

**Structure of the essential yeast telomerase protein Est3 from *S. cerevisiae* and structure-guided investigation of function**

by

**Timsi Rao**

M.Sc (IIT Bombay, India, 2006)

B.Sc (Kurukshetra University, Haryana, India, 2004)

A thesis submitted to the faculty of the Graduate School of the University of Colorado in partial fulfillment of the requirement for the degree of Doctor of Philosophy

Department of Molecular, Cellular and Developmental Biology

University of Colorado, Boulder

2012

This thesis entitled:  
Structure of the essential yeast telomerase protein Est3 from *S. cerevisiae* and structure-guided investigation of function

written by Timsi Rao  
has been approved by the Department of Molecular, Cellular and  
Developmental Biology

---

Deborah S. Wuttke

---

Mark Winey

Date\_\_\_\_\_

The final copy of this thesis has been examined by the signatories and we find that both the content and form meet the acceptable presentation standards of scholarly work in the above-mentioned discipline.

Rao, Timsi (Ph.D., Molecular Cellular and Developmental Biology)

Structure of the essential yeast telomerase protein Est3 from *S. cerevisiae* and structure-guided investigation of function

Thesis directed by Professor Deborah S. Wuttke

Telomeres are the nucleoprotein complexes at the ends of linear chromosomes that protect the genome from degradation and chromosomal fusions. Telomeres are replicated by the specialized enzyme telomerase. The telomerase holoenzyme in *S. cerevisiae* contains an RNA template and three known protein subunits, Est1, Est2 and Est3 (est = ever shorter telomeres for the phenotype observed upon their deletion). The reverse transcriptase Est2 and the RNA template TLC1 constitute the catalytic core of the telomerase holoenzyme. While Est1 and Est3 are not required for catalysis *in vitro*, they are strictly required for activity *in vivo*. The function of Est3 has remained elusive, although genetic data suggests that one mode of Est3's function might be carried out via its interaction with Est2.

To provide insights into Est3 function, we have solved its high-resolution structure. Because Est3 is a difficult protein target for structure elucidation, the structure was solved using a relatively novel strategy of combining minimal NMR experimental data (chemical shifts, RDCs and NOEs) with Rosetta *de novo* structure prediction. The structure is an OB-fold, with five-stranded  $\beta$ -barrel,

capped with an  $\alpha$ -helix and has some specialized features that distinguish it from other OB folds. The canonical loop L45 is quite unusual in the case of Est3, in that it is long and highly structured and plugs on top of the OB-fold canonical ligand binding face. Even in the absence of appreciable sequence relationship, the Est3 structure shows remarkable similarity to *Hs*TPP1-OB structure, not only in the core  $\beta$ -barrel, but also in the positioning of the  $\alpha$ -helix at the base and placement of C-terminal tail partially covering the canonical OB-fold binding face. Mapping residues involved in telomerase-association onto the structure reveals a novel protein interaction surface at the base of the  $\beta$ -barrel for Est3 and TPP1-OB. *In vivo* analysis, using structure-guided mutagenesis of Est3 surface also identified several new, functionally relevant, residues of Est3. The structure also served as a validation tool for an *in vivo* guided *in vitro* study that showed a direct correlation of *in vivo* dominant-negative mutants with their structural retention *in vitro*.

## Acknowledgments

I thank my advisor Dr. Deborah Wuttke, for tremendous support and guidance during the course of this project. The encouragement I received from her allowed me to keep persevering on this difficult project. Without her support, the completion of this project would not have been possible. I also thank the NMR facility at CU, especially Dr. Geoffrey Armstrong (NMR facility manager and my collaborator on the NMR side of the Est3 structure project) whose help and guidance throughout the project was invaluable. I am grateful to the Wuttke lab members especially Johnny Croy, Karen Lewis, Robert Hom, Sarah Altschuler, Thayne Dickey and Danielle Pfaff for discussions and suggestions during the different stages of my project. I will also like to thank my thesis committee members Dr. Mark Winey, Dr. Tom Blumenthal, Dr. Shelley Copley, Dr. Arthur Pardi and Dr. Andreas Hoenger for their suggestions and guidance. In addition, I thank our collaborators Dr. Vicki Lundblad and Johnathan Lubin for *in vivo* studies on the Est3 protein. I will also like to acknowledge some of the present and past Cech lab members Dr. David Zappulla, Dr. Anne Gooding, Dr. Elaine Podell, Dr. Andrea Berman and Dr. Jayakrishnan Nandakumar for help with reagents and experiments related suggestions. Funding for this work was provided by the National Institutes of Health grants T32 GM08759 (to T.R) and R01 GM05941 (to D.S.W). Last, but not the least, I thank my family and friends for support during the course of my thesis work.

# Table of Contents

<b>Chapter 1. Introduction</b> .....	<b>1</b>
<b>1.1 Telomeres</b> .....	<b>1</b>
1.1.1 Location and Role .....	1
1.1.2 Telomeric DNA: structure, length and sequence variability from different organisms .....	5
<b>1.2 Telomere end-capping</b> .....	<b>8</b>
1.2.1 Telomere-binding proteins in budding yeast .....	9
1.2.2 Telomere-binding proteins in mammals .....	13
1.2.2a Shelterin complex protein TPP1 is a key player for telomerase recruitment and activity in humans .....	15
1.2.2b CST-like complexes in mammals .....	17
1.2.3 End-capping proteins share common structural themes .....	18
<b>1.3 Telomere length homeostasis</b> .....	<b>19</b>
1.3.1 Telomere length homeostasis in budding yeast .....	21
1.3.2 Telomere length homeostasis in humans .....	23
<b>1.4 Telomerase</b> .....	<b>23</b>
1.4.1 Telomerase in humans .....	30
1.4.1a Telomerase dysfunction linked to human disease .....	31
1.4.2 Telomerase in the budding yeast .....	33
<b>1.5 Est3: Discovery and enigmatic function</b> .....	<b>38</b>
1.5.1 Est3 from <i>S. cerevisiae</i> .....	42
1.5.2 Est3 from <i>S. castellii</i> .....	47
<b>1.6 Thesis statement</b> .....	<b>52</b>
<b>Chapter 2. Structure of the <i>Saccharomyces cerevisiae</i> Est3 protein</b> .....	<b>53</b>
<b>2.1 Introduction</b> .....	<b>53</b>
2.1.1 Est3 is a diverged protein with predominantly uncharacterized structure and function .....	53
2.1.2 Est3 secondary structure elements and prediction of 3D models.....	54
2.1.2a Structural features of OB-fold proteins .....	57
<b>2.2 Materials and Methods</b> .....	<b>60</b>
2.2.1 Est3 expression, purification and construct optimization.....	60
2.2.2 Crystallization screening .....	63
2.2.3 Solubility screening.....	65
2.2.4 Buffer for NMR studies of Est3 <sup>AN</sup> .....	69
2.2.5 Screening for expression and solubility of Est3 in deuterated media .....	70
2.2.6 Minimal media growth for expression of labeled Est3 for protein backbone-assignment NMR experiments .....	74
2.2.7 Screening for growth conditions for selective methylation of ILV side-chain in otherwise deuterated medium .....	75
2.2.8 Minimal medium growth for expression of selectively methyl-protonated Est3 <sup>AN</sup> and sample purification .....	78
2.2.9 Sample preparation and NMR experiments for backbone assignment.....	80
2.2.10 Sample preparation and NMR experiments for amide-amide and methyl-methyl NOESY.....	81
2.2.11 Sample preparation and NMR experiment for Residual Dipolar Coupling (RDC) measurement .....	82

2.2.12 RASREC Rosetta structure calculation of Est3 <sup>ΔN</sup> .....	84
2.2.13 HetNOE experiment for assessment of dynamics .....	89
2.2.14 H/D-exchange experiment for Est3 .....	90
2.2.15 Structure figures preparation .....	90
<b>2.3 Results .....</b>	<b>91</b>
2.3.1 Est3 protein sample optimization: challenges and solutions .....	91
2.3.1a Protein-yield optimization for unlabeled and isotopically labeled Est3 .....	91
2.3.1b Est3 construct optimization .....	93
2.3.1c Est3 storage buffer optimization .....	99
2.3.2 Est3 crystallization trials .....	100
2.3.3 NMR experimental data .....	101
2.3.3a Greater than 97% of the Est3 <sup>ΔN</sup> HSQC peaks assigned .....	101
2.3.3b Long range and orientational structural restraints: NOEs and RDCs .....	103
2.3.3c <sup>15</sup> N- <sup>1</sup> H-heteronuclear NOE experiment in parallel with backbone chemical shift based S <sup>2</sup> prediction helped elucidate dynamic regions of the protein .....	109
2.3.3d H/D-exchange data identified solvent protected regions of the Est3 protein and was used in structure validation .....	112
2.3.4 Relatively novel strategy, RASREC Rosetta, was used for Est3 <sup>ΔN</sup> structure calculation .....	114
2.3.5 Est3 structure evaluation .....	117
2.3.5a Est3 is an OB-fold protein .....	117
2.3.5b Mapping the conserved residues on the Est3 structure .....	121
2.3.5c Electrostatic surface map of Est3 structure .....	124
2.3.5d Structural homolog search for Est3 .....	126
2.3.5e Structure quality assessment and validation indicates the calculated structure of Est3 <sup>ΔN</sup> is within the permitted structure quality range and is correct .....	128
2.3.5f Comparison of predicted secondary and tertiary elements of Est3 to the calculated Rosetta Structure reveals robustness of secondary element prediction .....	132
<b>2.4 Discussion .....</b>	<b>135</b>
2.4.1 Est3 structure reveals novel features previously unidentified in other OB-fold proteins .....	135
2.4.1a An extended Loop 45 (L45) blocks the canonical binding surface of the OB - fold in Est3 .....	139
2.4.1b C-terminal of the Est3 partially sits on the canonical ligand-binding site and in conjunction with the structured L45 might be regulated to allow or block the access of Est3 ligands .....	140
<b>Chapter 3. Structure-function analysis of Est3: <i>in vitro</i> and <i>in vivo</i> studies .....</b>	<b>143</b>
<b>3.1 Introduction .....</b>	<b>143</b>
<b>3.2 Materials and Methods .....</b>	<b>147</b>
3.2.1 Est3 expression and purification .....	147
3.2.2 Expression and purification of Est2 N-terminal domain constructs .....	148
3.2.3 Construct design and cloning for a stable L45 deletion mutant of <i>S. cerevisiae</i> and <i>S. castellii</i> Est3 for DNA binding studies .....	152
3.2.4 Electrophoretic mobility shift assays .....	154
3.2.5 Chemical shift perturbation analysis in the <sup>15</sup> N HSQC spectra for protein-DNA binding .....	156
3.2.6 Affinity pull-down assays .....	158
3.2.7 Multi-Angle Light Scattering (MALS) study for purified Est3 protein .....	161
3.2.8 <i>In vitro</i> telomerase assay .....	161

3.2.9 <i>In vivo</i> analysis of est3 <sup>-</sup> mutants .....	162
<b>3.3 Results</b> .....	<b>163</b>
3.3.1 <i>In vitro</i> studies.....	163
3.3.1a Unambiguous binding of Est3 to N-terminal domain of Est2 was not observed .....	163
3.3.1b Purified Est3 does not bind to telomeric ssDNA <i>in vitro</i> .....	172
3.3.1c Est3 protein solution is comprised of an equilibrium between dimer and monomer forms.....	178
3.3.1d <i>S. cerevisiae</i> Est3, does not enhance yeast telomerase activity <i>in vitro</i> .....	179
3.3.2 Structure based <i>in vivo</i> analysis of Est3 mutants led to identification of residues important for Est3's function .....	181
3.3.2a Mutational analysis of L45 .....	181
3.3.2b Conservation-independent, complete surface mutagenesis of Est3 identifies functionally important residues .....	184
<b>3.4 Discussion</b> .....	<b>188</b>
3.4.1 Direct binding of ScEst3 with the TEN domain of Est2 was not established <i>in vitro</i> , but the existence of a common surface for telomerase association points to functional homology between Est3 and human TPP1-OB .....	188
3.4.2 Isolated ScEst3 does not bind telomeric ssDNA or enhance telomerase activity <i>in vitro</i> .....	191
3.4.3 Long and structured L45 is a unique feature, not involved in Est3's telomere length regulation function .....	192
3.4.4 Region between $\beta$ 3 and $\beta$ 4 is the hub for telomere length regulation function of Est3 .....	194
<b>Chapter 4. Correlation of the <i>in vivo</i> over-expression dominant-negative phenotype with <i>in vitro</i> stability measurements: a new strategy for <i>in vivo</i> protein characterization</b> .....	<b>196</b>
4.1 Introduction.....	196
4.1.1 Loss-of-function assays fail to distinguish separation-of-function alleles from alleles that encode disrupted protein structure.....	196
4.1.2 Est3 as a model protein for identification of loss-of-function mutants that also exhibit overexpression dominant negative phenotype <i>in vivo</i> .....	199
4.1.3 Telomerase-association versus separation-of-function mutants.....	204
4.1.4 CD as a tool for protein structure assessment.....	205
<b>4.2 Materials and Methods</b> .....	<b>208</b>
4.2.1 Generating Est3 mutants for the CD stability study .....	208
4.2.2 Expression and solubility test of Est3 and its point mutants.....	210
4.2.3 Purification of Est3 and its point mutants .....	212
4.2.4 CD melting curves: acquisition and analysis.....	213
<b>4.3 Results</b> .....	<b>214</b>
4.3.1 Point-mutants of Est3, tested for loss-of-function and over-expression dominant negative effect, reveal importance of several residues in Est3's role in telomere-length regulation .....	214
4.3.2 Point mutations in <i>S. cerevisiae</i> Est3 affect expression and solubility of the recombinant proteins .....	218
4.3.3 Secondary structure analysis of Est3 mutants by CD correlates well with their <i>in vivo</i> dominant negative phenotypes .....	222
<b>4.4 Discussion</b> .....	<b>230</b>
4.4.1 Mutant proteins with dominant-negative effect <i>in vivo</i> are structurally stable.	230

4.4.2 Predictions from the <i>in vivo</i> dominant-negative and <i>in vitro</i> protein stability correlation study hold true in the context of the Est3 structure .....	231
<b>References.....</b>	<b>236</b>
<b>APPENDIX .....</b>	<b>254</b>
<b>I. Chemical shift assignments for Est3<sup>ΔN</sup> .....</b>	<b>254</b>
<b>II. NOEs used for structure calculation .....</b>	<b>258</b>
<b>III. Orientational restraints (RDC data) for structure calculation .....</b>	<b>266</b>

## List of Tables

Table 1. 1 Telomeric DNA: sequences and lengths of some example organisms.....	7
Table 1. 2 Summary of telomere-associated proteins in budding yeast, fission yeast and mammals.....	12
Table 1. 3 Comparison of the counter-part mutants from <i>C. albicans</i> and <i>S. cerevisiae</i> Est3.....	51
Table 2. 1 Sequence identity and similarity of Est3 proteins to <i>S. cerevisiae</i> Est3 protein sequence .....	54
Table 2. 2 Structural statistics for the Est3 <sup>ΔN</sup> RASREC Rosetta structures .....	117
Table 2. 3 DALI search for structurally similar proteins to Est3 in the PDB.....	126
Table 2. 4 Internal validation for convergence of RASREC Rosetta.....	129
Table 3. 1 Primers for cloning <i>S. cerevisiae</i> and <i>S. castellii</i> Est2 domains.....	149
Table 3. 2 Overexpression dominant negative (ODN) phenotypes of point-mutants in solvent exposed residues of Est3 .....	185
Table 4. 1 Primers for point-mutants of Est3 for CD study .....	210
Table 4. 2 Result of loss-of-function and dominant-negative assays for the hydrophobic to –alanine and –charged residue (Glu/E) mutations (from John Lubin, Lundblad Lab, unpublished data).....	217
Table 4. 3 Comparison of point-mutants of <i>S. cerevisiae</i> Est3, recombinantly expressed from <i>E.coli</i> cells .....	220
Table 4. 4 Summary of the <i>in vivo</i> and <i>in vitro</i> data for protein stability analysis of Est3 mutants.....	234
Table A 1 <sup>1</sup> HN <sup>15</sup> N <sup>13</sup> CO <sup>13</sup> Ca and <sup>13</sup> Cβ protein backbone resonances assignments for the Est3 <sup>ΔN</sup> protein.....	254
Table A 2 All NOESs (amide-amide and methyl-methyl) for Est3 <sup>ΔN</sup> structure calculation .....	258
Table A 3 RDC restraints for Est3 <sup>ΔN</sup> .....	266

