,** W531–7 (2007).

171. Kragelj, J., Ozenne, V., Blackledge, M. & Jensen, M. R. Conformational propensities of intrinsically disordered proteins from NMR chemical shifts. *Chemphyschem* **14,** 3034–45 (2013).

172. Wishart, D. S. & Sykes, B. D. Chemical shifts as a tool for structure determination. *Methods Enzymol.* **239,** 363–92 (1994).

173. Schwarzinger, S. *et al.* Sequence-dependent correction of random coil NMR chemical shifts. *J. Am. Chem. Soc.* **123,** 2970–8 (2001).

174. De Simone, A., Cavalli, A., Hsu, S.-T. D., Vranken, W. & Vendruscolo, M. Accurate random coil chemical shifts from an analysis of loop regions in native states of proteins. *J. Am. Chem. Soc.* **131,** 16332–3 (2009).

175. Tamiola, K., Acar, B. & Mulder, F. A. A. Sequence-specific random coil chemical shifts of intrinsically disordered proteins. *J. Am. Chem. Soc.* **132,** 18000–3 (2010).

176. Sibley, A. B., Cosman, M. & Krishnan, V. V. An empirical correlation between secondary structure content and averaged chemical shifts in proteins. *Biophys. J.* **84,** 1223–7 (2003).

177. Weinstock, D. S., Narayanan, C., Baum, J. & Levy, R. M. Correlation between 13Calpha chemical shifts and helix content of peptide ensembles. *Protein Sci.* **17,** 950–4 (2008).

178. Marsh, J. A., Singh, V. K., Jia, Z. & Forman-kay, J. D. Sensitivity of secondary structure propensities to sequence differences between a - and g -synuclein : Implications for fibrillation. 2795–2804 (2006). doi:10.1110/ps.062465306.conditions

179. Tamiola, K. & Mulder, F. a a. Using NMR chemical shifts to calculate the propensity for structural order and disorder in proteins. *Biochem. Soc. Trans.* **40,** 1014–20 (2012).

180. Camilloni, C., De Simone, A., Vranken, W. F. & Vendruscolo, M. Determination of secondary structure populations in disordered states of proteins using nuclear magnetic resonance chemical shifts. *Biochemistry* **51,** 2224–31 (2012).

181. Senn, H. *et al.* Stereospecific assignment of the methyl lH NMR lines of valine and leucine in polypeptides by nonrandom 13C labelling. **249,** 113–118 (1989).

182. Melacini, G., V, G. D. & Jolla, L. Separation of Intra- and Intermolecular NOEs through Simultaneous Editing and J -Compensated Filtering : A 4D Quadrature-Free Constant-Time J -Resolved Approach. 9735–9738 (2000).

183. López-Méndez, B. & Güntert, P. Automated protein structure determination from NMR spectra. *J. Am. Chem. Soc.* **128,** 13112–22 (2006).

184. Güntert, P. Automated NMR structure calculation with CYANA. *Methods Mol. Biol.* **278,** 353–78 (2004).

185. Güntert, P. Automated NMR protein structure calculation. *Prog. Nucl. Magn. Reson. Spectrosc.* **43,** 105–125 (2003).

186. Linge, J. P., Habeck, M., Rieping, W. & Nilges, M. ARIA: automated NOE assignment and NMR structure calculation. *Bioinformatics* **19,** 315–6 (2003).

187. Linge, J. P. & Nilges, M. Influence of non-bonded parameters on the quality of NMR structures: a new force field for NMR structure calculation. *J. Biomol. NMR* **13,** 51–9 (1999).

188. Linge, J. P., Williams, M. A., Spronk, C. A. E. M., Bonvin, A. M. J. J. & Nilges, M. Refinement of protein structures in explicit solvent. *Proteins* **50,** 496–506 (2003).

189. Laskowski, R., Rullmann, J. A., MacArthur, M., Kaptein, R. & Thornton, J. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8,** (1996).

190. Coutandin, D., Ou, H. Der, Löhr, F. & Dötsch, V. Tracing the protectors path from the germ line to the genome. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 15318–25 (2010).

191. Lane, D. P. *et al.* Conservation of all three p53 family members and Mdm2 and Mdm4 in the cartilaginous fish. *Cell Cycle* **10,** 4272–4279 (2011).

192. Cowan, R. & Whittaker, R. G. Hydrophobicity indices for amino acid residues as determined by high-performance liquid chromatography. *Pept. Res.* **3,** 75–80

193. Hautbergue, G. M. & Golovanov, A. P. Increasing the sensitivity of cryoprobe protein NMR experiments by using the sole low-conductivity arginine glutamate salt. *J. Magn. Reson.* **191,** 335–9 (2008).

194. Mora, P., Carbajo, R. J., Pineda-Lucena, A., Sánchez del Pino, M. M. & Pérez-Payá, E. Solvent-exposed residues located in the beta-sheet modulate the stability of the tetramerization domain of p53--a structural and combinatorial approach. *Proteins* **71,** 1670–85 (2008).

195. Joerger, A. C., Wilcken, R. & Andreeva, A. Tracing the Evolution of the p53 Tetramerization Domain. *Structure* **22,** 1301–1310 (2014).

196. Roth, C. *et al.* Evolution After Gene Duplication: Models, Mechanisms, Sequences, Systems, and Organisms. **73,** 58–73 (2007).

197. Force, A. *et al.* Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* **151,** 1531–45 (1999).

198. Lynch, M. & Force, A. The Probability of Duplicate Gene Preservation by Subfunctionalization. *Genetics* **154,** 459–473 (2000).

199. Bellomaria, A., Barbato, G., Melino, G., Paci, M. & Melino, S. Recognition of p63 by the E3 ligase ITCH: Effect of an ectodermal dysplasia mutant. *Cell Cycle* **9,** 3730–9 (2010).

200. Leng, R. P. *et al.* Promotes p53 Degradation. **112,** 779–791 (2003).

201. Sheng, Y., Laister, R., Lemak, A. & Wu, B. Molecular basis of Pirh2-mediated p53 ubiquitylation. *… Struct. Mol. …* **15,** 1334–1342 (2008).

202. Jung, Y.-S., Qian, Y. & Chen, X. The p73 tumor suppressor is targeted by Pirh2 RING finger E3 ubiquitin ligase for the proteasome-dependent degradation. *J. Biol. Chem.* **286,** 35388–95 (2011).

203. Manni, L. & Burighel, P. Common and divergent pathways in alternative developmental processes of ascidians. *Bioessays* **28,** 902–12 (2006).

