

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Structural characterization of the C-terminal domains in the p53 protein family

Permalink

<https://escholarship.org/uc/item/8818g5zt>

Author

Ou, Horng Der

Publication Date

2007-10-01

Peer reviewed|Thesis/dissertation

Structural Characterization of the C-terminal Domains in the p53 Protein Family

by

Hong Der Ou

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biophysics Graduate Group

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Acknowledgement

The completion of this dissertation is not possible without the following people who have guided my thinking, lent a helping hand to my work, supported me during the low point, and enriched my life experience through these years.

The most influential person in my graduate school is Volker Dötsch, my thesis advisor and a great mentor. True to his words in my first year that I should see the world, he brought me to Frankfurt, Germany in 2003. It was a great experience that I would never forget. Of course, Volker's generous support, boundless patience, contagious enthusiasm and optimism, and encouragement for free thinking have developed my confidence and matured my independent thinking. Holly Ingraham and Dan Minor, my thesis committee advisors, were invaluable in correcting my ideal thinking and steered me toward graduation in our annual thesis meeting.

I am also indebted to two amazing people who have helped me in paper works and in settling into a new place. Julie Ransom, the biophysics program coordinator, has helped me in many occasions whether I was in San Francisco or in Germany. In Germany, Siggi Fachinger has helped me in apartment searching and smoothed out my transition to Germany.

I would also like to thank Frank Löhr and Vitali Vogel in collecting data for NMR spectra and analytical ultracentrifugation respectively. Without them, I could not complete a major portion of my dissertation.

Felician Dancea has answered many of my questions regarding structure calculations and general computer knowledge. I could always count on his prompt reply. He was the best interim network administrator in the lab, even though it was above his

duty besides being a graduate student.

Birgit Schaffer is the foundation of our lab in Germany. She made sure everything was running smoothly so all of us could do our experiments. Her articulate German has allowed me to practice my German with her.

My first three years of graduate school were filled with good memories of our small lab on the Parnassus Ave in room S1234. All the good and bad times were shared among Zach Serber, Alex Kelly, Helen Lai, and me when we were trying to establish some orders to the lab. Later Michael Reese has joined our lab to contribute some more chaos, but everything seemed to work out at the end in the midst of all disorders.

Our two adjunct lab members, Anson Nomura and Nobuhisa Shimba from Charles Craik's lab, have brought in their intelligence and humor to our group meeting and sparked stimulating discussions. In addition, they are also great poker players.

Outside of the lab, I have also enjoyed immensely my brainstorm lunch sessions with Alex Fay and Helen to discuss anything on our mind from each other's research ideas and problems to gossips. I thank them for their patience in listening to some of my naïve and crazy ideas.

Our understanding of p63 was greatly advanced in Germany by Wesley McGinn-Straub and Florian Durst who have further characterized the protein and shared the vision that a full length structure is a possibility. Tobias Weber is invaluable in testing many of my incorrect hypotheses on p63 in cell culture assays.

My life in grad school could not be sustained if there were no badminton and running to release my stress and tension. In the US, I was very lucky to have a group of players consists of Chao Zhang, Agnes Wong, Darice Chan, and Wei-Wei for our

Saturday ritual: two hours of badminton and a dinner afterward. In Germany I was able to find Friederike Junge and Stephe Körber who would play badminton every Saturday.

I want to thank Vladimir and Natasha Rogov for their generosity in inviting me to their family Christmas dinner, and their encouragement and inspiration for me to run the Frankfurt marathon and taking the ballroom dancing course. Our 4 hours training runs for the marathon were grueling but fun in our exploration around the Taunus. It was sad I had to leave Germany when I found my recent fellow runners Friederike in long distance and Daniel Coutandin in short distance.

Stephe has also enriched my experience in Germany by introducing me to clubbing, which I have never done in the US. It was fun to dance the night away till 5 a.m. and wait for the first U-Bahn to get home. The trip to the University of Erlangen was also a memorable one.

Mehdi Talebzadeh and Solmaz Sobhanifar were my late night and weekend friends for our teatime and dinner gatherings. It was fun to talk about any topics from wine to feet and broaden my perspective on world politics and new fictional authors, and so it goes. To Solmaz, I am especially grateful for her careful reading of my dissertation.

Whenever I drink of Kölsch, I will always remember Birgit Schnider from Köln who believes it is the best beer on the Rhine.

Yes, Daniel Schwarz, I will not forget the moment when Germany won the penalty shootout against Argentina in the 2006 World Cup in Germany.

Meichen Shi has helped me a great deal during my transition to Germany, and helped me searching for an apartment. I really enjoyed all the coffee time we had and our

discussion on what her childhood and life was like in China.

My lunch partner and lab-bay mate, Kerstin Schmöe, has made my last year of stay in Germany a very enjoyable experience. I enjoyed greatly my daily lunch and discussion with her. I thank her for her tolerance to my messiness.

Daniel Basting and Ines Lehner were two friends who always hosted great parties and enjoyed their works despite of the difficult topics in their research.

Lastly, my seemingly never-ending pursuit for my degree could not have started if my parents had not made the decision to emigrate from Taiwan and allowed me to choose my own career. Whenever I need something fun to do, I can always count on my brother, Bin, to keep me entertained. My grandma, my aunts and uncles (number 4, 5, 6, 7, 9, 10), and all my cousins, have also supported me throughout all these years with their care and all the great parties they had hosted to celebrate the love within the family and recharged me for my next endeavor.

Structural characterization of the C-terminal domains in the p53 protein family

Horng Der Ou

The identification of many homologs and paralogs of the most famous tumor suppressor, p53, has expanded its role from tumor suppression to epidermal development, neuronal development, and protection of germ cells. Of all the members in the p53 protein family, the DNA binding domain (DBD) is the most conserved domain. In contrast, the C-terminal region is diverse in sequence and composition of protein domains that include the oligomerization domain (OD), the sterile alpha motif (SAM) domain, and the transcription inhibitory domain (TID). While the function of each domain has been delineated, the SAM domain is the least functionally characterized, yet it is conserved between vertebrate and invertebrate p53. Furthermore, mutations of this domain in p63, a homolog of p53, cause defects in epidermal development in human.

The diversity of the C-terminus is most apparent in two p53 like proteins in *C. elegans* (CEP-1) and *Drosophila* (Dmp53). Neither of these proteins have any domains in the C-terminus found in other p53 protein family members, yet the DBD in both proteins recognize the DNA consensus motif *in vitro*. Interestingly, CEP-1 and Dmp53 could only elicit an apoptotic response, but not both cell cycle arrest and apoptosis like in human. The variation of the C-terminal end by each member of the p53 protein family may account for the discrepancy between identical *in vitro* DNA specificity, and distinct promoter specificity *in vivo*.

By using bioinformatics and structural determination by nuclear magnetic resonance, the domain architecture of the C-termini of CEP-1 and Dmp53 was revealed.

In CEP-1, an OD and a SAM domain were identified, in which the stability of the OD depends on its interaction with the SAM domain, thus suggesting an early function for the SAM domain. In Dmp53, the OD displays an unconventional fold that requires an additional helix to stabilize the OD. Structural and biochemical investigations into the human SAM domain in p63 also reveal that the SAM domain has interactions with the OD. Mutations in the SAM domain may disrupt this interaction and cause a change in the conformation of p63 that results in its abnormal function.

Table of Contents

| | |
|--|-----------|
| Chapter 1 | 1 |
| An introduction of the p53 protein family in vertebrates and invertebrates | |
| Introduction | 2-8 |
| References | 9-13 |
| Figures | 14-15 |
| Chapter 2 | 16 |
| Structural evolution of C-terminal domains in the p53 family | |
| Abstract | 17 |
| Introduction | 18-21 |
| Results | 22-31 |
| Discussion | 32-36 |
| Materials & methods | 37-40 |
| References | 41-44 |
| Figures & tables | 45-67 |
| Chapter 3 | 68 |
| Hay-Wells syndrome mutations alter the tertiary structure of the oligomerization domain, but not the SAM domain | |
| Abstract | 69 |
| Introduction | 70-72 |
| Results | 73-78 |
| Discussion | 79-82 |
| Materials & methods | 83-85 |

Table of Contents

| | |
|--|---------|
| References | 86-88 |
| Figures | 89-104 |
| Chapter 4 | 105 |
| Conclusions and Future Directions | |
| Conclusions and Future Directions | 106-109 |
| References | 110 |
| Figures | 111-118 |

List of Tables

| | | |
|-----------|---|----|
| Table 2-1 | Velocity sedimentation of the C-terminal domain of CEP-1 and Dmp53 | 65 |
| Table 2-2 | Structural Statistics of CEP-1 | 66 |
| Table 2-3 | Structural Statistics of Dmp53 | 67 |

List of Figures

| | | |
|-------------|---|-------|
| Figure 1-1 | Domain layout of the p53 protein family | 15 |
| Figure 2-1 | Domain architecture of the p53 protein family | 46 |
| Figure 2-2 | Gel filtration curves of the C-terminus of CEP-1 and Dmp53 | 48 |
| Figure 2-3 | Sequence divergence of the OD in CEP-1 and Dmp53 relative to other p53 protein family members | 50 |
| Figure 2-4 | The tetrameric interface in the oligomerization domain of human p53, CEP-1 and Dmp53 | 52 |
| Figure 2-5 | Oligomerization domain of CEP-1 and Dmp53 are stabilized by additional structural elements | 54 |
| Figure 2-6 | Different hypotheses on the evolution of the p53 protein family | 56 |
| Figure 2-7 | Sequence alignment of CEP-1 in <i>C. elegans</i> and <i>C. briggsae</i> | 58 |
| Figure 2-8 | CD spectra of the C-terminus of CEP-1 in full length and the OD alone | 60 |
| Figure 2-9 | Structural alignment of individual domain in CEP-1 and Dmp53 | 62 |
| Figure 2-10 | CD spectra of the Dmp53 and different deletion mutants | 64 |
| Figure 3-1 | A schematic diagram of mutations found in AEC and RHS | 90 |
| Figure 3-2 | Structural alignment of the p63SAM domain vs. p73SAM domain vs. L514F mutant | 92 |
| Figure 3-3 | Thermal denaturation curve of the wild type SAM domain and mutants | 94 |
| Figure 3-4 | Chemical shift deviations between the wild type SAM domain and mutants | 96-98 |

List of Figures

| | | |
|------------|--|-----|
| Figure 3-5 | The L514F mutation causes rearrangement of residues in helix 3 and helix 5 | 100 |
| Figure 3-6 | Gel filtration curve of the MBP fusion protein of the wild type SAM domain and mutants | 102 |
| Figure 3-7 | Velocity sedimentation of the MBP fusion protein of the wild type SAM domain and mutants | 104 |
| Figure 4-1 | The electron microscopy image of the TAp63 α with negative staining | 112 |
| Figure 4-2 | Crystal of TAp63 α and the first X-ray diffraction pattern | 114 |
| Figure 4-3 | The OD of TAp63 γ has an additional helix C-terminal to the OD | 116 |
| Figure 4-4 | SDS-PAGE of OD-SAM and MBP-OD-SAM purified proteins | 118 |

Chapter 1

An introduction of the p53 protein family in vertebrates and invertebrates

Ever since its discovery in 1979 in cells infected by the SV40 virus (1, 2), p53 has fascinated generations of scientists into the study of its cellular, biochemical, and biophysical properties. It took nearly 10 years to elucidate the role of p53 in cancer biology: from a 53 kilo-daltons (kD) protein thought to function as an oncogene due to its association with the large T antigen of the simian virus, to a tumor suppressor in which the protein inhibits tumor growth by arresting cells in the G1/S state of the cell cycle (3) or as an inducer apoptosis (4). The role of p53 as a guardian of the genome was further affirmed when sequencing data showed that 50% of tumors had mutations in p53, thus implying its role as a tumor suppressor. The finding that p53 is not essential for development is demonstrated by p53 homozygous knock-out mice, which have normal development but are susceptible to tumorigenesis within 6 months after birth (5). Thus, it provides a possibility that extraneous intervention of abnormal p53 activity or protein level could act as a therapeutic tool in cancer without disrupting normal development.

Biochemical, cell culture, and structural studies have delineated p53 as a multi-domain protein consisting of a transcriptional activation domain (TA) at the N-terminus, a DNA binding domain (DBD), an oligomerization domain (OD), and a short fragment of 30 amino acids that consists mainly of serine and lysine at the C-terminus (Figure 1-1). The oligomerization domain forms a tetramer in isolation, and is composed of two dimeric subunits (6). It has been shown that p53 co-translationally formed a dimer, which then assembled into a tetramer (7). DNA recognition is carried out by the DBD (8), in which two DBDs recognize the RRRCWWGYYY DNA motif (R is purine, Y is pyrimidine, and W is either A or T), and four DBDs recognize two pairs of motif that could have 0-14 base pairs between them (9). However, Chip-on-Chip experiments

showed most p53 motifs have a gap of 0-1 base pairs (10). The TA domain interacts with the core transcriptional machinery factors like the TATA binding protein and p62/Tfb1, as well as a p53 regulator, MDM2, a E3 ubiquitin ligase (11-13). The C-terminal 30 amino acids fragment interacts non-specifically with DNA and is also a site of multiple phosphorylations, acetylations, and methylations that regulate the activity of p53 (14).

Two mechanisms that contribute to the tumor suppressor function of p53 are its ability to activate transcription of genes that elicit cell cycle arrest or apoptosis, thus maintaining genomic stability of cells and preventing proliferation of abnormal cells. Some known p53 downstream target genes such as p21, GADD45, and 14-3-3- σ play a role in cell cycle arrest, while BAX, PUMA, PERP, and NOXA can activate the apoptotic pathways (15). p53 can also activate apoptosis independent of transcription by translocating from the nucleus to the mitochondria and facilitating cytochrome C release (16). Although the network of proteins that associate with each mechanism has been identified, the question remains how p53 determines which set of genes to transcribe upon damage to the cellular genome, since each set of genes has similar DNA binding motifs (15, 17). Mutations in p53 have been found that specifically activate the apoptotic pathways, but not cell cycle arrest (18, 19). Post-translational modifications of p53 also contribute as determinants in p53 decision (20). The ASPP protein family has also been identified, which may stimulate p53 to initiate the apoptotic pathway, but not cell cycle arrest (21). Nevertheless, understanding modifications of p53 that differentiate transcription of cell cycle arresting genes and apoptotic genes still remains an area of active research.

In 1997 and 1999, two paralogs of p53 were discovered, called p73 and p63,

which have a divergent function from p53 (22, 23). Knock-out mice studies showed p63 is involved in epidermal and limb development, and p73 is involved in neuronal development (24-26). Unlike p53, which is not an essential developmental gene, both p63 and p73 homozygous knock-out mice result in a lethal phenotype, further distinguishing their roles from a tumor suppressor. In addition, p63 has six different splice variants created from a combination of two different N-termini (the presence and the absence of a TA domain) and three different C-termini (α , β , γ). In p73, a combination of four different N-termini and seven different C-termini yields a total of 28 isoforms (Figure 1-1) (27). The C-terminus of both p63 and p73 contains three additional domains, QP, SAM and TID. The QP domain consists mostly of methionine, glutamine, and proline residues and has been shown to be the second transactivation domain in p63 (28). The SAM (sterile alpha motif) domain is an 8 kD protein domain, whose homolog was first identified as a protein interaction domain that can form homotypic- heterotypic-dimerization in yeast (29). The TID (transcription inhibitory domain) was shown to be an inhibitory domain of transcriptional activity in p63 through interaction with the TA domain, with deletion of the TID resulting in constitutive transcriptional activity (30). Of all the isoforms in p63, the $\Delta N\alpha$ isoform is highly expressed in the basal layer of the epidermis, although its functional role is not clear since it could function as a terminal differentiation initiator, or as a means to maintain the proliferative property of the basal epidermal cells (31). The $TA\alpha$ isoform is found in female oocytes and has been shown to induce apoptosis upon damage to the genome (32); however, its role in epidermal development remains controversial since data for its necessity or dispensability in commitment to differentiation of the epidermis have both been reported (33, 34). Each

isoform of p63 shows a different transcriptional activity in cell culture assays, but the *in vivo* function of each isoform remains to be determined.

The role of p63 in developmental function is further affirmed by three human genetic diseases cause by mutations in the p63 gene. Mutations in the DBD of p63 are linked to the EEC (Ectrodactyly, Ectodermal dysplasias, and Cleft lip/palate) syndrome, characterized by split hands and feet, sparse hair, thin and dry skin, and cleft lip palate (35). Mutations in the SAM domain of p63 are linked to the Hay-Wells Syndrome (AEC), characterized by cleft lip and palate, fused eyelids, dry skin, and parsed hair (36). The RHS (Rapp-Hodgkin Syndrome), which has similar phenotypes as the Hay-Wells Syndrome, is linked to mutations in both the SAM and the TID domains (37, 38). These human diseases have spontaneous heterozygous mutations in p63, in which the mutated gene has a dominant phenotype. The skin histology of patients with Hay-Wells Syndrome shows the mesenchymal layer of the skin epidermis is not differentiated. Based on the function of other SAM domains, it is believed that the SAM domain in p63 recruits essential proteins during its normal function, and the mutants abrogate this association. Two proteins, RACK-1 and ABBP1, were found to associate with the SAM domain through two independent yeast two hybrid assays. RACK-1 is an adaptor protein that recruits protein kinase C to phosphorylate p63 at unknown residues, and ABBP1 can control the splicing activity of keratin-1 by p63 (39). In addition, lipids have been shown to bind to the SAM domain of p63 (40). These new *in vitro* data will need further study to elucidate the role of the SAM domain within p63 in epidermal development.

The existence of p63 and p73 in mice and humans as well as the growing genomic database have led to the identification of the p53 protein family members in many

species, suggesting the ancient origin of this protein family. A protein homology search in the NCBI genomic database reveals p53 like molecules are ubiquitous proteins found in all vertebrates, present in many invertebrates, and even found in protozoa (41). Of all the domains in the p53 protein family members, the DBD is the most conserved. In comparison with the DBD of human p53, the sequence identity of DBD ranges from 13% in invertebrates to 62% in vertebrates. This conserved DBD motif allows all members of the p53 protein family to recognize the p53 consensus motif in *in vitro* assays, even in species having diverged over 550 millions years ago. In contrast, domains C-terminal to the DBD vary greatly between vertebrates and invertebrates; some invertebrate p53-like molecules found in nematodes, fruit flies, and beetles do not even have recognizable domains found in vertebrate p53, p63, and p73. Studies of nematodes and fruit flies show that invertebrate p53 is involved in maintaining genomic stability in germ cells. Knock-out studies of p53 in nematodes and fruit flies result in resistance to apoptosis upon damage to the genome, and no lethal effect in development except for defects in meiosis observed in some nematodes (42-44). Similarly p63 has been reported to maintain genomic stability in the oocytes of mice by inducing apoptosis of oocytes with a damaged genome (32). From these studies, a model could be proposed, in which the tumor suppressor function of p53 and the developmental functions of p63 and p73 in humans are later evolved functions that originate from the protein's role to protect the genome of germ cells.

The SAM domain is another shared domain found in both invertebrate and vertebrate p63/p73 like proteins. The presence of the SAM domain in p63/p73 like molecules in early invertebrates like mollusks suggests the domain has been conserved

within the p63/p73 like proteins. Although two proteins, RACK1 and ABBP1, have been found to associate with the SAM domain in humans (39), these two proteins are not conserved in evolution. Thus, it remains a possibility that the SAM domain has an alternate function or associates with interaction partners that are evolutionarily conserved. In contrast, the C-termini of the p53 homologues in *C. elegans* and *Drosophila* have no recognizable domains that are common among the p53 protein family, for example, the OD and the SAM domains, yet the function of both proteins is conserved. The divergent sequences in the C-terminus of the p53 like proteins in nematodes and fruit flies may reveal conserved structural motifs that retain function. The focus of this study is the characterization of the C-terminal domains of the two p53 like proteins in *C. elegans* and *Drosophila*, and the investigation of the function of the SAM domain in human p63 through structural and biochemical analysis (Figure 1-1).

Chapter two of this thesis describes the structural determination of the C-terminal end of the p53 homologues in *C. elegans* and *Drosophila*. Through sequence comparison and structural determination by NMR, the study identifies an OD and a SAM domain in *C. elegans* p53 (CEP-1), and a novel tetramerization domain in *Drosophila* p53 (Dmp53) that has not been observed in other p53 proteins. Both oligomerization domains of CEP-1 and Dmp53 are similar to the canonical fold of the OD in mammalian p53; however, each species has additional features that are necessary in stabilizing the OD. Without these structural elements, the OD either becomes structurally unstable in CEP-1, or becomes defective in tetramerization, and forms a dimer in Dmp53. The results from this study suggest the evolutionary relationship of CEP-1 and Dmp53 in the p53 protein family, and propose a putative function for the SAM domain within the p53 protein

family.

Chapter three of this thesis investigates the effects of mutations identified in the Hay-Wells syndrome on the property of the SAM domain in p63. Structural and temperature denaturation studies show that mutations do not prevent folding, but change the stability of the protein and its putative protein-protein interaction area. In addition, the SAM and the QP domains seem to have a stabilization effect on the OD of p63. In the presence of the QP and the SAM domains, the OD is soluble in *E. coli* expression. Without the QP and the SAM domain or when the SAM domain carries mutations identified from the Hay-Wells syndrome, the OD becomes insoluble. The results present here suggest the SAM domain of p63 may serve a similar role as the SAM domain of CEP-1 in contributing to the stability of the OD. Mutations in the SAM domain would alter the gross conformation of p63, which can lead to its aberration in transcriptional activity and result in the Hay-Wells syndrome.

The p53, p63, and p73 field is expanding rapidly. The evolutionary structural approach taken in this study was conducted with the belief that the most conserved element of a protein, either its function or structure, is preserved in ancestral organisms. By studying simple organisms like nematodes, yeast, and fruit flies, fundamental molecular mechanisms of apoptosis, the cell cycle, and embryonic development were revealed respectively. The structures presented in this study bring forth two hypotheses: the SAM domain has an intrinsic functional role (a domain that stabilizes the whole protein molecule) in addition to an extrinsic functional role (a protein-protein interaction domain which remains to be tested) in the p53 protein family, and the oligomeric state of p53 proteins (dimer or tetramer) may have distinct functional consequences.

References

1. Lane, D. P., and Crawford, L. V. (1979) *Nature* 278, 261-3.
2. Linzer, D. I., and Levine, A. J. (1979) *Cell* 17, 43-52.
3. el-Deiry, W. S., Tokino, T., Velculescu, V. E., Levy, D. B., Parsons, R., Trent, J. M., Lin, D., Mercer, W. E., Kinzler, K. W., and Vogelstein, B. (1993) *Cell* 75, 817-25.
4. Symonds, H., Krall, L., Remington, L., Saenz-Robles, M., Lowe, S., Jacks, T., and Van Dyke, T. (1994) *Cell* 78, 703-11.
5. Donehower, L. A., Harvey, M., Slagle, B. L., McArthur, M. J., Montgomery, C. A., Jr., Butel, J. S., and Bradley, A. (1992) *Nature* 356, 215-21.
6. Lee, W., Harvey, T. S., Yin, Y., Yau, P., Litchfield, D., and Arrowsmith, C. H. (1994) *Nat Struct Biol* 1, 877-90.
7. Nicholls, C. D., McLure, K. G., Shields, M. A., and Lee, P. W. (2002) *J Biol Chem* 277, 12937-45.
8. Cho, Y., Gorina, S., Jeffrey, P. D., and Pavletich, N. P. (1994) *Science* 265, 346-55.
9. el-Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W., and Vogelstein, B. (1992) *Nat Genet* 1, 45-9.
10. Wei, C. L., Wu, Q., Vega, V. B., Chiu, K. P., Ng, P., Zhang, T., Shahab, A., Yong, H. C., Fu, Y., Weng, Z., Liu, J., Zhao, X. D., Chew, J. L., Lee, Y. L., Kuznetsov, V. A., Sung, W. K., Miller, L. D., Lim, B., Liu, E. T., Yu, Q., Ng, H. H., and Ruan, Y. (2006) *Cell* 124, 207-19.
11. Haupt, Y., Maya, R., Kazaz, A., and Oren, M. (1997) *Nature* 387, 296-9.