## List of Figures

Figure 1.1 Telomeres help differentiate chromosome ends from double-strand breaks .....	2
Figure 1.2 The end-replication problem .....	5
Figure 1.3 Chromosomal telomeric ends .....	6
Figure 1.4 Formation of the t-loop at telomeric ends.....	8
Figure 1.5 Telomere end-protection complexes from humans and yeast.....	11
Figure 1.6 Schematic of the domain map of human TPP1 and structure of human TPP1-OB.....	16
Figure 1.7 A canonical OB-fold structure .....	19
Figure 1.8 Protein counting model for telomere length homeostasis.....	21
Figure 1.9 Domain maps of TERT proteins from different organisms .....	25
Figure 1.10 Structure of the <i>T. thermophila</i> TEN domain and <i>T. castaneum</i> TERT26	
Figure 1.11 Cell-cycle dependent lengthening of telomeres by telomerase .....	28
Figure 1.12 Nucleotide and repeat addition processivity of telomerase .....	29
Figure 1.13 Schematic representation of human telomerase holoenzyme .....	30
Figure 1.14 Schematic representation of yeast telomerase holoenzyme .....	34
Figure 1.15 Alignment of Est3 protein sequences.....	39
Figure 1.16 Secondary structure prediction of <i>S. cerevisiae</i> Est3 .....	43
Figure 2.1 pET28b vector map with His <sub>10</sub> -SUMO-Est3 coding sequence.....	60
Figure 2.2 Hanging-drop and sitting-drop set-ups for solubility/ crystallization screens .....	64
Figure 2.3 Comparison of NDSB-201 and NDSB-195 additives .....	66
Figure 2.4 Cell-line and post-induction growth duration screening for deuterated Est3 <sup>ΔN</sup> expression.....	72
Figure 2.5 Western-blot for cell growth conditions screening for selective methyl protonation of Est3 <sup>ΔN</sup> .....	78
Figure 2.6 Rosetta <i>de novo</i> modeling pipeline.....	85
Figure 2.7 Purification scheme for Est3 proteins.....	92
Figure 2.8 Pre-induction cell-growth enhancement by presence of Isogro in minimal deuterated growth media .....	93
Figure 2.9 HSQC spectrum of GB1-Est3 fusion protein, highlighting peak-intensity differential.....	95
Figure 2.10 Summary of Est3 constructs screened for protein yield and solubility ..	96
Figure 2.11 <sup>15</sup> N HSQC spectrum of wild-type Est3 shows a cluster of sharp peaks..	98
Figure 2.12 Superposition of <sup>15</sup> N HSQC spectra of wild-type and Est3 <sup>ΔN</sup> proteins ..	99
Figure 2.13 C $\alpha$ and C $\beta$ connectivities established from corresponding three-resonance experiments .....	102
Figure 2.14 <sup>15</sup> N HSQC spectrum of Est3 <sup>ΔN</sup> showing residue assignments.....	103
Figure 2.15 <sup>13</sup> C HMQC spectrum showing methyl proton assignments.....	104
Figure 2.16 Schematic of secondary structure of Est3 <sup>ΔN</sup> showing long-range amide-amide NOE connectivities used for structure calculation .....	105

Figure 2.17 Schematic of secondary structure of Est3 <sup>ΔN</sup> showing long-range methyl-methyl NOE connectivities used for structure calculation.....	106
Figure 2.18a Spectra for residual dipolar coupling calculations.....	108
Figure 2.18b Correlation plot between the back-calculated and experimental RDCs .....	109
Figure 2.19 HetNOE and RCI-S <sup>2</sup> relaxation data for Est3 <sup>ΔN</sup> .....	111
Figure 2.20 <sup>15</sup> N HSQC spectra for H/D exchange of Est3 <sup>ΔN</sup> .....	113
Figure 2.21 Improvement of structure ensemble convergence upon step-wise addition of NMR experimental data to the RASREC Rosetta structure calculation of Est3 <sup>ΔN</sup> .....	115
Figure 2.22 Ribbon representation of Est3 <sup>ΔN</sup> structure ensemble .....	116
Figure 2.23a Ribbon representation of single structure of Est3 <sup>ΔN</sup> .....	119
Figure 2.23b Long range NOEs define the packing of L45 against the face of β-barrel. ....	119
Figure 2.24 Sequence alignment of Est3 proteins highlighting features of the structure .....	121
Figure 2.25 Conserved residues mapped on the Est3 <sup>ΔN</sup> structure .....	122
Figure 2.26 Threaded models of some Est3 homologs based on <i>S. cerevisiae</i> Est3 <sup>ΔN</sup> structure .....	123
Figure 2.27 Electrostatic surface representation of the Est3 <sup>ΔN</sup> structure .....	124
Figure 2.28 Electrostatic surface representation of the Est3 structure with loop L45 removed by modeling .....	125
Figure 2.29 Structure comparisons of Est3 to its closest structural homologs <i>Hs</i> TPP1-OB and <i>Sn</i> TEBPβ .....	128
Figure 2.30 CYANA 2.1 calculated structure of Est3 <sup>ΔN</sup> .....	131
Figure 2.31 H/D exchange protected residues mapped on the structure of Est3 <sup>ΔN</sup> .....	132
Figure 2.32 Comparison of Est3 secondary structure predictions from various sources .....	133
Figure 2.33 Comparison of Est3 calculated structure with the predicted 3D models in literature .....	134
Figure 2.34 Comparison of Est3 calculated structure to the LOMETS threaded 3D model .....	135
Figure 2.35 Comparison of acidic patches on the base and C-tail of <i>Sc</i> Est3 and homologous proteins .....	142
Figure 3.1 Schematic of the <i>S. cerevisiae</i> Est2 domains .....	149
Figure 3.2 Purified <i>S. cerevisiae</i> and <i>S. castellii</i> Est3 and Est2 proteins used for affinity pull-down studies.....	165
Figure 3.3 Affinity pull-down assay with SUMO- <i>Sc</i> Est2 <sup>173</sup> and <i>Sc</i> Est3 .....	166
Figure 3.4 Affinity pull-down assay with GST- <i>Sc</i> Est2 <sup>173</sup> and <i>Sc</i> Est3 .....	167
Figure 3.5 Affinity pull-down assay with SUMO- <i>Scas</i> Est2 <sup>176</sup> and <i>Scas</i> Est3.....	168
Figure 3.6 Affinity pull-down assay with GST- <i>Scas</i> Est2 <sup>176</sup> and <i>Scas</i> Est3.....	169
Figure 3.7 <sup>15</sup> N TROSY HSQC spectrum of <i>Scas</i> Est3.....	170
Figure 3.8 Chemical shift perturbation of <sup>15</sup> N-labeled Est3 in the presence of small peptide of Est2 .....	171

Figure 3.10 No Est3-ssDNA binding is detected from minimal NMR chemical shifts analysis.....	173
Figure 3.11 Electrostatic mobility shift assay to test direct binding of ScEst3 and telomeric Tel17 DNA .....	175
Figure 3.12 Filter binding assay to test direct binding of ScEst3 and its mutants to telomeric DNA.....	177
Figure 3.13 ScEst3 purification on size-exclusion chromatography and multi-angle light scattering.....	179
Figure 3.14 Addition of untagged-Est3 to the telomerase activity assay <i>in vitro</i> showed no enhancement in telomerase activity.....	180
Figure 3.15 Mutants in L45 region of Est3 do not show dominant-negative phenotypes .....	183
Figure 3.16 Mutants in L45 region of Est3 do not show dominant-negative phenotype .....	183
Figure 3.17 Est3 residues important for telomere-length maintenance, picked from the extensive point-mutant screening of the entire Est3 surface .....	187
Figure 3.18 Est3 residues important for telomere-length maintenance, picked from conserved residues of Est3 .....	187
Figure 3.9 Mapping residues of ScEst3 and HsTPP1-OB that are potentially involved in binding to telomerase.....	190
Figure 4.1 Cartoon representations of a loss-of-function and an overexpression dominant-negative case.....	197
Figure 4.2 Simplified scheme of the loss-of-function and overexpression dominant negative experiments.....	200
Figure 4.3 Telomere-length analysis of the Est3 mutants in a loss-of-function assay .....	201
Figure 4.4 Synthetic lethality test for mutants in Est3 protein.....	203
Figure 4.5 Dominant-negative effects of Est3 mutants on telomere-length.....	204
Figure 4.6 Representative far-UV circular-dichroism spectra for secondary structure elements of proteins.....	207
Figure 4.7 Length defect in loss-of-function mutations in hydrophobic residues of Est3.....	215
Figure 4.8 Telomere-length defects in a subset of overexpression dominant-negative mutants in hydrophobic residues of Est3.....	217
Figure 4.9 Western-blot of recombinantly expressed His <sub>10</sub> -SUMO-Est3 wild type and its mutants.....	219
Figure 4.10 Purity of the Est3 wild type and mutants used for CD analysis.....	222
Figure 4.11 Secondary structure analysis from the circular dichroism spectra of wild-type Est3 (WT) and its point mutants as a function of temperature .....	223
Figure 4.12 Melting curve at 208nm and 222nm for wild type and mutant S3 protein .....	225
Figure 4.13 Circular Dichroism spectra of Est3 mutants recorded from 30°C to 50°C .....	227

Figure 4.14 Secondary structure analysis from the circular dichroism spectra of point mutants of Est3 as a function of temperature.....	229
Figure 4.15 Mapping of residues from Est3 mutagenesis screen that do not show dominant-negative effect .....	233
Figure 4.16 Mapping of Est3 residues, identified from dominant-negative and loss-of-function assay, to be important for Est3's function .....	234

# Chapter 1. Introduction

## 1.1 Telomeres

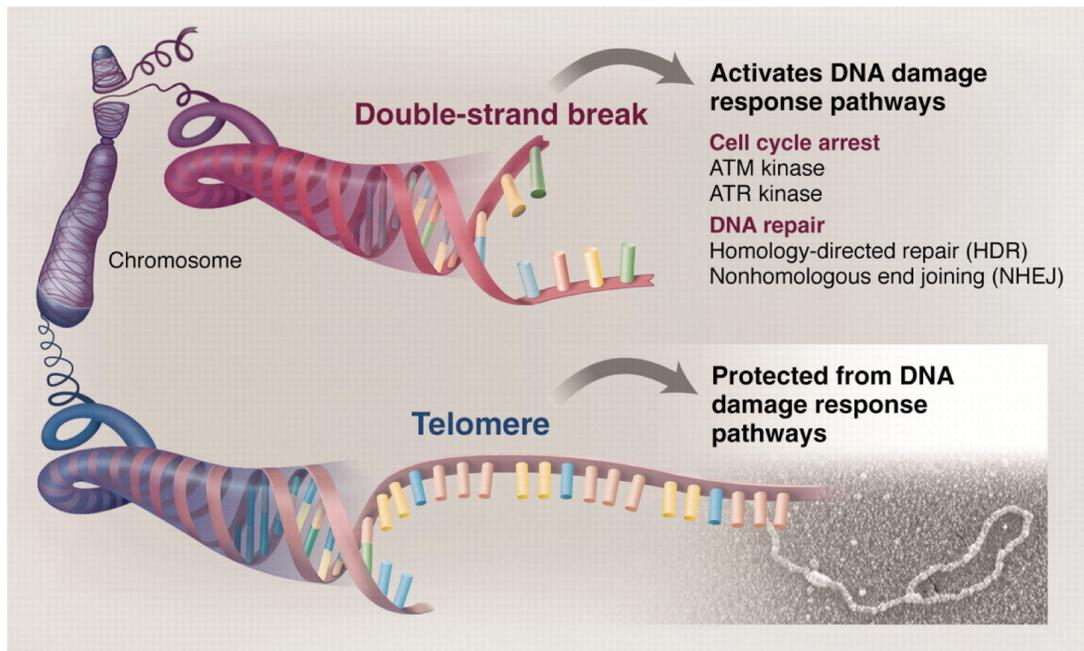
### 1.1.1 Location and Role

The term “telomere” originates from the Greek words meaning “end” (telos) and “part” (meros). The word was coined by Herman Muller in the 1930s [1]. Muller and Barbara McClintock [2] independently discovered the difference in behavior of double strand chromosome breaks from intact chromosomes, ends of which need telomeres as protective caps for structural and functional integrity.

Telomeres are the ends of linear chromosomes and features of telomeres and associated machinery are conserved in evolution between distantly related organisms. Studies in late 1970s and early 1980s by Blackburn and Gall shed light on the telomeric DNA sequence from *Tetrahymena* [3] and its use as model organism for telomere biology. *T. thermophila* has hundreds of chromosomes [4] produced during sexual reproduction of the micronucleus by subdivision of the chromosomes into hundreds of smaller molecules and further amplification in copy number of these molecules [5,6]. Therefore, due to the abundance of telomeric ends to study, *T. thermophila* provides a good biochemical model system for telomere biology [7]. In an important species cross-talk study, telomere sequence from *Tetrahymena* could be preserved in the context of a different organism, *Saccharomyces cerevisiae* (*S. cerevisiae*) [8], indicating structural and functional

conservation in evolution between even distantly related organisms. Discovery of telomeric DNA sequences from multiple organisms during the 1980s further established the presence of telomeres as a universally conserved phenomenon among eukaryotes [9,10].

Critically short telomeres trigger DNA damage checkpoints [11] and ultimately cell senescence (Figure 1.1). Thus, shortening of linear chromosomes corresponds to the limited number of times a cell can replicate and hence to cellular aging [12,13]. Telomeric DNA and associated proteins, therefore, are key players in cell's replicative senescence.



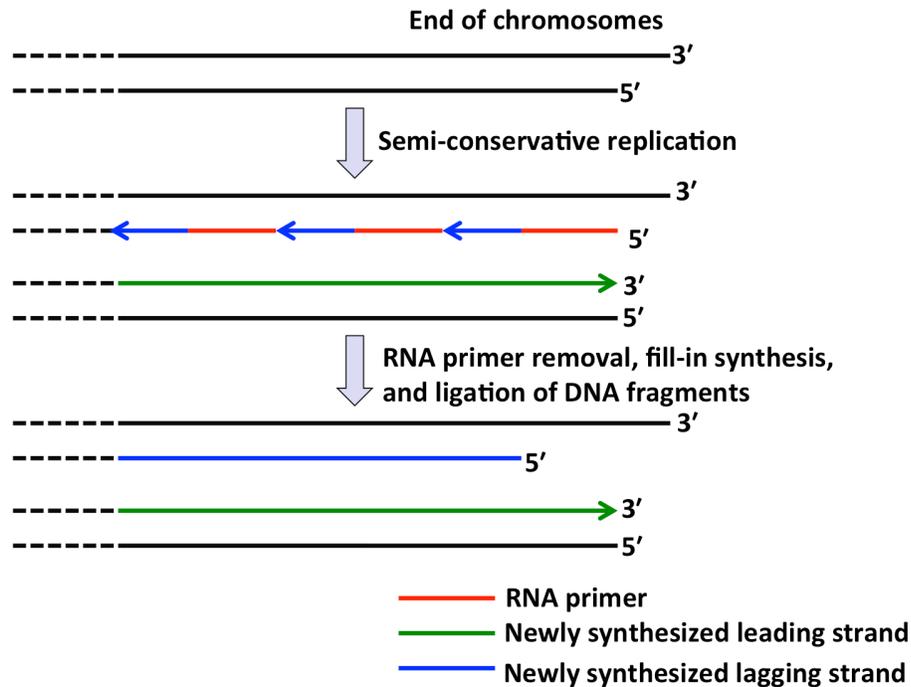
**Figure 1.1 Telomeres help differentiate chromosome ends from double-strand breaks**

Chromosomal double-strand breaks are recognized by DNA damage machinery and this triggers DNA repair and cell cycle arrest. Telomeres prevent the ends from being recognized as breaks. They do so by the presence of end-capping complexes and formation of lariat-like structures at the very ends (from [14]).