204. Stoner, D. S. & Weissman, I. L. Somatic and germ cell parasitism in a colonial ascidian: Possible role for a highly polymorphic allorecognition system. *Proc. Natl. Acad. Sci.* **93,** 15254–15259 (1996).

205. Stoner, D. S., Rinkevich, B. & Weissman, I. L. Heritable germ and somatic cell lineage competitions in chimeric colonial protochordates. *Proc. Natl. Acad. Sci.* **96,** 9148–9153 (1999).

206. Laird, D. J., De Tomaso, A. W. & Weissman, I. L. Stem cells are units of natural selection in a colonial ascidian. *Cell* **123,** 1351–60 (2005).

207. Voskoboynik, A. *et al.* Identification of the endostyle as a stem cell niche in a colonial chordate. *Cell Stem Cell* **3,** 456–64 (2008).

208. Rinkevich, Y. *et al.* Repeated, long-term cycling of putative stem cells between niches in a basal chordate. *Dev. Cell* **24,** 76–88 (2013).

7. Appendix

7.1. Protein Sequences

7.1.1. *Ciona intestinalis* p53/p73-a isoforms with numbering according to T. Noda [140]

_Ci_-p53/p73-a isoform 1

aniseed transcript Id: KH2012:KH.C1.573.v2.A.SL1-1

UniProtKB/Swiss-Prot: Q4H301

```
  1 MDRPPEGSFD PDCFSDIWNN QLSQQSEITS SQQIYTFNEN TDGLTNSLTI
 51 DLLTEFQQFE QGPTSDANAG ENQAGNTEKH DGNEGSYQCL SNFPTSNPPM
101 RSYSNDMTQQ NTINQHHPAN MANLVTTSSH PINQHNLVYH STSPNTGDAP
151 NSTQNSPYAP SPCGGYEYPT SPLSSKPTIP PSTDYPGEWD FQINFGEATE
201 SAPKSAQYTY SPIINKLFVK MNVTCPIKFK CARPPPNGCV VRVMPVFKRP
251 EHVTDIVTRC PNHKIPDQAQ HIPHSQHLIR AEMPGENPAI YNVAMDGREN
301 VAVMFERPQI GAEYTTVLYK FMCLSSCVGG INRRPLNAVF NLENAEGQVL
351 GRRVVEVRVC SCPGRDRSQE EKRKRTAESE VKSGSKRTYK CMQGSMNIVQ
401 TAPAKRKCAM NGNDDQEEFT LKIRGRDKFE MLKKIKEALD LAEVAAHQQV
451 ENYRDEDART FKVEPSSIHG KYPKQIMTSD LLSQVGYQE
```

_Ci_-p53/p73-a isoform 2

aniseed transcript Id: KH2012:KH.C1.573.v5.A.SL1-2

UniProtKB/Swiss-Prot: no reference name available

UniProtKB/Swiss-Prot reference of a fragment encompassing the protein sequence starting within the last third of the DBD and ranging to the C-terminus: Q4H2Z9 (fragment)

```
  1 MDRPPEGSFD PDCFSDIWNN QLSQQSEITS SQQIYTFNEN TDGLTNSLTI
 51 DLLTEFQQFE QGPTSDANAG ENQAGNTEKH DGNEGSYQCL SNFPTSNPPM
101 RSYSNDMTQQ NTINQHHPAN MANLVTTSSH PINQHNLVYH STSPNTGDAP
151 NSTQNSPYAP SPCGGYEYPT SPLSSKPTIP PSTDYPGEWD FQINFGEATE
201 SAPKSAQYTY SPIINKLFVK MNVTCPIKFK CARPPPNGCV VRVMPVFKRP
251 EHVTDIVTRC PNHKIPDQAQ HIPHSQHLIR AEMPGENPAI YNVAMDGREN
301 VAVMFERPQI GAEYTTVLYK FMCLSSCVGG INRRPLNAVF NLENAEGQVL
351 GRRVVEVRVC SCPGRDRSQE EKRKRTAESE VKSGSKRTYK CMQGSMNIVQ
401 TAPAKRKCAM NGNDDQEEFT LKIRGRDKFE MLKKIKEALD LAEVAAHQQV
451 ENYRDEDART RQSTSSAAPP REESSASTSR AGQASTSALC MTEHVSHGSQ
501 GFSLTPTHSS HPLTVTPTLE PPIVEGKTPY THNTSAYPGM VDQSYWSTES
551 DLGYDIPDTK QPTTTGNGLP NWPNQESLDS NPKLLRNISG LSSVSSLTSS
601 QLSSQTGRFV TRVTLRQTVA LNPDMPNSQR WDHDTTSDNE DVTRNDVKRE
651 HDNVTNTDSF DFL
```

_Ci_-p53/p73-a isoform 3

aniseed transcript Id: none

UniProtKB/Swiss-Prot: no reference available

NCBI GenBank, cDNA clone: cima833F15 (translation)

```
  1 MDRPPEGSFD PDCFSDIWNN QLSQQSEITS SQQIYTFNEN TDGLTNSLTI
 51 DLLTEFQQFE QGPTSDANAG ENQSGNTEKH DGNEGSYQCL SNFPTSNPPM
101 RSYSNDMTQQ NTINQHHAAN MANLVTTSSH PINQHNLVYH STSPNTGDAP
151 NSTQNSPYAP SPCGGYEYPT SPLSSKPTIP PSTDYPGEWD FQINFGEATE
201 SAPKSAQYTY SPIINKLFVK MNVTCPIKFK CARPPPNGCV VRVMPVFKRP
251 EHVTDIVTRC PNHKIPDQAQ HIPHSQHLIR AEMPGENPAI YNVAMDGREN
301 VAVMFERPQI GAEYTTVLYK FMCLSSCVGG INRRPLNAVF NLENAEGQVL
351 GRRVVEVRVC SCPGRDRSQE EKRKRTAESE VKSGSKRTYK CMQGSMNIVQ
401 TAPAKRKCAM NGNDDQEEFT LKIRGRDKFE MLKKIKEALD LAEVAAHQQV
451 ENYRDEDART
```