12. Chang, J., Kim, D. H., Lee, S. W., Choi, K. Y., and Sung, Y. C. (1995) *J Biol Chem* 270, 25014-9.
13. Di Lello, P., Jenkins, L. M., Jones, T. N., Nguyen, B. D., Hara, T., Yamaguchi, H., Dikeakos, J. D., Appella, E., Legault, P., and Omichinski, J. G. (2006) *Mol Cell* 22, 731-40.
14. Ahn, J., and Prives, C. (2001) *Nat Struct Biol* 8, 730-2.
15. Vousden, K. H., and Lu, X. (2002) *Nat Rev Cancer* 2, 594-604.
16. Mihara, M., Erster, S., Zaika, A., Petrenko, O., Chittenden, T., Pancoska, P., and Moll, U. M. (2003) *Mol Cell* 11, 577-90.
17. Wahl, G. M., and Carr, A. M. (2001) *Nat Cell Biol* 3, E277-86.
18. Johnson, T. M., Hammond, E. M., Giaccia, A., and Attardi, L. D. (2005) *Nat Genet* 37, 145-52.
19. Haupt, Y., Rowan, S., Shaulian, E., Vousden, K. H., and Oren, M. (1995) *Genes Dev* 9, 2170-83.
20. Xu, Y. (2003) *Cell Death Differ* 10, 400-3.
21. Samuels-Lev, Y., O'Connor, D. J., Bergamaschi, D., Trigiante, G., Hsieh, J. K., Zhong, S., Campargue, I., Naumovski, L., Crook, T., and Lu, X. (2001) *Mol Cell* 8, 781-94.
22. Kaghad, M., Bonnet, H., Yang, A., Creancier, L., Biscan, J. C., Valent, A., Minty, A., Chalon, P., Lelias, J. M., Dumont, X., Ferrara, P., McKeon, F., and Caput, D. (1997) *Cell* 90, 809-19.
23. Yang, A., Kaghad, M., Wang, Y., Gillett, E., Fleming, M. D., Dotsch, V., Andrews, N. C., Caput, D., and McKeon, F. (1998) *Mol Cell* 2, 305-16.

24. Yang, A., Schweitzer, R., Sun, D., Kaghad, M., Walker, N., Bronson, R. T., Tabin, C., Sharpe, A., Caput, D., Crum, C., and McKeon, F. (1999) *Nature* 398, 714-8.
25. Yang, A., Walker, N., Bronson, R., Kaghad, M., Oosterwegel, M., Bonnin, J., Vagner, C., Bonnet, H., Dikkes, P., Sharpe, A., McKeon, F., and Caput, D. (2000) *Nature* 404, 99-103.
26. Mills, A. A., Zheng, B., Wang, X. J., Vogel, H., Roop, D. R., and Bradley, A. (1999) *Nature* 398, 708-13.
27. Yang, A., Kaghad, M., Caput, D., and McKeon, F. (2002) *Trends Genet* 18, 90-5.
28. Ghioni, P., Bolognese, F., Duijf, P. H., Van Bokhoven, H., Mantovani, R., and Guerrini, L. (2002) *Mol Cell Biol* 22, 8659-68.
29. Ponting, C. P. (1995) *Protein Sci* 4, 1928-30.
30. Serber, Z., Lai, H. C., Yang, A., Ou, H. D., Sigal, M. S., Kelly, A. E., Darimont, B. D., Duijf, P. H., Van Bokhoven, H., McKeon, F., and Dotsch, V. (2002) *Mol Cell Biol* 22, 8601-11.
31. Koster, M. I., Dai, D., and Roop, D. R. (2007) *Cell Cycle* 6, 269-73.
32. Suh, E. K., Yang, A., Kettenbach, A., Bamberger, C., Michaelis, A. H., Zhu, Z., Elvin, J. A., Bronson, R. T., Crum, C. P., and McKeon, F. (2006) *Nature* 444, 624-8.
33. Koster, M. I., Kim, S., Mills, A. A., DeMayo, F. J., and Roop, D. R. (2004) *Genes Dev* 18, 126-31.
34. Senoo, M., Pinto, F., Crum, C. P., and McKeon, F. (2007) *Cell* 129, 523-36.
35. Celli, J., Duijf, P., Hamel, B. C., Bamshad, M., Kramer, B., Smits, A. P.,

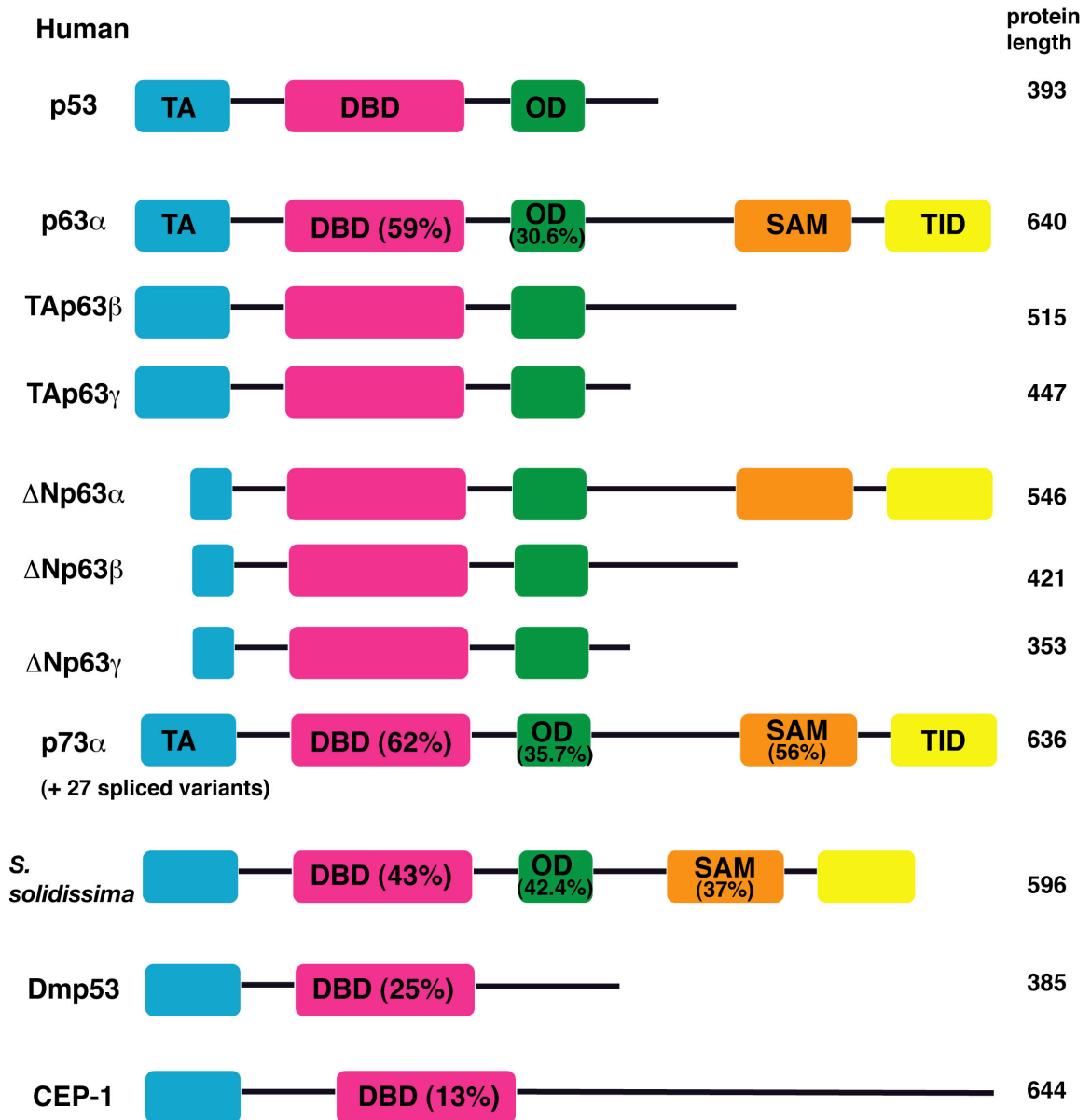
- Newbury-Ecob, R., Hennekam, R. C., Van Buggenhout, G., van Haeringen, A., Woods, C. G., van Essen, A. J., de Waal, R., Vriend, G., Haber, D. A., Yang, A., McKeon, F., Brunner, H. G., and van Bokhoven, H. (1999) *Cell* 99, 143-53.
36. McGrath, J. A., Duijf, P. H., Doetsch, V., Irvine, A. D., de Waal, R., Vanmolkot, K. R., Wessagowit, V., Kelly, A., Atherton, D. J., Griffiths, W. A., Orlow, S. J., van Haeringen, A., Ausems, M. G., Yang, A., McKeon, F., Bamshad, M. A., Brunner, H. G., Hamel, B. C., and van Bokhoven, H. (2001) *Hum Mol Genet* 10, 221-9.
37. Kantaputra, P. N., Hamada, T., Kumchai, T., and McGrath, J. A. (2003) *J Dent Res* 82, 433-7.
38. Bougeard, G., Hadj-Rabia, S., Faivre, L., Sarafan-Vasseur, N., and Frebourg, T. (2003) *Eur J Hum Genet* 11, 700-4.
39. Fomenkov, A., Huang, Y. P., Topaloglu, O., Brechman, A., Osada, M., Fomenkova, T., Yuriditsky, E., Trink, B., Sidransky, D., and Ratovitski, E. (2003) *J Biol Chem* 278, 23906-14.
40. Barrera, F. N., Poveda, J. A., Gonzalez-Ros, J. M., and Neira, J. L. (2003) *J Biol Chem* 278, 46878-85.
41. Mendoza, L., Orozco, E., Rodriguez, M. A., Garcia-Rivera, G., Sanchez, T., Garcia, E., and Gariglio, P. (2003) *Microbiology* 149, 885-93.
42. Derry, W. B., Putzke, A. P., and Rothman, J. H. (2001) *Science* 294, 591-5.
43. Brodsky, M. H., Nordstrom, W., Tsang, G., Kwan, E., Rubin, G. M., and Abrams, J. M. (2000) *Cell* 101, 103-13.
44. Ollmann, M., Young, L. M., Di Como, C. J., Karim, F., Belvin, M., Robertson, S.,

Whittaker, K., Demsky, M., Fisher, W. W., Buchman, A., Duyk, G., Friedman, L., Prives, C., and Kopczynski, C. (2000) *Cell* 101, 91-101.

Figure 1-1. Domain layout of the p53 protein family

In addition to p53, there are two other paralogs in human called p63 and p73. In p63, there are six isoforms, and in p73, there are 28 isoforms. Here only one isoform of p73 is shown. The SAM domain, the QP domain, and the TID are not found in p53. Numbers inside parenthesis are percentage of identical sequence in comparison with human p53 for the DBD and the OD, and human p63 for the SAM domain. p53 like molecules are also found in invertebrates, for example in *S. solidissima*, *Drosophila* (dmp53), and *C. elegans* (CEP-1). The p63/p73 like protein in *S. solidissima* has similar domain architecture as human p63 and p73, but the C-terminal end of Dmp53 and CEP-1 has no recognizable domains. This thesis explores the domain structures in these two proteins and investigates the function of the SAM domain in human p63.

Figure 1-1.



Chapter 2

Structural evolution of C-terminal domains in the p53 family

Hong Der Ou^{1,2}, Frank Löhr¹, Vitali Vogel³, Werner Mäntele³ and Volker Dötsch^{1*}

¹ Institute of Biophysical Chemistry, Centre for Biomolecular Magnetic Resonance (BMRZ), J.W. Goethe University of Frankfurt, Max-von-Laue Str. 9, 60438 Frankfurt/Main

² Graduate Group in Biophysics, University of California San Francisco, 600 – 16th Street, San Francisco, CA 94143

³ Institute of Biophysics, J.W. Goethe University of Frankfurt, Max-von-Laue Str. 1, 60438 Frankfurt/Main

* The content of this chapter is published in the EMBO Journal: 26(14):3463-73, 2007

Abstract

The tetrameric state of p53, p63, and p73 has been considered one of the hallmarks of this protein family. While the DNA binding domain is highly conserved among vertebrates and invertebrates, sequences C-terminal to the DBD are highly divergent. In particular, the oligomerization domain of the p53 from the model organisms *C. elegans* and *Drosophila* cannot be identified through sequence analysis. Here we present the solution structures of their oligomerization domains and show that they both differ significantly from each other as well as from human p53. CEP-1 contains a composite domain of an oligomerization domain and a SAM domain and forms dimers instead of tetramers. The Dmp53 structure is characterized by an additional N-terminal β -strand and a C-terminal helix. Truncation analysis in both domains reveals that the additional structural elements are necessary to stabilize the structure of the oligomerization domain, suggesting a new function for the SAM domain. Furthermore, these structures show a potential path of evolution from an ancestral dimeric form over a tetrameric form, with additional stabilization elements, to the tetramerization domain of mammalian p53.

Introduction

Since its discovery in 1979, the role of p53 within cells has been intensely investigated by many research groups. Its importance in cancer biology is highlighted by the fact that more than 50% of tumors have mutations in p53, which makes it one of the most important proteins with respect to human disease. p53 is a tetrameric transcription factor that suppresses tumor formation by activating a set of genes inducing either cell cycle arrest or apoptosis (1-3). Which one of the two pathways is chosen depends on the nature of the cellular stress that triggers the p53 response such as DNA damage by UV- or γ -irradiation, hypoxia, or oncogene activation (4). However, the question of how p53 distinguishes these diverse cellular stress signals and determines which set of genes to be turned on is still not well understood, although recent reports on the role of ASPP (5) and acetylation of K120 in the DNA binding domain of p53 by Tip60 suggest a mechanism in which the apoptotic pathway is preferred over the cell cycle arrest pathway (6, 7).

During the past ten years, p53 has expanded into a protein family with complexity and diversity exemplified by the discovery of different splice variants of p53 and two mammalian paralogues, p63 and p73 (8, 9). Sequence analysis of the two mammalian paralogues has revealed major differences to p53. Structurally, p63 and p73 have additional domains not found in p53, and exist in various isoforms created by a combination of C-terminal splicing and different N-terminal promoters (10-12). While p53 has an N-terminal transactivation domain (TA) for recruitment of core transcriptional factors, a central DNA-binding domain (DBD) for recognition of promoter sequences, an oligomerization domain (OD) for tetramerization, and a short basic stretch of 30 amino acids for regulation of transcriptional activity, the C-terminus of p63 and p73 contains

(depending on the splice form) a sterile alpha motif (SAM) domain and a transcriptional inhibitory domain (TID). SAM domains are small protein-protein interaction modules that are found in a wide variety of different proteins, ranging from kinases and transcriptional regulators to cell surface receptors (13). The TID, an unstructured region C-terminal to the SAM domain, was shown to inhibit the transcriptional activity of p63 by its interaction with the TA domain (14). Functional analyses through knock-out mice studies have shown that both proteins do not seem to be tumor suppressors like p53, instead they play important roles in development and maintenance of epithelial tissue (p63) or maintenance of certain neurons (p73) (15-17). The identification of mammalian p53, p63, and p73 allowed rapid classification of p53 homologues in other species. Vertebrate species like frogs, zebrafish and chicken possess all three paralogues, while invertebrate species such as squids, clams, and molluscs have p53 homologues that are more closely related to the mammalian p63 and p73 than to mammalian p53. Interestingly, sequences from other invertebrate species like nematodes, fruit flies, and beetles cannot be classified as either p53- or p63/p73 like and show similarity only within the DBD.

According to the current hypothesis on the evolution of the p53 protein family, p63/p73 resemble the ancestral form while p53 evolved later (8). The identification of many new p53 protein family members in invertebrates could further elucidate the evolutionary development of this important protein family. The structural differences between p53 and p63/p73 also reflect their different functions in surveillance of the genetic integrity of a cell and in tissue development, respectively. Since some of the invertebrate p53 forms exhibit significant differences to both p53 and p63/p73, studying

these family members might shed some new light onto the evolution of p53's function. In particular, the recent discovery of p53 homologues in the two important model organisms *Drosophila melanogaster* (Dmp53) and *Caenorhabditis elegans* (CEP-1) provides new genetic tools for the investigation of their function (18-20). Both proteins show only a low degree of sequence homology to p53 in the DNA binding domain, yet they exhibit very similar DNA binding specificity (19, 21). Studies have shown that both proteins are functionally distinct from mammalian p53. While the hallmark of p53 is its dual function in inducing either cell cycle arrest or apoptosis, both CEP-1 and Dmp53 can only induce apoptosis, but not cell cycle arrest upon irradiation. These observations led to the idea that induction of cell cycle arrest was a later evolutionary development since *Drosophila* and *C. elegans* have long diverged from the vertebrates.

The lack of any recognizable domain C-terminal to the DBD, including the highly conserved OD and the distinct functions of CEP-1 and Dmp53 prompted us to investigate the domain organization of the C-terminus in both proteins. We have determined the NMR structure of the C-terminus of CEP-1 and Dmp53. The CEP-1 C-terminus is composed of two sub-domains, an OD followed by a SAM domain that closely interacts with the OD. Surprisingly, the OD domain of CEP-1 forms a dimer instead of the usual tetramer observed in human p53. Equally surprising, the Dmp53 C-terminus reveals a unique oligomerization fold not found in other p53 proteins in which an extra β -strand and an extra helix complement the canonical p53 oligomerization fold. Furthermore, deletion studies and NMR data show that protein domains C-terminal to the OD are necessary to maintain the tertiary fold. The findings from this study suggest that the ancestral p53 form had a similar domain organization as p63/p73, and that the OD of

vertebrate p53 that is conformationally stable without additional structural elements evolved later in evolution.

Results

Identification of a conserved domain in the C-terminus of CEP-1

Based on the protein sequence length, CEP-1 with 644 amino acids seems more closely related to p63 and p73, thus, we suspected the existence of additional domains C-terminal to the DBD in CEP-1. However, sequence analysis failed to reveal significant homology to domains frequently found in p63/p73, such as the SAM domain or even the OD identified so far in all members of the p53 family (Figure 2-1A).

Sequence comparison of CEP-1 from two Nematode species, *C. elegans* and *C. briggsae*, identified a conserved region C-terminal to the DNA binding domain (after residues 510, *C. elegans* numbering) (Figure 2-7). After several rounds of optimization, we identified a core domain between residues 528-644 in the C-terminus of CEP-1 with a molecular weight of 13.6 kD. A melting curve monitored by circular dichroism spectroscopy revealed a sharp transition point at 41°C, an indicator of a folded domain with a two-state unfolding transition (Figure 2-8). Since this identified domain is C-terminal to the DNA binding domain, it should contain the oligomerization domain responsible for tetramerization, a hallmark of the p53 protein family (22). In contrast to p53, however, the protein elutes off a size exclusion column as a dimer (Figure 2-2A). This dimeric state was further confirmed by velocity sedimentation measurements, in which the data fit the expected molecular weight of a dimer (27.2 kD) (Table 2-1). Since neither a classical oligomerization domain nor any other domains could be identified based on sequence comparison, we used standard heteronuclear NMR spectroscopy to investigate the structure of the C-terminus.

Overall structure of the C-terminus of CEP-1

The three-dimensional structure of the C-terminus of CEP-1 reveals that it consists of two interacting domains. Despite the lack of a significant sequence homology, one of these two domains is an oligomerization domain (OD) as found in all members of the p53 protein family and the second domain is a SAM domain (Figures 2-1B and 2-1C). The dimerization interface of the OD consists mostly of interactions between the β -strand and the α -helix of one OD packing against the other OD in an anti-parallel fashion as observed in the human p53 OD structure (22). The superimposed structures of the OD in human p53 dimer and CEP-1 have an rmsd of 2.0 Å, with the α -helix in CEP-1 being two turns shorter than the corresponding helix in human p53 (Figure 2-9). The SAM domain is composed of five helices that adopt the same topology as other SAM domain structures, e.g. in EphB2 and p73 (23, 24). However, the SAM domain in CEP-1 resembles the EphB2 structure more closely than the p73 SAM domain as indicated by its smaller rmsd values: 1.45 Å between EphB2 and CEP-1, and 1.83 Å between p73 and CEP-1 (Figure 2-9). The OD and the SAM domain are connected by a 16 amino acid linker of which some resonances are either missing or could not be assigned due to significant overlap in the spectra. This linker region contains a short helix in its C-terminus that makes contacts with the first helix of the SAM domain. Direct interactions between the OD and the SAM domain are observed between the α -helix of the OD and the last α helix of the SAM domain within the same monomer and the β sheet of the OD from the other monomer. The relative angle between the helix of the OD and the last helix of the SAM domain measures 125°.

The Oligomerization domain of CEP-1

The Oligomerization domain is a well conserved domain among vertebrate p53 forms due to its necessity in tetramer formation (25). In contrast, the OD of CEP-1 shows a poor sequence homology, including key residues that are strictly conserved in all vertebrate p53 sequences (Figure 2-3A). The signature glycine residue that permits a sharp turn between the β -strand and the α -helix in the structure of the OD of p53 is replaced by a serine, threonine di-peptide in CEP-1 (Figures 2-3A and 2-3B). This substitution causes the angle between the α -helix and the β -strand within a monomer to widen from 26° as seen in the p53 OD to 50° in CEP-1. Further substitutions involve the dimer interface. In p53 the dimer interface is stabilized by hydrophobic interactions between F338 from the α -helix of one monomer with F328 of the β -strand of the other monomer. In CEP-1, this dimer interface is maintained by interaction between Y542 and the aliphatic part of the side chain of R533, confirmed by NOEs between the β , γ , and δ protons of R533 and the δ and ϵ protons of Y542.

The most significant difference between CEP-1 and other members of the p53 protein family is its oligomerization property. While all other p53 protein family members form a tetramer through the OD, CEP-1 forms a dimer (Figure 2-2A; Table 2-1). The typical OD of p53 family members is a dimer of dimers, involving two separate and distinct interfaces (26) (Figure 2-4A). The first dimerization interface consists of an anti-parallel β -sheet formed by the β -strand of each monomer with additional contacts to the α -helix. This dimerization interface is, as described above, conserved in CEP-1. The second dimerization interface in p53 is formed by interaction of the two helices of each dimer with the corresponding helices of the second dimer, thus creating a four-helix

bundle. Key residues of the hydrophobic core of this tetramerization interface in p53 are M340, L344, A347, L348, and L350 from each helix (25) (Figure 2-4A). A structural alignment of the CEP-1 OD with other members of the p53 protein family shows that M340, A347, and L348 are replaced by K544, R551, and E552 in CEP-1 (Figures 2-3A and 2-3B). This substitution creates a ring of charged residues that encircles F548 from each monomer in the center of the ring (Figure 2-4B). If the OD of CEP-1 formed a tetramer with similar packing arrangement as the human p53 OD, then this electrostatic charged ring would form unfavorable interactions in the tetrameric interface due to charge repulsion. Thus, the lack of charge complementarity for residues K544, R551, and E552 explains the inability of the OD in CEP-1 to form a tetramer.