Because of direct implications in replicative senescence, understanding telomere shortening mechanisms and rates have been important questions in

telomere biology. One such shortening mechanism is associated with the inherent inability of the DNA replication machinery to replicate the very ends of linear chromosomes. The genetic information in a cell's DNA is passed on to its progeny by DNA replication, which is carried out by DNA polymerase. All known DNA polymerases synthesize DNA in 5' to 3' direction and require a free 3' -OH group to add a new nucleotide (nt). This 3' -OH group provided by RNA primers, small fragments of RNA that hybridize with parent DNA strands, is essential for the synthesis of new DNA. The RNA fragments are later removed and filled in by new DNA [15,16]. For circular chromosomes (of bacteria) the newly synthesized DNA strand leaves a 3'-OH immediately upstream of the primer for new DNA bases to be added onto. However, for linear chromosomes (of eukaryotes) replication of the lagging strand of DNA does not leave any 3' -OH group upstream of the last primer (Figure 1.2). At the end of semi-conservative replication, leading strands end in blunt ends with the parental strand. In contrast the lagging strands end up having RNA primers at the 5' ends that can not be replaced by DNA [17] (Figure 1.2). The lagging strand therefore leaves behind a 3' overhang that is digested by endonucleases and lead to shortening of the new chromosome end. In a recent study on human telomeres, it was found that the positioning of the last primer of the lagging strand is not at the very end of the parental strand being replicated but ~70-100 nts from the end and this positioning of the primer acts as an important determinant of telomere shortening [17]. Besides this placement of the last RNA primer away from the ends, the 5' strands (or C-strands) at both the chromosome

ends also undergo resectioning by putative exonucleases towards the end of S phase of the cell cycle to give rise to long 3' overhangs. Apollo exonuclease has been found to be involved in 5' resectioning at both the ends, whereas Exo1 exonuclease only acts at the 5' end of the leading strand in a study conducted in mouse cells [18,19]. This leads to the problem of shortening of chromosome ends in every replication cycle. Known as the end replication problem (ERP) (Figure 1.2), this shortening of linear chromosomes has been associated with the limited number of times a somatic cell can replicate [20] and thus to its senescence. This replication limit was discovered by Leonard Hayflick in the 1960's and is termed as the Hayflick limit [20]. Composed of non-coding regions of DNA, telomeres act as buffers that are lost at the end of every replication cycle because of the ERP and C-strand resection [12,14,21-23], aiding in cell's genetic integrity.



**Figure 1.2 The end-replication problem**

Short RNA primers (in red) are required for daughter strand synthesis (green and blue) during replication. RNA primers are later removed and replaced by DNA. Although, the very last RNA primer is also removed, it cannot be replaced, leading to DNA loss at the very ends. This loss at the end of every replication is known as the end-replication problem.

### **1.1.2 Telomeric DNA: structure, length and sequence variability from different organisms**

Telomeric DNA is composed of dsDNA repeats followed by a single-stranded (ss) 3' overhang (Figure 1.3). The repeats are composed of G-rich DNA sequences that are non-coding, allowing it to be lost at the end of replication cycles without loss of genetic information. In addition, both ds- and ssDNA telomeres are associated with proteins that aid in chromosomal capping and telomere length homeostasis.



**Figure 1.3 Chromosomal telomeric ends**

Telomeric DNA comprises of the double- and single-stranded regions of DNA (from [24]).

The length of the telomeric DNA is highly variable, not only between the different organisms but also between different cell-lines in multi-cellular organisms. The length of the telomeric dsDNA varies from exceptionally long 30-150 kilobases (kb) in laboratory maintained inbred mice [25] to just 130-175 bp for the yeast *C. tropicalis* [24,26]. The 3' overhang length also varies from 50-500 bases (b) in mammals [27] to just 12-14 b in budding yeasts [28].

Telomere sequence, like length, also varies significantly between organisms, although the theme of TG-rich tracts is universal. Whereas many organisms, like humans, mice, ciliates, etc., have precise repetition of the telomeric sequence, some others like *S. cerevisiae* budding yeast and *Schizosaccharomyces pombe* (*S. pombe*) fission yeast have more variable sequence and length of the repeat (degenerate sequence) [29] (Table 1.1).

**Table 1. 1 Telomeric DNA: sequences and lengths of some example organisms**

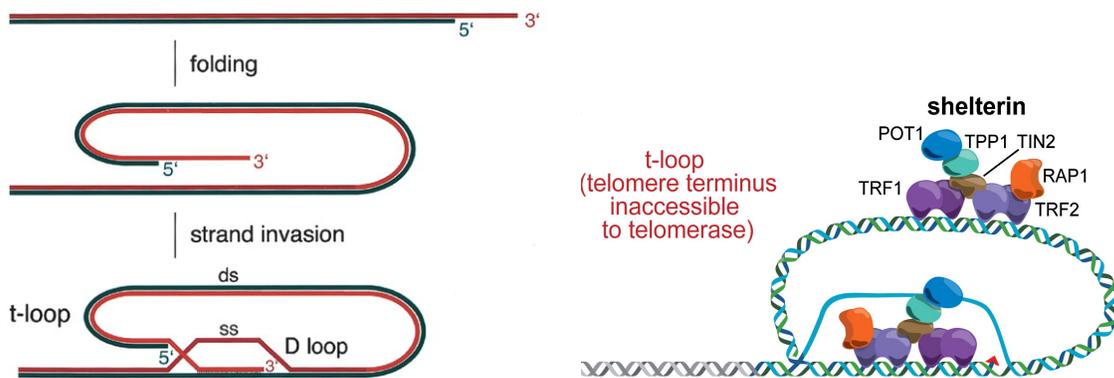
Organism name	Telomere Sequence <sup>a</sup>	Telomere length <sup>b</sup>	Overhang length in bases <sup>c</sup>
<i>H. sapiens</i> (human)	(TTAGGG) <sub>n</sub>	5-15 kbp	50-500
<i>M. musculus</i> (mouse)	(TTAGGG) <sub>n</sub>	30-150 kbp (inbred) 8-10 kbp (wild)	50-500
<i>T. thermophila</i> (ciliate protozoa)	(TTGGGG) <sub>n</sub>	250-350 bp	14-21
<i>S. cerevisiae</i> (budding yeast)	TG <sub>2-3</sub> -(TG) <sub>1-6</sub>	275-400 bp	12-14
<i>S. castellii</i> (budding yeast)	(TCTGGGTG) <sub>n</sub>	100-500 bp	unknown
<i>C. albicans</i> (budding yeast)	(GGTGTACGGATG TCTAACTTCTT) <sub>n</sub>	250-600 bp	unknown
<i>S. pombe</i> (fission yeast)	G <sub>2</sub> -8TTAC(A)	200-300 bp	unknown

## References:

*H. sapiens* a: [30] b: [31] c: [27]  
*M. musculus* a: [30] b: [25] c: [27]  
*T. thermophila* a: [3] b: [32] c: [33]  
*S. cerevisiae* a: [26] b: [34] c: [28]  
*S. castellii* a: [35] b: [36]  
*C. albicans* a: [37] b: [26]  
*S. pombe* a: [38] b: [39]

Telomeric DNA ends have been suggested to be not just linear DNA lined with telomeric proteins but to be folded upon themselves in defined structures, such that the free 3' end is tucked in. One piece of evidence for higher-order structures at the end of telomeres comes from electron microscopy images of human and mouse telomeric DNA that reveals large duplex lariat structures called telomeric loops or t-loops [40]. t-loops are formed upon invasion of the telomeric overhang into the

duplex region of the telomeric DNA [27,41]. t-loop like structures have also been reported in some ciliate and yeast mitochondrial DNAs [42,43]. In other instances telomeric circles or t-circles have been observed that comprise the single-stranded circular DNA molecules [41]. A common evolutionary origin of t-circles and t-loops for stabilization of terminal repeats, prior to the origin of specific telomeric DNA binding proteins, has been proposed as initiator of ancient telomeric tracts (Figure 1.4) [41,43].



**Figure 1.4 Formation of the t-loop at telomeric ends**

Left panel shows the schematic of for the 3' overhang insertion into the duplex region of telomeric DNA to form the telomeric loops or the t-loops (from [44]). The right panel shows attachment of the shelterin complex proteins to the t-loop, to prevent telomerase access (from [24]).

## 1.2 Telomere end-capping

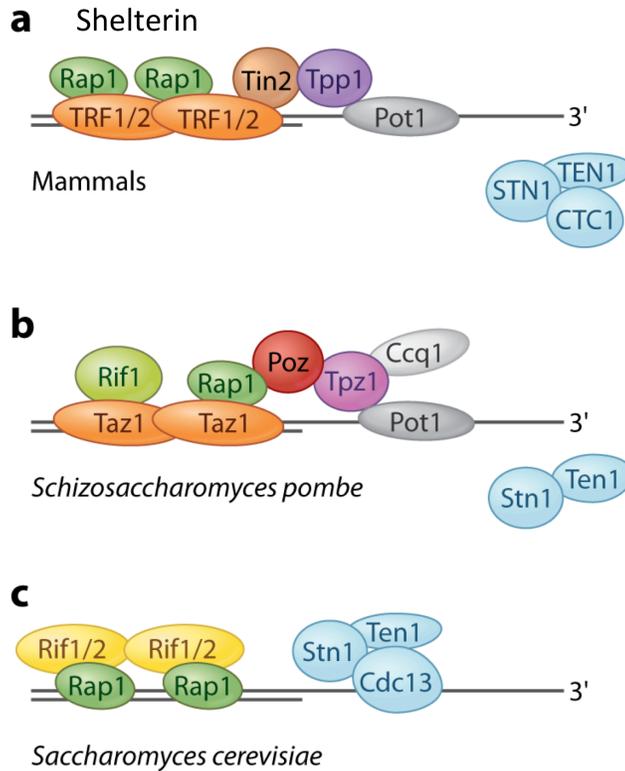
Chromosomal end-capping is necessary to prevent the ends from being recognized as dsDNA breaks and triggering DNA repair response [45]. Telomere-binding proteins are composed of those that associate with ss telomeric DNA and those that associate with ds telomeric DNA. Telomere end binding proteins, TEBP $\alpha$  and TEBP $\beta$  from the ciliate *Sterkiella nova* (*S. nova*) (formerly known as *Oxytricha*

*nova*) were the first telomere-specific proteins to be discovered [44,46]. TEBP $\alpha$  and TEBP $\beta$  form a heterodimeric complex that recognizes, binds and caps the very 3' ends of the chromosomes [24,47,48]. On the ds-telomeric DNA side, *S. cerevisiae* RAP1 (repressor activator protein 1) was the first protein discovered to bind ds-telomeres [49,50]. Over the years, new ss- and ds-telomeric DNA associating proteins have been discovered, some orthologous to *Sn*TEBPs and *Sc*Rap1 that are part of end capping factors in different organisms.

### 1.2.1 Telomere-binding proteins in budding yeast

A complex containing ss-telomeric DNA binding protein Cdc13 and inhibitors of telomerase recruitment Stn1 and Ten1 (referred to subsequently as CST) form the end-capping complex in *S. cerevisiae* (Figure 1.5 and table 1.2) [51-53]. Removal of individual subunits of the CST complex leads to long G-overhangs due to telomeric C-strand degradation and activation of DNA damage response pathways [53-56]. Cdc13 is essential for the cell and two separation-of-function mutants of Cdc13 have played major role in elucidation of its function. *cdc13-2* mutant results in shortened telomeres and the cells eventually senesce due to gradually shortening telomeres [57]. This is different from the immediate and severe response upon complete deletion of Cdc13. Hence, the end-capping function of Cdc13 is retained in the *cdc13-2* mutant. In contrast to *cdc13-2*, *cdc13-5* mutant results in elongated telomeres [58]. Thus, these two mutants define Cdc13's role as both positive and negative regulator of telomere length. *cdc13-5* encodes the Cdc13 protein lacking 230 residues from the C-terminal end. It has been proposed that this C-terminal

region of Cdc13 interacts with Stn1 for affecting its function as a negative regulator [59]. Besides ssDNA binding and interaction with Stn1, the Cdc13 protein, through its N-terminal domain, has been shown to interact with Pol1 (the catalytic subunit of DNA polymerase Pol $\alpha$ ) [60,61]. This interaction has been proposed to be important for Pol $\alpha$  primase recruitment to the telomere to initiate lagging-strand synthesis [60,62]. Interestingly, CST has been proposed to be evolutionarily close to the ssDNA binding replication protein A (RPA) complex but unlike RPA, that acts at chromosomal breaks, CST is telomere specific and required for telomere integrity [53,63]. This provides a model for analogous mechanisms in operation at double strand breaks and telomere ends and insights into their possible evolutionary origins.



**Figure 1.5 Telomere end-protection complexes from humans and yeast**

(a) Telomere-specific proteins in mammalian cells showing components of shelterin and CST complexes. In the six-protein complex, shelterin, TRF1 and TRF2 recruit TIN2 and Rap1 respectively (as indicated by black lines). (b) Telomere-specific proteins in fission yeast *S. pombe*. (c) Telomere-specific proteins in budding yeast *S. cerevisiae*. Homologous proteins are colored similarly for the three organisms. (from [64]).

The complex on telomeric dsDNA comprises of Rap1 and Rap1-interacting factors, Rif1 and Rif2 (Figure 1.5 and table 1.2). Rap1 is an essential protein that plays a varied number of roles in the cell and therefore deletion of Rap1 is lethal [65]. *S. cerevisiae* Rap1 (*ScRap1*) is the ortholog of *S. pombe* (*SpRap1*) and human Rap1 (*HsRap1*), however, unlike the fission yeast and human Rap1, *ScRap1* binds directly to telomeric DNA [50,66,67]. Besides DNA binding, the *ScRap1* is also associated with other proteins. *ScRap1* associates with Rif1 and Rif2 and deletion of

these factors gives extremely elongated telomeres, a case similar to Rap1 C-terminus deletion [68,69]. ScRap1, Rif1 and Rif2 are important mediators of telomere length maintenance in budding yeast and will be discussed in that context in section 1.3.

**Table 1. 2 Summary of telomere-associated proteins in budding yeast, fission yeast and mammals**

<b>Telomere-associated protein</b>	<b>Organism</b>	<b>telomeric DNA binding</b>	<b>Function</b>
Rap1	Budding yeast, Fission yeast, Mammals	ds (only in budding yeast)	Homeodomain topology, inhibitor of telomere length extension, associated directly/ indirectly with ds telomeric DNA
Rif1	Budding yeast, Fission yeast	-	Inhibitor of telomere length extension
Rif2	Budding yeast	-	Inhibitor of telomere length extension
Cdc13	Budding yeast	ss	End-capping protein, positive and negative regulator of telomere length
Stn1	Budding yeast, Fission yeast	-	Associates with Cdc13, negative regulator of telomere length
Ten1	Budding yeast, Fission yeast	-	Associates with Cdc13, negative regulator of telomere length
POT1	Fission yeast, Mammals	ss	Part of shelterin (telomere capping complex), negative regulator of telomere length
TPP1	Fission yeast (Tpz1), Mammals	-	Multi-domain protein, binds POT1, TIN2 and telomerase, important for recruitment and RAP
TIN2	Mammals	-	Bridge between ds and ss telomere binding proteins in shelterin complex, with TPP1 important for telomerase recruitment
TRF1	Mammals	ds	Homeodomain topology, negative regulator of telomere length
TRF2	Mammals	ds	Homeodomain topology, negative regulator of telomere length
CTC1	Mammals	-	Putative OB-fold protein, part of CST in mammals
Poz	Fission yeast	-	Bridge between ds and ss telomere binding proteins in shelterin complex
Taz1	Fission yeast	ds	Homeodomain topology, similar to TRF1/TRF2 in mammals
TEBP $\alpha$	Ciliate	ss	3' end capping
TEBP $\beta$	Ciliate	ss	3' end capping

## 1.2.2 Telomere-binding proteins in mammals

In humans, the ortholog of  $TEBP\alpha$ , POT1 (protection of telomeres 1) protein forms a heterodimeric complex with TPP1, the human ortholog of  $TEBP\beta$  and this association is important for localization of POT1 to telomeres [70,71]. POT1 specifically binds the telomeric ssDNA sequence. However, POT1-TPP1 assembly on telomeric ends is different from the orthologous  $TEBP\alpha/\beta$  complex as POT1-TPP1 does not form a tight-cap at the very 3' terminus of the overhang, unlike  $TEBP\alpha/\beta$  in ciliates. TPP1 interacts with TRF2- and TRF1-Interacting Nuclear protein 2 (TIN2) (Figure 1.5 and table 1.2) [27,72] that bridges TPP1 to the protein complex on the telomeric dsDNA. This TPP1-TIN2 interaction has been proposed to be important for telomerase recruitment [73,74].