_Ci_-p53/p73-a isoform 4

aniseed transcript Id: KH2012:KH.C1.573.v3.R.ND3-1

UniProtKB/Swiss-Prot: Q4H300

```
  1 MEQLMQQRIM GSYQCLSNFP TSNPPMRSYS NDMTQQNTIN QHHPANMANL
 51 VTTSSHPINQ HNLVYHSTSP NTGDAPNSTQ NSPYAPSPCG GYEYPTSPLS
101 SKPTIPPSTD YPGEWDFQIN FGEATESAPK SAQYTYSPII NKLFVKMNVT
151 CPIKFKCARP PPNGCVVRVM PVFKRPEHVT DIVTRCPNHK IPDQAQHIPH
201 SQHLIRAEMP GENPAIYNVA MDGRENVAVM FERPQIGAEY TTVLYKFMCL
251 SSCVGGINRR PLNAVFNLEN AEGQVLGRRV VEVRVCSCPG RDRSQEEKRK
301 RTAESEVKSG SKRTYKCMQG SMNIVQTAPA KRKCAMNGND DQEEFTLKIR
351 GRDKFEMLKK IKEALDLAEV AAHQQVENYR DEDARTFKVE PSSIHGKYPK
401 QIMTSDLLSQ VGYQE
```

7.1.2. *Ciona intestinalis* p53/p73-b

Single isoform aniseed transcript Id: KH2012:KH.C3.713.v1.A.SL1-1

UniProtKB/Swiss-Prot: Q4H2Z8

```
  1 MAVADTSELN FPDSQESFSD FWMNTLSENN ELPSWQTDLN QEYDQCKETV
 51 DVLQLDTTKA NDIEFPVSEF LTSSQASQQS IGDLFAQSLP STQCGGNSQV
101 TTVSKHEYPD SGYVMTLPNN SVLAELQTSS LQTEVSLLPN NEYPGIYNFE
151 INFGEKTESA PKSAPFTYSY SLQKLFVKMN ENCPIKFRCS PQPPSGCVIR
201 AIPVFEKPNN VTEIVTRCFN HRNECRTESS DSNTPNSHLI RVESKSNNIQ
251 YCLTHEGREC VVVPYEPPHS GSEYMALLYR FMCLSSCRTE TGINRRPLLT
301 IFNLESETGE LLGKRVVSTR ICACPGRDRT QEEEKKNVTS QNKSRKRLCK
351 SATNSKSIPV VQKNEENKVD NDDDGDVVYT LNIRGKRKFE KVKEYKEALD
401 LLDYVQPDVK KACCQRNQI
```

7.1.3. *Ciona savignyi* p53/p73-a isoforms

The numbering of the isoforms was inferred from the homologous gene from *Ciona intestinalis* (*Ci*-p53/p73-a).

Csa-p53/p73-a isoform 1

ensembl transcript Id: ENSCSAVT00000017123

UniProtKB/Swiss-Prot: H2ZH74

```
  1 MQRGGEGSYD PDCFSDIWSN LSQTEPIYAF DENNERLSNS ISIDLLSEFQ
 51 QFEPTTPGSD ANGSGNQAGN TENQGGNEGS YQCIGNFPSV NPTMRQYSTD
101 MTQQNSINQH QSGNMQNSVT SASHHINQHN LAYHSTSPNT GDAPNSTQNS
151 PFAPSPCGGY EYPTSPPIMS SKATLPPSTD YPGDWDFQIN FGEATESAPK
201 SAQYTYSPVL NKLFVKMNVT CPIKFKCSRP APNGCVVRVI PVFKRPEHVT
251 DLVTRCPNHK IPDEAQGMVN SRYLVRAEMP GENQALYGVG IDGRENVAVM
301 LERPQIGAEY TTVLYKFMCL SSCVGGINRR PLNVVFNLEN AEGQIIGRRV
351 VEVRVCSCPG RDRSQEEKRK RSESGSKEAT KRPYKSLQGT MNIMQTSAKR
401 KCMMAGGDDQ EEFTLKIRGR DKFEMLKKIK EALDIAEVAA QQQMETYREE
451 EPRQFKVESN LPSKYPKMMT TELLAPVGFH E
```

Csa-p53/p73-a isoform 2

Because sufficient gene models for the genome of *Ciona savignyi* are mostly missing, this isoform was not annotated in any database. I inferred the following protein sequence from homology to *Ci*-p53/p73-a isoform 2 based on a most likely incomplete transcript, which is coding for the extended C-terminus unique to this isoform, and the protein sequence of isoform 1. The protein sequence translated from the transcript matches isoform 1 aa 300 to 454, but is missing about 60% of DBD, hence would not be functional.

ensembl Id of incomplete: ENSCSAVT00000017122

UniProtKB/Swiss-Prot: H2ZH73 (fragment)

```
  1 MQRGGEGSYD PDCFSDIWSN LSQTEPIYAF DENNERLSNS ISIDLLSEFQ
 51 QFEPTTPGSD ANGSGNQAGN TENQGGNEGS YQCIGNFPSV NPTMRQYSTD
101 MTQQNSINQH QSGNMQNSVT SASHHINQHN LAYHSTSPNT GDAPNSTQNS
151 PFAPSPCGGY EYPTSPPIMS SKATLPPSTD YPGDWDFQIN FGEATESAPK
201 SAQYTYSPVL NKLFVKMNVT CPIKFKCSRP APNGCVVRVI PVFKRPEHVT
251 DLVTRCPNHK IPDEAQGMVN SRYLVRAEMP GENQALYGVG IDGRENVAVM
301 LERPQIGAEY TTVLYKFMCL SSCVGGINRR PLNVVFNLEN AEGQIIGRRV
351 VEVRVCSCPG RDRSQEEKRK RSESGSKEAT KRPYKSLQGT MNIMQTSAKR
401 KCMMAGGDDQ EEFTLKIRGR DKFEMLKKIK EALDIAEVAA QQQMETYREE
451 EPRQNLPSSS TTPPSDIKSS PNLQENPQIA STFTVSNPIT PDSQGFSLTP
501 THSSHPLTAT PNLDLPTGNS PYWSVDADMG CDVPDTNQPS TSAGLTQWAE
551 QPNTDCTPRL VRDLSGLSTT SSLSSSQMNS SGRFVTRVTL RQRVTLNQEP
601 TTDFWKPNAS EATNGVKVER KDDVESPPIT GEESFEFL
```