To test this hypothesis, we made two series of mutations toward converting the OD of CEP-1 from a dimer into a tetramer. Upon mutation of R551 and E552 to leucine, we observed an equilibrium of two oligomeric forms, with roughly 33% of CEP-1 in a tetrameric and 67% in a dimeric form (Figure 2-2A). Additional mutation of K544 to methionine transforms the OD of CEP-1 completely into a tetrameric state, as indicated by gel filtration and confirmed by analytical ultracentrifugation (Figure 2-2A; Table 2-1). It is also worth noting that our constructs contained the whole C-terminus of CEP-1, which includes the SAM domain. Thus it excludes the possibility that the SAM domain could contribute to tetramerization, since tetramerization can be achieved solely by mutations in the OD of CEP-1.

The SAM domain of CEP-1

Another significant difference between the OD of CEP-1 and the OD of other p53

family members is its close interaction with the C-terminal SAM domain. SAM domains are found in many different proteins including ephrin receptors, transcriptional repressors, as well as the p53 homologues p63 and p73. They are composed of four short helices followed by a long helix that completes its hydrophobic core. SAM domains are protein-protein interaction modules that have the ability to homo- and hetero-dimerize as well as oligomerize. In the ephrin receptor, the SAM domain forms a dimer in the crystal structure involving the last helix and loop 3 (between helix 3 and 4) (23). In CEP-1, the equivalent loop 3 region is exposed to the solvent, and the C-terminus of the last helix interacts with the helix in the OD domain, thus it is unlikely that this site constitutes a dimerization surface with another SAM domain. In the p63/p73 protein family, the SAM domain is a monomer in isolation and, to date, no homo-dimerization tendency has been identified (24). Similarly, the SAM domain of CEP-1 exists as a monomer in isolation (Table 2-1), and no contacts are made between the SAM domains in the full length construct of the C-terminus of CEP-1. Recently, SAM domains in the Smaug family, a translational repressor, have also been found as part of RNA binding modules that recognizes RNA hairpins with the loop sequence CUGGC (27, 28). The RNA interaction surface on the SAM domain is composed of the N-terminus of helix 5 and loop 1 (between helix 1 and 2) which has a positively charged surface. In CEP-1, this region does not have a highly positive charged surface, so it is unlikely that the SAM domain in CEP-1 interacts with RNA as in Smaug.

Characterization of C-terminus of Dmp53

In addition to CEP-1, the *Drosophila* homolog of p53, Dmp53, shows low

sequence homology to other p53 protein family members, again mainly confined to the DNA binding domain. The protein sequence length suggests it is more related to p53. Dmp53 induces apoptosis through the transcription of genes, such as *hid*, *reaper*, and *sickle* (29). The functional similarity of Dmp53 and CEP-1 and their low sequence homology to other p53 protein family members prompted us to investigate the oligomeric state of Dmp53 as well. We have purified the OD of Dmp53 (315-361, Dmp53 numbering in Ollmann et al, 2000), and shown by gel filtration as well as analytical ultracentrifugation that it forms a dimer (Figure 2-2B; Table 2-1). NMR investigations of this dimeric form, however, have revealed that it exhibits conformational heterogeneity characterized by broad peaks in the NMR spectrum (Figure 2-5). Similar to human p53, Dmp53 also contains a stretch of highly basic residues C-terminal to the OD. Others have shown that this basic region in human p53 serves as a regulatory tail that controls DNA binding affinity (30). To test whether this region has any structural function in Dmp53 and in particular if it stabilizes the dimer, we purified a construct that includes the OD and this C-terminal region (additional 24 residues). Surprisingly, this extended OD forms a tetramer and its NMR spectra indicate the existence of a well-folded single conformation (Figure 2-5; Table 2-1). Structural determination by NMR spectroscopy shows that the C-terminal 24 amino acids form an α -helix that packs against the α -helix that forms the canonical oligomerization domain (Figures 2-1B and 2-1D). In addition, the N-terminus is extended relative to the p53 OD and contains an additional β -strand that forms together with the canonical β -strand and the corresponding elements of another monomer an anti-parallel four-stranded β -sheet. Thus the Dmp53 OD domain contains both an N-terminal extension (the β 1-strand) as well as a C-terminal extension

(the H2 α -helix) relative to human p53 and can best be described as a four helical hairpin bundle sandwiched between two four-stranded β -sheets. In Dmp53, the conserved glycine residue in the canonical OD is replaced by proline and asparagine (Figures 2-3A and 2-3B). The tetramerization interface is built by the inner four helices that pack against each other similar to the four helices of human p53 (Figure 2-4C). In contrast to human p53, the tetramerization interface of Dmp53 contains a central cluster of charged residues made up of lysine 352 (K352) and glutamic acid 353 (E353) from each helix, which can form four salt bridges that constitute the core of the tetramerization interface. Mutating E353 of this cluster to lysine disrupts the electrostatic pairings and destroys the tetramerization interface, thus creating a stable dimer (Figure 2-2B). The core of the tetramerization interface is shifted by one helical turn towards the C-terminus in comparison with human p53 (Figure 2-3A). In addition, the inner helix of each monomer (H1:C) also packs across the tetrameric interface to the C-terminal helix (H2:B) of another monomer, with contacts between L359 from the inner helix and A371 and L375 of the C-terminal helix (Figure 2-4C).

Further contacts of the C-terminal helix (H2:B) that contribute to the dimer interface involve a surface built by the inner helix (H1:B) of the same monomer and the N-terminal β -strand (β 1:A) of another monomer within a dimeric unit. In addition this structural arrangement of the dimeric unit is capped by interaction of H370 (H2:B) with two tryptophans from the β -strand (W321, β 1:A) and the inner helix (W342, H1:A) of the other monomer (Figure 2-5). The involvement of the C-terminal helix in both the dimeric as well as the tetrameric interface explains why its deletion destabilizes the structure and results in a conformationally unstable dimer. It also predicts that removing the first β -

strand should have a similar destabilizing effect as deleting the C-terminal helix. Indeed, the loss of strand β 1 leads to a dimer that shows broad peaks in NMR spectra as well as a non-sigmoidal transition in CD-melting experiments suggesting that it is conformationally unstable (Figures 2-2B, 2-5, and Figure 2-10). These results demonstrate that the minimal OD consisting of one β -strand followed by one helix as identified in human p53 does not form a stable tetrameric structure in Dmp53. The additional N-terminal (β -strand) and C-terminal (α -helix) extensions play a key role in stabilizing Dmp53.

The SAM domain stabilizes the OD in CEP-1

Based on these observations with Dmp53, we asked if the SAM domain in CEP-1 might have a similar stabilizing function as the C-terminal helix in Dmp53. The absence of the SAM domain in the C-terminus of CEP-1 (528-555) indeed forms a less stable structure as indicated by the lack of a sharp sigmoidal transition in the temperature denaturation curve as well as the absence of an alpha helical pattern (two minima at 208 and 222 nm) in the wavelength scan as observed by CD spectroscopy (Figure 2-8). Since the expression level of the isolated OD decreased significantly relative to the full length C-terminus of CEP-1, we were not able to record an NMR spectrum.

Evolution of the p53 protein family

The structural analysis of the C-termini of CEP-1 and Dmp53 provides an opportunity to trace the evolution of the p53 protein family based on the presence of the SAM and structure of the OD domains in the C-terminus. While the positioning of

nematodes in a phylogenetic tree remains controversial (ectysozoa vs. coelomata) (31), we placed the currently available sequences of the p53 protein family members from the NCBI library into a phylogenetic tree constructed from the ectysozoa perspective which is the current favored view. As shown in figure 2-6, the SAM domain has been integrated early into the p53 protein family since it can be identified in the protostome branch represented by nematodes and molluscs as well as in the early deuterostomes such as echinoderm and urochordate. In the Urochordata phylum, a gene duplication seems to have occurred that results in a p53 form without a SAM domain which resembles vertebrate p53 in addition to a p63/p73-like form. The loss of the SAM domain occurred also within the protostome branch in the arthropoda phylum represented by Dmp53. However, in arthropods a C-terminal helix replaces the SAM domain and serves a similar functional role in stabilizing the OD. While the functional role of the SAM domain in human p63 and p73 remains elusive, the existence of the SAM domain in p53 like molecules in the protostome phylum underscores its important role within the p53 protein family, such as its contribution in stabilizing the OD. The current study provides additional evidence that the vertebrate p53 is a recent evolutionary development, and p63 and p73 are the more ancestral family members.

However, this ectysozoa based phylogenetic tree does not permit a direct correlation of the oligomeric state and the evolutionary development of p53. Instead this tree would either suggest that the dimeric CEP-1 has de-evolved from an ancestral tetrameric p53 or that tetramerization has evolved twice independently. Given the high sequence identity between the oligomierzation domains of vertebrates and molluscs, this last hypothesis seems less likely. Alternatively, a phylogenetic tree that is based on a

coelomata view places nematodes in an earlier phylogenetic branch than both *Drosophila* and vertebrates (Figure 2-6B). This hypothesis supports the view that p53 evolved from a dimeric molecule into a tetrameric form with further divergence between molluscs and arthropods. A variation of this phylogenetic tree places the arthropoda phylum before the divergence of molluscs and vertebrates and would be most consistent with our structural data (32, 33).

Discussion

The low homology of the C-terminus of CEP-1 and Dmp53 with other p53/p63/p73 proteins suggests that significant differences in both structure and function might exist in comparison with other p53 protein family members (8). Indeed, this study shows that the C-terminus of CEP-1 is a dimeric molecule, making the *C. elegans* form of p53 the first native dimer of the entire protein family. The existence of a natural dimeric form of p53 reaffirms previous identification of residues that are essential for tetramerization (34, 35), since the C-terminus of CEP-1 can be converted into a tetrameric form by mutations of residues at the tetrameric interface (K544, R551, E552) into hydrophobic residues.

The structure of the OD of CEP-1 illustrates a conserved strategy in the p53 protein family to utilize the electrostatic nature of the tetrameric interface of the OD to control the oligomeric state of the protein. McCoy *et al.* had reported that mutating three key residues of the tetrameric interface of the OD in human p53 (M340K, F341I, L344Y) results in a dimeric molecule with the orientation of the helices switched from anti-parallel to parallel (36). Interestingly, in CEP-1, the corresponding residues (K544, V545, and F548) show greater similarity to the p53 mutant sequence than to the p53 wild type sequence, yet the structure of the CEP-1 OD resembles the structure of the p53 wild type OD more closely by adopting an anti-parallel packing of helices (angle between the helices of one dimeric unit: p53 156°; CEP-1 125°; mutant 78°). Mapping CEP-1 residues onto the structure of the p53 mutant shows that CEP-1 retains the anti parallel orientation due to unfavorable charge repulsion by the two lysines (K554) located at the C-terminus of the helix. In the mutant p53 molecule these lysines are replaced by two leucines

(L350) that form favorable hydrophobic packing, thus enabling a parallel orientation of the helices. The structural and mutational analysis of CEP-1 show that controlling the electrostatic nature of the interface allows the formation of both dimers as well as tetramers without changing the anti-parallel topology of the OD.

The presence of the SAM domain at the C-terminus of CEP-1 is a surprise discovery since sequence alignment does not reveal its existence. The SAM domain of CEP-1 is C-terminal to the OD and it makes contacts with the OD through the last helix, but not to the SAM domain from the other monomer. In other SAM domain containing proteins, the SAM domain functions as a homo- or hetero- dimerization domain, for example in the ephrin receptor B2 and the SAM domains of Ste11 and Ste50 in *Saccharomyces cerevisiae* (23, 37). Based on our structure, however, the SAM domain of CEP-1 does not seem to play a direct role in oligomerization. This result also has implications for p63 and p73 which both contain SAM domains at their C-terminus. The structure of the C-terminus of CEP-1 as well as the incapability of the isolated CEP-1/p63/p73 SAM domains to oligomerize suggests that SAM domains in the p53 protein family also do not form homo dimers or oligomers in the context of the full length protein.

Despite the observation that CEP-1 is a dimer in solution, the possibility exists that two CEP-1 dimers can form a functional tetramer through cooperative binding to their promoter sites. Analyzing the promoter sequences of known targets of CEP-1 and Dmp53 should reflect the difference in the oligomeric state of both proteins *in vivo*. The promoter region of *egl-1* (38), a CEP-1 inducible gene, contains only a half-site (AAACAAGCTT), which satisfied the p53 consensus sequence motif for a dimeric p53

molecule (CEP-1) (39). In contrast, the *Drosophila* p53 responsive gene, *reaper*, has two half sites (TGACATGTTT/GAACAAGTCG) (19), which allows the binding of a tetrameric p53 molecule (Dmp53).

The hypothesis that a dimeric p53 is the ancestral form is further supported by a study on the formation of human p53 in rabbit reticulate lysate (40). This study showed that human p53 first forms a dimer co-translationally and that tetramers are only formed at a later stage post-translationally. It was further shown that when mutant and wild type p53 are co-expressed, there is only one form of heterotetramer, which consists of a dimer of wild type and a dimer of mutants. Thus, even after millions of years of evolution, the mammalian p53 still retains a dimeric building block as its basic unit.

Dmp53 and perhaps the entire arthropod phylum have adopted a different tetramerization mode by utilizing an additional helix C-terminal to the canonical mammalian p53 OD and an additional β -strand before the OD. By sequence alignment with CEP-1, one would conclude that Dmp53 would be a dimer due to the presence of the charged residues K352 and E353 in Dmp53. However, the helix and the additional β -strand transform Dmp53 into a tetramer. The additional β -strand provides a stable dimeric unit, which positions helix 2 of Dmp53 in the correct orientation to interact with the helix 1 of another dimeric unit, and concomitantly allows attractive electrostatic interactions along the tetrameric interface between the two charged residues. The composition of this unique mode is necessary for tetramerization, since a deletion of either additional element results in incorrect topology of helices, thus reverting to the dimeric state as in CEP-1.

It is interesting to note that only the C-terminus of the p53 protein family has

undergone significant evolutionary changes (dimer or tetramer, presence or absence of the SAM domain), while the DNA specificity and the structure of the DNA binding domain has remained basically unchanged. The X-ray crystal structure of the DBD of CEP-1 was recently solved (21). It demonstrates that despite a low sequence homology of only 15%, the core structure is very similar, and the DNA binding sequence specificity of CEP-1 is virtually identical to human p53, even though loop L1, a part of the DNA binding interface in the DBD of human p53, adopts a different conformation in CEP-1. In Dmp53, the DBD is 25% identical to human p53, which is higher than in the case of CEP-1. This higher sequence identity in Dmp53 and the high conservation of the structure and function of the DBD of CEP-1 predict that the structure of the Dmp53 DBD and its DNA binding specificity are also highly conserved. Despite this high conservation in the DBD, the biological function of the individual members of the p53 protein family is distinct, and sequences C-terminal to the DBD show very significant divergence, thus suggesting that these C-termini play an important role in specifying the biological function of the individual family members.

The availability of structural and biochemical data for CEP-1, Dmp53, and human p53 in combination with p53 sequence data from many species sheds some light on the ancestral form of p53 and its evolutionary development. It was estimated that *C. elegans* and *Drosophila* have diverged from vertebrates 550 million years ago, thus CEP-1 and Dmp53 could resemble the ancestral form of p53. Despite the lack of a SAM domain in Dmp53, the additional structural elements (the first β -strand and the last α -helix) serve a similar stabilizing function as the SAM domain in CEP-1. Oligomerization in the ancestral p53 forms most likely required either the fusion of two domains (like in CEP-1)

or additional structural elements (like in Dmp53). The minimal OD found in vertebrate p53 that can form stable tetramers by itself is probably a later evolutionary result, while the C-terminal tail that is a necessary structural element in Dmp53 became an important regulatory region including many sites for posttranslational modifications.

In summary, we have determined the structure of the C-terminal domain of CEP-1 and Dmp53 and show nature's multifaceted means to achieve oligomerization in p53 besides the canonical form found in mammalian p53. Furthermore we have shown that additional structural elements identified in CEP-1 and Dmp53 that are not present in human p53 are necessary for the integrity of the oligomerization domain. A loss of these elements results in conformationally unstable structures, and in Dmp53, leads to a change in the oligomeric property. The structural investigations described here suggest an evolutionary path from an ancestral dimeric form over tetrameric forms that need additional stabilization elements to the minimal tetramerization domain known from mammalian p53.

Acknowledgement

A plasmid with a partial CEP-1 sequence was a gift from Brent Derry. We thank Felician Dancea, Wesley McGinn-Straub, Zach Serber, Michael Reese, Florian Durst, Meichen Shi for discussions. This work was supported by the Centre for Biomolecular Magnetic Resonance at the University Frankfurt (BMRZ), the DFG (DO 545/2-1), EU-Grant EPISTEM (LSHB-CT-019067) and Philip Morris USA Inc. and by Philip Morris International.

Materials & Methods

Protein Expression and Molecular Clonings

All CEP-1 and Dmp53 constructs used for structural studies were cloned into plasmid pGEX-6P-2 (Amersham Bioscience) or pBH4 (gift from Wendell Lim lab) using BamHI and XhoI sites. Mutagenesis constructs of both CEP-1 and Dmp53 were made by the QuickChange protocol from Stratagene. BL21 cells were grown to an $OD_{600} = 0.8$ and induced with 500 μ M IPTG at 25° Celsius for 8 hours. The proteins were purified as described in the manufacturer's protocol, cleaved by precision protease for the pGEX plasmid or by TEV protease for the pBH4 plasmid, and further purified on a Superdex-75 gel filtration column. Protein samples were stored in a buffer consisting of 20mM sodium phosphate (pH 7.0), 100mM sodium chloride, and 0.03% sodium azide. For the expression of ^{15}N and $^{15}\text{N}/^{13}\text{C}$ labeled proteins bacteria were first grown in LB media to an $OD_{600} = 0.8$, then transferred to M9 minimal media with the appropriate isotopic components, and induced under the same condition as described above. Protein samples used to obtain inter-monomer NOEs consist of an equal ratio of ^{15}N and ^{13}C labeled proteins. For CEP-1, 2 weeks of equilibration time were needed to obtain signals in the experiment. For Dmp53, equal molar ratio of proteins were mixed, denatured in 6M guanidium hydrochloride, then refolded in the buffer described above.

NMR Experiments and Structure Calculations

Backbone residues of CEP-1 and Dmp53 were assigned using the TROSY version of HNCA and HNCOCA. For CEP-1, specific labeling of lysine, tyrosine, and leucine were

used to confirm assignments. Distance constraints were derived from ^{15}N -NOESY-HSQC, and ^{13}C -NOESY-HSQC. Aromatic protons were assigned based on 2D- D_2O -NOESY, 2D- D_2O -TOCSY, and non-constant time 3D- ^{13}C -NOESY. Inter-monomer NOEs were obtained through a 4D constant time J-Resolved NOESY (41) measured with a 1:1 mixture of ^{12}C - and ^{13}C -labeled proteins in both cases. In the case of Dmp53 the protein had to be denatured first with guanidium hydrochloride prior to mixing of the ^{12}C - and ^{13}C -labeled proteins and subsequent refolding by dialysis. For CEP-1 44 unambiguous inter-monomer distance constraints all located in the β -sheet were identified from the 4D J-Resolved NOESY. Overall, 114 inter-monomer NOEs were assigned. In the case of Dmp53 184 NOEs obtained from the 4D J-Resolved NOESY and an overall of 240 inter-monomeric NOEs were used. Of these inter-monomeric NOEs 46 are located in the β -sheet, 130 between a β -strand of one monomer and an α -helix of the other monomer within one dimeric unit and 31 between helices of different monomers within a dimer were observed. In addition, 33 NOEs across the tetrameric interface were identified. Dihedral angle constraints were derived from TALOS based on chemical shifts of N, CA, HA, and CB (42). Hydrogen bond constraints of secondary structure elements were based on TALOS calculations and confirmed by characteristic NOE patterns for α -helices and β -sheets as well as deuterium hydrogen exchange measurements. Structure calculations were carried out with Aria 1.2 with modified protocols that imposed a C_2 symmetry for CEP-1 and a D_2 symmetry for Dmp53 throughout every stage of calculation (43). For the structure calculation of CEP-1 unambiguous inter-monomer constraints obtained from the 4D J-Resolved NOESY were included in every iteration of the calculation. For Dmp53 unambiguous NOEs between

the β 2-strands of two monomers established the existence of an anti-parallel inter-monomeric β -sheet. Using these unambiguous NOEs as a starting point we first calculated dimer structures. Based on these dimer structures the other distance constraints obtained from the 4D J-Resolved NOESY were evaluated for their consistency with the dimer structure. Several distance constraints that could not be satisfied within the dimer were assumed to be constraints across the tetrameric interface and were used as those in the following structure calculations. 20 structures were calculated in 7 iterations, and 100 structures were calculated in the last iteration. Initial assigned peaks were separated into unambiguous and ambiguous peaks and ambiguous peaks with multiple assignments and violated restraints were manually inspected. The new ambiguous peak list included peaks with multiple assignments and inter-monomer assignments (not obtained from the 4D constant time J-Resolved NOESY). After multiple iterations of peak inspections and structure calculations, 100 structures were calculated and the best 20 structures were used for water refinements and analysis. The structural statistics of CEP-1 and Dmp53 can be found in table 2-2 and 2-3. Figures were prepared with Pymol (44). MOLMOL was used for structural alignment of NMR derived models and generation of the electrostatic map of the molecule (45). LSQMAN was used for structural alignment of different proteins (46) and the structure validation program of the PDB server for further structural analysis.

Circular Dichroism Experiments

Temperature scans (20 °C to 90 °C for CEP-1, 20 °C to 100 °C for Dmp53) were measured in a Jasco 810 CD spectrometer. For CEP-1 OD alone construct (528-555), it

contains additional his-tags and TEV protease cleavage site. Due to the unusual circular dichroism spectra in Dmp53, the observed wavelength for temperature scan was chosen based on signal intensity. Thus for wild type Dmp53, CD was observed at 234 nm, Dmp53 (326-385) at 228 nm, and Dmp53 (315-361) at 220 nm. For temperature denaturation scans, ellipticity of each protein sample was converted into fractional ellipticity with respect to the signal at 100% denatured state in order to normalize the data for all samples. Protein concentration of each construct ranged from 50 μ M to 300 μ M, in 20 mM sodium phosphate (pH 7.0) and 100 mM NaCl buffer.

Analytical Ultracentrifugation Experiments

Analytical ultracentrifugation runs were conducted on an Optima XL-A centrifuge (Beckman Coulter Instruments, CA). The data were collected at a wavelength of 280 nm.

Sedimentation Velocity (SV). SV experiments were conducted with 200 μ L- 300 μ l samples in 20mM sodium phosphate pH 7.0, 100mM NaCl at protein concentrations of 0.5-1mg/ml. Absorbance data were acquired at rotor speeds of 35,000- 40,000 rpm and at a temperature of 20° C. The buffer density of 1,005g/mL and viscosity of 1.031 cPoise and the protein partial-specific volumes were calculated using the software SEDNTERP, kindly provided by Dr. J. Philo. Data were analyzed using the c(s) continuous distribution of Lamm equation solutions with the software SEDFIT (47, 48).