The complex on telomeric dsDNA comprises of the homodimeric proteins, dsDNA Telomeric Repeat binding Factor 1 and 2 (TRF1 and TRF2), both of which directly interact with and recruit TIN2 and TRF2 has also been implicated in formation of t-loop structures at telomeric ends [44,67]. TRF2 also interacts with *HsRap1* through Rap1's C-terminal region [27]. Unlike *ScRap1*, however, *HsRap1* does not bind ds-telomeric DNA directly. In this activity, TRF1 and TRF2 proteins in humans and TRF-related protein, Taz1, in fission yeast are more similar to the *ScRap1* [75]. All of these proteins bind DNA through homeodomain DNA-binding motif, characterized by helix-turn-helix structure [76,77]. Whereas TRF1 and TRF2 bind telomeric repeats as homodimers [78,79] and that disruption of dimerization prevents TRF1/2 localization to telomeres [80], *ScRap1* binds as a monomer [81,82].

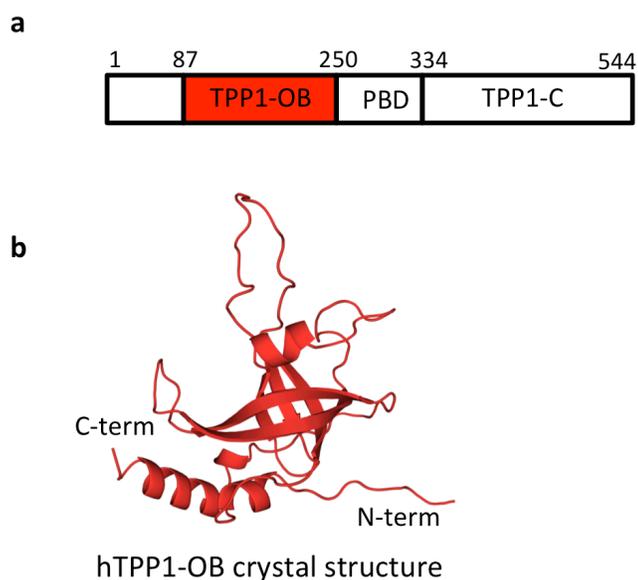
Interestingly, the TRFs carry single DNA binding homeodomain in each monomer and therefore dimerization leads to two DNA binding domains, whereas *ScRap1* monomer itself is composed of two DNA binding domains [76,77]. Thus, the overall architecture between these budding yeast and human telomere binding proteins is conserved and another example of common evolutionary themes between yeast and higher eukaryotes for telomere length homeostasis.

The six-protein complex, formed by TRF1, TRF2, Rap1, TIN2, POT1 and TPP1 is called shelterin and forms the end-capping complex on human telomeres [27,44,64,83]. Deletion of individual subunits of the shelterin complex leads to activation of various DNA damage response pathways [14]. For example, deletion of POT1 triggered ataxia telangiectasia and Rad3 related (ATR)-dependent DNA damage response, whereas deletion of TRF2 triggered the ataxia telangiectasia mutated (ATM)-dependent DNA-damage response. ATM and ATR are related kinases that are activated by DNA damage like double-strand breaks (DSBs) and ssDNA generated from those DSBs, respectively. Activation of either kinase triggers signaling cascades that ultimately leads to stalled cell-cycle progression, apoptosis and senescence [84,85]. Thus, presence of shelterin components, POT1 and TRF2, prevents the recognition of telomeres as ssDNA or double-strand breaks, thereby protecting cells from unnecessary DNA damage response. In addition, complete removal of the shelterin complex revealed its protective role in preventing six different DNA damage response pathways of ATM and ATR signaling, classical and alternate NHEJ (non-homologous end joining), homologous recombination and

resection [86]. The de-protected chromosomes showed aberrations like end-to-end fusions, sister telomere associations and sister-chromatid exchanges, indicating severe defects in chromosomal integrity and the importance of shelterin in preventing these aberrations. How the complex interplay of shelterin proteins within themselves and with other telomere-associated factors assert telomere capping and telomere length maintenance functions remain active questions in telomere biology.

### **1.2.2a Shelterin complex protein TPP1 is a key player for telomerase recruitment and activity in humans**

Human TPP1 was identified from a mass-spectrometry analysis of the protein complex co-immunoprecipitated with TRF1 and TIN2 and from two-hybrid screen with TIN2 [71,87,88]. It is a multi-domain protein (Figure 1.6a) with domain specific functions. The C-terminal domain binds to N-terminus of TIN2 and helps stabilize the TRF1-TIN2-TRF2 interaction [74]. A central domain (PBD) binds to C-terminal half of POT1 [88]. Finally, an OB-fold domain in the N-terminal half of TPP1, TPP1-OB, has been proposed to interact with telomerase [89-92] and its interaction with telomerase is important for recruitment of telomerase onto the telomeres [73]. The structure of TPP1-OB has been solved (Figure 1.6b), and the structure is similar to that of *SnTEBP* $\beta$  [90]. The identified telomerase-associating residues of TPP1 map to a loop on the base of the OB-fold (chapter 3, Figure 3.9), identifying a new binding surface for *Hs*TPP1-OB [92,93].



**Figure 1.6 Schematic of the domain map of human TPP1 and structure of human TPP1-OB**

(a) Schematic map of the multi-domain protein, human TPP1. (b) Crystal structure of human TPP1-OB [90].

Functionally, in addition to telomerase recruitment, TPP1 associates with POT1 and recruits it to telomeres. This TPP1-POT1 association has been shown to enhance POT1's ssDNA binding affinity by ~9-fold *in vitro* [94]. Upon shelterin complex dissociation, POT1-TPP1 complex plays a role in increasing the repeat addition processivity (RAP) of the human telomerase enzyme and this RAP enhancement has been corroborated in *in vitro* telomerase activity experiments [90]. This processivity enhancement by POT1-TPP1 complex is likely due to slowing down the dissociation of telomerase from the telomeric substrate and the aided translocation of telomerase along the extending 3' end of the substrate [91]. Interestingly, deletion of the OB-fold domain of TPP1 abrogates this telomere

elongation, indicating that TPP1-OB domain is a positive regulator of telomere length and this is supported by the TPP1-OB's direct involvement in telomerase recruitment [94]. As TPP1-OB has been previously hypothesized to be structural homolog of the yeast telomerase subunit Est3, it would be interesting to see if this putative structural homolog is functionally similar as well [95,96]. However, in the absence of a structure of Est3 protein and low sequence identity (15%) between Est3 and TPP1-OB, this hypothesis remains unsupported. The Est3 structure elucidation from this thesis (chapter 2) will be used for consolidating the structural and functional information known about these two proteins, one protein being part of the end-capping complex in humans and another protein being part of the telomerase holoenzyme itself in budding yeast.

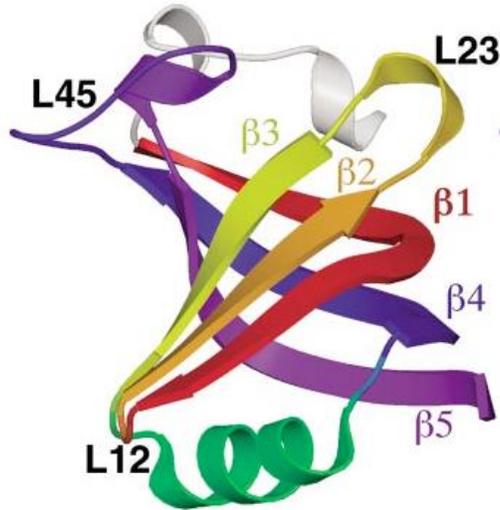
### **1.2.2b CST-like complexes in mammals**

Recent studies have proposed presence of yeast CST-like complexes in mice and humans [53]. The human CST complex comprises of the ssDNA binding protein CTC1 (conserved telomere maintenance component 1), STN1 and TEN1. STN1 of humans (hSTN1) is a homolog of yeast STN1 and has been shown to specifically bind telomeric ssDNA *in vitro* [97]. *HsSTN1* also interacts with the shelterin component *HsTPP1*, indicating co-existence of CST and shelterin complex [97]. Although CTC1 has little sequence similarity to yeast Cdc13, it likely contains putative OB-fold domains like Cdc13 consistent with the OB-fold theme in telomere end-binding proteins. However, what role CST plays in the context of mammalian cells that also have shelterin end-capping complex is unclear. Conflicting models

have arisen from different studies on mammalian CSTs and it remains to be seen what role CST plays and how much cross talk occurs between CST and shelterin complex proteins.

### **1.2.3 End-capping proteins share common structural themes**

Despite budding yeast Cdc13 not being orthologous or sharing any sequence similarity to the end-binding factors TEBP $\alpha/\beta$ , Cdc13, POT1 and TEBP $\alpha/\beta$  all recognize telomeric ssDNA by a common motif called the oligonucleotide/ oligosaccharide/ oligopeptide-binding (OB) fold (Figure 1.7) [98]. OB-fold topology is also observed for *Hs*TPP1 and yeast Stn1 and Ten1 proteins of the end-capping complexes [63,90], indicating a conserved theme for binding in telomere-associated proteins. As OB-folds accommodate highly variant amino acid sequences and telomere-associated proteins are also highly diverged in sequence, sequence-based search for homologs is difficult. However, as new structures of telomere-associated proteins get solved, they shed light not only on the structural and functional conservation between proteins but on evolutionary relationships as well [99].



**Figure 1.7 A canonical OB-fold structure**

OB-folds are oligonucleotide, oligopeptide, oligosaccharide binding proteins and it is characterized by a five-stranded  $\beta$ -barrel, capped by a helix at the base (green). The grooved region between L12 and L45 is especially important as the binding surface for the ligands (from [81]).

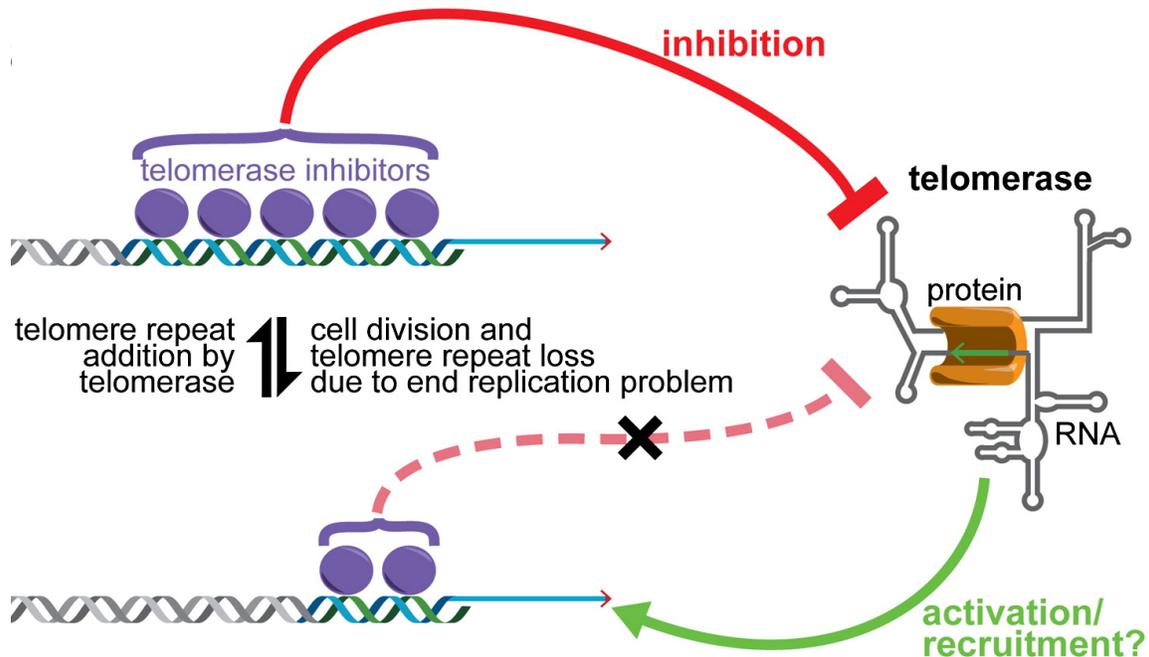
### **1.3 Telomere length homeostasis**

Telomeric DNA is lost due to ERP at the end of every replication cycle and because of end-resection [100]. On the other hand, telomeres are replicated by the specialized ribonucleoprotein (RNP) complex, telomerase (section 1.4 for details). Telomere length, therefore, is the result of a balance between attrition and lengthening processes.

In the absence of telomerase, telomeres shorten at a constant rate until the cells reach replicative senescence; the point at which they no longer proliferate. Replicative senescence is triggered when the telomeres have reached a critically short size ( $< 1.5$  kb), leading to a peak in dicentric chromosomes due to chromosomal fusions [101]. It has also been shown, both in mammalian and yeast cells, that the shortest telomere, rather than the average telomere length,

determines the triggering of cell cycle arrest [102,103]. Somatic cells in humans lose ~150 bp of telomeric DNA per doubling [104], and in yeast this rate is 3-6 bp per generation [105,106]. The differences in rates have been proposed to be due to the different telomere extension rates, which parallel the overall shortening rates [28]. Another justification for this difference in shortening rates comes from the recent finding in humans that the last RNA primer during the lagging strand synthesis is placed ~70-100 nt from the very end and thus the placement of the last RNA primer itself drives telomere shortening [17]. This placement of the RNA primer might be species-specific and might be one of reasons for differences in shortening rates.

In the presence of telomerase, telomere shortening is countered by telomere elongation (see section 1.4 for overview of telomere replication). Regulation of telomere length, *i.e.*, the process in place to maintain an average length characteristic of a species, requires a feedback mechanism that senses the length of the telomeric DNA and triggers appropriate response to either extend the end, if it is short, or to let it shorten, if it is long [52,107]. In support of the feedback mechanism, protein-counting model for telomere length regulation has been proposed for yeast [68,108,109] and humans [110,111] (Figure 1.8).



**Figure 1.8 Protein counting model for telomere length homeostasis**

A representative scheme for the proposed protein-counting model for telomere length regulation in humans and yeast is shown. When telomere length is long, there are more negative regulators (shown as purple blobs) of the telomerase recruitment present on the telomeres. These include, in yeast, the double-stranded DNA binding protein Rap1 and its associating proteins Rif1 and Rif2. In humans, the negative regulators TRF1 and TRF2 (both double-stranded DNA binding proteins) are part of the shelterin complex. Upon reduction of telomere length, the number of inhibitors occupying the telomeres is reduced and this reduction in telomerase inhibition ultimately leads to telomerase recruitment and telomere elongation (from [24]).