***Csa*-p53/p73-a isoform 4**

ensembl transcript Id: ENSCSAVT00000017124

UniProtKB/Swiss-Prot: H2ZH75

```
  1 MEQLMQTHIR GSYQCIGNFP SVNPTMRQYS TDMTQQNSIN QHQSGNMQNS
 51 VTSASHHINQ HNLAYHSTSP NTGDAPNSTQ NSPFAPSPCG GYEYPTSPPI
101 MSSKATLPPS TDYPGDWDFQ INFGEATESA PKSAQYTYSP VLNKLFVKMN
151 VTCPIKFKCS RPAPNGCVVR VIPVFKRPEH VTDLVTRCPN HKIPDEAQGM
201 VNSRYLVRAE MPGENQALYG VGIDGRENVA VMLERPQIGA EYTTVLYKFM
251 CLSSCVGGIN RRPLNVVFNL ENAEGQIIGR RVVEVRVCSC PGRDRSQEEK
301 RKRSESGSKE ATKRPYKSLQ GTMNIMQTSA KRKCMMAGGD DQEEFTLKIR
351 GRDKFEMLKK IKEALDIAEV AAQQQMETYR EEEPRQFKVE SNLPSKYPKM
401 MTTELLAPVG FHE
```

7.1.4. *Ciona savignyi* p53/p73-b

Single isoform ensembl transcript Id: ENSCSAVT00000016840

UniProtKB/Swiss-Prot: H2ZGE3

```
  1 MADSTEPNLP DSQESFSEFW CSSLQTNDFP NIVIDESALP ATSNWQTYTT
 51 MAPVCSLLDI GAQINHGDDT GLFDFNCSQH SADDLFDQSL PSTQASITCH
101 TSVGAQQECS ISQNTLLVQK IETPSIISIL PKFSEFPGVY NFQVDFGKKT
151 DGAPKSAPYT YSHRLHKLYV KMNENCPIRF TCSLSPPTGC IIRAIPVFER
201 ANNVTEMVTR CSLHSNENLS ENVAPKSHLI RVEANHHVTY TLTSDGRESV
251 CTSYEGPQAG CEYVAVLYKY MCLSSCSSAG GINRRPLLTV FNLEKKCGEL
301 LGRRVVSTRI CTCPGRDRTQ EEEKKGVLEN KKSSKRPLLP KKDNLNIDKG
351 ESEVYSLTII GKRKFEKVKE YKEALDLMDF VQPEVKKACC DKNQM
```

7.1.5. Preliminary protein sequence of *Botryllus schlosserie* p53/p73

The following protein sequence was predicted by spliced alignment using Spaln2 with *C.int.* p53/p73-a isoform 1 as reference. The underlying DNA sequence is chromosome 5: 27091149-27094658 (+); botctg 012291.

```
  1 MKNALRLMSD FTHYREIPSC GAAYRIETDD LLFAVNDEPA TIAERAAIND
 51 LISQSSQREI ASRYIVHRRI SLIGNTPLAN CPRGGFYPRT PTPHPDYAHR
101 THQPYVTYQP CHTEQQYQQT DHPAQRNIHI AEQTNYDAYI HSASPDSGDR
151 ASISNSPYTP TDFSVSPPSN PPLQNTKANS IPSMSDYAGD NNFEIYFGNH
201 TESAVKSAQY TYSSMLGKLF VKMNVLCPIQ FRCTQQPPAG CYIRALPVYT
251 RPEHVTEVVT RCPNHRIADN SNNDHLSRHL MRGEQPGNDS VQYLTAPDGR
301 ESVTITYVPP EVGSEYITVF YTYTCLSSCV GGINRRPIMT IFTLEAPDGK
351 ILARRCVEVR ICCCPGRDRA QEEKRRRKPG LVVLPKKKKM SSLPSTSTAE
401 PESGDGQEFF LRIKGKEKYE ILRKIKEALD LKELVTKEQK EEYEHRSATE
451 EESAMKQILS NYFKKGDGDG PGPFPT
```

7.1.6. Reference numbers of protein sequences from other invertebrate species as listet in the uniprot database

7.1.6.1. *Branchiostoma floridae* (Florida lancelet; amphioxus) protein sequences

p53	UniProtKB/Swiss-Prot:	C3YXH3
p63/p73	UniProtKB/Swiss-Prot:	C3XPU2

7.1.6.2. *Drosophila melanogaster* (fruit fly) protein sequence

p53	UniProtKB/Swiss-Prot:	Q9N6D8

7.1.6.3. *Caenorhabditis elegans* (roundworm) protein sequence

cep-1	UniProtKB/Swiss-Prot:	Q20646	(p63/p73-like with SAM domain)

7.1.7. Reference numbers of protein sequences from vertebrate species as listet in the uniprot database

7.1.7.1. *Callorhynchus milii* (elephant shark) protein sequences

p53	UniProtKB/Swiss-Prot:	G9J1L8	
p63	UniProtKB/Swiss-Prot:	G9J1L9	
p73	UniProtKB/Swiss-Prot:	G9J1M0	

7.1.7.2. *Danio rerio* (zebrafish) protein sequences

p53	UniProtKB/Swiss-Prot:	P79734	
p63	UniProtKB/Swiss-Prot:	A7YYJ7	
p73	UniProtKB/Swiss-Prot:	B0S576	

7.1.7.3. *Xenopus tropicalis* (western clawed frog) protein sequences

p53	UniProtKB/Swiss-Prot:	F7A9U0	
p63	UniProtKB/Swiss-Prot:	F6ZGN7	
p73	UniProtKB/Swiss-Prot:	F6TKT0	

7.1.7.4. *Mus musculus (mouse) protein sequences*

p53	UniProtKB/Swiss-Prot:	P02340	
p63	UniProtKB/Swiss-Prot:	O88898-1	isoform TA p63 alpha
p73	UniProtKB/Swiss-Prot:	Q9JJP2-1	isoform TA p73 alpha

7.1.7.5. *Homo sapiens* (human) protein sequences

p53	UniProtKB/Swiss-Prot:	P04637-1	
p63	UniProtKB/Swiss-Prot:	Q9H3D4-1	isoform TA p63 alpha
p73	UniProtKB/Swiss-Prot:	O15350-1	isoform TA p73 alpha

7.1.8. Alignments

7.1.8.1. Percentage amino acid identity within DBD between p53 family proteins of vertebrate and Ciona species

Table 7-1 Pairwise percentage amino acid identity within DNA binding domain of p53 family proteins from different deuterostome species.