Sedimentation Equilibrium. Sedimentation equilibrium experiments were conducted at 20°C at rotor speeds of 15,000 rpm at an optical density of 0.283. Global nonlinear regression of the experimental absorbance profiles was performed using the software SEDPHAT, kindly provided by Dr. P.Schuck.

References

1. Kastan, M. B., Zhan, Q., el-Deiry, W. S., Carrier, F., Jacks, T., Walsh, W. V., Plunkett, B. S., Vogelstein, B., and Fornace, A. J., Jr. (1992) *Cell* 71, 587-97.
2. el-Deiry, W. S., Tokino, T., Velculescu, V. E., Levy, D. B., Parsons, R., Trent, J. M., Lin, D., Mercer, W. E., Kinzler, K. W., and Vogelstein, B. (1993) *Cell* 75, 817-25.
3. Symonds, H., Krall, L., Remington, L., Saenz-Robles, M., Lowe, S., Jacks, T., and Van Dyke, T. (1994) *Cell* 78, 703-11.
4. Levine, A. J., Hu, W., and Feng, Z. (2006) *Cell Death Differ* 13, 1027-36.
5. Samuels-Lev, Y., O'Connor, D. J., Bergamaschi, D., Trigiant, G., Hsieh, J. K., Zhong, S., Campargue, I., Naumovski, L., Crook, T., and Lu, X. (2001) *Mol Cell* 8, 781-94.
6. Tang, Y., Luo, J., Zhang, W., and Gu, W. (2006) *Mol Cell* 24, 827-39.
7. Sykes, S. M., Mellert, H. S., Holbert, M. A., Li, K., Marmorstein, R., Lane, W. S., and McMahon, S. B. (2006) *Mol Cell* 24, 841-51.
8. Lu, W. J., and Abrams, J. M. (2006) *Cell Death Differ* 13, 909-12.
9. Yang, A., and McKeon, F. (2000) *Nat Rev Mol Cell Biol* 1, 199-207.
10. Yang, A., Kaghad, M., Wang, Y., Gillett, E., Fleming, M. D., Dotsch, V., Andrews, N. C., Caput, D., and McKeon, F. (1998) *Mol Cell* 2, 305-16.
11. Kaghad, M., Bonnet, H., Yang, A., Creancier, L., Biscan, J. C., Valent, A., Minty, A., Chalon, P., Lelias, J. M., Dumont, X., Ferrara, P., McKeon, F., and Caput, D. (1997) *Cell* 90, 809-19.
12. De Laurenzi, V., Costanzo, A., Barcaroli, D., Terrinoni, A., Falco, M.,

- Annicchiarico-Petruzzelli, M., Levrero, M., and Melino, G. (1998) *J Exp Med* 188, 1763-8.
13. Schultz, J., Ponting, C. P., Hofmann, K., and Bork, P. (1997) *Protein Sci* 6, 249-53.
 14. Serber, Z., Lai, H. C., Yang, A., Ou, H. D., Sigal, M. S., Kelly, A. E., Darimont, B. D., Duijf, P. H., Van Bokhoven, H., McKeon, F., and Dotsch, V. (2002) *Mol Cell Biol* 22, 8601-11.
 15. Mills, A. A., Zheng, B., Wang, X. J., Vogel, H., Roop, D. R., and Bradley, A. (1999) *Nature* 398, 708-13.
 16. Yang, A., Schweitzer, R., Sun, D., Kaghad, M., Walker, N., Bronson, R. T., Tabin, C., Sharpe, A., Caput, D., Crum, C., and McKeon, F. (1999) *Nature* 398, 714-8.
 17. Yang, A., Walker, N., Bronson, R., Kaghad, M., Oosterwegel, M., Bonnin, J., Vagner, C., Bonnet, H., Dikkes, P., Sharpe, A., McKeon, F., and Caput, D. (2000) *Nature* 404, 99-103.
 18. Ollmann, M., Young, L. M., Di Como, C. J., Karim, F., Belvin, M., Robertson, S., Whittaker, K., Demsky, M., Fisher, W. W., Buchman, A., Duyk, G., Friedman, L., Prives, C., and Kopczynski, C. (2000) *Cell* 101, 91-101.
 19. Brodsky, M. H., Nordstrom, W., Tsang, G., Kwan, E., Rubin, G. M., and Abrams, J. M. (2000) *Cell* 101, 103-13.
 20. Derry, W. B., Putzke, A. P., and Rothman, J. H. (2001) *Science* 294, 591-5.
 21. Huyen, Y., Jeffrey, P. D., Derry, W. B., Rothman, J. H., Pavletich, N. P., Stavridi, E. S., and Halazonetis, T. D. (2004) *Structure (Camb)* 12, 1237-43.

22. Lee, W., Harvey, T. S., Yin, Y., Yau, P., Litchfield, D., and Arrowsmith, C. H. (1994) *Nat Struct Biol* 1, 877-90.
23. Thanos, C. D., Goodwill, K. E., and Bowie, J. U. (1999) *Science* 283, 833-6.
24. Chi, S. W., Ayed, A., and Arrowsmith, C. H. (1999) *Embo J* 18, 4438-45.
25. Chene, P. (2001) *Oncogene* 20, 2611-7.
26. Jeffrey, P. D., Gorina, S., and Pavletich, N. P. (1995) *Science* 267, 1498-502.
27. Aviv, T., Lin, Z., Lau, S., Rendl, L. M., Sicheri, F., and Smibert, C. A. (2003) *Nat Struct Biol* 10, 614-21.
28. Green, J. B., Gardner, C. D., Wharton, R. P., and Aggarwal, A. K. (2003) *Mol Cell* 11, 1537-48.
29. Brodsky, M. H., Weinert, B. T., Tsang, G., Rong, Y. S., McGinnis, N. M., Golic, K. G., Rio, D. C., and Rubin, G. M. (2004) *Mol Cell Biol* 24, 1219-31.
30. Ahn, J., and Prives, C. (2001) *Nat Struct Biol* 8, 730-2.
31. Aguinaldo, A. M., Turbeville, J. M., Linford, L. S., Rivera, M. C., Garey, J. R., Raff, R. A., and Lake, J. A. (1997) *Nature* 387, 489-93.
32. Sidow, A., and Thomas, W. K. (1994) *Curr Biol* 4, 596-603.
33. Hedges, S. B. (2002) *Nat Rev Genet* 3, 838-49.
34. Mateu, M. G., and Fersht, A. R. (1999) *Proc Natl Acad Sci U S A* 96, 3595-9.
35. Mateu, M. G., and Fersht, A. R. (1998) *Embo J* 17, 2748-58.
36. McCoy, M., Stavridi, E. S., Waterman, J. L., Wieczorek, A. M., Opella, S. J., and Halazonetis, T. D. (1997) *Embo J* 16, 6230-6.
37. Kwan, J. J., Warner, N., Maini, J., Chan Tung, K. W., Zakaria, H., Pawson, T., and Donaldson, L. W. (2006) *J Mol Biol* 356, 142-54.

38. Hofmann, E. R., Milstein, S., Boulton, S. J., Ye, M., Hofmann, J. J., Stergiou, L., Gartner, A., Vidal, M., and Hengartner, M. O. (2002) *Curr Biol* 12, 1908-18.
39. el-Deiry, W. S., Kern, S. E., Pietenpol, J. A., Kinzler, K. W., and Vogelstein, B. (1992) *Nat Genet* 1, 45-9.
40. Nicholls, C. D., McLure, K. G., Shields, M. A., and Lee, P. W. (2002) *J Biol Chem* 277, 12937-45.
41. Melacini, G. (2000) *J. Am. Chem. Soc.* 122, 9735-9738.
42. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J Biomol NMR* 13, 289-302.
43. Linge, J. P., O'Donoghue, S. I., and Nilges, M. (2001) *Methods Enzymol* 339, 71-90.
44. DeLano, W. L. (2002), DeLano Scientific, San Carlos, California.
45. Koradi, R., Billeter, M., and Wuthrich, K. (1996) *J Mol Graph* 14, 51-5, 29-32.
46. Kleywegt, G. J. (1996) *Acta Crystallogr D Biol Crystallogr* 52, 842-57.
47. Schuck, P. (2000) *Biophys J* 78, 1606-1619.
48. Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J., and Schubert, D. (2002) *Biophys J* 82, 1096-111.

Figure 2-1. Domain architecture of the p53 protein family

(A) The p53 protein family can be categorized into two classes by the number of individual domains: p53-like proteins have three domains, and p63-like protein have four domains. TA, transactivation domain; DBD, DNA binding domain; OD, oligomerization domain; SAM, sterile alpha motif domain. Question marks indicate domains with low sequence homology to known domains that however retain the same fold. Numbers inside the DBD box represent sequence identity in comparison with human p53. (B) Secondary structure elements of the C-terminal domains of CEP-1 and Dmp53. The colored letters correspond to the domain color designation in A. (C) The overall structure of the C-terminus of CEP-1 reveals its dimeric structure, containing an oligomerization domain (green), a SAM domain (orange), and a linker region (blue). The second monomer is shown in yellow. (D) The overall structure of the C-terminal domain of Dmp53. The monomers are colored in magenta (chain A), green (chain B), blue (chain C), cyan (chain D).

Figure 2-1.

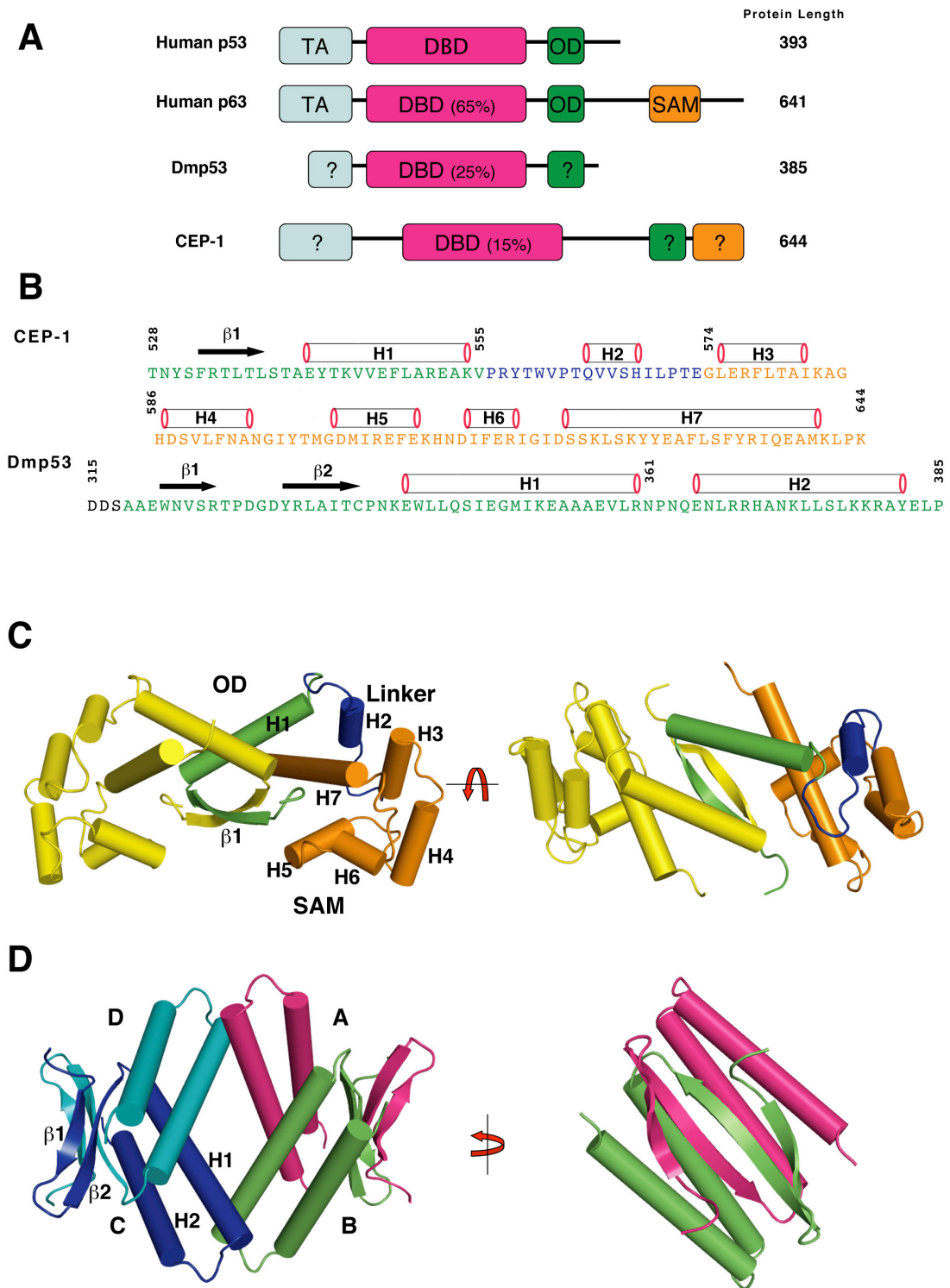
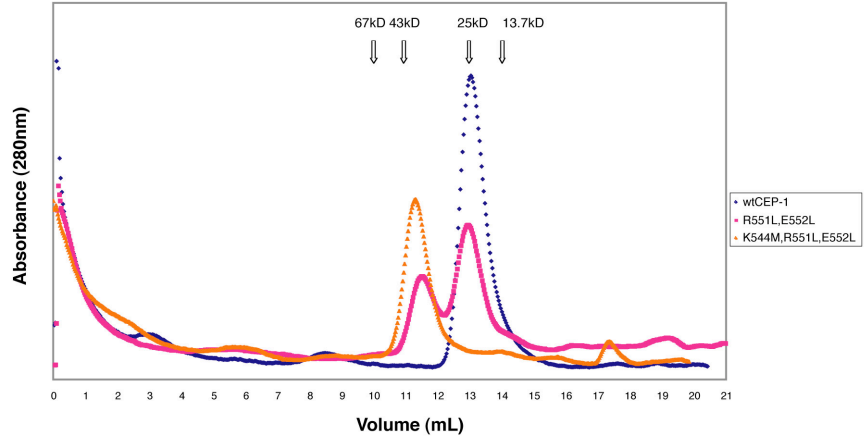


Figure 2-2. Gel filtration curves of the C-terminus of CEP-1 and Dmp53

(A) Gel filtration curve of the C-terminus of CEP-1 and two mutant proteins. The R551L/E552L mutant displays an equilibrium between dimer and tetramer. In the K544M/R551L/E552L triple mutant, this equilibrium is strongly shifted towards the tetramer. (B) Gel filtration experiments with the C-terminus of Dmp53 and different mutants. Mutants with deletion of strand β 1 (326-385) or helix H2 (315-361) or carrying the point mutation E353K all convert the tetramer into a dimer.

Figure 2-2.

A



B

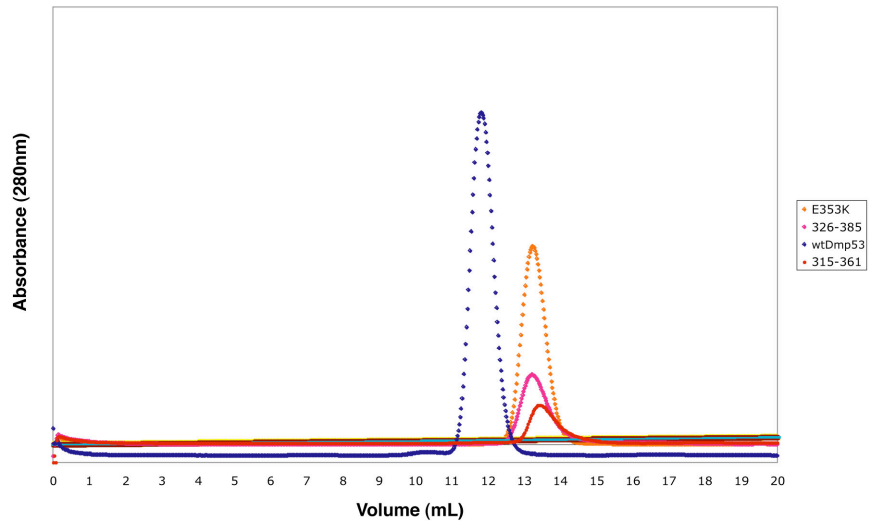
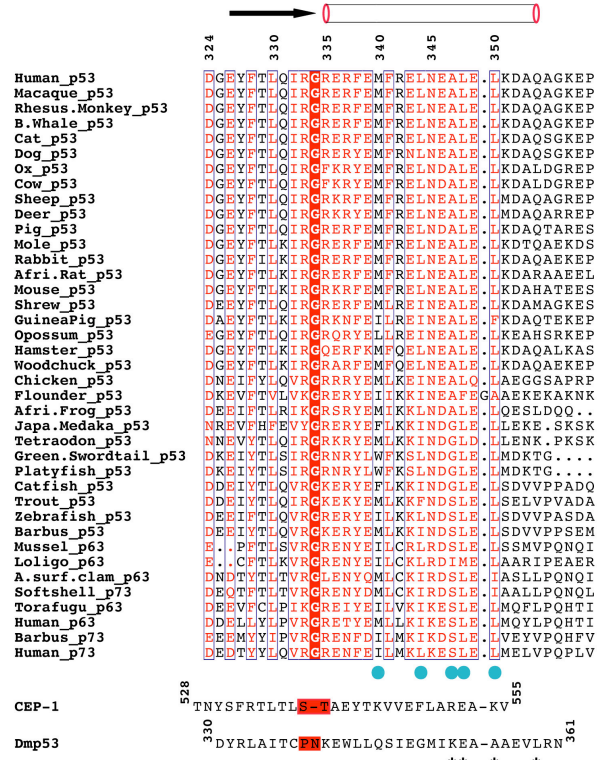


Figure 2-3. Sequence divergence of the OD in CEP-1 and Dmp53 relative to other p53 protein family members

(A) The signature hinge residue G334 (human p53 numbering) is conserved across many different species, as well as the five residues, marked by cyan dots, that are involved in the tetrameric interface. In both CEP-1 and Dmp53, the sequence similarity is very low in comparison to other members of the p53 protein family, especially at the tetrameric interface. The di-peptides in CEP-1 and Dmp53 that replaced the glycine residue are shaded in red. Residues marked by an asterisk in the Dmp53 sequence are important for the tetramerization interface. (B) The highly conserved glycine residue (colored in red) of the p53 OD is replaced with the di-peptides, S538 / T539 in CEP-1 and P338 / N339 in Dmp53.

Figure 2-3.

A



B

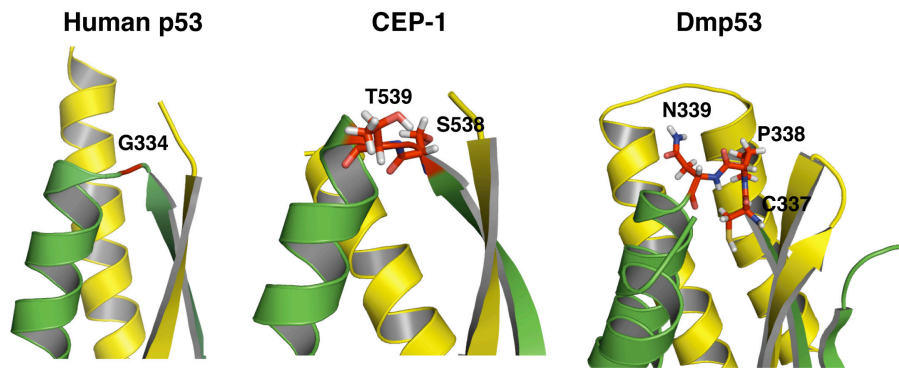


Figure 2-4. The tetrameric interface in the oligomerization domain of human p53, CEP-1 and Dmp53

(A) In human p53, M340, L344, A347, L348, L350 constitute the tetrameric interface. The electrostatic map shows the tetrameric interface is mostly hydrophobic. Positive charged surface is colored blue, negative charged surface is colored red, and non-polar surface is colored white. The second dimer subunit is removed for clarity. (B) The electrostatic map of the potential tetramerization interface of the CEP-1 OD shows the charged residues (K544, R551, and E552) surrounding F548 from each monomer. (C) The tetrameric interface in Dmp53 consists of a charged cluster in the center, and hydrophobic contacts on the outer edge of the interface. K352 and E353 from one monomer form salt bridges with K352 and E353 across the tetrameric interface. At the outer edge, L359 from helix H1 makes contacts with A371 and L375 in helix H2 across the tetrameric interface.

Figure 2-4.

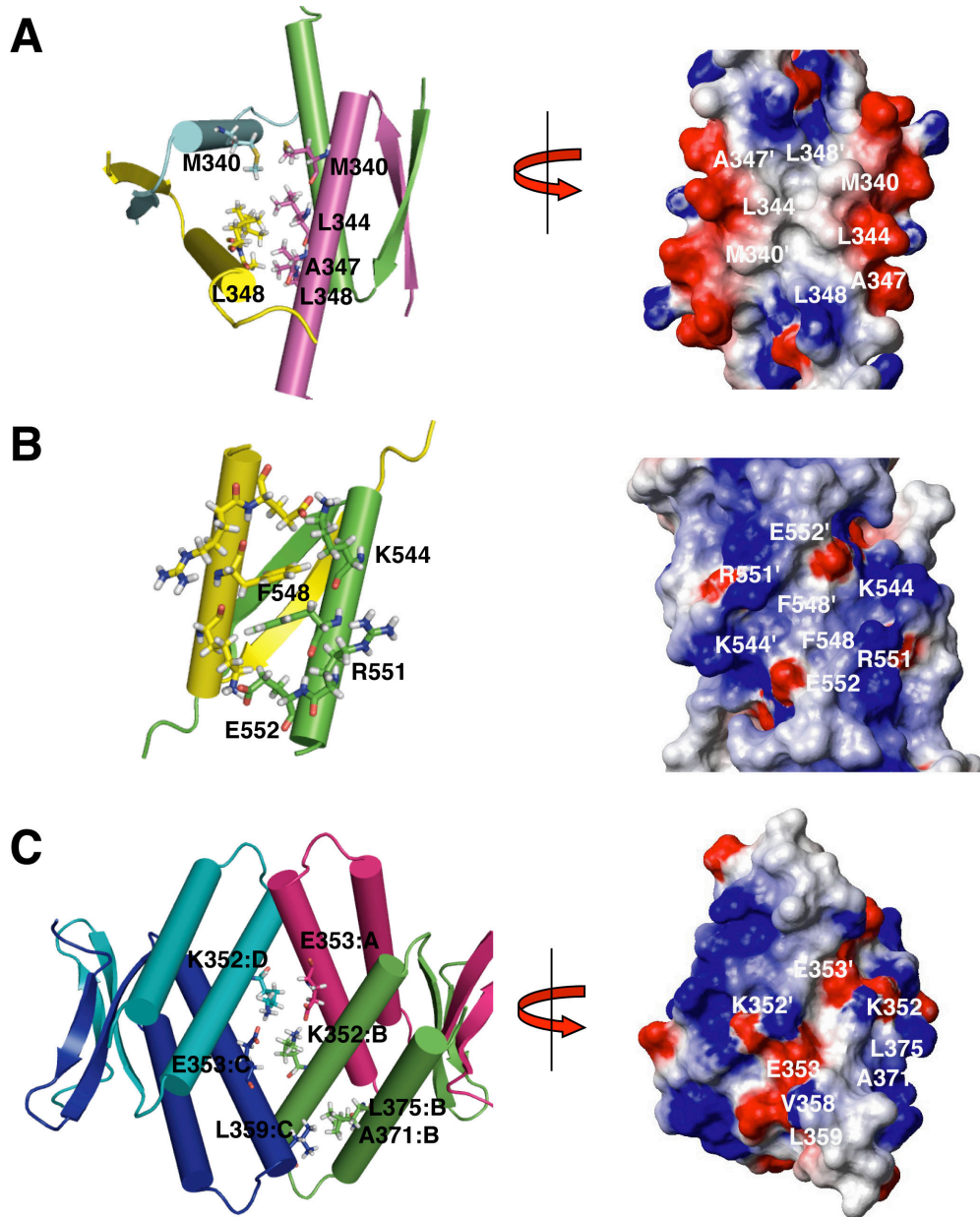


Figure 2-5. Oligomerization domain of CEP-1 and Dmp53 are stabilized by additional structural elements

In Dmp53, extensive contacts between the additional structural elements (strand $\beta 1$ and helix H2) and the canonical OD of p53 are present with 1902 Å² of buried solvent accessible area. The ¹⁵N-HSQC spectra show deletion of either structural element destabilizes the OD and results in multiple conformations as shown by broad peaks in the NMR spectra. Left panel: wild type OD of Dmp53. Middle panel: deletion of the H2 helix. Right panel: deletion of the $\beta 1$ strand.