### 1.3.1 Telomere length homeostasis in budding yeast

In budding yeast, the telomeric dsDNA-binding protein Rap1 has been shown to negatively regulate telomere elongation and to be a key component of the protein counting model [112]. The extent of telomere shortening has been found to be directly proportional to the number of Rap1 molecules targeted to telomeric DNA and that this activity of Rap1 is present in its C-terminus [108,112]. Deletion of Rap1 C-terminus leads to extremely elongated telomeres. Rap1 C-terminus has been reported to recruit another set of telomere-elongation inhibitors, Rif1 and Rif2,

and that deletion of either Rif1 or Rif2 leads to ~600bp to ~150 bp longer telomeres [68]. Deletion of both Rif1 and Rif2 gives extremely elongated telomeres, a case similar to Rap1 C-terminus deletion. Increased Rif1 and Rif2 present on longer telomeric tracts inhibit association of another protein Tel1 (homolog of mammalian ATM kinase). Tel1 kinase or the MRX complex (MRE11-Rad50-Xrs2) are proposed to promote the generation of the single-stranded TG-rich 3' overhangs by recruitment of exonucleolytic proteins and is more active/ abundant at shorter telomeres [52]. The overhangs generated by Tel1 mediation promote binding of Cdc13, ssDNA binding protein that is both a negative and positive regulator of telomerase [24]. Generation of overhangs and recruitment of Cdc13 are important steps that subsequently lead to telomerase recruitment and overhang extension.

In another set of studies, telomeres have been proposed to switch between non-extendible and extendible states [109]. Assessing the number of nucleotides added to a telomere in a single cell cycle, telomerase exhibited an increased preference for telomeres as their lengths decline (extendible state); however, the number of nucleotides added was not dependent on telomere length [109]. Deletion of Rif1 and Rif2 in these experiments, led to longer telomeres, supporting the proposal that Rif1 and Rif2 are negative regulators of telomerase. In a similar study, the deletion of Tel1 reduced the overall frequency of telomere elongation, and the preference to short telomeres was further lost in the absence of a sub-telomeric sequence [113]. This preference for short telomeres was restored by tethering of subtelomeric binding protein Tbf1 to upstream sequence. Thus, the effectors of

length maintenance, preferentially for shorter telomeres, were identified in these studies, looking at single chromosome levels.

### **1.3.2 Telomere length homeostasis in humans**

In humans, the telomere repeat binding proteins TRF1 and TRF2 (proteins of the shelterin complex) have been proposed to mediate the protein-counting for telomere length regulation [110]. Number of shelterin complexes, more on longer telomeres and *vice versa*, has been proposed to regulate the recruitment of telomerase in a telomere tract length dependent manner [44] (Figure 1.8). However, much of this is derivative from the yeast data and the steps and effectors involved in length maintenance in humans are not well understood.

## **1.4 Telomerase**

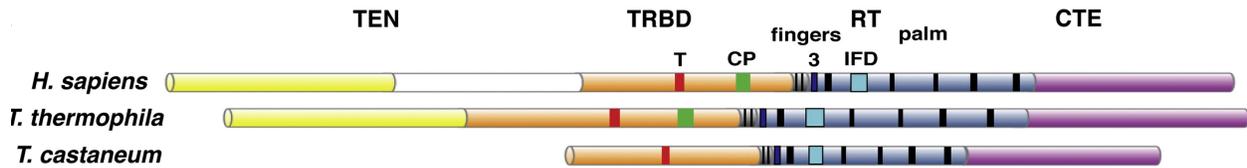
Telomerase was first discovered by Elizabeth Blackburn and Carol Greider as a telomere terminal transferase activity in *Tetrahymena* extracts, capable of extending synthetic telomeric primers [114]. For this seminal discovery, they were awarded the 2009 Nobel Prize. They identified the telomere terminal transferase to be an RNP complex with necessary RNA and protein components and named this complex “telomerase” [115]. A general discussion of telomerase is presented here, with discussion of telomerase in yeast and humans in the succeeding sections (1.4.1 and 1.4.2).

Telomerase prevents telomere shortening of the chromosomes in germline and adult stem cells. However, it is turned off in human somatic cells making the

cells prone to replicative senescence [116]. Consistently, ectopic overexpression of telomerase catalytic subunit in telomerase-negative human cells in culture is sufficient to maintain the telomeres at a steady length and allow cells to overcome senescence [117]. In unicellular organisms like yeast, telomerase is constitutively active - a requirement for long-term proliferation [83,118].

Telomerase is an RNP particle having two conserved and essential core components: a reverse transcriptase (*HsTERT* in humans and *Est2* in yeast) and an associated telomerase RNA (TR) component (*HsTERC* in humans and *TLC1* in yeast). The TERT/*Est2* and TER/*TLC1* components are essential for telomerase activity *in vitro* [35,119]; however, *in vivo* the telomerase complex also requires associated proteins for its assembly, localization and regulation at telomeres [120].

The reverse transcriptase component in telomerase has similar active site domain (RT) to other known reverse transcriptases like HIV RT [121,122], but telomerase RTs, unlike other known RTs, have a constitutively associated multi-functional RNA component [123]. To date, complete structure of any telomerase RT has not been solved, but structures of some individual domains are now available that help guide the understanding for overall structural organization of TERTs. The domains defining the human TERT protein from N- to C-terminus are the telomerase essential N-terminal domain (TEN)/GQ, TR binding domain (TRBD) containing QP, CFP and T motifs, catalytic RT domain containing telomerase-specific domain insertions IFD and motif 3, and a C-terminal extension (CTE) (Figure 1.9) [122-125].

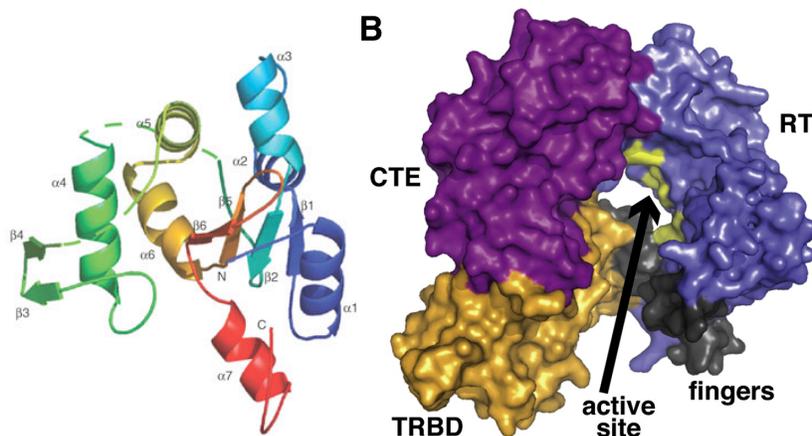


**Figure 1.9 Domain maps of TERT proteins from different organisms**

TERT consists of 4 major domains from N- to C-terminus (left to right): telomerase essential N-terminal domain (TEN) in yellow and white, telomerase RNA binding domain (TRBD) in orange, catalytic reverse transcriptase domain (RT) in blue, and a C-terminal extension (CTE) in magenta (from [125]). T, CP, 3 and IFD are some of the conserved motifs in TRBD and RT domains. As seen in this comparison, the *T. castaneum* TERT does not have the TEN domain, which is essential for activity of other known TERTs.

This scheme of domain arrangement also generally holds true for the budding yeast telomerase reverse transcriptase Est2 [123]. Some of the architectural insights into the three-dimensional assembly of the TERT proteins have come from the structure of the putative TERT from *Tribolium castaneum* *TcTERT*, which curiously lacks the TEN domain and with no reported telomerase activity or RNA component reported thus far [126]. However, RT domain in *TcTERT* does contain the telomerase-specific domain insertions IFD and motif 3, indicating *TcTERT* might be an active or evolutionary intermediate of known telomerases [125]. *TcTERT* forms a ring like structure with RT making a "palm and fingers" motif and CTE, the "thumb", wraps around it to make extensive contacts with the TRBD that specifically recognizes the TR and completes the ring (Figure 1.10b) [126,127]. TRBD is an essential and conserved domain of TERTs and its structure from *T. thermolphila* is composed of mostly helices that come together to form a nucleic-acid binding fold [128]. TRBD makes extensive contacts with the TR and these contacts promote assembly of the

TERT/TR complex. In addition, TRBD has been proposed to enhance telomerase RAP by restricting the RT domain from traveling to the end of the template RNA and dissociating during telomere-length extension [128]. The ring of the reverse transcriptase allows docking of the nucleic acid chain [127]. As mentioned previously, most known telomerases also contain the TEN domain, structural insights into which comes from crystal structure of *T. thermophila* TEN (Figure 1.10a) [129]. This structure is a mixed  $\alpha$ - $\beta$  structure composed of four anti-parallel  $\beta$ -strands, surrounded by seven  $\alpha$ -helices. This domain facilitates telomerase RAP via its interaction with both TR and telomeric ssDNA [125,130].



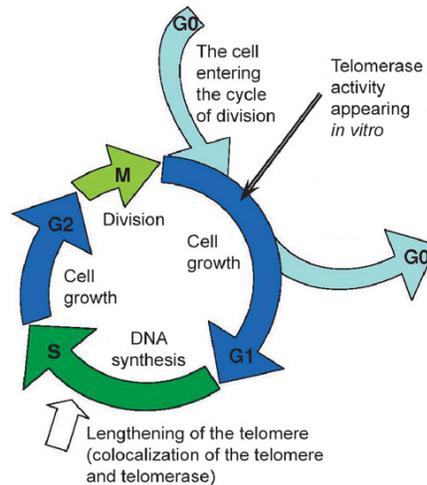
**Figure 1.10 Structure of the *T. thermophila* TEN domain and *T. castaneum* TERT**

(A) The crystal structure of the *Tetrahymena* TEN domain (from [129]). The color code is blue to red from N- to C-terminus. (B) Surface representation of *T. castaneum* TERT (from [125]). The RT domain forms the palm and finger motif, with CTE as the thumb. Both motifs make extensive contacts with the telomerase RNA binding domain (TRBD). Active site is pointed by an arrow and yellow colored surface.

The RNA moiety provides the template for extension of telomeric DNA repeats and serves as a scaffold for binding of telomerase associated proteins that regulate the reverse transcription reaction [11,131]. Overall, the reported length of

the TRs have been found to be highly variable, ranging from the smallest (147-209 nucleotides (nts)) in ciliates, moderately long (312-559 nts) in vertebrates [132], considerably long (779-1817 nts) [133] in yeasts, to the longest recorded 2.2 kb long TR in *Plasmodium falciparum* [134]. In addition to the enormous variability in length, the sequence of the TR is also highly diverged [135,136]. Despite the length and sequence diversity, TRs are characterized by some universal structural motifs like the template region defining the sequence of the telomere repeat, a pseudoknot that interacts with TERT, and a stem terminus element [135-137]. Some species specific elaborations/ motifs are also present that allow RNP assembly, for *e.g.*, budding yeast contains various extended arms for binding the holoenzyme subunit Est1, the telomerase recruitment and localization protein Ku, and the RNP biogenesis protein Sm complex [135].

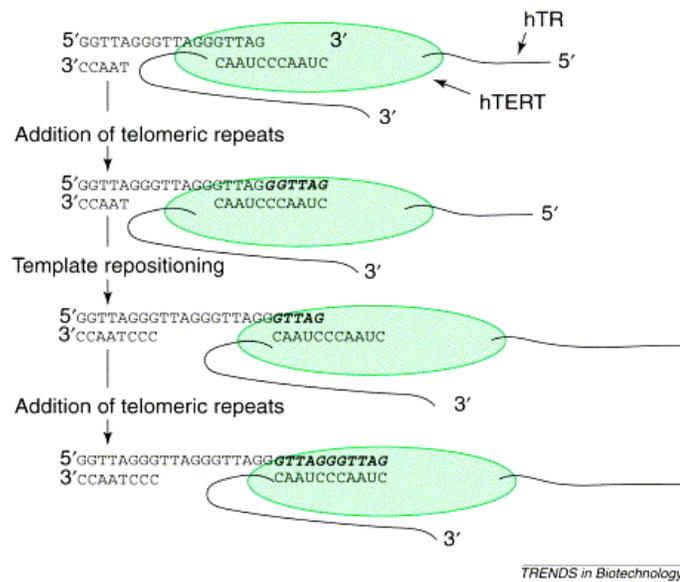
Telomere replication by telomerase requires the recruitment of telomerase RNP preferentially onto the short telomeres most in need of elongation [103,109]. This is a regulated, but still not well-understood, process that involves varied number of proteins and is cell-cycle dependent (Figure 1.11) [24,129,138]. Extension of telomeres by telomerase is restricted to the late S phase in yeast [22,125], however, in humans the telomeres replicate throughout the S phase [139] and telomerase extends telomeres within 30 minutes of the conventional chromosomal replication [140].



**Figure 1.11 Cell-cycle dependent lengthening of telomeres by telomerase**

Cell-cycle progression is shown. Human leukocytes, having reduced but detectable levels of telomerase activity show presence of that activity at G1, but actual telomere lengthening is reported from the S phase [141] (image from [142]).

Extension of the 3' telomeric end by telomerase involves the base pairing of the RNA template with the terminal telomeric repeat in the initiation step. Elongation proceeds in the catalytic core of telomerase, such that only one repeat of the telomere is added in single elongation and synthesis ends when 5' terminal boundary element of the template is reached. Telomerase can continue to extend the same telomeric end with multiple repeats by unwinding the RNA:DNA duplex and sliding along to base pair the template RNA with the new 3' end for another round of extension. This addition of multiple repeats by telomerase, before dissociation, is known as repeat addition processivity (RAP) (Figure 1.12) [143].



**Figure 1.12 Nucleotide and repeat addition processivity of telomerase**

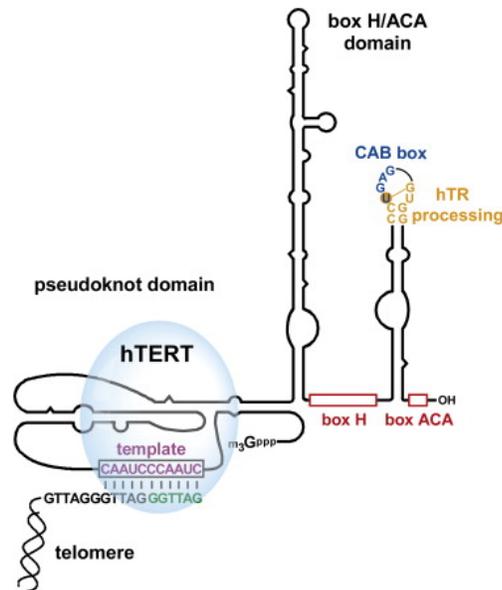
Telomerase RNA (*HsTR*) template hybridizes with the telomeric substrate DNA and *HsTERT* reverse transcriptase then catalyzes the addition of nucleotides till the end of the template is reached. This translocation of telomerase on the DNA-RNA hybrid as single nucleotides are added is known as type I translocation and it mediates nucleotide addition processivity. When the end of the template is reached, the DNA-RNA hybrid separates and the template RNA is translocated to end of the newly formed 3' terminus. This way, the *HsTERT* is ready to catalyze the addition of another set of repeats. This template repositioning is called type II translocation and is important for addition of multiple repeats onto the telomeric end before telomerase dissociation. Capability to add multiple repeats, is thus, called repeat addition processivity (image from [144]).

On the other hand, nucleotide addition processivity involves forward translocation of the telomerase catalytic site along the template after each nucleotide incorporation step until the 5' template boundary element. Whereas most telomerases, with some exceptions, are nucleotide addition processive *in vitro* [143], the RAP varies widely from different organisms and at different telomere lengths. A number of telomerase-associated proteins have been identified thus far that promote RAP in different organisms, however, a unifying picture of the mechanisms

involved and the effects on telomere length maintenance due to defects in processivity are not well understood [145].

### 1.4.1 Telomerase in humans

In humans, the core telomerase RNP is composed of the reverse transcriptase TERT and telomerase RNA *HsTERC*. RNA-binding protein dyskerin, with specificity to H/ACA sequence motif is also part of the holoenzyme [146] (Figure 1.13). Dyskerin plays important role in RNP biogenesis and stability of the *HsTERC* and because of these requirements dyskerin forms an integral part of the catalytically active human telomerase complex [147,148].