DBD % amino acid identity	human p53	mouse p53	frog p53	zebrafaish p53	e. shark p53	human p63	mouse p63	frog p63	zebrafaish p63	e. shark p63	human p73	mouse p73	frog p73	zebrafaish p73	e. shark p73	lancelet p53	lancelet p63/p73	*Ciona int.* p53/p73-a	*Ciona sav.* p53/p73-a	*Ciona int.* p53/p73-b	*Ciona sav.* p53/p73-b
human p53 aa 102-292		88	67	71	67	56	56	48	58	59	60	59	59	58	60	34	49	40	38	41	40
mouse p53 aa 96-286			68	73	66	58	58	48	59	61	59	59	60	59	60	37	52	42	40	43	42
frog p53 aa 76-267				73	73	65	65	53	65	65	61	61	61	60	62	38	56	44	43	43	43
zebrafish p53 aa 70-260					69	64	64	54	63	64	61	62	63	64	64	37	54	46	45	42	42
e. shark p53 aa 92-283						69	69	58	69	67	68	68	69	67	69	38	55	44	45	46	44
human p63 aa 170-362							=	80	96	95	86	85	84	84	85	46	64	50	48	45	46
mouse p63 aa 170-362								80	96	95	86	85	84	84	85	46	64	50	48	45	46
frog p63 aa 76-273									78	78	73	72	71	72	73	39	52	42	44	39	39
zebrafish p63 aa 74-266										94	84	84	84	83	85	44	63	48	46	46	46
e. shark p63 aa 193-385											85	84	85	83	87	45	64	49	49	46	46
human p73 aa 120-312												98	93	91	93	45	61	50	48	45	45
mouse p73 aa 112-304													92	91	93	45	62	51	49	47	45
frog p73 aa 120-312														93	93	44	62	51	48	46	46
zebrafish p73 aa 127-319															92	44	62	50	46	47	45
e. shark p73 aa 125-317																44	62	51	48	47	44
lancelet p53 aa 99-284																	43	38	36	34	37
lancelet p63/p73 aa 169-361																		52	51	48	47
***C. int.* p53/p73-a** aa 184-375																			88	55	53
***C. sav.* p53/p73-a** aa 180-371																				54	52
***C. int.* p53/p73-b** aa 142-337																					65
***C. sav.* p53/p73-b** aa 135-326																					

Protein sequences belonging to the p53 family of transcription factors from the tunicate species *Ciona intestinalis* (*C.int.*) and *Ciona savignyi* (*C. sav.*) were aligned against those of the vertebrate species *Homo sapiens* (human), *Mus musculus* (mouse), *Xenopus tropicalis* (frog), *Danio rerio* (zebrafish) and *Callorhynchus milii* (e. shark), and the invertebrate species *Branchiostoma floridae* (lancelet). The full length protein sequences were aligned using clustalW2. The sections corresponding to the DNA binding domain (DBD) were identified as those spanned by the secondary structure elements in the NMR solution structure of the DBD of human p63 (pdb 2RMN). Percent amino acid identity was determined for pairwise alignments using Jalview and rounded. The accession numbers are listed in the appendix in section 7.1 and the amino acid (aa) numberings correspond to the respective canonical isoforms.

7.1.8.2. Percentage amino acid identity within TD between p53 family proteins of vertebrate and Ciona species

Table 7-2 Pairwise percentage amino acid identity within tetramerization domain of p53 family proteins from different deuterostome species.

TD % amino acid identity	human p53	mouse p53	frog p53	zebrafaish p53	e. shark p53	human p63	mouse p63	frog p63	zebrafaish p63	e. shark p63	human p73	mouse p73	frog p73	zebrafaish p73	e. shark p73	lancelet p53	lancelet p63/p73	*Ciona int.* p53/p73-a	*Ciona sav.* p53/p73-a	*Ciona int.* p53/p73-b	*Ciona sav.* p53/p73-b
human p53 aa 324-367		84	41	43	36	32	32	30	32	30	34	30	30	23	32	30	27	41	39	30	27
mouse p53 aa 318-361			38	39	34	32	32	30	30	27	32	30	30	23	30	30	25	43	41	30	30
frog p53 aa 302-338				43	38	32	32	35	41	32	35	27	30	27	30	27	19	38	38	24	22
zebrafish p53 aa 300-343					57	48	48	45	50	50	52	48	48	45	50	30	41	41	39	25	23
e. shark p53 aa 313-356						43	43	45	48	45	48	45	48	39	48	39	39	36	34	30	27
human p63 aa 397-440							=	89	71	86	59	57	61	55	66	36	59	41	43	27	23
mouse p63 aa 397-440								89	71	86	59	57	61	55	66	36	59	41	43	27	23
frog p63 aa 304-347									82	82	59	57	61	57	66	32	59	39	39	25	27
zebrafish p63 aa 298-341										75	57	57	59	57	68	36	52	43	43	25	30
e. shark p63 aa 421-464											64	64	68	61	73	39	55	39	39	23	25
human p73 aa 351-394												89	77	75	82	30	57	39	36	30	27
mouse p73 aa 343-386													80	75	84	34	55	39	36	25	23
frog p73 aa 351-394														86	86	32	57	43	41	27	25
zebrafish p73 aa 358-401															80	25	52	36	34	23	25
e. shark p73 aa 356-399																32	59	43	43	25	25
lancelet p53 aa 321-364																	39	34	34	25	25
lancelet p63/p73 aa 399-442																		39	32	25	23
***C. int.* p53/p73-a** aa 415-458																			84	34	34
***C. sav.* p53/p73-a** aa 409-452																				32	30
***C. int.* p53/p73-b** aa 375-418																					75
***C. sav.* p53/p73-b** aa 351-394																					

Protein sequences belonging to the p53 family of transcription factors from the tunicate species *Ciona intestinalis* (*C.int.*) and *Ciona savignyi* (*C. sav.*) were aligned against those of the vertebrate species *Homo sapiens* (human), *Mus musculus* (mouse), *Xenopus tropicalis* (frog), *Danio rerio* (zebrafish) and *Callorhynchus milii* (e. shark), and the invertebrate species *Branchiostoma floridae* (lancelet). The full length protein sequences were aligned using clustalW2. The sections corresponding to the tetramerization domain (TD) were identified as those spanned by the secondary structure elements in the NMR solution structure of the TD of human p73 (pdb 2KBY). Percent amino acid identity was determined for pairwise alignments using Jalview and rounded. The accession numbers are listed in the appendix in section 7.1 and the amino acid (aa) numberings correspond to the respective canonical isoforms.

7.1.8.3. Percentage amino acid identity within SAM domain between vertebrate p63 and p73 proteins

Table 7-3 Pairwise percentage amino acid identity within SAM domain of p63 and p73 proteins from different vertebrate species.