Figure 2-5.

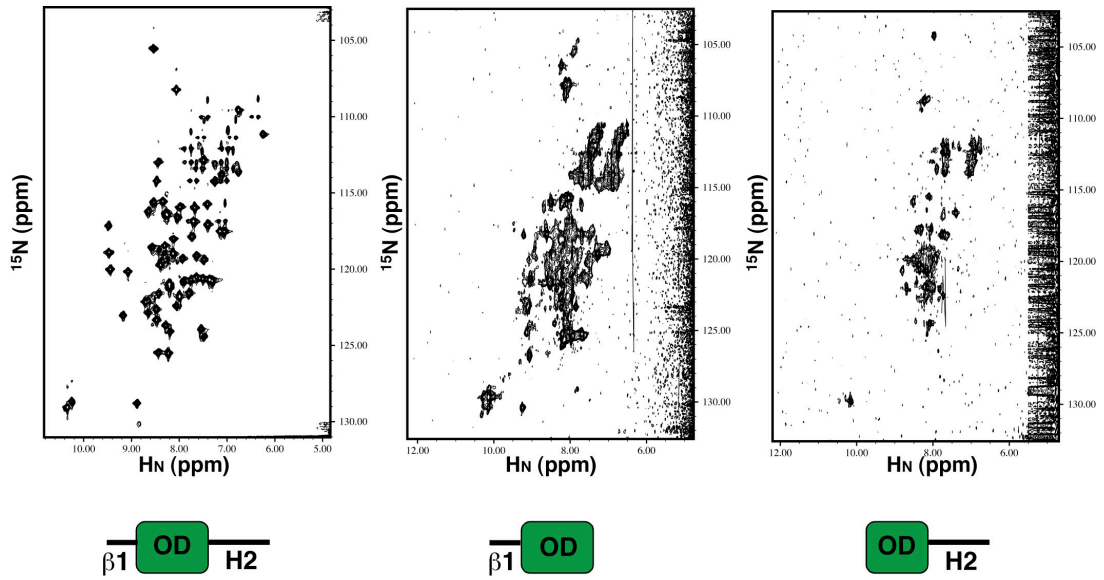
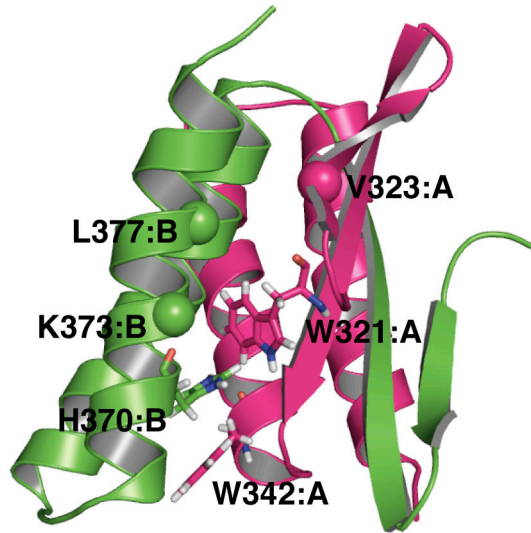


Figure 2-6. Different hypothesis on the evolution of the p53 protein family

(A) A phylogenetic tree of the p53 protein families based on the ecdysozoa topology shows that the SAM domain has appeared in all protostomes identified to date that have p53 like molecule, except in the arthropoda phyla, in which a helix has replaced the SAM domain in stabilizing the OD. In the deuterostome branch of the Urochordate phylum, the p53 protein first appeared in two isoforms: one form with the SAM domain, and another form without the SAM domain. (p63/p73) denotes the protein could be classified as either p63 or p73. Question mark indicates no significant sequence homology by sequence alignment.

(B) A comparison of a phylogenetic tree based on the ecdysozoa topology and on the coelomata topology predicts different oligomeric states for the ancestral p53 form. Each red circle represents a monomer and each bar symbolizes the C-terminal tail. In the coelomata topology, a dimeric form of p53 represents the ancestral form which evolved into the current vertebrate and mollusc p53 tetrameric form as well as into the *Drosophila* form that uses a different mode of tetramerization. In the ecdysozoa topology, the ancestral form was either a tetramer and the dimeric CEP-1 form evolved from this tetrameric state or the ancestral form was a dimer and tetramerization in the vertebrate branch and in the mollusc branch has evolved twice independently.

Figure 2-6.

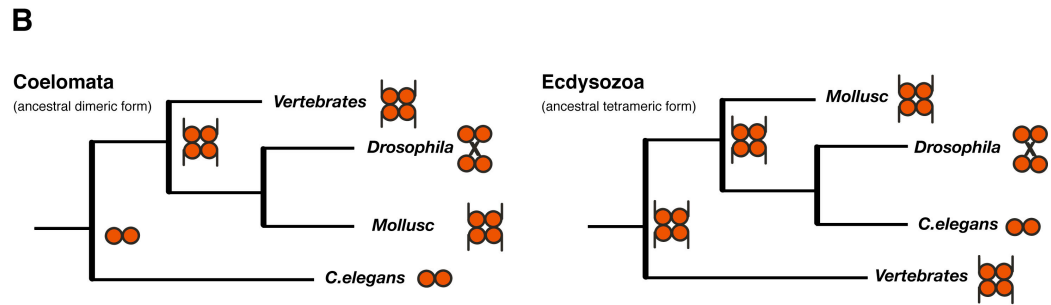
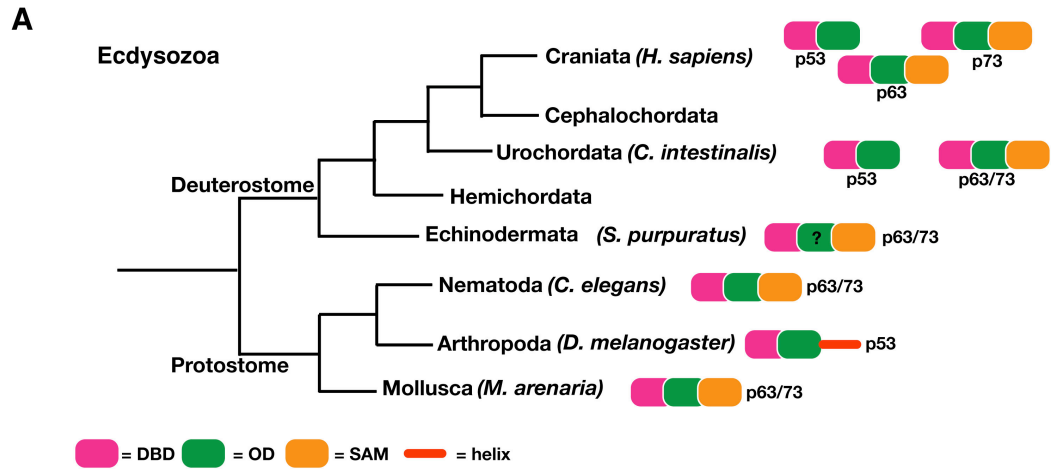


Figure 2-7. Sequence alignment of CEP-1 in *C. elegans* and *C. briggsae*

Sequence comparison between two different species of Nematode reveals two conserved regions after residue 510. The oligomerization domain of CEP-1 is colored green, and the SAM domain of CEP-1 is colored orange. Identical residues are marked by “*”, conserved substitutions are marked by “:”, and semi-conserved substitutions are marked by “.”.

Figure 2-7.

```

                    506                               528                               555
C.elegans  CVPPIETEHEHCQSPSMKRSRCTNYSFRTLTLSTAEYTKVVEFLAREAKVPRYTWVPTQV
C.briggsae  ---FYRKFKENEDSLSNKRPRSQYGLQRQVKLSEKEYSKFVAFFAKEGENEISKYASAHC
              .. :** :* * **.*. * :.** **:*. * *:**.*.: .....:

                    574                               620
C.elegans  VS-----HILPTEGLERFLTAKAGHDSVLFNANGIYTMGDMIREFEKHNDIFERIGIDS
C.briggsae  LTPAQASRLDPSDKIEKFLAFVGDESAADNFRKHGLFTMLDLDKYFQVYDSAFETIGVDS
              ::      :: *:: :*:**:: :      : *. :*:** * : : * : :. ** **:*

                    641
C.elegans  SKLSKYYEAFLSFYRIQEAMLPK
C.briggsae  SKMEKYYDLFLHYHRVQENIR---
              ** :.***: ** :*:** ::

```

Figure 2-8. CD spectra of the C-terminus of CEP-1 in full length and the OD alone

The top graph shows a wavelength scan of CEP-1 with just the OD (528-555). The lower graph shows the CD spectrum of the temperature denaturation curve of the full length C-terminus of CEP-1(528-641) and OD alone. The full length C-terminus of CEP-1 displays a sigmoidal transition at 41°C. In contrast, the OD alone seems to undergo aggregation during temperature denaturation as shown by the downward CD curve. CD signals are absolute values from measurements.

Figure 2-8.

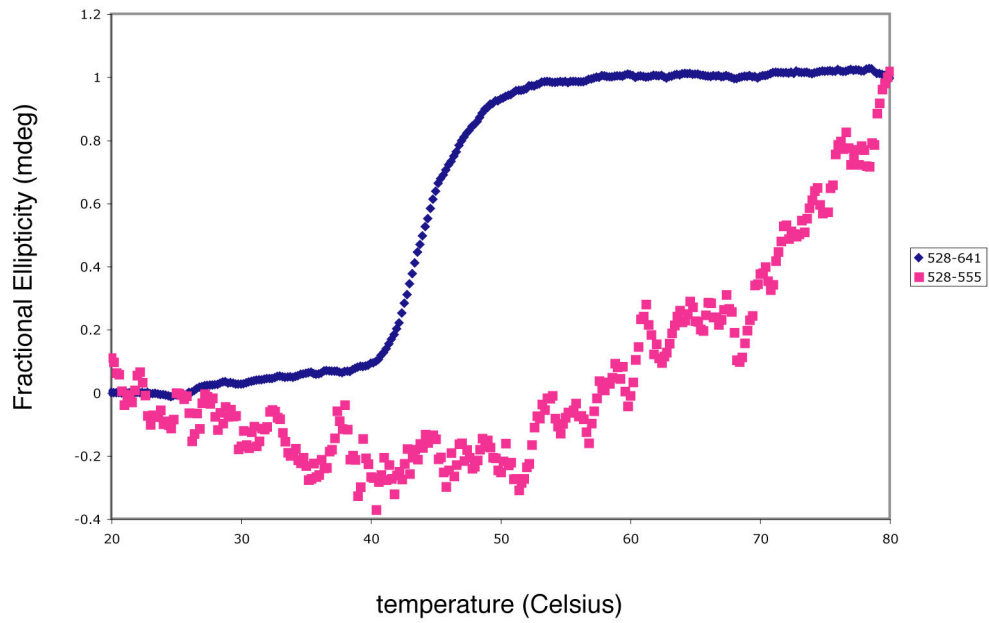
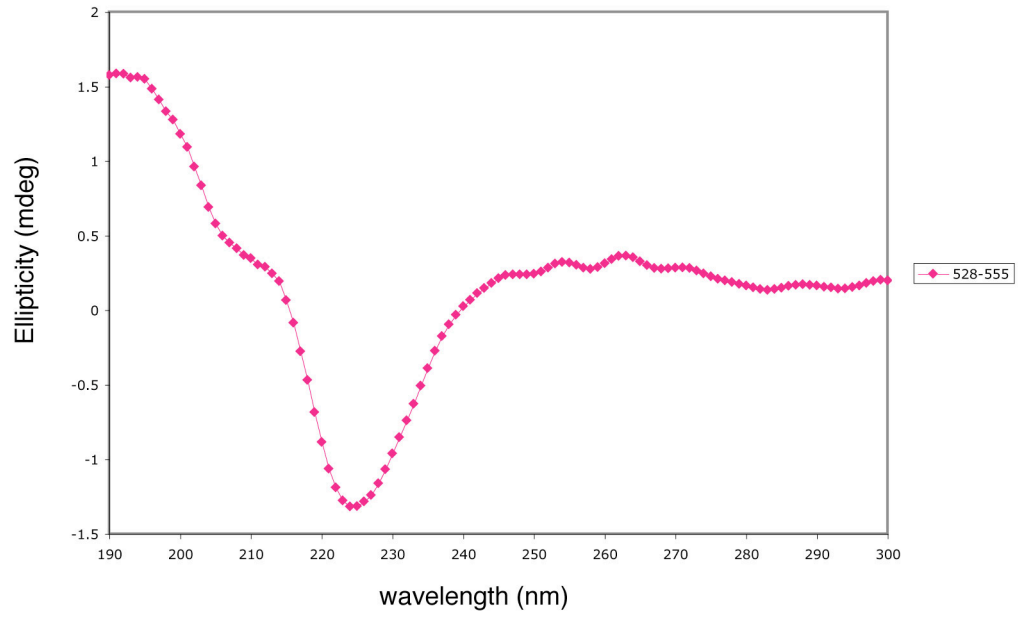


Figure 2-9. Structural alignment of individual domain in CEP-1 and Dmp53

Top figures show the superimposed structures of the human p53 OD (magenta) and CEP-1 OD (yellow) with an rmsd of 2.0 Å. An alignment of human p53 OD (magenta) and Dmp53 OD (yellow) has a rmsd of 1.8 Å. Bottom figures show the superimposed structures of the EphB2 (green) and the CEP-1 SAM domain (red) with an rmsd of 1.45 Å. Superimposed structures of the p73 (blue) and the CEP-1 SAM domain (red) have an rmsd of 1.83 Å.

Figure 2-9.

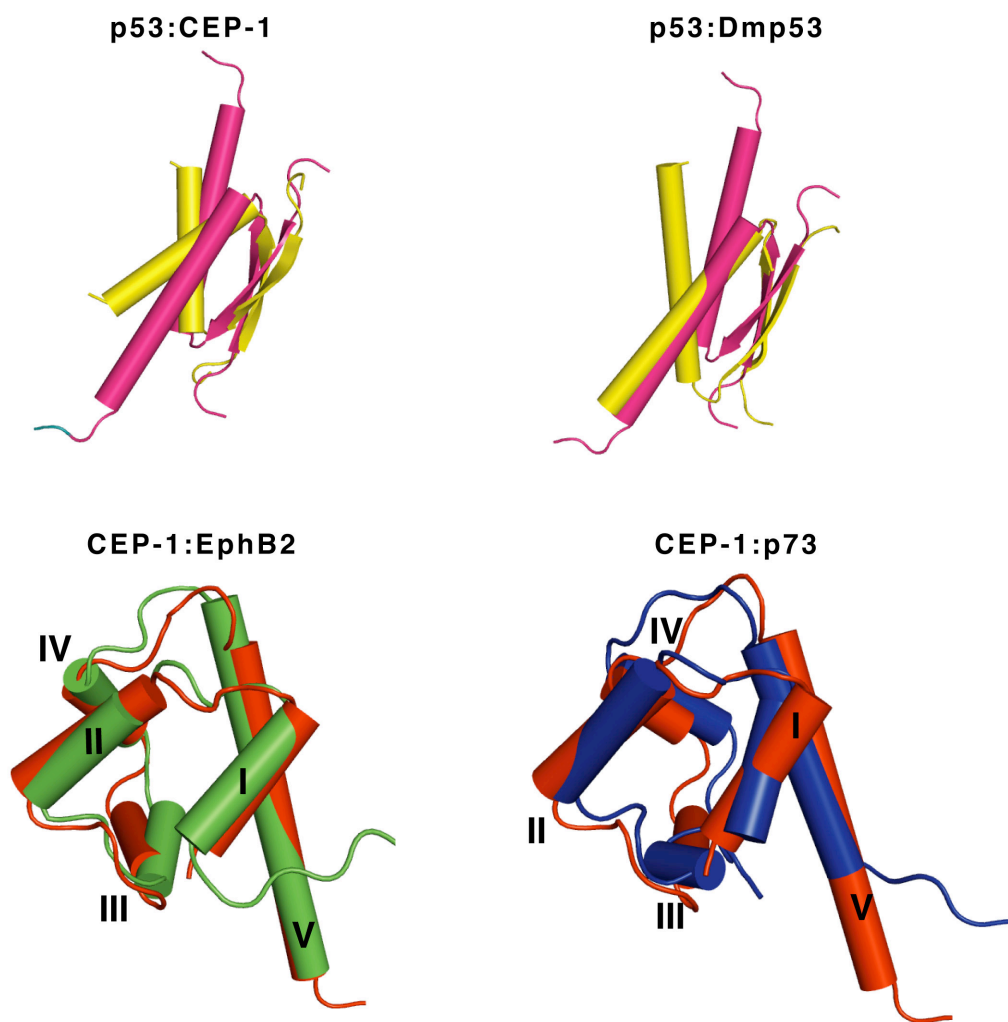


Figure 2-10. CD spectra of the Dmp53 and different deletion mutants

The top graph shows the wavelength scan of the full length C-terminus of Dmp53 (wt), a mutant with deletion of strand b1 (326-385), a mutant with deletion of helix H2 (315-361), and a E353K mutant. The bottom graph shows the CD temperature denaturation of the same proteins as in the top graph. The full length C-terminus of Dmp53 exhibits extreme stability with a transition point at 96°C. While the other deletion mutants have a gradual melting transition, an indication of multiple conformational states of the protein. CD signals are absolute values from measurements.

Figure 2-10.

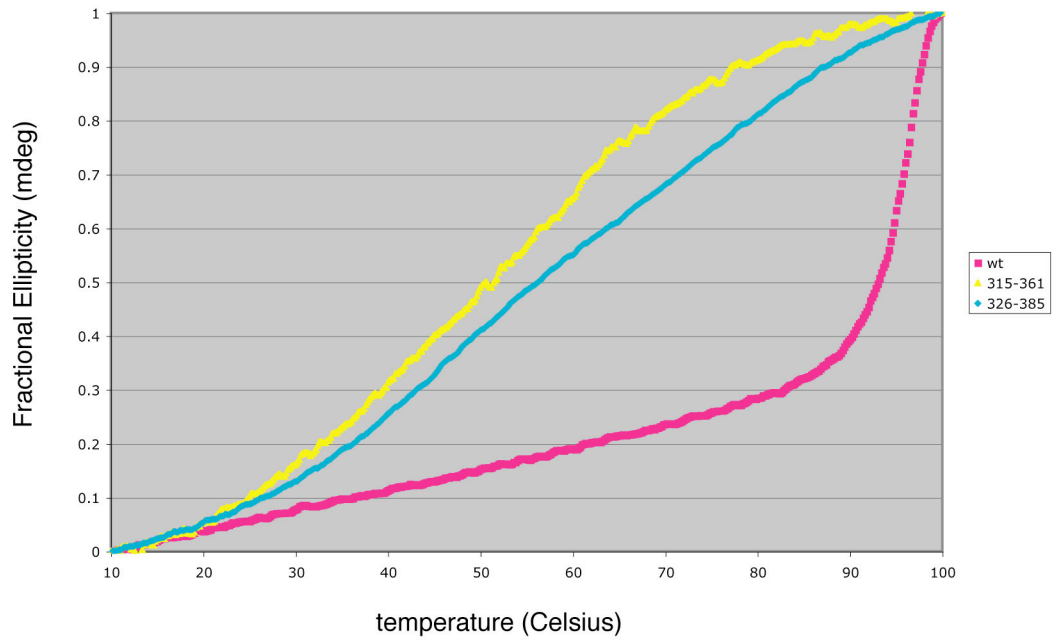
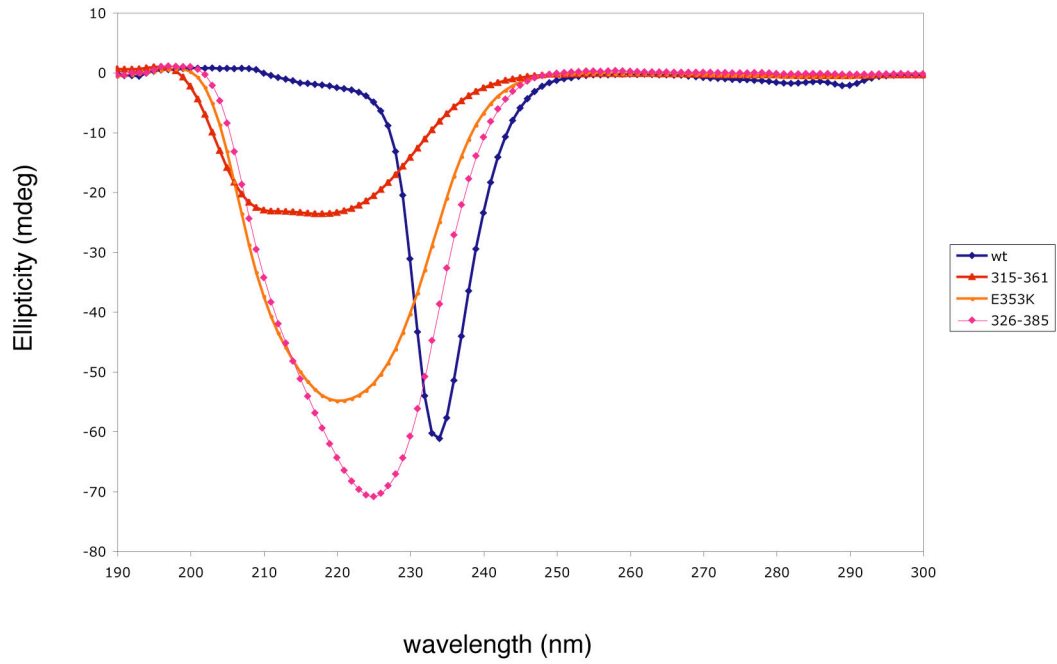


Table 2-1. Velocity sedimentation of the C-terminal domain of CEP-1 and Dmp53

| Protein | Variants | Theoretical Molecular Weight (Dalton) | Sed. Coeff. (S) | Experimental Molecular Weight (Dalton) | Oligomeric State |
|---------------------|-----------------|--|------------------------|---|-------------------------|
| CEP-1 C-terminus | 528-644 | 14004 | 2.39 | 28000 | dimer |
| | 552-644 | 11210 | 1.67 | 15200 | monomer |
| | 561-644 | 10079 | 1.45 | 9900 | monomer |
| | triple mutants | 13948 | 3.75 | 62600 | tetramer |
| Dmp53 C-terminus | 315-385 | 8560 | 3.07 | 32000 | tetramer |
| | 315-361 | 5422 | 1.33 | 12500 | dimer |

Triple mutants: K544M, R551L, E552L

Table 2-2. Structural Statistics of CEP-1

| | | | |
|---|----------------------------------|-------------|-------------|
| Distance restraints | | | |
| Total NOE | 2183 | | |
| Intra-residue | 838 | | |
| Inter-residue | | | |
| Sequential (i-j = 1) | 465 | | |
| Medium-range (i-j < 5) | 321 | | |
| Long-range (i-j > 5) | 445 | | |
| Intermolecular (ambiguous) | 114 (44 experimental determined) | | |
| Hydrogen bonds | 40 | | |
| Dihedral angle restraints | | | |
| Φ | 79 | | |
| Ψ | 80 | | |
| NOE Assignment Completeness | | | |
| Backbone amide protons | 93.8% | | |
| Sidechain protons | 70.0% | | |
| R.M.S. deviation from experimental restraints | | | |
| Distance restraints (Å) | 0.035 ± 0.003 | | |
| Dihedral angle restraints (°) | 1.22 ± 0.08 | | |
| Deviations from idealized geometry | | | |
| Bond lengths (Å) | 0.004 | | |
| Bond angles (°) | 0.559 | | |
| Improper (°) | 0.461 | | |
| Averaged pairwise r.m.s.d. (Å)^a | | | |
| | <u>OD+SAM</u> | <u>OD</u> | <u>SAM</u> |
| Backbone | 1.16 ± 0.49 | 0.52 ± 0.20 | 0.51 ± 0.11 |
| Heavy | 1.52 ± 0.43 | 0.92 ± 0.19 | 1.10 ± 0.17 |
| Number of conformers (out of) | 20 (100) | | |
| Ramachandran plot (percentage of residues)^b | | | |
| Core region: | 89.1 % | | |
| Allowed region: | 8.1 % | | |
| Generous region: | 2.4 % | | |
| Disallowed region: | 0.5 % | | |

a. Residues from 532-555 (OD) and 575-642 (SAM) were used for alignment.

a. Residues from 528-642 were used for the Ramachandran plot.