**Figure 1.13 Schematic representation of human telomerase holoenzyme**

A model of the human telomerase RNA *HsTR* with predicted secondary structure, is shown in association with a subset of the known interactors of the human telomerase holoenzyme (from [149]).

Apart from these three proteins, other proteins potentially part of the telomerase holoenzyme are the evolutionarily conserved dyskerin-associated

proteins reptin and pontin, nucleolar protein 10 (NOP10), non-histone protein 2 (NHP2), glycine/arginine-rich domain containing protein 1 (GAR1) [150-152], and ever-shorter telomeres 1A (EST1A), a homolog of the yeast telomerase protein Est1p [153,154]. Unlike yeast Est1 that binds TLC1 RNA [155], *HsEST1A* binds *HsTERT* independent of *HsTERC* [153]. However, like yeast Est1, hEST1A also shows direct interaction with telomeric ssDNA [153]. Over-expression or depletion of *HsEST1A* leads to cellular problems like telomere-shortening, chromosomal fusions, apoptosis and nonsense mediated mRNA decay [153,154,156]. Another telomerase component, TCAB1 (telomerase Cajal body protein 1), was identified in co-immunoprecipitation complex with tagged-dyskerin [146]. Association of TCAB1 with the holoenzyme is through direct binding to *HsTERC* and TCAB1 has been proposed to be important for recruitment of telomerase to telomeres [157]. Thus, human telomerase associates with multiple accessory factors. The functional significance of many of these associations is still unknown.

#### **1.4.1a Telomerase dysfunction linked to human disease**

Mutations in telomerase components or associated machinery have been linked to telomerase dysfunction and consequently to a number of human diseases. Abnormally short telomere lengths in bone-marrow cells, due to mutations in telomerase subunit genes, have been linked to Dyskeratosis congenita (DC) [158]. DC is a disease characterized by abnormal skin pigmentation, nail dystrophy and mucosal leukoplakia and patients have short telomeres [159]. Multiple mutations from the dyskerin subunit are reported to be linked with DC and majority of these

are due to single amino acid substitutions. It is implicated that the substitutions lead to disruption of some protein-protein or protein-RNA associations being mediated by dyskerin, important for telomerase stability [160,161]. In another report, some DC patients have been shown to carry mutations in the TIN2 subunit of shelterin that lead to disruption of TIN2-TPP1 association and likely problems in telomerase recruitment [73]. Other, related and non-related diseases are also known, all stemming from mutations in one or the other component of telomerase holoenzyme. Some of these diseases are enumerated in the following paragraph.

In a related disease to DC, aplastic anemia patients also show short telomere lengths and the primary defect is proposed to be at the stem cell level [162]. Aplastic anemia has been reported in some cases to be caused by mutations in telomerase and *HsTERC* genes. Additionally, recently identified mutations that severely affect RAP have been reported to manifest as a familial telomere-mediated syndrome [145], indicating important role of RAP in human telomere length extension. Finally, pulmonary fibrosis is a fatal lung disease that affects individuals age 50 years old or more [163]. This disease is also characterized by mutations in TERT and TERC components of telomerase [164]. In a related disease individuals develop liver cirrhosis, due to mutations in the TERT component [165]. Thus, overall, understanding telomerase regulation and thereby modes of its dysregulation is important for developing therapeutics to counter the effects of such deleterious diseases as DC and other telomerase-dysfunction related diseases.

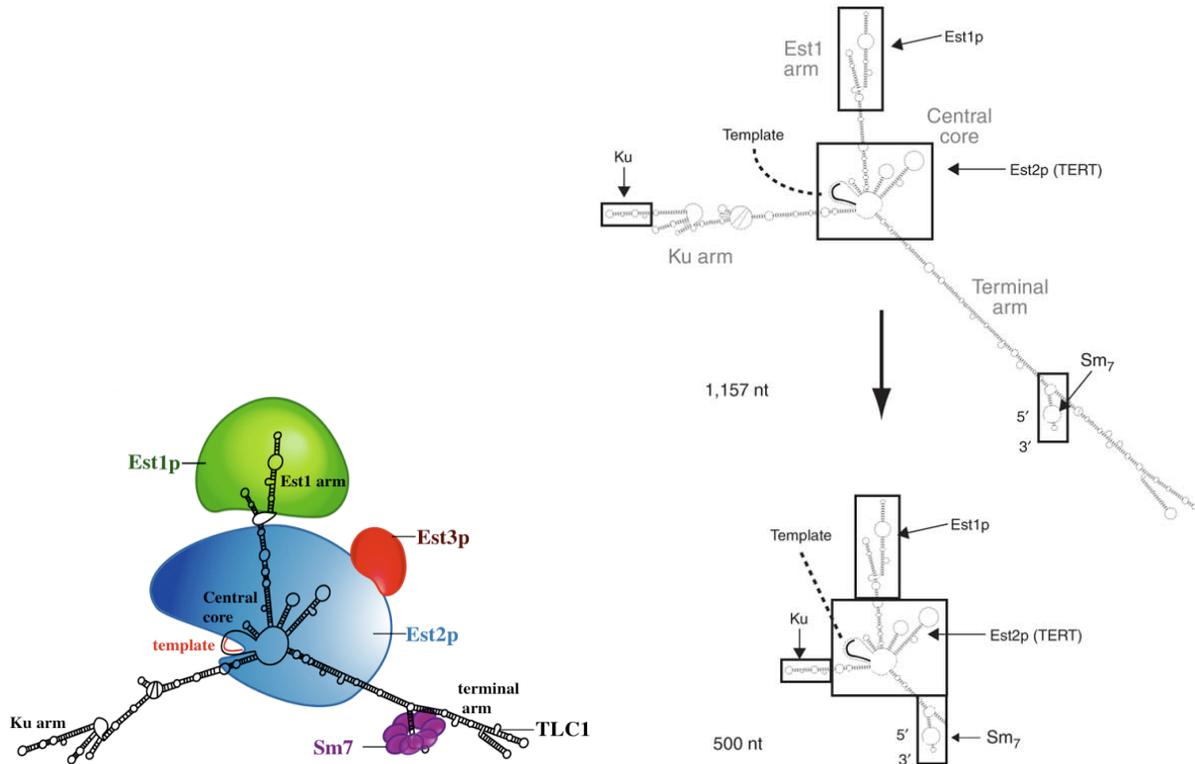
Apart from the telomerase dysfunction related diseases listed above, telomerase activation has been reported in somatic cells in ~90% of human cancers [116]. The activation of telomerase in cancerous somatic cells leads to circumvention of the proliferative limit placed on the cell leading to cellular immortality. This makes telomerase and its regulatory proteins and pathways key areas of study for cancer therapy. This requires a better understanding of the role that telomerase and its accessory proteins play in telomere length homeostasis in humans and model organisms like the budding yeast.

#### **1.4.2 Telomerase in the budding yeast**

Yeast telomerase activities were first identified from two budding yeast species - *S. cerevisiae* and *S. castellii* [35]. Fractionated yeast extracts from both species were used in telomeric substrate extension reactions *in vitro* and visualization of extension was through a series of <sup>32</sup>P-labeled products separated on denaturing sequencing gels. Whereas extension by *S. castellii* fractions showed multiple rounds of translocation with 8-nt periodicity, extension by *S. cerevisiae* telomerase extract was not as robust without any product extension beyond the first pass of the RNA template. Thus, it was concluded that *S. castellii* telomerase is repeat-addition processive *in vitro*, but no RAP has been detected for *S. cerevisiae* telomerase.

The budding yeast telomerase holoenzyme RNP (Figure 1.14) is composed of Est1, Est2, Est3 (Ever Shorter Telomeres: EST phenotype) and TLC1 RNA. Est1, Est2, Est3 and Cdc13 (also known as Est4 because of EST phenotype of *cdc13-2*

allele) proteins were identified from genetic screens looking for mutagenized colonies with defective telomere length and growth phenotypes [57,166].



**Figure 1.14 Schematic representation of yeast telomerase holoenzyme**

Panel on the left shows a model of the *S. cerevisiae* telomerase RNA TLC1 with predicted secondary structure, is shown with other known subunits of the yeast telomerase holoenzyme (from [11]). Panel on the right shows full-length TLC1 RNA on top and the mini-T RNA at the bottom. Scaffolds for protein binding are boxed in black. (from [167])

Est1 was the first to be identified from a screen devised in 1989 [166] to identify telomerase components using mutants that satisfy three criteria. Mutants had to (i) be defective in telomere addition, (ii) null mutant of the identified gene should have a senescence rather than inviable phenotype and (iii) show progressive increase in frequency of chromosome loss. The assay involved a 3-tier approach starting with transformation of the yeast with a circular plasmid containing inverted repeats of telomeric sequence around a URA3 gene and selectable LEU2

marker. This circular plasmid resolves into a linear molecule at low frequency, such that the URA3 marker is disrupted and cells become resistant to 5-FoA. After mutagenesis with EMS (ethyl methane-sulfonate), cells were scored for increased or decreased resistance to 5-FoA and also for slower growth compared to wild-type colonies. Mutants that had colony number smaller than the wild type were selected and tested for telomere shortening and senescence phenotypes. Two out of the 32 clones were defective in telomere length, detected by TRF (telomere restriction fragment) analysis with only one of the mutant showing shorter than wild-type telomere length and this mutation was discovered to be carried in a new gene that was named Est1, from the ever shorter telomeres phenotype. A genomic yeast library was used to complement the telomere length defect identified in this last clone. Complementing activity was mapped to a gene originating from a fragment of chromosome VIII and sequencing of the fragment led to the sequence of the newly coined Est1 gene. Little could be surmised at this point about the function of this novel protein.

The TLC1 gene that encodes the yeast TR was identified in a separate screen in 1994 [168] as having the same set of characteristics as Est1. This assay was based on the fact that telomeres are capable of silencing genes in their vicinity [169], therefore those genes were chosen from the screen that when over-expressed suppressed telomeric silencing. TLC1 was found to contain *S. cerevisiae* telomere templating sequence 5' -CACCACACCCACACAC- 3'. TLC1 RNA in yeast *S. cerevisiae* is ~1.2 kb long and acts as a scaffold for association of protein subunits

directly and indirectly associated with telomerase holoenzyme [170]. A number of studies have attempted to predict the functionally relevant secondary structure elements of this long RNA using sequence conservation and RNA secondary structure prediction programs (Mfold [171], Alifold [141,172]) and to map TLC1 associated proteins to the relevant structure elements [142,155,170,173]. Modeling and genetic analysis led to the creation of a miniature TLC1 (mini-T) RNA composed of all the functional domains of TLC1 [167]. mini-T was able to reconstitute telomerase activity *in vitro* and was functional *in vivo*, indicating that most of the RNA is not essential and that scaffolding arms are sufficient for basal telomerase activity (Figure 1.14).

In 1996 [57], in a screen similar to the 1989 screen [166], 22 new mutants with similar phenotypes to Est1 and TLC1 were found. Thirteen of these mapped to the Est1 gene locus. The remaining 9 mutations were identified to be in Est2, Est3 and Est4 (subsequently identified as Cdc13 [57,144]). Epistatic analysis revealed that Est1, Est3 and Est4 function of Cdc13 function in the same *in vivo* pathway as telomerase [120] and therefore these proteins might function either as components of the holoenzyme or as regulatory factors for telomerase activity [174]. Est3 has been shown to be the most abundant of the three subunits of the *S. cerevisiae* telomerase holoenzyme with  $84.3 \pm 13.3$  molecules per cell, compared to  $71.1 \mp 19.2$  for Est1 and  $37.2 \mp 6.5$  for Est2 [175]. In another study, the copy number of TLC1 RNA was reported to be  $\sim 37$  molecules per diploid cell [176]. Thus, TLC1 and Est2 levels are comparable from the two studies.

The reverse transcriptase Est2 and the TR TLC1 constitute the catalytic core of the telomerase holoenzyme and this core is sufficient for telomeric substrate extension *in vitro* [11,35,120]. Telomerase activity was present in extracts prepared from *est1-Δ*, *est3-Δ* or *cdc13-2<sup>est</sup>* yeast strains but was absent from *est2-Δ* strain indicating that Est1, Est3 and Cdc13 are not required for catalysis *in vitro*. Additionally, a direct association of Est1 and Est3 with the telomerase catalytic core is implicated due to co-immunoprecipitation of TLC1 TR by HA<sub>3</sub>-tagged Est1, Est2 and Est3 but not by Cdc13 [167,177]. Although Est1 and Est3 are not required for catalytic activity *in vitro*, they are absolutely essential *in vivo*, implicating a regulatory role for these accessory proteins [120,174].

Est1 is a large, 82 kDa, protein that has been genetically well characterized, however, until now Est1 protein has not been amenable to *in vitro* biochemical studies. Est1 binds to a bulged stem loop of TLC1 RNA [155,178] but the Est1 domain involved in this interaction is not well characterized. Est1's association with telomeric ssDNA binding protein Cdc13 [179] is important for recruitment of the telomerase onto the telomeres [180]. Binding of Est1 to Cdc13 has been shown to include an electrostatic component, as suggested by the interaction between E252 of Cdc13 and K444 of Est1 [59]. Essential interaction of Est1 and Cdc13 is further confirmed by telomerase-recruitment deficient alleles of Cdc13 and Est1, *cdc13-1* and *est1-60*, which exhibit reciprocal co-suppression in a double mutant strain. ScEst1 has also been shown to have weak but sequence-specific telomere ssDNA binding activity *in vitro*, but it needs to be further characterized [181]. Finally, the

two regulatory proteins, Est1 and Est3, have been proposed to interact and this allows Est1 to recruit Est3 to telomeres in cell-cycle dependent manner [175,182], however, this direct interaction is also not well established.

Est3 is an important regulatory factor for yeast telomere length homeostasis. To date Est3 has been ascribed no conserved biochemical activity, although roles in telomerase activity stimulation have been alluded to from different yeast species (for further elaboration on Est3 see section 1.5) [96,183,184]. This thesis aims to elucidate the high-resolution structure of Est3 and use structural insights towards functional exploration. Towards that goal, a more detailed discussion of important discoveries reported for Est3 follows here.

## **1.5 Est3: Discovery and enigmatic function**

Est3 was first discovered in a genetic screen in budding yeast *S. cerevisiae* and later shown to be an obligate subunit of the yeast telomerase holoenzyme [57,177] (section 1.5.1). *S. cerevisiae* Est3 gene was also identified to carry the programmed +1 frameshifting function, that is required for complete translation of the protein [185]. The significance of this +1 frameshifting in Est3 is not clear and the Est3 genes not belonging to budding yeast taxa do not carry this frameshift function [186,187].

To date Est3 gene has been identified in multiple fungal species, however, no obvious Est3 gene has been discovered in non-fungal organisms. Like most telomere-associated proteins, sequence of the Est3 proteins is highly diverged. This makes identification of Est3 genes from genomes difficult based on sequence

conservation alone. Sequence alignment of annotated Est3 proteins from different fungal species is shown in Figure 1.15. Despite overall sequence divergence, some regions of the alignment indicate sequence conservation and the conserved residues are likely important either for structure or function of the protein. The length of the thus far identified Est3 proteins averages around 200 amino acid residues indicating that Est3 is a relatively small protein (~20 – 26 kDa) and likely defined by a single domain. Refer to Table 2.1 for sequence similarity between ScEst3 and its homologs. As such, Est3 is also the smallest known subunit of the telomerase holoenzyme complex.

**Figure 1.15 Alignment of Est3 protein sequences**

Est3 sequences from 17 different yeast species were aligned using the PROMALS multiple sequence alignment [188]. PROMALS denomination is used for showing conservation on top of the alignment, with 9 being the most conserved and anything below 5 (not marked) is variable. Residue numbering is shown on the left and right ends of the sequence panels. Variable cysteines in the *S. cerevisiae* sequence are pointed by red inverted-triangles with numbering on top of each triangle.