SAM % amino acid identity	**human p63**	**mouse p63**	**frog p63**	**zebrafaish p63**	**e. shark p63**	**human p73**	**mouse p73**	**frog p73**	**zebrafaish p73**	**e. shark p73**
human p63 aa 549-607	=	100	83	68	78	51	46	47	54	53
mouse p63 aa 549-607		=	83	68	78	51	46	47	54	53
frog p63 aa 461-519			=	71	71	53	51	46	54	54
zebrafish p63 aa 455-519				=	64	46	42	49	47	46
e. shark p63 aa 564-622					=	49	46	53	53	49
human p73 aa 493-551						=	93	81	81	83
mouse p73 aa 487-545							=	78	75	76
frog p73 aa 493-551								=	73	78
zebrafish p73 aa 499-557									=	80
e. shark p73 aa 503-561										=

The sequences of the p63 and p73 proteins from the vertebrate species *Homo sapiens* (human), *Mus musculus* (mouse), *Xenopus tropicalis* (frog), *Danio rerio* (zebrafish) and *Callorhynchus milii* (e. shark) were aligned using clustalW2. The sections corresponding to the sterile alpha motif (SAM) were identified as those spanned by the secondary structure elements in the NMR solution structure of the SAM domain of human p63 (pdb 1RG6). Percent amino acid identity was determined for pairwise alignments using Jalview and rounded. The accession numbers are listed in the appendix in section 7.1 and the amino acid (aa) numberings correspond to the respective TA alpha isoforms.

7.2. Prediction of secondary structure propensity by SSP and ncSPC

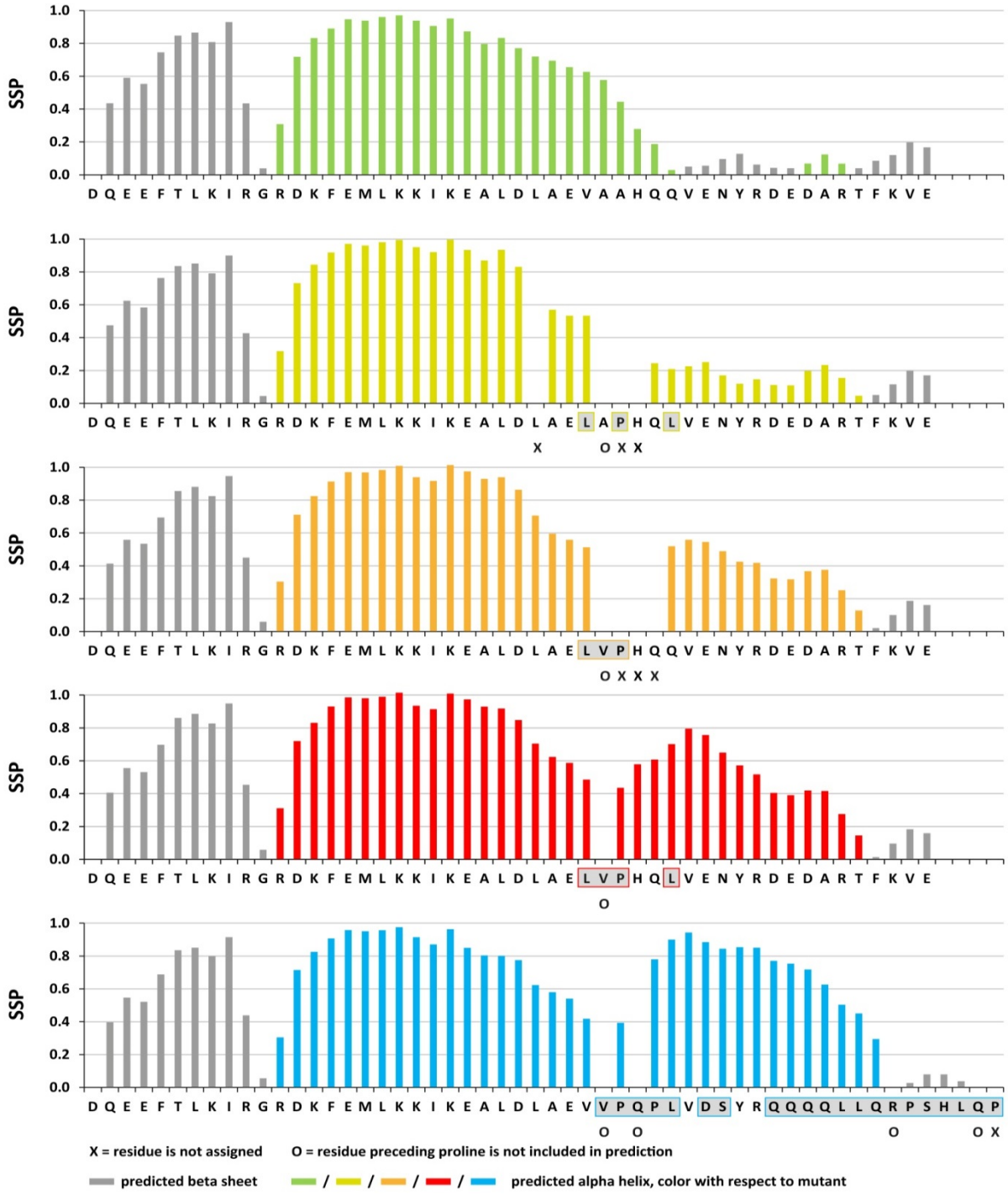


Figure 7-1 Propensities of adaptation of β-strand and α-helical conformation as calculated with the program SSP for several partially mutated constructs encompassing aa 415-464 of *C.int.* p53/p73-a isoform 1 and in comparison those for the chimeric protein in which aa 415-444 from *C.int.* p53/p73-a were fused to aa 381-404 from human TAp73α. The probability of occupation of secondary structure (secondary structure propensity) was calculated using the program SSP as described in 3.3.7 with the chemical shifts of CA, CB, HN and N as input. The chemical shift values were derived from spectra recorded at 25°C. Residues preceding prolines were not included in the prediction and are labeled with O. Those residues with incomplete assignment are labeled with an X. Each bar diagram shows the propensities of adaptation of β-strand or α-helical conformation determined for the respective protein sequence, written below. Bars depicted in dark grey represent the propensity of the respective residue to adopt β-strand conformation. The bars representing the propensity towards α-helical conformation have the color, which was assigned to the respective construct. Residues, which differ from the *C.int.* wild type sequence, are underlined in light grey and boxed in the color assigned to the respective mutant.