Table 2-3. Structural Statistics of Dmp53

| | |
|---|-----------------------------------|
| Distance restraints | |
| Total NOE | 1519 |
| Intra-residue | 561 |
| Inter-residue | |
| Sequential (li-jl = 1) | 268 |
| Medium-range (li-jl < 5) | 211 |
| Long-range (li-jl > 5) | 235 |
| Intermolecular (ambiguous) | 240 (184 experimental determined) |
| Hydrogen bonds | 46 |
| Dihedral angle restraints | |
| Φ | 52 |
| Ψ | 52 |
| NOE Assignment Completeness | |
| Backbone amide protons | 96.9% |
| Sidechain protons | 71.6% |
| R.M.S. deviation from experimental restraints | |
| Distance restraints (Å) | 0.055 ± 0.002 |
| Dihedral angle restraints (°) | 1.221 ± 0.052 |
| Deviations from idealized geometry | |
| Bond lengths (Å) | 0.007 |
| Bond angles (°) | 0.655 |
| Improper (°) | 0.522 |
| Averaged pairwise r.m.s.d. (Å)^a | |
| Backbone | 0.47 ± 0.15 |
| Heavy | 0.94 ± 0.18 |
| Number of conformers (out of) | 20 (100) |
| Ramachandran plot (percentage of residues)^b | |
| Core region: | 90.2 % |
| Allowed region: | 9.6 % |
| Generous region: | 0.2 % |
| Disallowed region: | 0.0 % |

a. Residues from 321-325, 330-335, 341-359, and 366-383 were used for alignment

b. Residues from 315-385 were used for the Ramachandran plot.

Chapter 3

Hay-Wells syndrome mutations alter the tertiary structure of the oligomerization domain, but not the SAM domain

Horng Der Ou^{1,2}, Jörg Rinnenthal¹, Frank Löhr¹, Vitali Vogel³, Werner Mäntele³ and Volker Dötsch^{1*}

¹ Institute of Biophysical Chemistry, Centre for Biomolecular Magnetic Resonance (BMRZ), J.W. Goethe University of Frankfurt, Max-von-Laue Str. 9, 60438 Frankfurt/Main

² Graduate Group in Biophysics, University of California San Francisco, 600 – 16th Street, San Francisco, CA 94143

³ Institute of Biophysics, J.W. Goethe University of Frankfurt, Max-von-Laue Str. 1, 60438 Frankfurt/Main

Abstract

Mutations in the SAM domain of p63 lead to the Hay-Wells and the Rapp-Hodgkin syndromes, characterized by fused eyelids (ankyloblepharon), ectodermal dysplasia, and cleft lip/palate. The molecular mechanisms of the syndromes due to the mutations remain unknown. We have solved the structure of the p63 SAM domain and the L514F mutant and, surprisingly, these mutations (L514F, G530V, T533P, Q536L) in the SAM domain do not cause gross structural changes, but rather decrease the stability of the protein. However, when the SAM domain is in a tetrameric context with the oligomerization domain (OD), the mutations disrupt the quaternary structure of the OD-SAM and cause aggregation. Deletion studies further show that the SAM domain is required to stabilize the OD. Thus, mutations in the SAM domain may perturb the overall structure of p63 and result in the misregulation of genes involved in the development of the epidermis.

Introduction

The SAM (sterile alpha motif) domain was first identified in 1995 in the yeast proteins, Byr2, Ste11, Ste50, and Ste4, as a small domain of 65-70 amino acids (1). The domain was named after its two characteristics: its alpha helical content, and mutations of the SAM domain in yeast proteins such as Ste11, Ste50, and Ste4 that prevent sexual differentiation and result in sterility. Since then, SAM domains have been found in a diverse array of proteins in surface receptors, post-synaptic proteins, polyhomeotic proteins, ser/thr kinases, RNA binding proteins, and transcription factors (2). The position of the SAM domain among these multi-domain proteins is either at the N-terminus or at the C-terminus, and one, two, or sometimes three tandem copies of the SAM domain are possible within a protein (3). The first crystal structure of the SAM domain in Ephrin receptor B2 (EphB2) revealed a topology of five helices, in which two pairs of helical hairpins are connected by a short helix (4).

The function of the SAM domain is mainly to mediate protein-protein interactions through homotypic- or heterotypic- SAM domains. Different oligomeric states of proteins can form from SAM mediated interactions, such as a dimer (Ste11 and Ste50) (5), a tetramer (three Ste4 and one Byr2) (6), or oligomers (TEL and Ph) (7, 8). Structural and biochemical studies have defined two areas of interaction: a mid-loop surface (ML) that surrounds helix 3, and an extended helix surface (EH) that involves helix 5. These two regions are located opposite from each other. SAM domains that form oligomeric interactions use both surfaces, and those that form a dimeric complex use only one of the interaction areas. SAM domains can also interact with non-SAM domain containing proteins, although structural data for such complexes are not presently

available (2). In addition to proteins, the SAM domain in the Smaug protein family can recognize RNA stem loop CNGGN (N represents any nucleotide) with nanomolar affinity through the EH region (9, 10).

Hay-Wells or Ankyloblepharon-Ectodermal defects-Cleft lip/palate (AEC) syndrome and a related Rapp-Hodgkin syndrome (RHS) are human autosomal dominant diseases caused by missense mutations in the SAM domain of p63 or insertion/deletion of nucleotide bases (Figure 3-1) (11-13). Patients of AEC usually have fused eyelids (ankyloblepharon), dry skin, cleft lip with or without palate, and coarse and sparse hair. Biopsy of these patients shows abnormal localization of p63 throughout all epidermal layers, in contrast to the normal restriction to the basal layer, where p63 is normally expressed. RHS patients exhibit abnormal development of teeth, sparse hair, cleft lip/palate, narrow nose, and small mouth. p63 is a paralog of the tumor suppressor p53, and knock-out mice of p63 have shown that p63 is involved in limb, craniofacial, and epidermal development (14, 15). The architecture of protein domains in p63 is similar to p53, but contains different C-termini. Besides the transactivation domain (TA), the DNA binding domain (DBD), and the oligomerization domain (OD), there is a poly-glutamine/proline region (QP), the SAM domain, and the transcription inhibitory domain (TID) (16). There are six isoforms of p63, which are combinations of two different translational starts (TA or Δ N) and three C-terminal variants (α , β , γ) (17). In epidermal cells, the Δ N form of p63 is predominant, while in oocytes, the TA form of p63 is highly expressed (17, 18).

Recent reports have identified RACK-1 and ABBP1 as two candidate partners of the SAM domain in p63 from a yeast two hybrid assay (19). ABBP1 plays a role in Fgf

splicing, in which its interaction with p63 is necessary for correct splicing; however, there is no direct biophysical evidence to support the interaction. The second protein interaction partner is RACK-1, the receptor for activated protein kinase C, but the interaction has not been fully characterized. A different study has also found RACK-1 to interact with the C-terminus of p73, another member of the p53 protein family with similar structural domain layout as p63, but it was not made clear whether the SAM domain was the sole interaction domain in p63 that associates with RACK-1 (20).

In our previous study, we have identified a SAM domain at the C-terminus of CEP-1, a p53 homolog in *C. elegans*. The structural study revealed that the last helix of the SAM domain has close contacts with the β -sheet and the α -helix of the OD domain in CEP-1. It was further shown that the SAM domain was necessary for the stability of the OD domain, in which a deletion of the SAM domain resulted in an unstable OD. The presence of the SAM domain in ancestral p63 suggests there may be a conservation of function that also applies to the SAM domain in mammalian p63. In this study, we characterize four mutants of the SAM domain identified in the Hay-Wells syndrome, and show that each mutation has minimal effect on the tertiary structure of the SAM domain in isolation, although each decreases the stability of the protein in comparison with the wild type SAM domain. However, when the SAM domain is placed in context with the OD of p63, each mutation disrupts the integrity of the OD-SAM molecule and transforms the oligomeric state of the OD from a tetramer into soluble aggregates. Thus, mutations in the Hay-Wells syndrome may lead to a gross conformational change in p63 and may cause aberrant transcriptional and repressive activities during epidermal development that result in the syndrome.

Results

Through genotyping of patients of AEC and RHS, 12 missense mutations and one in-frame insertion of phenylalanine were identified in the SAM domain of p63 (Figure 3-1) (21). The substitutions of amino acids are found in the first three helices of the domain at a mutational hot spot in helix 3. During the course of the study, the structure of the SAM domain of p73 was determined by NMR (22). p73, another paralog of p53, has a similar structural domain layout and splice variants as p63, although p73 is involved in the development of pheromone receptors, hypothalamus, and other neurons (23). Since the SAM domain of p63 and p73 shares 53% sequence identity, the structure of the SAM domain in p63 should be similar to that of p73 (p73SAM). Thus, a mapping of mutated residues from AEC and RHS on p73SAM could predict the structural consequence of mutations. Mutations in AEC and RHS can be classified into three categories: structural distortion, surface binding, and elimination of the TID. In the first category, a mutation introduces either a larger residue (L514F), a non-polar to a polar residue (I510T), or a disruption in the topology of the fold by unusual backbone dihedral angles (G530V, T533P). The second category belongs to the mutational hot spot, which has multiple mutated residues (Q536L, I537T, S541Y) in helix 3. Since the main function of the SAM domain is protein-protein interaction, mutations in the first two categories could lead to defects in protein interaction by the loss of the binding interaction surface. In order to understand the structural effects of these mutations, we used NMR to characterize these mutants. Despite the close homology, the chemical shifts of the SAM domain of p73 were not readily transferable to the p63 SAM domain. Therefore, we have undertaken the structural determination of the p63 SAM domain (p63SAM) and have compared the

chemical shifts of L514F, G530V, and Q536L mutants.

The SAM domain of p63 has identical topology layout to the SAM domain of p73 with a root mean square deviation of 2.68 Å (Figure 3-2). The p63SAM has two pairs of hairpin helices linked by a 3-10 helix. The domain is eluted as a monomer on the gel filtration column. This oligomeric property is similar to p73SAM, which is also a monomer in isolation, and differs from other SAM domains, which can form spontaneous dimer or oligomers (22). The purified SAM domain mutants of L514F, G530V, T533P, and Q536L also elute from the gel filtration column as monomers. Thus, mutations identified in the AEC do not seem to change the oligomeric property of the protein in isolation.

Although all examined AEC mutants behave as monomers, the behavior of each mutant is clearly different from the wild type since the protein yield of each after purification is significantly less, with a majority of proteins in inclusion bodies rather than in the supernatant. Based on the p63 structure, some mutations may disrupt the domain structure and unfold the protein. We investigated the stability of each mutant with circular dichroism (CD) spectroscopy by temperature denaturation. Even though the results presented here are not under equilibrium conditions since all proteins undergo irreversible precipitation, the data provide the relative kinetic stability of each mutant in comparison with the wild type SAM domain. As expected, the SAM domain exhibits the highest stability by having the highest melting temperature (78 °C). The Q536L and T533P mutants exhibit a similar denaturation profile (melting temperature of 63 °C), follow by L514F (53 °C) and G530V being the least stable mutant (42 °C) (Figure 3-3). Surprisingly, all mutants display a sigmoidal transition curve that suggests all retain a

folded structure albeit with different stabilities. Thus, mutations found in the Hay-Wells syndrome do not seem to cause a loss of tertiary structure even though mutants like the L514F, G530V, and T533P would suggest a significant disruption to the protein fold due to inefficient packing (L514F) and drastic changes to the protein backbone geometry (G530V and T533P) after amino acid substitutions (11).

To confirm the data from the CD denaturation curves and investigate the tertiary structure of the mutants, we recorded an HSQC spectrum of L514F, G530V, and Q536L mutants. In each mutant, the spectrum has a wide dispersion in chemical shifts, indicating a folded structure. An introduction of a different chemical environment due to substitution of amino acids in proteins can cause chemical shift changes between a normal and a mutated protein. As expected, a comparison of the G530V and Q536L mutants with the wild type SAM domain shows larger chemical shift changes in the local area around each mutation (in both ΔHN and ΔN frequencies) in comparison with the rest of the protein, which shows minor chemical shift changes, an indication that both mutants still retain a similar structure as the wild type protein (Figures 3-4B and 3-4C). Thus, both G530V and Q536L mutations do not seem to disrupt the native fold of the protein, but do affect the stability of the protein as observed from CD. In contrast to G530V and Q536L, the L514F mutant displays significant chemical shift changes throughout the whole domain, to the extent that most residues could not be identified due to large changes in their chemical shifts (Figure 3-4A). L514 plays a structural role in the SAM domain by making hydrophobic contacts with residues from helix 1 and 2. A phenylalanine substitution at this position may disrupt the packing between helix 1 and 2 because of a larger sidechain volume (from 124 \AA^3 to 135 \AA^3), and may consequently

perturb the whole structure. The drastic changes in chemical shifts of this mutant preclude any definitive conclusion about its tertiary structure; therefore, a full structural determination of L514F SAM domain was completed to answer this question.

The structure of the L514F mutant was solved by standard heteronuclear triple resonance pulse sequences for backbone and side chain assignment, and ^{15}N -edited HSQC and D_2O -NOESY spectra were recorded to obtain distance constraints. Surprisingly, when the mutant structure is compared with the wild type SAM domain, it has a rmsd of 2.75 Å, in which the differences in the structure mostly reside in the first three helices. The substituted phenylalanine residue occupies the same position as leucine. The structure accommodates the extra volume with no dramatic change to the overall fold. However, the increase in residue volume by phenylalanine allows closer contact with F526, and induces F554 to change its conformation from solvent exposed to interaction with F514 in the hydrophobic core of the domain (Figure 3-5). The core of the L514F mutant now constitutes three phenylalanines within 4 Å of each other. The interaction between phenylalanines may induce ring currents, that cause changes in the chemical shifts of most residues in the domain even though the overall structure does not change from the normal SAM domain of p63. The L514F mutant also shows slight structural differences from the wild type in the mutational hot spot of the Hay-Wells syndrome that includes T533, Q536, and I537. In the wild type domain, T533 is exposed to the solvent with no contact with I537, but in the mutant, T533 is rotated inward toward the core of the protein and interacts with I537 (Figure 3-5). If T533 were part of the SAM domain protein interaction surface, then the L514F mutant would distort this region and prevent successful binding.

In a previous study on the C-terminus of CEP-1 (*C. elegans* p53-like protein-1), the SAM domain of CEP-1 was shown to be necessary for the stability of the OD. Without the SAM domain, the OD is unstable and could disrupt the overall structure of CEP-1. A similar effect is observed in Dmp53 (*Drosophila* p53), in which the OD requires a C-terminal helix to be tetrameric, and the loss of the helix results in a dimer. Thus, the C-terminal helix seems to play an analogous role as the SAM domain of CEP-1 in stabilizing the OD. To investigate whether the SAM domain in p63 also plays an identical role, a fusion protein of maltose binding protein (MBP) along with all residues between the OD and the SAM domain of p63 (OD-SAM, residues 352-570) was constructed. Besides the wild type OD-SAM, four fusion proteins with different mutations in the SAM domain (L514F, G530V, T533P, and Q536L) were made. Each fusion protein was purified and tested for its oligomerization property. The wild type MBP-OD-SAM fusion protein exhibits a different behavior from the mutants in oligomerization. The wild type fusion protein elutes as a single peak with an apparent molecular weight of 440 kD, and all four mutants elute with two broad peaks: one peak at an apparent molecular weight of 669 kD, and the second peak at the void volume. The larger apparent molecular weights of mutants in comparison with the wild type suggest aggregation in the mutants (Figure 3-6). All of the apparent molecular weights from each protein exceed the estimated tetrameric molecular weight; thus, the gel filtration data is an indication of the stoke's radius of each molecule rather than its molecular weight. To assess the actual oligomeric state of the protein, the wild type, L514F, and Q536L fusion proteins were measured by velocity sedimentation (Figure 3-7). As expected, the wild type MBP-OD-SAM is mono-disperse with a molecular weight that corresponds to a

tetramer. In contrast, the L514F and Q536L OD-SAM fusion proteins have multiple species with the majority of protein centers around 1.8 MDa, which is a sign of aggregation. Attempts to purify the OD alone, or the OD-SAM constructs with mutations, were not successful since all these constructs result in insoluble inclusion bodies. To further test whether the SAM domain is necessary for stabilization of the OD, a construct from the OD to the QP (OD-QP, residues 352-499) was cloned and expressed, in which only the MBP fusion protein was soluble, but not the isolated protein. Gel filtration study of the MBP fusion of the OD-QP construct shows it also elutes like the OD-SAM fusion protein as a tetramer in gel filtration; however in velocity centrifugation, the majority of the protein behaves like a tetramer with a minor fraction as a dimer, in contrast to the mono-species behavior of the OD-SAM fusion protein (Figures 3-6 and 3-7). Thus, mutations in the Hay-Wells syndrome destabilize the OD-SAM complex, and the SAM domain is necessary for the stability of the OD in p63.

Discussion

Although the role of p63 in epidermal development awaits further clarification of whether it functions as a stem cell regenerator or an effector of commitment in differentiation, the existence of human patients with mutations in p63 and the resulting epidermal and limb abnormalities highlight its essential function during development (24). In the current study, we investigated the effect of mutations in the AEC syndrome on the structure of the SAM domain and found that two surface mutants (G530V, Q536L) show only localized chemical shift perturbation around the mutated residue, an indication that both mutants still retain the SAM domain fold. Unexpectedly, the structural mutant (L514F) maintains the SAM domain fold as well, except for small structural changes in the first two helices and the 3-10 helix, the site of hot spot mutations in AEC. Further investigation on each mutant in the context of the quaternary structure with the OD reveals a different behavior from the wild type SAM domain, which suggests a putative interaction between the OD and the SAM domain when the latter is in a tetrameric context. The similar behavior observed in gel filtration for all the mutants studied here seem to indicate a possible interaction area that involves the 3-10 helix of the SAM domain since it is an area that shows minute structural changes in all three mutants, and mutations of this area would disrupt the OD and SAM interaction.

The SAM domain has thus far been prevalently thought of as an interaction domain because of its ability to mediate homotypic or heterotypic protein interactions. In addition, a recent study has shown that the SAM domain is a RNA interaction domain (25). It was reported that RACK-1, ABBP1, and Scaf4/rA4 are interaction partners of the p63 SAM domain through a yeast two hybrid screen; however there is no direct

experimental evidence with in vitro purified proteins to verify these interactions. Indeed, the SAM domain does not interact with RACK-1 since the interaction still persists with the L514F SAM mutant, and another report has shown it is the extreme C-terminus of p73 α , a highly homologous protein to p63 α , that interacts with RACK-1 (20). Thus, more experiments are needed to test whether ABBP1 and Scaf4/rA4 are true SAM domain interaction partners. Furthermore, the yeast two hybrid screen used only the SAM domain as the bait, yet p63 α exists as a tetrameric state through the OD. The screen might not reveal a true partner since there could be a new interaction surface when the SAM domain is in the context of a tetramer. Besides being a protein interaction partner, the SAM domain can also be a protein scaffold domain that stabilizes the OD as shown in the CEP-1 C-terminus structure. The gel filtration data of a MBP fusion to the OD-SAM construct of the wild type and the four mutants show each mutant can disrupt the quaternary structure of the OD-SAM, although the exact nature of disruption remains to be elucidated. However based on the cell culture data of co-expression assays of wild type TA γ and Δ N α form of each mutant, each SAM domain mutant lost the dominant negative effect, or the ability to oligomerize with TA γ to suppress its activity (11).

Skin histology of a patient that carries the G530V mutation shows abnormal intense p63 protein staining that extends beyond the basal layer where p63 is normally observed. In addition, filaggrin, a marker of intermediate filament in terminally differentiated cells, also shows intense staining in the nucleus of the cells. These data in combination with the findings that each mutant can disrupt the tetrameric quaternary structure and a loss of the dominant negative effect suggest that the SAM domain mutation is a gain of function mutation or exhibits haploid insufficiency, rather than a

loss of function mutation. If the SAM domain mutations were a loss of function mutation, then patients of AEC and RHS syndrome would not exhibit any phenotype since they still retain a wild type copy of p63 because all AEC and RHS mutations are heterozygous. However, a gain of function or a reduced population of normal functional p63 by the SAM mutants could activate additional genes or fail to repress or activate genes at the appropriate time as demonstrated by microarray analysis of the TAp63 α Q536L mutant (26). The IKK α gene is activated by Δ N α to initiate epidermal differentiation (27), and misregulation of this gene due to mutations in the SAM domain could result in abnormal development of the epidermis. Genes such as p21 and 14-3-3 σ were repressed by Δ N α , but mutations in the SAM domain abolished the repression (28). The loss of the dominant negative effect in p63 by the SAM domain mutations could also cause AEC since Koster *et al* have shown the proliferative effect of TAp63 α was counter-balanced by the inhibition of Δ N α for proper epidermal differentiation, and an imbalance in the ratio of TAp63 α to Δ N α resulted in mice that have skin fragility similar to ectodermal dysplasia observed in AEC (29).