**S.cerevisiae**  
**S.mikatae**  
**S.bayanus**  
**S.castellii**  
**S.dairensis**  
**S.paradoxus**  
**A.gossypii**  
**K.lactis**  
**P.stipitis**  
**P.pastoris**  
**P.guilliermondii**  
**D.hansenii**  
**C.albicans**  
**C.tropicalis**  
**C.lusitanae**  
**C.glabrata**  
**C.dubliniensis**

9 HSKPTDSVFLQPWIKALI--EDNSEHQ--YHP-SGHVIPSIL-----TKODIALPHMSPTILNPNCHFAKIKRFYNVCDY--KYVASIRDS-SHQI-----LVEFSO-ECVSNFERTH 107  
 9 HSKPTDSIFLQPWIKALV-EDNSEHQ--YHP-SDHVIVPL-----TEQDLKLPMSAKILNPNCHFAKIKRFYNVCDY--KYVASIRDS-SHQI-----LVEFSP-ECVSNFERTH 108  
 9 NSRQVDSIFLQPWIKALI--DDNSEHH--HIP-SDHVIVPAL-----AQQDLALPHMCQOLLNPNFHAKIRFRYNVCDY--RVVASVRS-SHQI-----LVEFSE-KCVSDFERIN 96  
 9 KLSQTSIFLQPWIEGLLRESLQKTY-LPGN-QOREVPSL-----NEADLRAPQCSPKVLTNHCFTKVKFKINNY-AISASIRDS-RFQI-----LVEFTP-KCVSNFERRH 109  
 9 KQTVSDSVFLQPWIANIIRKEIOAKTY-LPGN-QRLSARL-----SRDDAPEISETILNNSHFAKVTKFVSNNY--KIFASIRDS-TYQI-----LVEFTP-KCVLEFERVY 109  
 9 HSKPADSVFLQPWIKALI--ENNSEHQ--HHP-SDHVIVPIL-----TKODIALPHMSPKILNPNCHFAKIKRFYNVCDY--KYCASIRDS-SHQI-----LVEFSP-KCVSNFERTH 116  
 9 RAHKADSIFLREWLVDAVVPALERS-----GACAPWAGVECFIPALPPATATLSL-EPTVIQPKRFRIRVFRTRVHDF--AVCAVARDA-GCCI-----LVEFTP-HCVSNFERRY 110  
 9 RLTNNSVFLQEWIKPSVRYLKNKTKFWEQRELVTDL-----LEHDIIIESAFOQTALNAPAFORVIRVFKFRVNDY--TWAVATIRS-TALI-----LCYFTV-DCVLDVETIN 111  
 5 -NDKFPFLVFLQSWLWESVEKCIDFS-----RRYTSIL-----IKKNFVDETP-----RKAQPIVRISKFLKVDN--KDLTVLILSDSTNLM-----FAIFFKPTIIVNFEDKY 94  
 7 -----SNIFVCSWLLKAL---NEV-----KDETVSK-----IPKF-----NRTPRVMKRLRLITVSRD--LIRLIIIDS-ANSI-----LAEVSA-DAIAIEIYNE 80  
 5 -TYWEWELVLTQTTWIOKAVTRCLQHH-----LRFSAVD-----VEHFIHFSSH--KHVRETPILRVRYLTKTDQLALTAIVSDT-RHLI-----FAKFPYDPTIILKFFQLY 97  
 3 VNIKYPLVLHDNWLLESITGOINKN-----VKYTNLSL-----IKKDFIIPSS-----ISKTPILRILRFLKTKSSDITAILSDS--THTV-----IAIFFPDPFAIINFERY 93  
 24 PENDVPMIMQSSWLVNEVIKSINS-----GOYINAL-----IKKDFQPTV-----LPVSFILRLIKFTATTDSDKITAVLADS--THKI-----FAILFFPAIVDFENKY 113  
 7 DNTGVPTRVLSWLSKDVINSIQGN-----TNYVNPV-----VRKNFKPTL-----MKVSFLRILELFIPTESMDITAILSDS--THKI-----LSIFKDFPAIVDFENRY 96  
 21 IAMDYPIVLLNSWLGVLGTCIRDQ-----SIYSTNL-----LVKNF IKFESPTCNLNTPLRLVFAFTKSDSGETSIVLHDS--SHKI-----LVLFTK-ECIERFESRY 115  
 9 RSNAVESVYLHGWRDMLLESKTSO-----NIAVIPR-----VDPEASIFLLSRRIYANRHFVKIKTFQVHNY--SVYASVKDS-OHOI-----LSQFTP-KCVSEFERN 103  
 4 SENDVPMIMQSSWLVNEVIKSINS-----GOYINSI-----IKKDFQPTV-----LPVSFVLRLIKFTATTDSDKITAVLADS--THKI-----FAILFFPAIVDFENKY 93

**S.cerevisiae**  
**S.mikatae**  
**S.bayanus**  
**S.castellii**  
**S.dairensis**  
**S.paradoxus**  
**A.gossypii**  
**K.lactis**  
**P.stipitis**  
**P.pastoris**  
**P.guilliermondii**  
**D.hansenii**  
**C.albicans**  
**C.tropicalis**  
**C.lusitanae**  
**C.glabrata**  
**C.dubliniensis**

108 NCRITSETNCLMIIIGDADLVYVNTSRAMSHFKICL-----SNISS--KEIVPVLNVNQATIFDIDOVGSLSTFP---FVYKYL----- 181  
 109 NCRITAEVTNCLMIIIGDASLINATSRVSKSHFKIRL-----SSISS--NDLVPVLOINQVTFIDIDOVGSLSAFP---FVYKYL----- 182  
 97 NFRITSETNCLMIIIGDADLAYVTSQALARMFKRL-----SSIST--SETPVLIINQATIFDIDOVGSLNPNFP---FVYKYL----- 180  
 110 HRRITSETNCLLIVIGDAAIYKSRDQITTFQGNID-----FIISKV--SPLVPLQINQASLFDGDOVQHLRSFP---FVYSTL----- 185  
 110 RSRITSETNCLFVIGDCTVIYKTRSQIRSSPKFDLQSLAGEKNKNG---FGLFPVLIQINQASLFDSDQVQLLFEPP---FIYNKLGHTV----- 196  
 117 NCRITSETNCLMIIIGDADLVYVNTSRAMSHFKIRL-----SNISS--REIVPVLINQATIFDIDOVGSLSSFP---FVYKYL----- 190  
 111 HORITSTVNSLFIQNTSLLFYARSDAAAFAEVA-----LMNG--SSTPLVLRVGDCAIFDQDOVESHRFP---LVQEPFRFVQSLDMAMOGS-----VLRS 196  
 112 NDRITLNTLNTLFIQNTVTLQFVNHRECKLWFNQDF-----LXEIFAM-----PG--LRMVPVLIKIEKARMDRQOISSNQVE---WVYDTLHG----- 184  
 95 QRITFTHTQNLIIHKKANLRFIERKDYP-LYEIPEAM-----LSHVLEVLIDLEIFQRDQIMLSNKNVESTLKFJHNEINLYELLGRKWNNAE--ENNDR-----EDYDD-VVVF 195  
 81 RORFSKGMNTEILVVESSLVFTSLHQLMPWYSYLPPR-ITKTNGEDYV--ISKYAVLEITKLSIFNRDOVKELVDIP---LVYLEKYSW-----PKLVTSHVDD-EPTF 180  
 98 QRRLTYNTAGCLFVIKQAKLRFAPGEVAKDFDSI-----SP--NIQIVILEIDDFVRSRQAILPHEALSKVNMIIYDREYDLCGRDCYIPE--YSKDVKEQVSDDYGD-VVSV 205  
 94 KHRITYHTPNLSLIIKQANLRFVNNVLETFEWGITI-----ED--EIDVAVLEVLIDLEIFQRDQIMLVGNIENNLQIYYDKYSYLNLCGRQRSDSLSGEKEVILQSYDD-VVSI 202  
 114 HRRMTYNTSRCLRIRTHKANLRFMDKSTVKKCYGRKS-----DG--GLAIAVLEVLIDLEIFQRDQIMLVGNIENNLQIYYDKYSYLNLCGRQRSDSLSGEKEVILQSYDD-VVSI 209  
 97 HQRTYNTVNRNIIHKKANLRFMTISVKNDFKLN-----KG--ALDVIVLEILDELFELIDEFVFNSTESLEFVYDDEPQOLCRQTKNT-----EIFNYDDGLINS 195  
 116 QRITTYHTVHSLLLVKQANLRFLLFQLRSKTFVYGG-----LRI SP--KVALVYLETSDVDFQRDQIMVSWFLAEKMLHMLVYDVEYFKKYGOQNVSP---QGIADVFLVENDDG-LISD 224  
 104 RSRITSDVTNLTFLMIGDQAKLGMVVDLHRVFGEXIV-----SILFNLGMPYIPIVLIINQAFILLVDYQEAEMKTP---FVYQYI----- 180  
 94 HRRMTYNTSRSSVIRIHKANLRFMDQSTVNNKCYNMKT-----NG--NLAIAVLEILELDFIFLKDQYSYFITSYENRKLKXVVEDTRYDQLCREKMHK---FEYDD-LMCD 189

Genetic and biochemical studies allude to a direct interaction of Est3 and catalytic core component Est2 [95,189-191], and this might be one mode of its association with the holoenzyme. Other studies also allude to a putative direct interaction between Est1 and Est3 [175,182], however, this interaction is not well established. Therefore, one mode of Est3's function is likely mediated via interaction with Est2. What that function is remains elusive.

Telomerase isolated from yeast extracts has not been reported to show RAP *in vitro* [167] and nucleotide addition is also non-processive *in vitro* [192]. No processivity factor has been discovered yet for the *in vivo* reaction as well. For human telomerase enzyme, binding of TPP1 to ssDNA binding protein POT1 *in vitro* increases telomerase's RAP [90]. The POT1-TPP1 complex was found to be structurally similar to the ciliate TEBP- (Telomere-End Binding Protein) complex and therefore a widespread presence of such heterodimeric complexes, including in yeast, was predicted. As Est3 and TPP1-OB have been hypothesized to be structurally similar (section 1.2.2a) one of the regulatory roles of Est3 might be to enhance the activity of the yeast telomerase. This, in part, was shown to be the case for *S. castellii* and *C. albicans* Est3 proteins *in vitro* (see sections 1.5.2 and 1.5.3) [183,184]. It remains to be seen if the stimulation of telomerase activity is a conserved feature of Est3 proteins and if there are other, species-specific or non-specific, functions associated for Est3's role as a regulator of telomere-length.

Analysis of Est3 from several yeast species reveals some conserved and non-conserved functions.

### **1.5.1 Est3 from *S. cerevisiae***

*S. cerevisiae* is a good model for telomere biology because of established genetic manipulation techniques and conserved telomeric components to higher eukaryotes. Consequently, a lot of research on Est3 has been done in this system. Secondary structure prediction of *S. cerevisiae* Est3 (ScEst3) using the PSIPred protein secondary structure prediction server [193], reveals secondary structure elements and arrangement similar to those for OB-fold motifs (five-stranded  $\beta$ -barrel, sometimes capped with a helix) in the core of the protein (Figure 1.16). However, without structural elucidation the OB-fold motif cannot be conclusively established for Est3.



TLC1 RNA, but HA<sub>3</sub>-tagged Cdc13 does not [120,177]. As *est3-Δ* strain shows reduced telomere-length *in vivo*, as well as, a senescence phenotype [57], a regulatory function of Est3 in telomere length homeostasis is implied.

Multiple lines of evidence indicates that one mode of Est3's interaction with telomerase is likely through its direct interaction with the Est2 subunit. HA<sub>3</sub>-tagged Est3 can immunoprecipitate TLC1 RNA, however, similar attempt from an *est2-Δ* strain failed to co-immunoprecipitate TLC1, strongly suggesting that an interaction with Est2 is required for Est3's association with the telomerase core [177]. Deletion of the TLC1 RNA stem-loop bulge that specifically binds Est2, also fully disrupts Est3's association with the RNA, whereas deletion of the RNA bulge that binds Est1 does not affect Est3's association with the RNA. This supports the model of Est2-dependent association of Est3 with the telomerase complex, without any contribution from Est1 [183]. This interaction of Est3 with telomerase has been ascribed to N-terminal (TEN) domain of Est2 [189]. Multiple temperature-sensitive (*ts*) alleles in the N-terminus of Est2 were identified, such that 25°C was permissive and 35°C was the restrictive temperature for colony growth. Rescue of temperature-sensitiveness was attempted by transformation of the *est2<sup>ts</sup>* strain with overexpression of Est1, Est3 or Smd3 (one of the seven Sm proteins that bind the telomerase complex via an arm of TLC1 RNA). Overexpression of Est1 could rescue growth at 35°C for all *est2<sup>ts</sup>* alleles from the N-terminus of Est2, however, Est3 overexpression rescued only the *est2<sup>ts</sup>* alleles, at either N80, G85, G112, A113, T135 or F143 position, from the TEN domain. This suggested that Est3 either directly or

indirectly associates with this region of Est2. In the same study [189], another mutant in TEN domain of Est2, *est2-ala1* (residues 40-49 mutated to alanines), showed EST phenotype with gradual telomere-shortening and cell senescence, but at the same time *in vitro* telomerase activity was retained, indicating that this mutant was functional and therefore folded. Strains carrying HA<sub>3</sub>-tagged Est1 and Est3 at the endogenous locus were transformed with plasmid containing coding sequence for protein A-tagged *est2-ala1* or just protein A-tagged *EST2*. Protein A-tagged Est2-ala1 was unable to co-immunoprecipitate either Est1 or Est3, supporting the proposal that this region of the N-terminus of Est2 might be involved in interaction with Est1 and Est3. As the co-immunoprecipitation was from cell-extracts, whether the binding between the bait and co-immunoprecipitated proteins was direct or indirect cannot be deduced.

A set of mutants on Est3's putative surface have also been identified that abrogate ability of Est3 to associate with telomerase core components [95]. In this study, putative surface residues that were also conserved were tested for telomere length, synthetic growth defect and association with telomerase (tested by probing for TLC1 RNA from the co-immunoprecipitated complex). Mutations in Est3 residues E104, E114, T115, N117 and D166 were found to be disruptive towards Est3's ability to bind telomerase, had reduced telomere lengths and showed growth defect in synthetic-lethality assay (Figures 4.3, 4.4). Another set of mutants that were also affected in telomere length and had growth-defect but that still allowed Est3's association with telomerase were also identified. These were termed

separation-of-function mutants and were in residues K68, K71, R110, D164 and V168 of ScEst3. It is unknown, however, what this function(s) of Est3 is(are).

To address the question whether Est3 and TEN domain of Est2 (Est2<sup>TEN</sup>) proteins directly interact with each other, binding with purified Est3 and Est2<sup>TEN</sup> was recently attempted and abrogation of binding tested for previously identified Est3 mutants [95,190]. The proteins in this study were recombinantly expressed from *E. coli* and His<sub>6</sub>-tagged Est3 and Mbp-tagged Est2<sup>TEN</sup> were incubated and captured on amylose resin. Retention of His<sub>6</sub>-Est3 on amylose resin was probed by Anti-His western-blot. Of the mutants tested, only ETN(114,115,117)AAK mutant of Est3 showed no association either by direct binding test to purified Mbp-tagged Est2<sup>TEN</sup> protein or to telomerase in the *in vivo* co-immunoprecipitation test. One caveat of this direct binding test was that Mbp-tagged Est2<sup>TEN</sup> was not very purified ( $\leq 50\%$ ), *i.e.*, had multiple other protein bands as viewed from the coomassie-stained SDS-PAGE gel. Impure protein solutions, sometimes, indicate less stable proteins and inadvertent co-purification of chaperone proteins for stability. Having chaperones in the purified fractions can, in turn, give misleading binding data with other proteins/ nucleic acids etc. Another caveat was that the wash fractions were not included in the western-blots and therefore it was not clear if sufficient washing to eliminate all non-specifically bound His<sub>6</sub>-Est3 was done. Hence, a direct binding between Est3 and TEN-domain of Est2 is still not well established and remains an open-ended question.