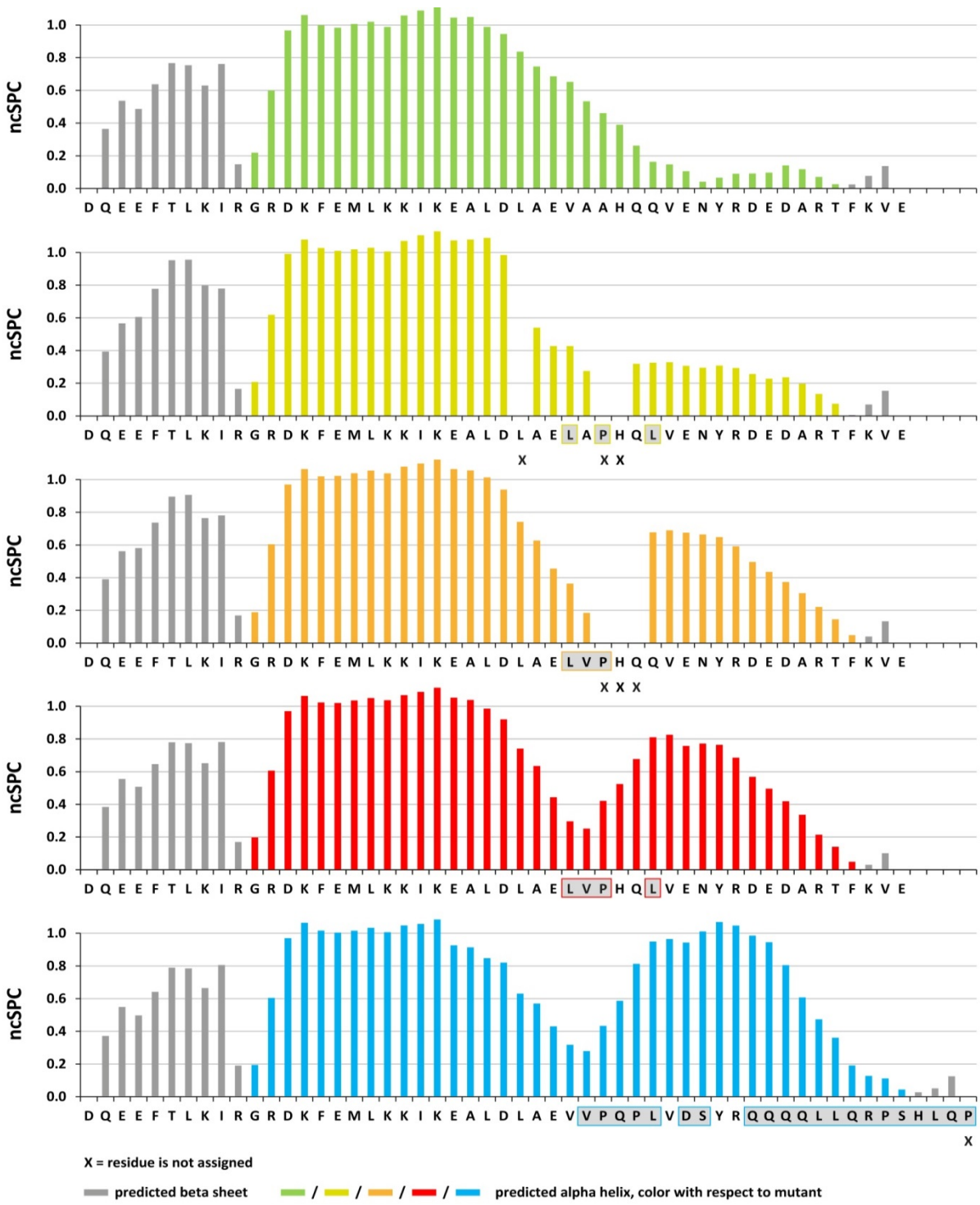


Figure 7-2 Propensities of adaptation of β-strand and α-helical conformation as calculated with the program ncSCP for several partially mutated constructs encompassing aa 415-464 of *C.int.* p53/p73-a isoform 1 and in comparison those for the chimeric protein in which aa 415-444 from *C.int.* p53/p73-a were fused to aa 381-404 from human TAp73α. The probability of occupation of secondary structure (secondary structure propensity) was calculated using the program SSP as described in 3.3.7 with the chemical shifts of CA, CB, HN and N as input. The chemical shift values were derived from spectra recorded at 25°C. Residues with incomplete assignment are labeled with an X. Each bar diagram shows the propensities of adaptation of β-strand or α-helical conformation determined for the respective protein sequence, written below. Bars depicted in dark grey represent the propensity of the respective residue to adopt β-strand conformation. The bars representing the propensity towards α-helical conformation have the color, which was assigned to the respective construct. Residues, which differ from the *C.int.* wild type sequence, are underlined in light grey and boxed in the color assigned to the respective mutant.

Acknowledgements

First of all, I would like to thank Volker Dötsch for the supervision and the support during my PhD. I am very grateful that I had the opportunity to do my PhD. in such an uncomplicated, inspiring and challenging atmosphere. Thank you for always having an open door for new ideas, for helpful discussions and for initiating great new collaborations.

A big thanks goes to Frank Löhr. I am very grateful for his explanations, his patience and all the NMR experiments he recorded for me. This thesis greatly relied on his expertise in getting the best possible spectra recorded even from samples he had been routinely not too optimistic about.

The calculation of the NMR solution structure presented in this thesis was performed by Henry Jonker whom I want to thank for his reliable work and this very productive cooperation.

Beside my main topic I had the pleasure to participate in a cooperation project initiated by Krishnaraj Rajalingam. Gergely Imre had been the responsible post doc in his lab and the discussions with both of them have been very inspiring. I would like to thank them for this fruitful cooperation.

I would like to thank Sigrid Oğuzer-Fachinger, Manfred Stumpf and Birgit Schäfer how are steadily keeping everything well administered, running and organized, very often silently in the background. Without them things would be way more stressful and one could less concentrate on science.

I would like to thank my present and former “labmates” Tobias Weber, Robert Hänsel, Alena Busche, Elisabeth Zielonka, Christopher Hein, Sebastian Kehrlößer, Davide Proverbio, Solmaz Sobhanifar and Sebastian Richers for the great working atmosphere in our lab coined by good discussions, constant laughter and hilarious jokes.