One reason for the difficulty to dissect the mutational effects of the AEC in epidermal development was the complexity of the p63 due to its six splice variants, in which the TAp63 α and the Δ N α are the dominant forms in keratinocytes that play a role in epidermal development (29). Both isoforms have a SAM domain, and mutations from AEC could have an effect on the normal functions of both proteins (24). This study has shown that the SAM domain may not just be a passive domain that recruits other proteins, but rather a crucial domain for the overall structural integrity of p63. This alteration in the structure may change the DNA binding specificity of the DBD, and upset

normal gene expression in epidermal development.

Materials and Methods:

Cloning and Protein Expression

Wild type and mutant SAM domain proteins

PCR fragment that corresponds to the SAM domain of p63 was cloned into pGEX vector (Amersham) as a GST fusion protein. PCR fragments of each SAM domain mutant: L514F, G530V, T533P, and Q536L were derived from plasmids provided by Dr. Hans van Bokhoven and cloned into either pGEX 6P-2 or pGEX. Each protein was expressed in BL21 cells, which were grown in 2X YT media to an OD₆₀₀ between 0.6-0.8, and induced with 500 μ M of IPTG at 24 °C for 5-6 hours. After harvesting the culture by centrifugation, the pellet was resuspended in standard PBS (phosphate buffered saline) buffer and stored at -20 °C overnight. Protein purification was done with GST sepharose beads and closely followed the manufacturer's protocol. The major exception was that 0.1% sarkosyl was added to the cell lysate of each mutant before sonication for better solubility of each mutant. After binding to a GST column and extensive washing with PBS buffer, the fusion protein was cleaved on column overnight at 4 °C in protease cleavage buffer of either the precision protease or the thrombin. 50 units of protease were used for digestion. Next day, the cleaved protein was eluted, concentrated and further purified on a superdex 75 gel filtration column in a buffer consists of 100 mM NaCl and 20 mM HEPES at pH 7.0. Labeling of ¹⁵N and ¹⁵N/¹³C proteins for NMR study was done by first growing 2L of bacteria cells in 2X YT media to an OD₆₀₀ between 0.6-0.8, and cell pellet was harvested by centrifugation. Next, the pellet was resuspended in minimal media supplemented with either ¹⁵N labeled ammonium chloride or ¹⁵N labeled

ammonium chloride and ^{13}C labeled glucose before induction. Purification of labeled proteins followed the same protocol as described above.

MBP fusion of OD-SAM proteins

A PCR fragment of p63 from the OD to the SAM domain (352-570) was cloned into a MBP fusion vector. The protein was expressed in DH5 α cells and grew to OD₆₀₀ between 0.6-0.8 in 2X YT media. Before induction, a final concentration of 300 mM NaCl and 20 mM proline was added to the media to aid solubility of the fusion protein *in vivo*. 500 μM of IPTG was added to induce protein expression for 8 hours at 24 $^{\circ}\text{C}$. The protein was purified over an amylose column in a wash buffer of 300 mM NaCl, 1 mM EDTA, and 50 mM Tris, pH 8.0. After extensive washing, the protein was eluted in 10 mM maltose with the wash buffer.

NMR Experiments and Structure Calculations

The ^{15}N -HSQC spectra of the wild type and each mutant were recorded on a 500 MHz triple resonance magnet with cryo-probe. The backbone and sidechain resonances of the wild type SAM domain were assigned with spectra from HNCA, HNCACB, CBCACONH, (H)CCOHN-TOCSY, and H(C)COHN-TOCSY. NOE distance constraints were obtained from ^{15}N -NOESY, ^{13}C -NOESY, aromatic D₂O-NOESY, and D₂O-NOESY experiments. For the L514F mutant, ^{15}N -NOESY, aromatic D₂O-NOESY and D₂O-NOESY spectra were collected for distance constraints. Based on chemical shifts of N, CA, HA, and HB, backbone dihedral angles restraints were calculated using TALOS (30). NOE assignments and structure calculations were carried out first in CYANA (31, 32), and ARIA was used for water refinement (33). Structural alignment

was done with LSQMAN (34), and all figures were prepared with PyMOL (35).

Circular Dichroism Experiments

The wild type and each mutant of the SAM domain were tested for their relative stability by temperature denaturation from 10 °C to 90 °C and observed by circular dichroism at 220 nm. The protein was in a buffer consisting of 100 mM NaCl and 20 mM HEPES at pH 7.0 for each measurement. The mid transition of the melting curve was determined from the derivative of the experimental melting curve.

Velocity Sedimentation Experiments

Experiments were conducted with 200 μ L- 300 μ L samples in a buffer consisting of 100 mM NaCl and 20 mM sodium phosphate at pH 7.0, at protein concentrations of 0.5-1 mg/mL. Absorbance data were acquired at a rotor speed of 20,000 rpm and at a temperature of 20 °C. The buffer density of 1,005 g/mL and viscosity of 1.031 cPoise and the protein partial-specific volumes were calculated using the software SEDNTERP. Data were analyzed using the c(s) continuous distribution of Lamm equation solutions with the software SEDFIT (36, 37).

References

1. Ponting, C. P. (1995) *Protein Sci* 4, 1928-30.
2. Kim, C. A., and Bowie, J. U. (2003) *Trends Biochem Sci* 28, 625-8.
3. Schultz, J., Ponting, C. P., Hofmann, K., and Bork, P. (1997) *Protein Sci* 6, 249-53.
4. Thanos, C. D., Goodwill, K. E., and Bowie, J. U. (1999) *Science* 283, 833-6.
5. Kwan, J. J., Warner, N., Maini, J., Chan Tung, K. W., Zakaria, H., Pawson, T., and Donaldson, L. W. (2006) *J Mol Biol* 356, 142-54.
6. Ramachander, R., Kim, C. A., Phillips, M. L., Mackereth, C. D., Thanos, C. D., McIntosh, L. P., and Bowie, J. U. (2002) *J Biol Chem* 277, 39585-93.
7. Kim, C. A., Phillips, M. L., Kim, W., Gingery, M., Tran, H. H., Robinson, M. A., Faham, S., and Bowie, J. U. (2001) *Embo J* 20, 4173-82.
8. Kim, C. A., Gingery, M., Pilpa, R. M., and Bowie, J. U. (2002) *Nat Struct Biol* 9, 453-7.
9. Aviv, T., Lin, Z., Lau, S., Rendl, L. M., Sicheri, F., and Smibert, C. A. (2003) *Nat Struct Biol* 10, 614-21.
10. Green, J. B., Gardner, C. D., Wharton, R. P., and Aggarwal, A. K. (2003) *Mol Cell* 11, 1537-48.
11. McGrath, J. A., Duijf, P. H., Doetsch, V., Irvine, A. D., de Waal, R., Vanmolkot, K. R., Wessagowit, V., Kelly, A., Atherton, D. J., Griffiths, W. A., Orlow, S. J., van Haeringen, A., Ausems, M. G., Yang, A., McKeon, F., Bamshad, M. A., Brunner, H. G., Hamel, B. C., and van Bokhoven, H. (2001) *Hum Mol Genet* 10, 221-9.

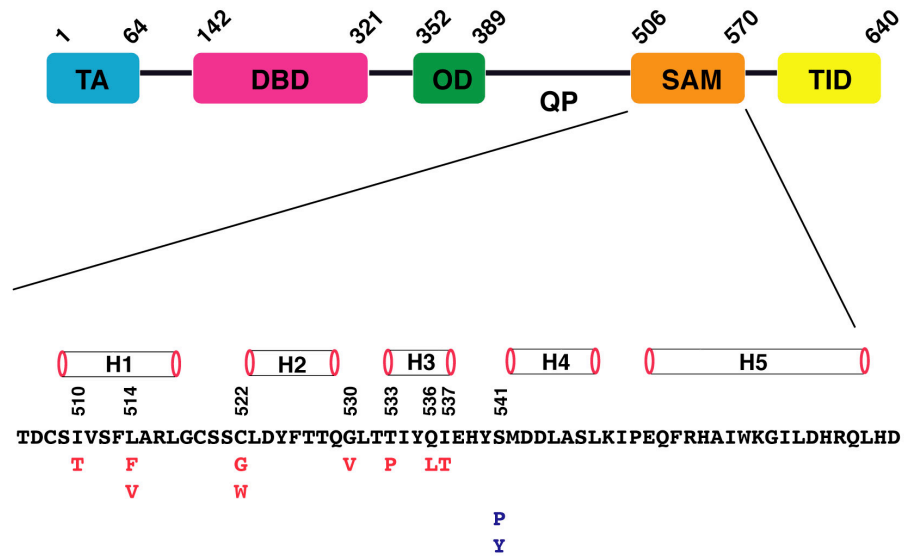
12. Kantaputra, P. N., Hamada, T., Kumchai, T., and McGrath, J. A. (2003) *J Dent Res* 82, 433-7.
13. Bougeard, G., Hadj-Rabia, S., Faivre, L., Sarafan-Vasseur, N., and Frebourg, T. (2003) *Eur J Hum Genet* 11, 700-4.
14. Yang, A., Schweitzer, R., Sun, D., Kaghad, M., Walker, N., Bronson, R. T., Tabin, C., Sharpe, A., Caput, D., Crum, C., and McKeon, F. (1999) *Nature* 398, 714-8.
15. Mills, A. A., Zheng, B., Wang, X. J., Vogel, H., Roop, D. R., and Bradley, A. (1999) *Nature* 398, 708-13.
16. Yang, A., and McKeon, F. (2000) *Nat Rev Mol Cell Biol* 1, 199-207.
17. Yang, A., Kaghad, M., Wang, Y., Gillett, E., Fleming, M. D., Dotsch, V., Andrews, N. C., Caput, D., and McKeon, F. (1998) *Mol Cell* 2, 305-16.
18. Suh, E. K., Yang, A., Kettenbach, A., Bamberger, C., Michaelis, A. H., Zhu, Z., Elvin, J. A., Bronson, R. T., Crum, C. P., and McKeon, F. (2006) *Nature* 444, 624-8.
19. Fomenkov, A., Huang, Y. P., Topaloglu, O., Brechman, A., Osada, M., Fomenkova, T., Yuriditsky, E., Trink, B., Sidransky, D., and Ratovitski, E. (2003) *J Biol Chem* 278, 23906-14.
20. Ozaki, T., Watanabe, K., Nakagawa, T., Miyazaki, K., Takahashi, M., and Nakagawara, A. (2003) *Oncogene* 22, 3231-42.
21. Rinne, T., Hamel, B., van Bokhoven, H., and Brunner, H. G. (2006) *Am J Med Genet A* 140, 1396-406.
22. Chi, S. W., Ayed, A., and Arrowsmith, C. H. (1999) *Embo J* 18, 4438-45.

23. Yang, A., Walker, N., Bronson, R., Kaghad, M., Oosterwegel, M., Bonnin, J., Vagner, C., Bonnet, H., Dikkes, P., Sharpe, A., McKeon, F., and Caput, D. (2000) *Nature* 404, 99-103.
24. Koster, M. I., Dai, D., and Roop, D. R. (2007) *Cell Cycle* 6, 269-73.
25. Qiao, F., and Bowie, J. U. (2005) *Sci STKE* 2005, re7.
26. Lo Iacono, M., Di Costanzo, A., Calogero, R. A., Mansueto, G., Saviozzi, S., Crispi, S., Pollice, A., La Mantia, G., and Calabro, V. (2006) *Cell Cycle* 5, 78-87.
27. Koster, M. I., Dai, D., Marinari, B., Sano, Y., Costanzo, A., Karin, M., and Roop, D. R. (2007) *Proc Natl Acad Sci U S A* 104, 3255-60.
28. Westfall, M. D., Mays, D. J., Sniezek, J. C., and Pietenpol, J. A. (2003) *Mol Cell Biol* 23, 2264-76.
29. Koster, M. I., Kim, S., Mills, A. A., DeMayo, F. J., and Roop, D. R. (2004) *Genes Dev* 18, 126-31.
30. Cornilescu, G., Delaglio, F., and Bax, A. (1999) *J Biomol NMR* 13, 289-302.
31. Herrmann, T., Guntert, P., and Wuthrich, K. (2002) *J Mol Biol* 319, 209-27.
32. Guntert, P., Mumenthaler, C., and Wuthrich, K. (1997) *J Mol Biol* 273, 283-98.
33. Linge, J. P., O'Donoghue, S. I., and Nilges, M. (2001) *Methods Enzymol* 339, 71-90.
34. Kleywegt, G. J. (1996) *Acta Crystallogr D Biol Crystallogr* 52, 842-57.
35. DeLano, W. L. (2002), DeLano Scientific, San Carlos, California.
36. Schuck, P. (2000) *Biophys J* 78, 1606-1619.
37. Schuck, P., Perugini, M. A., Gonzales, N. R., Howlett, G. J., and Schubert, D. (2002) *Biophys J* 82, 1096-111.

Figure 3-1. A schematic diagram of mutations found in AEC and RHS

p63 is a multi-domains protein consists of transcriptional activation (TA) domain, DNA binding domain (DBD), oligomerization domain (OD), poly- glutamine/proline rich region (QP), sterile alpha motif (SAM) domain, and transcriptional inactive domain (TID). Most AEC missense mutations are found in the SAM domain, and most RHS mutations are deletion of a nucleotide that result in a shortening of the TID. Letters marked red indicate mutations found in AEC, and letters marked blue indicate mutations found in RHS. A column with more than two letters indicates two separate mutations to the same amino acid.

Figure 3-1.



Missense mutations

Deletion mutations

568 640

WT L HDFSSPPHLLRTPSGASTVSVGSSETRGERVIDAVRFTLRQTISFPFRDEWDFNFDMSRRNKQQRKEEGE

RHS

1709delA LHASPHLLIS

1721delC LHDFSSLLIS

1787delG LHDFSSPPHLLRTPSGASTVSVGSSETRVNV

1859delA LHDFSSPPHLLRTPSGASTVSVGSSETRGERVIDAVRFTLRQTISFPFRDEWMISTLTWLVATSSSVSKRKEN

AEC

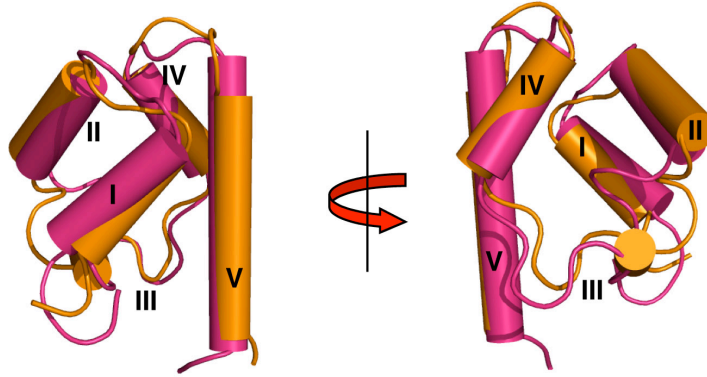
1742delC LHDFSSPHLLRTQVVPLPSVWAPVRPVNV

Figure 3-2. Structural alignment of the p63SAM domain vs. p73SAM domain vs. L514F mutant

The SAM domain of p63 is highly identical to the SAM domain of p73 except in the helix three region. Two structures have a RMSD of 2.68Å. The p63SAM domain is colored in magenta, and the p73SAM domain is colored in orange. In comparison with the L514F mutant, helices 1, 2 and 3 of p63SAM have different tilting angles than the mutant. The RMSD between p63SAM and L514F is 2.75 Å. The L514F mutant is colored in yellow. The C α atoms of the protein backbone were used for structural alignment.

Figure 3-2.

p63SAM : p73SAM



p63SAM : L514F

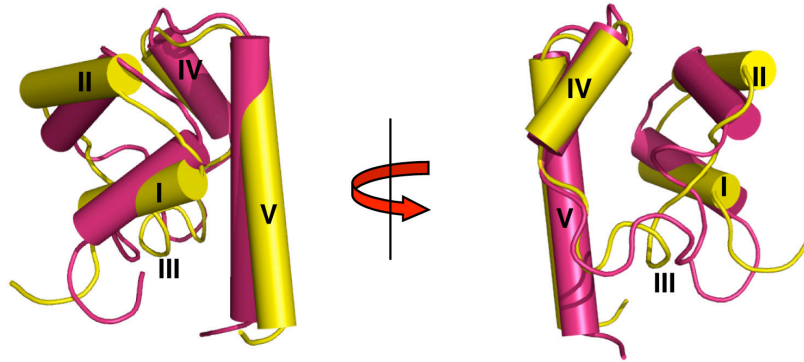


Figure 3-3. Thermal denaturation curve of the wild type SAM domain and mutants

Four mutants and the SAM domain undergo thermal denaturation to assess their relative stability. The mid transition point of the wild type SAM domain = 78 °C, Q536L and T533P = 63 °C, L514F = 53 °C, and G530V = 42 °C. The experimental ellipticity values of each construct have been converted to fractional ellipticity.

Figure 3-3.

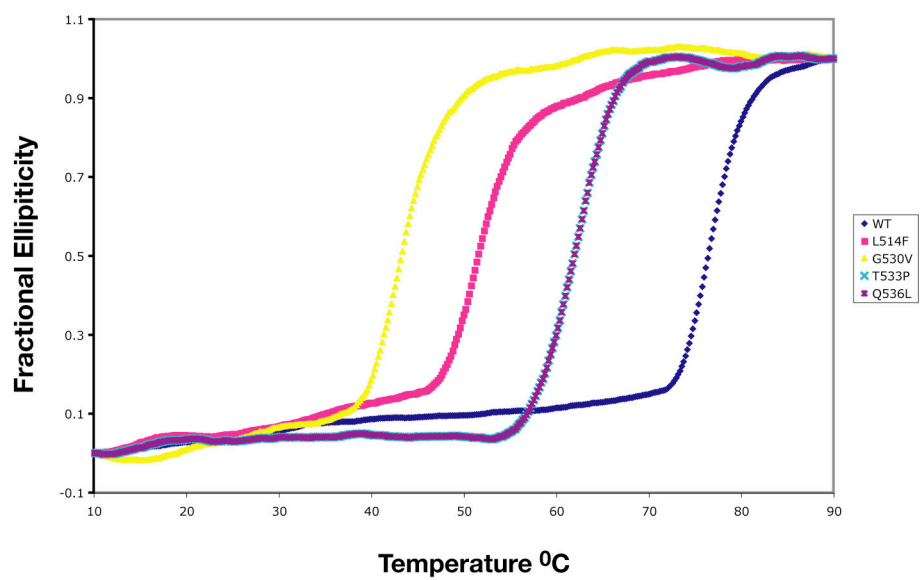


Figure 3-4. Chemical shift deviations between the wild type SAM domain and mutants

The chemical shifts of the nitrogen (N) and amide protons (HN) from each mutant were compared with the wild type SAM domain. Negative bars indicate residues for which assignments are not possible due to large deviation in chemical shifts compared to wild type SAM domain. A schematic diagram of the SAM domain is below each graph. Regions of the SAM domain that have significant deviations in chemical shifts due to the mutation are colored in red, and the mutated residue is marked by asterisk. (A) L514F mutant. (B) G530V mutant. (C) Q536L mutant.

Figure 3-4.

A

L514F

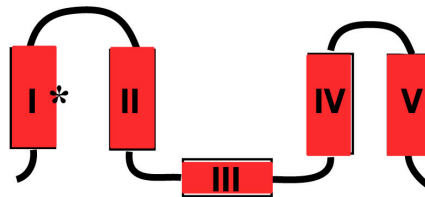
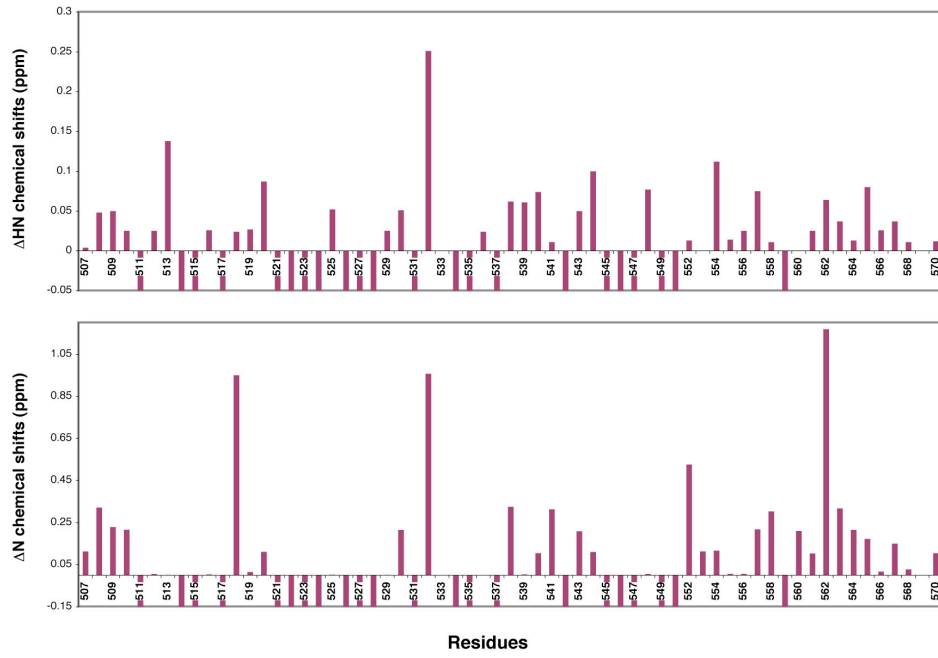


Figure 3-4.

B

G530V

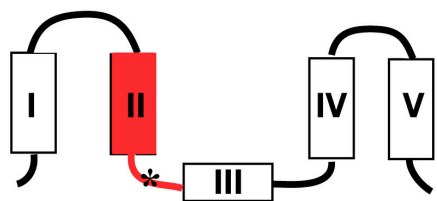
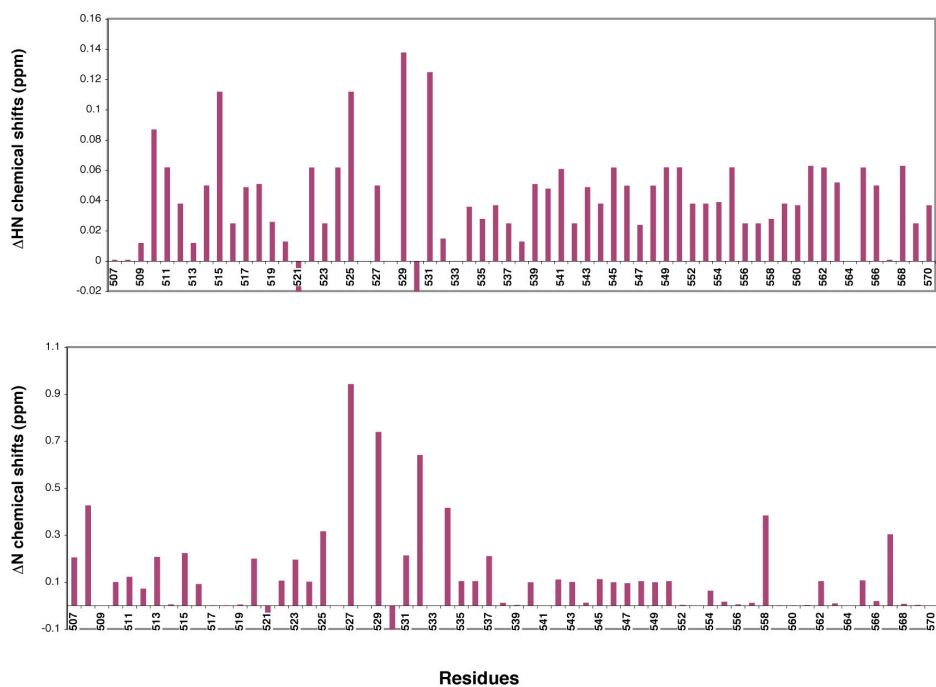


Figure 3-4.

C

Q536L

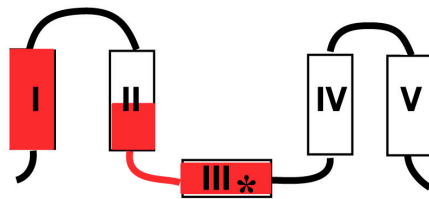
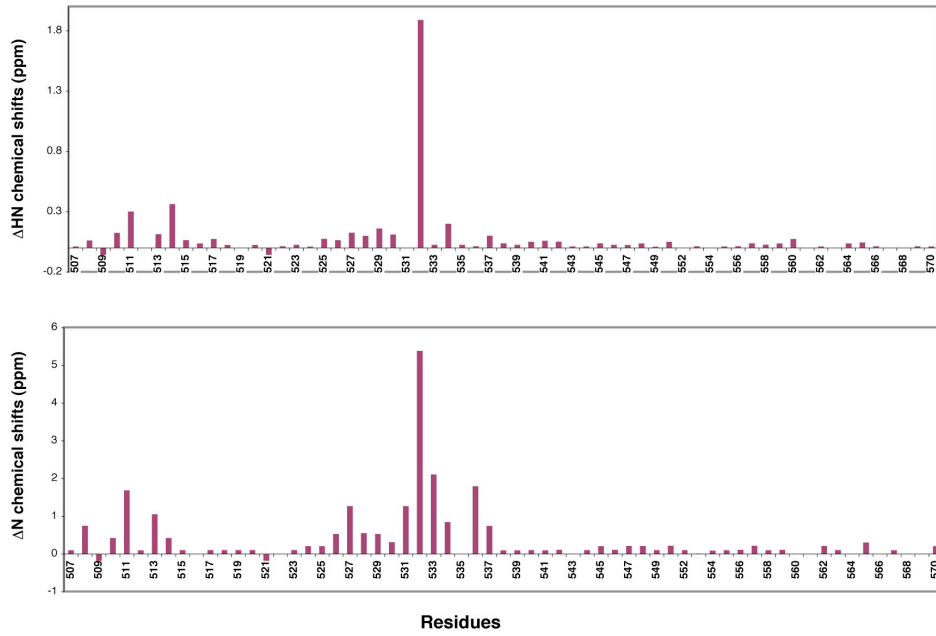
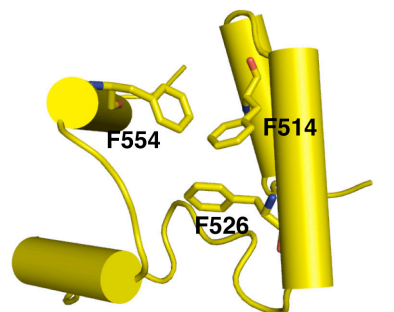


Figure 3-5. The L514F mutation causes rearrangement of residues in helix 3 and helix 5

F554 from helix 5 changes its conformation from solvent exposed in the wild type SAM domain to interacting with F514 in the mutant. In helix 3, T533 also changes its conformation from solvent exposed in the wild type SAM domain to interacting with I537 in the mutant.

Figure 3-5.

L514F



WT

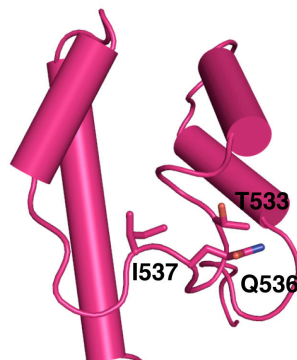
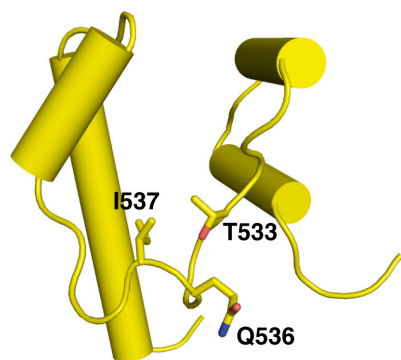


Figure 3-6. Gel filtration curve of the MBP fusion protein of the wild type SAM domain and mutants

The wild type, four mutants, and the QP are expressed as MBP fusion proteins and the oligomerization state was determined on a superose 6 gel filtration column. The MBP-OD-SAM (blue line) elutes as a tetramer (expected molecular weight of 270 kD) with an apparent molecular weight of 440 kD. A similar profile is seen with MBP-OD-QP (magenta line). In contrast to the wild type, fusion proteins that contain mutations in the SAM domain exhibit equilibrium at the void volume and at apparent molecular weight of 669 kD. Each mutant exhibits a larger apparent molecular weight than the wild type, an indication that mutations in the SAM domain can disrupt the quaternary structure of the OD-SAM.

Figure 3-6.

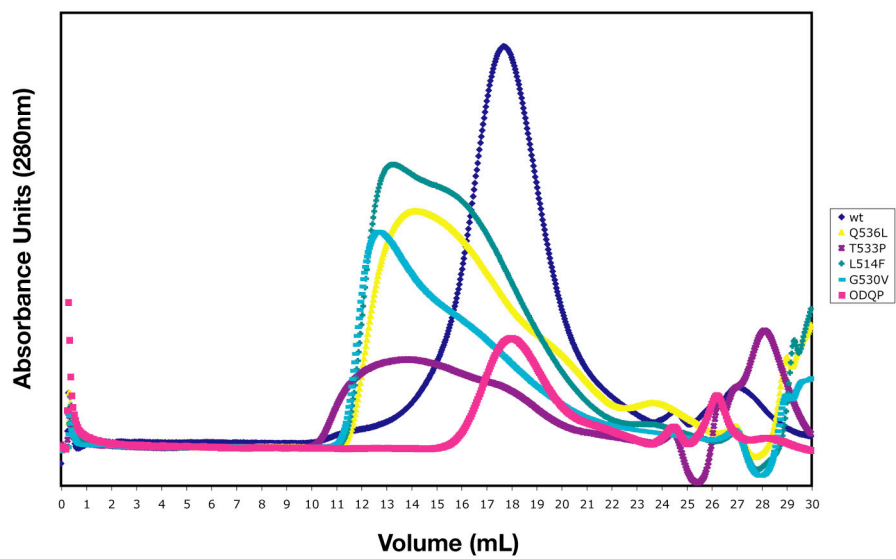
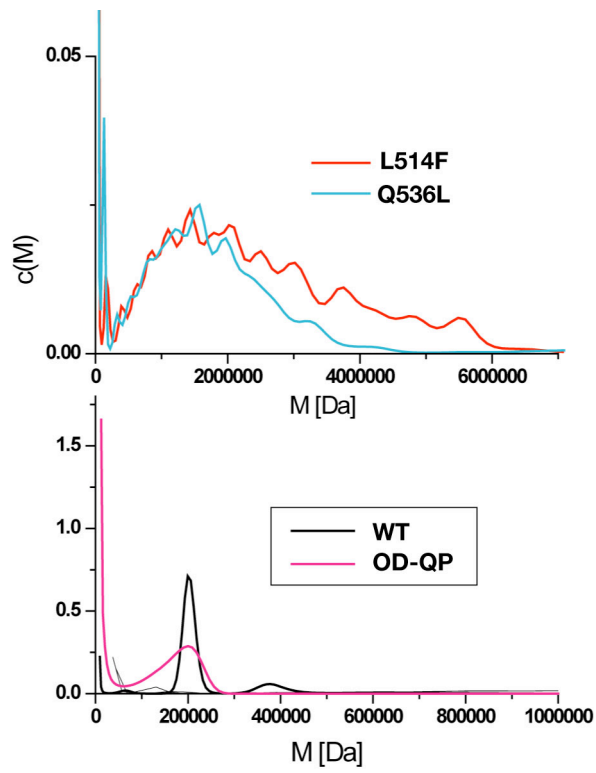


Figure 3-7. Velocity sedimentation of the MBP fusion protein of the wild type SAM domain and mutants

Distribution plots with respect to molecular weight after fitting with velocity sedimentation data are shown for the mutant fusion protein and the wild type fusion protein. The wild type SAM domain fusion protein sediments as a tetramer with a minor population being an octamer. In contrast, the QP fusion protein is a tetramer with a non-ideal peak shape that extends to a dimer, which signifies that the OD-QP is not a stable structure. Both the L514F and Q536L fusion proteins exhibit aggregation as shown by their large molecular weight of 1.8 MDa.

Figure 3-7.



Chapter 4

Conclusions and Future Directions

The first part of this thesis explores the evolution of the p53 protein family through structural investigation of the C-terminus of two p53 homologs in invertebrates, CEP-1 and Dmp53. The study reveals the OD and residues C-terminal to it may have an interdependent relationship in the stability of both domains, in which the loss of the C-terminal residues could have an effect on the oligomeric state of the protein. In CEP-1, the additional stabilizing element is the SAM domain and in Dmp53 is a helix. The function of the SAM domain in stabilizing the OD in CEP-1 suggests this function could be conserved in vertebrate p63. This hypothesis serves as the basis for the second part of this thesis, in which the effect of mutations identified in the Hay-Wells syndrome on the structure of the SAM domain and the OD are investigated. Structural study of the SAM domain with these mutations shows each mutation in the SAM domain has minor effects on the overall fold of the domain, except in the region surrounding the 3_{10} helix. However, the effect of each mutation becomes apparent when the SAM domain is in a tetrameric context with the OD of p63. A highly soluble fusion protein from the OD to the SAM domain of p63 becomes aggregated upon mutations in the SAM domain. The change in the behavior of the fusion protein suggests there may be an interaction between the SAM domain and the OD that could be disrupted by the mutations. A change in the quaternary structure between the OD and the SAM domain could consequently disrupt the overall quaternary structure of p63 and result in gain of function.

The surprise discovery that CEP-1 is a dimer rather than a tetramer provides a hint that the ancestral p53 molecule could be a dimer, which evolved into a tetramer, if one assumes that life evolves towards increasing complexity (coelomata). However based on the ecdysozoa view of the phylogenetic tree, CEP-1 is the only member of p53 that

exhibit this oligomerization property. One reason for the controversy in the phylogenetic tree topology between coelomata and the ecdysozoa is the ambiguous placement of the Nematoda phylum with regards to whether this phylum diverged before the split of protostome and deuterostome, or within the protostome branch (1). In order to determine the ancestral oligomeric state of p53, an investigation into the earliest species that has a p53 like molecule, such as Ehp53, found in *Entamoeba histolytica*, or the slowest evolving species in Nematoda, the *Trichnella*, would reveal an answer. If a dimer is found in both species, then it is most likely that the ancestral p53 was a dimeric unit, and if Ehp53 were found to be a tetramer, then the dimeric CEP-1 is most likely an independent evolutionary event in *C. elegans*.

In contrast to most transcription factors that function as a dimeric unit, p53 is unique in that its functional unit was assumed to be a tetramer due to the tetrameric property of the OD, but *in vitro* measurements have shown that p53 (from DBD to OD) has an equilibrium constant of 3 μM between dimers and tetramers (2). *In vivo*, the oligomeric state of p53 has not been clearly defined and consequences of different p53 oligomeric states on function are not known. The discovery of CEP-1 as a dimer and its ability to induce apoptosis but not cell cycle arrest in *C. elegans* shows that the dimeric state of p53 is a functional molecule *in vivo*. Furthermore, when CEP-1 is expressed in human Saos-2 cells, apoptosis is induced (3). This suggests a dimeric state of p53 is sufficient to induce apoptosis in human cells. If this were true, then the ultimate determinant in whether p53 would induce a cell cycle arrest or an apoptotic response would be the oligomeric state of p53: a dimer would induce apoptosis, and a tetramer would result in cell cycle arrest. A change in the oligomeric state of p53 could

concomitantly alter the specificity in promoter recognition for cell cycle arrest genes or apoptotic genes depending on a tetrameric or a dimeric state, respectively. This hypothesis may be wrong, but it is a clear, verifiable hypothesis that can be tested, in which a confirmation of this idea could bring new therapeutic approaches to cancer treatment based on p53.

Although the structures of individual domains have been solved for p53 and p63, the full length structure of either protein has been elusive. A complete model of p53 or p63 will settle not only many controversies arising from different publications, but also allow an explanation on how post-transcriptional modifications could affect the function of these important proteins. A recent electron microscopy (EM) structure of the full length p53 shows a skewed cube with a 26 \AA^2 hole in the middle of the cube (4). In order to fit all the domains into the electron density, the authors argued that p53 oligomerizes through interactions between the N- and the C-terminus, and not through the OD, whose structure has been determined in isolation by crystallography and NMR (5, 6). However, in light of the studies on CEP-1 and Dmp53, which have shown that the function of the OD has been preserved throughout evolution, the idea that the extreme termini are the oligomeric domain may not be sound and the model of p53 resulting from the EM data could be misinterpreted. Interestingly, our own analysis of the EM structure of p63 exhibits a similar negative staining image to p53, except for two extra lobes that could represent two additional domains in p63 that are not present in p53 (Figure 4-1). It will be of great interest to continue EM analysis of p63 at a higher resolution to reveal domain organization within p63. At the same time, a full crystallographic effort should be undertaken to solve the structure of p63. Unlike p53 which does not express well in

bacteria, p63 can be expressed in high quantity (~ 7 mg/L of culture) and obtained to a purity of 85-90% after isolation from bacteria lysates. An initial crystallization condition has resulted in a crystal with a diffraction of 35 Å (Figure 4-2). With advances in robotic crystal screening that can screen 20,000 conditions, in combination with automatic cloning, expression, and purification, a systematic search for the optimal construct and crystal could be undertaken in a reasonable time, and the full length structure of p63 could be in the near horizon.

While striving for the full length structure of p63, two more tangible structural targets in human p63 may also yield fruitful results. One is the structure of the OD of TAp63 γ by solution NMR analysis, and the other is the structure of the OD-SAM domain of p63. The OD of the TAp63 γ is the first soluble oligomerization domain of p63 in the absence of the SAM domain. Secondary structural prediction analysis indicates an additional helix after the OD, which coincides with the topology of OD in Dmp53 (Figure 4-3). The structure of this domain will reveal the stabilizing interaction in the C-terminus of the OD. A related question on how the SAM domain stabilizes the OD of p63 can be answered by the structure of the OD-SAM domain. This could explain the effect of Hay-Wells syndrome mutations in the context of a tetrameric structure and provide a partial picture of the full length molecule with the exact positioning of the SAM domain relative to the OD. This highly soluble and well behaved OD-SAM protein, either alone or as a MBP fusion protein, make this protein a highly desirable target for crystallization (Figure 4-4).

References

1. Aguinaldo, A. M., Turbeville, J. M., Linford, L. S., Rivera, M. C., Garey, J. R., Raff, R. A., and Lake, J. A. (1997) *Nature* 387, 489-93.
2. Weinberg, R. L., Veprintsev, D. B., and Fersht, A. R. (2004) *J Mol Biol* 341, 1145-59.
3. Bergamaschi, D., Samuels, Y., O'Neil, N. J., Trigianti, G., Crook, T., Hsieh, J. K., O'Connor, D. J., Zhong, S., Campargue, I., Tomlinson, M. L., Kuwabara, P. E., and Lu, X. (2003) *Nat Genet* 33, 162-7.
4. Okorokov, A. L., Sherman, M. B., Plisson, C., Grinkevich, V., Sigmundsson, K., Selivanova, G., Milner, J., and Orlova, E. V. (2006) *Embo J* 25, 5191-200.
5. Lee, W., Harvey, T. S., Yin, Y., Yau, P., Litchfield, D., and Arrowsmith, C. H. (1994) *Nat Struct Biol* 1, 877-90.
6. Jeffrey, P. D., Gorina, S., and Pavletich, N. P. (1995) *Science* 267, 1498-502.

Figure 4-1. The electron microscopy image of the TAp63 α with negative staining

1 μ L of 0.05 mg/mL of purified TAp63 α was placed on a carbon grid and stained with uranium acetate. The top panel shows one section of the carbon grid with the TAp63 α particle. The middle panel is a magnification of some particles with different orientations on the grid. The bottom panel is 8 classes of particles averaged from 5000 particles. One class has a clear tetrameric form as expected from TAp63 α , but a class with a three fold symmetry is also observed, possibly due to a different perspective of the particle.

Figure 4-1.

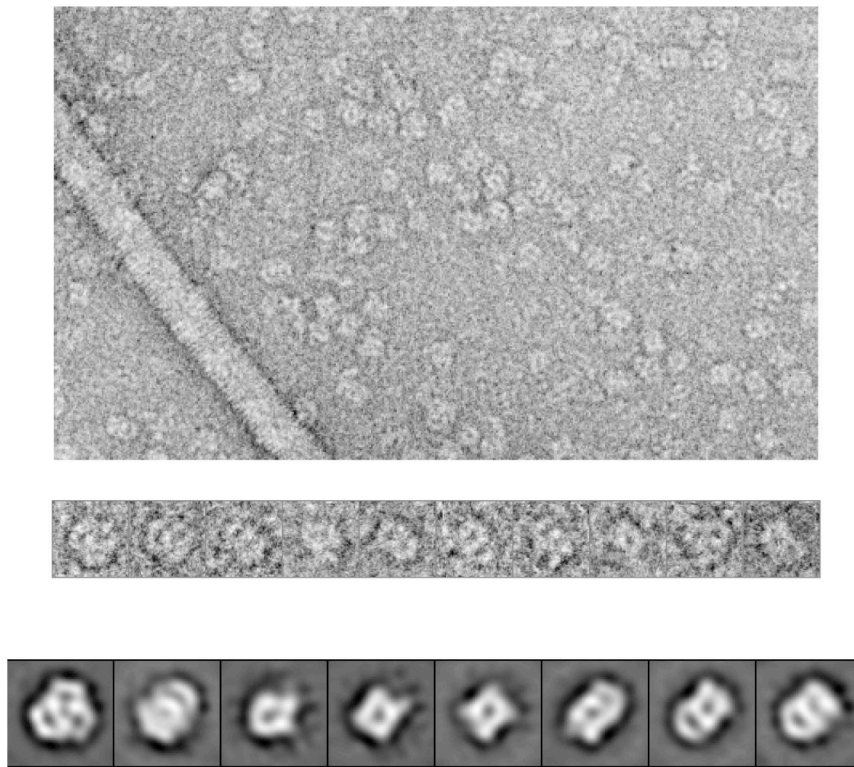


Figure 4-2. Crystal of TAp63 α and the first X-ray diffraction pattern

TAp63 α was co-crystallized with DNA by the hanging drop method in precipitants consists of 0.1 M KCl, 0.01 M MgCl₂, 0.05 M, 30% PEG 400, and Tris-Hcl pH 8,5. The protein to DNA ratio is 1:1.2 and 1 μ L of protein solution is mixed with 1 μ L of precipitants. The crystal diffracted to 35 Å with diffraction spots that resembled a protein crystal.

Figure 4-2.

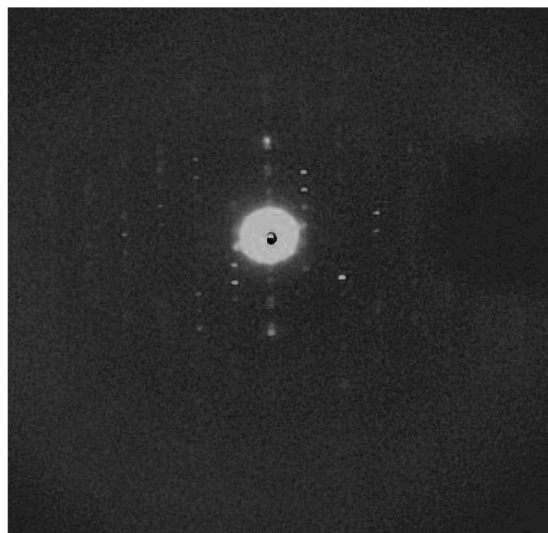
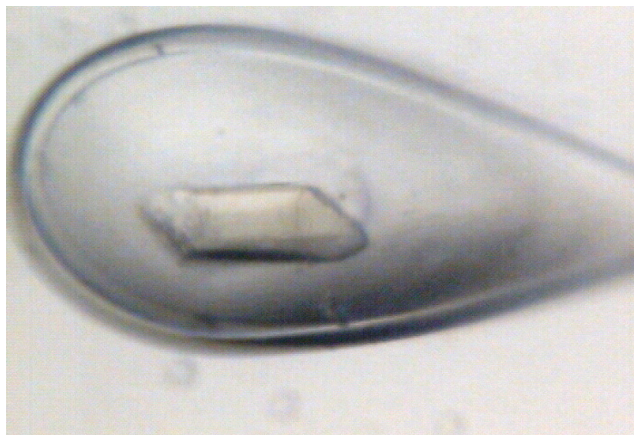


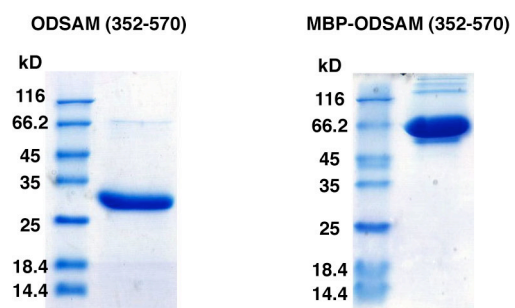
Figure 4-3. The OD of TAp63 γ has an additional helix C-terminal to the OD

Secondary structure prediction suggests an additional helix after the OD, which is similar to the layout of secondary structures in Dmp53. The predicted secondary structure is labeled below the sequence. Interestingly, a small fragment of the C-termini in TAp63 γ is conserved across p63 α and p63 β isoforms. The bottom panel shows the C-termini of p63 γ (343-428) with an expected molecular weight of 10,601 Dalton after elution from a nickel column. The double bands observed in the gel suggest there are unfolded region in this construct.

Figure 4-4. SDS-PAGE of OD-SAM and MBP-OD-SAM purified proteins

Both OD-SAM (expected molecular weight: 25 kD) and the MBP fusion protein of OD-SAM (expected molecular weight: 68 kD) elute as single specie in a gel filtration column with no apparent precipitation. The two proteins have high expression and solubility that make them suitable targets to obtain their structures by crystallography.

Figure 4-4.

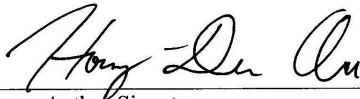


Publishing Agreement

It is the policy of the University to encourage the distribution of all theses and dissertations. Copies of all UCSF theses and dissertations will be routed to the library via the Graduate Division. The library will make all theses and dissertations accessible to the public and will preserve these to the best of their abilities, in perpetuity.

Please sign the following statement:

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis or dissertation to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

9-28-07

Date