Several other present and former colleagues likewise contributed to the good working atmosphere in our work group, which was characterized by a great attitude to help each other. Therefore special regards also go to Alexis Rozenknop, Birgit Schwarz, Christian Roos, Susanne Stefer, Stefan Haberstock, Sina Reckel, Laura Luh, Gregor Deutsch, Florian Durst, Wesley McGinn Straub, Aisha Laguerre, Christian Osterburg, Jakob Gebel, Daniel Schwarz, Peter Tufar as well as Vladimir Rogov and Natasha Rogova.

I would like to thank Florian Buhr for being a great friend and for equally sharing struggle as well as fun with me. Thank you for late night discussions on our projects and other concerns.

I would like to thank my love Stefanie Witt for her understanding, her patience, her continuous support and for simply having the right sense of humor to cheer me up at any time no matter what struggle I have to go through.

Mein Dank gilt des Weiteren meinen lieben Eltern für Ihre moralische und finanzielle Unterstützung, welche mir ein sorgenfreies Studium ermöglicht hat. Meiner Mutter danke ich dafür, dass Sie auch nach dem Tod meines Vaters immer die starke, positive und einfach großartige Frau geblieben ist, die sie immer schon war. Sie und meine liebe Schwester Inge haben mir immer den Rücken gestärkt und mir bei Rückschlägen Mut gemacht.

Personal information

Name:	Jan Peter Heering
Date of birth:	30. Oktober 1980
Place of birth:	Eckernförde, Schleswig-Holstein, Germany
Nationality:	german

Natural scientific education

10/2004 – 12/2006 Master course Molecular Life Science, University of Lübeck, Germany,
Master of Science (M.Sc.)
Master thesis at the Institute of Chemistry, University of Lübeck
(Supervisors: Prof. Dr. Thomas Peters and PD Dr. Thomas Weimar)
"Characterization of immunodominant epitopes in pemphigoid gestationis"

10/2001 – 07/2004 Bachelor course Molecular Biotechnology, University of Lübeck, Germany
Bachelor of Science (B.Sc.)
Bachelor thesis at the Institute of Molecular Biology, University of Lübeck
(Supervisors: Prof. Dr. Verena Gauss-Müller and Dr. Yuri Kusov)
"RNA interference mediated suppression of the expression of replicons from picorna viruses and cloning of shRNA plasmids for the expression of hepatitis-A virus siRNAs."

Publications

2014 **J. Heering**, H. Jonker, F. Loehr, H. Schwalbe, V. Doetsch: NMR solution structure of the tetramerization domain of Ciona intestinalis p53/p73-b. The NMR solution structure was submitted to the pdb and will be accessible upon release under the pdb entry code 2MW4**.**

2012 G. Imre, **J. Heering**, A.-N. Takeda, M. Husmann, B. Thiede, D. M. zu Heringdorf, D. R. Green, F. G. van der Goot, B. Sinha, V. Dötsch, and K. Rajalingam, "Caspase-2 is an initiator caspase responsible for pore-forming toxin-mediated apoptosis.," EMBO J., vol. 31, no. 11, pp. 2615–28, May 2012.

2011 S. Hahn, J. Achenbach, E. Buscató, F.-M. Klingler, M. Schroeder, K. Meirer, M. Hieke, **J. Heering**, E. Barbosa-Sicard, F. Loehr, I. Fleming, V. Doetsch, M. Schubert-Zsilavecz, D. Steinhilber, and E. Proschak, "Complementary screening techniques yielded fragments that inhibit the phosphatase activity of soluble epoxide hydrolase.," ChemMedChem, vol. 6, no. 12, pp. 2146–9, Dec. 2011.

2011 P. N. Kantaputra, S. Malaivijitnond, A. R. Vieira, **J. Heering**, V. Dötsch, T. Khankasikum, and W. Sripathomsawat, "Mutation in SAM domain of TP63 is associated with nonsyndromic cleft lip and palate and cleft palate.," Am. J. Med. Genet. A, vol. 155A, no. 6, pp. 1432–6, Jul. 2011.

2011 W. Sripathomsawat, P. Tanpaiboon, **J. Heering**, V. Dötsch, R. C. M. Hennekam, and P. Kantaputra, "Phenotypic analysis of Arg227 mutations of TP63 with emphasis on dental phenotype and micturition difficulties in EEC syndrome.," Am. J. Med. Genet. A, vol. 155A, no. 1, pp. 228–32, Jan. 2011.

Academic teachers

Goethe University Frankfurt: Prof. Dr. Volker Dötsch, Dr. Frank Loehr, Dr. Henry Jonker, Prof. Dr. Krishnaraj Rajalingam, Prof. Dr. Martin Zörnig

University of Lübeck and Research Center Borstel:
Dr. Thorsten Biet, PD Dr. Andreas Frey, PD Dr. med. Christoph Frohn, Prof. Dr. Verena Gauss-Müller, Prof. Dr. med. Siegfried Görg, Prof. Dr. Thomas Gutsmann, Prof. Dr. Enno Hartmann, Prof. Dr. Rolf Hilgenfeld, Prof. Dr. Christoph Hölscher, Prof. Dr. med. Wolfgang Jelkmann, Prof. Dr. Karsten Keller, Prof. Dr. med. Holger Kirchner, Dr. Rosel Kretschmer-Kazemi Far, PD Dr. Andrea Kruse, PD Dr. Bärbel Kunze, Dr. Yuri Kusov, Prof. Dr. Thomas Martinez, Prof. Dr. Berthold Matzanke-Markstein, Prof. Dr. Peter Müller, Prof. Dr. Holger Notbohm,Dr. Hannelore Peters, Prof. Thomas Peters, Prof. Dr. Jürgen Prestin, Prof. Dr. Tobias Restle, Prof. Dr. Georg Sczakiel, Prof. Dr. Ulrich Seydel, Dr. Christian Schmidt, Prof. Dr. Werner Solbach, Prof. Dr. Volker Schünemann (Prof. in Kaiserslautern, had been PD in Lübeck), Prof. Dr. Eberhard Schwinger, Prof. Dr. Alfred Trautwein, PD Dr. Thomas Weimar, Prof. Dr. Jürgen Westermann, Prof. Dr. Andreas Ziegler

School and military service

09/2000 – 07/2001	duty in the german armed forces
06/2000	university-entrance diploma
08/1991 – 06/2000	secondary school (Jungmannschule, Eckernförde)
08/1987 – 07/1991	primary school (Albert-Schweitzer-Schule, Eckernförde)