A direct association of Est1 and Est3 has also been implicated but not conclusively established [175]. In a recent study from the Zakian lab, chromatin immuno-precipitation (ChIP) experiments were used to show the recruitment of Est3 to the telomeres in the late S/G2 phase and the disruption of this recruitment in the absence of either Est1, Est2 or TLC1 [175]. Absence of Est3, however, did not affect the recruitment of Est1 and Est2 onto telomeres in the late S/G2 phase, indicating that Est3 works downstream of the Est1 and Est2/TLC1 assembly onto the telomeric ends and that Est3 is likely not required for this assembly. Whether this result is due to direct interaction with Est1 is unclear. This study implicated a direct interaction of Est3 with Est1, for promoting Est3's recruitment to telomerase but the binding data with purified Est1 and Est3 proteins was not conclusive.

### **1.5.2 Est3 from *S. castellii***

*S. castellii* is an attractive model system for telomerase study because, unlike *S. cerevisiae* telomerase, the isolated *S. castellii* telomerase is capable of multiple rounds of translocation with 8-nt periodicity, *i.e.*, RAP [35]. This allows more sensitive assessment of telomere-associated factors for effect on telomerase activity [183]. The periodicity is because *S. castellii* telomere has exact repeats of the 8 nt TCTGGGTG sequence, whereas *S. cerevisiae* telomere sequences are degenerate. *ScasEst3* shares a 48% identity with *ScEst3* (Table 2.1).

In contrast to what has been observed with the less robust *S. cerevisiae* enzyme, telomerase activity from *est3-Δ* strain of *S. castellii* was found to be 2- to 3-fold reduced in nucleotide addition processivity, relative to the telomerase activity

from the *EST3* strain and this reduction was primer-independent [183]. This reduced activity could be fully restored upon back-addition of purified *ScasEst3*, confirming Est3's contribution to enhancement of basal telomerase activity. Furthermore, a mutant of *ScasEst3*, R112E-Est3, was also able to restore the activity completely upon back-addition. This observation is interesting because the R112 position in *S. castellii* Est3 is same as the R110 position in *S. cerevisiae* Est3. R110E-Est3 mutant in *S. cerevisiae* is a critical mutant, equivalent to the null-phenotype, but still retaining telomerase association [95]. Restoration of complete telomerase activity by the corresponding R112E-*ScasEst3* indicates that the R112 residue is important for the function of Est3, but this function is independent of Est3's proposed role in telomerase binding. 1.5.3 Est3 from *C. albicans*.

Budding yeast, *C. albicans*, is a common fungal pathogen in humans. In that respect it is quite different from the budding yeast *S. cerevisiae* (also known as baker's yeast). These two species of budding yeasts diverged ~800 million years ago [194] and Est3 from *C. albicans* Est3 (*CaEst3*) shares <30% identity with *ScEst3* (Table 2.1). Comparing the Est3s from these evolutionarily diverged species can lead to insights into distinct features these proteins have adopted and the features that got retained because of some selection pressure in favor of that function. *C. albicans* telomerase activity, like that of *S. cerevisiae*, is non-processive *in vitro* and the isolated enzyme is capable of adding only a few nucleotides to the telomeric substrate [195]. This *in vitro* activity is further affected in nucleotide addition when the telomerase extracts were derived from either *est1*- $\Delta$  or *est3*- $\Delta$  strains,

implicating the role of Est1 and Est3 proteins in enhancement of telomerase activity [184,196]. This result was similar to the telomerase activity enhancement seen for *ScasEst3*. However, the activity enhancement for *CaEst1* and *CaEst3* was dependent on the primer (telomeric substrate) being extended, unlike the primer independent effect for *ScasEst3* [183]. This difference might be attributed to the unusually long (28 nt) RNA template present in *C. albicans* TR that might have led to development of specialized mechanism(s) different from other telomerases with smaller RNA templates (for *e.g.*, *S. castellii* with 13 nt RNA template and *S. cerevisiae* with 16 nt RNA template). Further, for *C. albicans*, the primer-specific enhancement of activity shows mutual dependency of Est1 and Est3 in telomerase activation [184].

The association of either Est1 or Est3 with the telomerase core was also found to be co-dependent, as protein A-tagged Est1 was unable to co-immunoprecipitate TR from an *est3-Δ* strain and conversely the protein A-tagged Est3 was unable to co-immunoprecipitate TR from an *est1-Δ* strain [184]. This is in contrast to *ScEst1*, association of which with the core telomerase is not Est3 dependent [175,178], and *ScEst3* also associates with the core-complex in an Est2-dependent manner even in the absence of *ScEst1* [183]. Barring experimental errors, the co-dependency of Est1 and Est3 for recruitment in *C. albicans* telomerase, therefore, might indicate a shift from *S. cerevisiae* Est1 and Est3. *S. cerevisiae* Est1 and Est3 do not require each other for recruitment (loss- or gain-of-

function), however, association between these two proteins might still be important for overall activity of the telomerase.

In a mutational screen of *CaEst3*, a number of key residues in Est3 were identified, based on alignment with human TPP1-OB (putative structural homolog of Est3) [96]. This study was similar to the mutagenesis screen in *ScEst3*, done around the same time in the Lundblad lab (section 1.5.1) [95]. Both *CaEst3* and *ScEst3* studies selected residues for mutational analysis by sequence conservation and protein model building [90]. See Table 1.3 for comparison of mutants from *CaEst3* and *ScEst3* studies. Most of the mutants matched in their effect in either of the species, with a few exceptions as noted in the table.

**Table 1. 3 Comparison of the counter-part mutants from *C. albicans* and *S. cerevisiae* Est3**

Mutant class	<i>CaEst3</i> mutant <sup>a</sup>	Effect	<i>ScEst3</i> mutant <sup>b</sup>	Effect
	D91A	Null	D86A	Null
	W36A	Protein expression like WT, Progressive loss of telomeres, Reduced telomerase association, Drastically Reduced telomerase activity	W21A	<u>Null</u>
i	D169A	Protein expression like WT, Progressive loss of telomeres, Reduced telomerase association, Drastically Reduced telomerase activity	D166A	Protein expression like WT, Progressive loss of telomeres, Reduced telomerase association
iii	F109A	Progressive loss of telomere, Normal telomerase association, Slightly reduced telomerase activity	--	
ii	R116A	Progressive loss of telomere, Normal telomerase association, Normal telomerase activity	R110A	Protein expression like WT, Progressive loss of telomeres, Normal telomerase association
ii	T121A	Progressive loss of telomere, Normal telomerase association, Normal telomerase activity	T115A	Protein expression like WT, Progressive loss of telomeres, <u>Reduced telomerase association</u>

Reference: a: [96] b: [95]

Thus, comparing and contrasting the two mutational studies on *ScEst3* and *CaEst3* residues, three classes of Est3 residues emerge: i) that are involved in association of Est3 with Est2, ii) that are involved in Est3's affect on telomerase activity, and iii) those that are involved in function(s) separate from the first two. However, because they were based on just the comparison to residues on *HsTPP1-OB* and sequence alignment, these mutagenesis efforts towards Est3's functional characterization are in no way complete, although they are a good starting point towards shedding light on Est3's role(s). In the absence of high sequence similarity between Est3 and its putative structural homolog, the threaded model is likely an

incomplete picture of Est3 structure and an inspection of Est3's high-resolution structure will help guide a more extensive structure-function analysis.

## 1.6 Thesis statement

Despite numerous genetic- and biochemistry-based experiments, the function of the yeast Est3 protein remains elusive. Est3 is not required for telomerase activity *in vitro*, but is absolutely essential for telomere length maintenance *in vivo*. Therefore, Est3 is an important regulatory factor in budding yeasts for telomere length homeostasis, but how it brings about that regulation is unknown. As structure of a protein can help to shed light on its function through structural similarity to other known proteins and guide the search for relevant mutations for *in vivo* studies, our goal for this thesis was to solve a high resolution structure of the Est3 protein and do structure-dependent and independent search for Est3 function. Est3 is not easily amenable to experiments for structure elucidation. Thus, intensive protein and sample condition optimizations were done to obtain relatively stable NMR sample and collect NMR data. Structure of the protein was solved by combining minimal NMR experimental data to the RASREC Rosetta structure calculation program (Chapter 2). Finally, this structure of Est3 was used for structure-guided *in vitro* and *in vivo* analysis (Chapter 3). The solved structure was instrumental in conclusively establishing human TPP1-OB as the structural homolog of yeast Est3. The structure was also used as a validation tool for an *in vivo* guided *in vitro* analysis study to test a set of dominant negative mutants of Est3 for retention of structural integrity (Chapter 4).

## **Chapter 2. Structure of the *Saccharomyces cerevisiae* Est3 protein**

### **2.1 Introduction**

#### **2.1.1 Est3 is a diverged protein with predominantly uncharacterized structure and function**

Est3 is an essential regulatory subunit of yeast telomerase. This protein was discovered in a screen in 1996 [57], that identified mutants that (i) showed progressive increase in frequency of chromosome loss, (ii) showed defects in transformed plasmid linearization, (iii) were defective in telomere addition, and (iv) led to senescence rather than inviable phenotype upon deletion of the identified gene. Rescue, by complementation analysis, of the senescence and telomere-length addition defect was achieved by transformation with plasmids containing either of the EST1, EST2, EST3 and EST4 genes. However, only Est2 was found to be required for telomerase activity *in vitro* [120], suggesting the remaining factor play central roles in telomerase regulation. Despite numerous genetics and biochemistry-based efforts to elucidate the role of this protein, its function has eluded discovery. Thus, in this study, we undertook a structural approach to characterize this protein, in the expectation that elucidation of a high-resolution structure of Est3 will help guide the search for its function as well.

Est3 is the smallest protein of the telomerase complex, with average estimated molecular weight around ~22 kDa between species. It has been identified

in several yeast species, however, Est3 has not been found in non-yeast organisms. The overall sequence similarity in the various Est3 proteins (Table 2.1) is low, and therefore, sequence-based identification of new Est3 proteins, in yeast or non-yeast organisms, is difficult. In addition, no known homologs have been structurally characterized, leading to a lack of structural information about this protein.

**Table 2.1 Sequence identity and similarity of Est3 proteins to *S. cerevisiae* Est3 protein sequence**

Organism	Identity	Similarity	Length of alignment*
<i>S. mikatae</i>	81%	88%	182
<i>S. paradoxus</i>	86%	92%	190
<i>S. bayanus</i>	73%	88%	181
<i>S. dairensis</i>	47%	65%	192
<i>S. castellii</i>	49%	67%	187
<i>A. gossypii</i>	40%	55%	185
<i>D. hansenii</i>	32%	48%	128
<i>K. lactis</i>	35%	51%	185
<i>P. stipitis</i>	27%	45%	157
<i>P. pastoris</i>	24%	45%	119
<i>C. lusitaniae</i>	26%	46%	156
<i>C. glabrata</i>	39%	56%	185
<i>C. dubliniensis</i>	26%	46%	149
<i>C. tropicalis</i>	32%	53%	104
<i>C. albicans</i>	30%	47%	148

Target sequence is *S. cerevisiae* sequence for matching with query sequences. Alignment was done with NCBI's BLAST program

\* Length is the number of amino acid sequences in the query sequence aligned to the 181 residue target sequence of *S. cerevisiae*

### 2.1.2 Est3 secondary structure elements and prediction of 3D models

Knowledge-based algorithms have been developed to predict secondary structure of macromolecules like proteins. Some programs freely available through [expasy.org](http://expasy.org) are PSIPred, PredictProtein, Jpred etc. PSIPred [193] is a fairly accurate

method and used by 3D modeling servers as well for secondary structure prediction. Prediction of Est3 protein secondary structure by PSIPred suggested that the Est3 protein has five  $\beta$ -strands with helices in between (Figure 1.15).

Like secondary structure prediction, programs have been developed for 3D structure predictions as well and many of these methods use template-based modeling. Two examples of template-based methods are HHpred/MODELLER[197] and I-TASSER/LOMETS [198,199], which were the top two multi-server methods in ninth edition of the Critical Assessment of Techniques for Protein Structure Prediction (CASP9) held in 2010 [200]. Another structure prediction server called ROBETTA [201] combines both the template-based and *de novo* prediction strategies. The template-based modeling relies on sequence homology between proteins with similar 3D structures. When proteins are more sequentially similar, template-based method finds a homologous structured-template from the Protein Data Bank (PDB) more accurately, leading to prediction of close-to-accurate model(s) of the target protein. However, protein targets become “hard” as the sequence homology becomes low (< 30%) [200]. As Est3 sequences are highly diverged, even from other known Est3 sequences, any structural matches from the PDB will be low in sequence similarity, thus reducing the accuracy of sequence homology based model prediction for Est3.

In the absence of a high-resolution structure, to get an idea about the topology of the Est3 protein, three-dimensional model of Est3 was predicted, using the Local Meta-Threading-Server (LOMETS,

<http://zhanglab.ccmb.med.umich.edu/LOMETS/>). LOMETS creates a sequence profile by PSI-BLAST and subsequently uses this profile to predict secondary structure using PSIPred. Using the constraints from PSI-BLAST and PSIPred, the meta-threading server then threads the target sequence over the structures in the PDB library. The template-structures are then ranked according to the best fit. One feature that was notable from the output of the modeling run for Est3 was that although the LOMETS uses PSIPred to predict the secondary structure of the protein, the best tertiary model(s) might not contain that exact secondary structure. The top “hit” for Est3 homologous structure was identified as the human TPP1 OB fold domain (*HsTPP1-OB*) and the predicted threaded-model reflects features quite similar to this “hit” (see section 2.3.5f for 3D model). However, as the sequence identity between Est3 and the top “hit” template protein, *HsTPP1-OB*, is low (15% identity), template-based homology modeling alone is likely not the best method to reliably predict tertiary model of Est3 completely. The improvement in recent years in 3D structure modeling, however, has provided a good starting strategy to obtain putative structure models of proteins, given just their primary amino acid sequence depending on the level of detail needed for analysis. A few studies have recently attempted to base genetic screenings for Est3 on such predicted 3D models [95,96]. Such strategies have their advantages, however, the *de novo* 3D modeling for low sequence conservation proteins (like most proteins found in telomere biology) is not very good and the best-hit models might not reflect the complete picture of the protein’s structure. Modeling results from LOMETS suggested that Est3 was a

"hard" target and over the years improvement in modeling algorithm has resulting in Est3 being classified as "medium" difficulty. This indicates that the predictions have to be viewed with the caveat in mind that some of major parts of the predicted model might be wrong or unassigned to any secondary structure element. However, even if the predicted structure is not accurate with active sites of the protein not correctly displayed, the overall topology predicted by secondary and tertiary prediction programs most likely holds true. This means that Est3, like many other telomerase-associated proteins, is derived from the common theme of OB-fold motifs.

### **2.1.2a Structural features of OB-fold proteins**

The OB-fold is the oligonucleotide-/ oligopeptide-/ oligosaccharide-binding fold found in proteins and it pervades through all three kingdoms of life [99]. It is a flattened  $\beta$ -barrel comprising of an antiparallel five-stranded Greek-key motif [81], first described by Murzin [98] (Figure 1.6). An  $\alpha$ -helix is often found between the  $\beta$ -strands 3 and 4 and this helix caps the bottom of the  $\beta$ -barrel. Anti-parallel sheet formed by strands 1,2 and 3 generally forms one face of the barrel and another anti-parallel sheet from strand 4 and 5 (and 1) forms a second face. Strand 1, often, is very long and bulged in the middle to allow  $\beta$ 1 to contribute to both the sheets of the barrel [81]. Despite the common structural features of the OB-fold, the sequence of the core  $\beta$ -strands, the helix and the loops is highly diverse [202]. This flexibility in sequence accommodation likely allows for divergent evolution and establishment of new features from the same protein fold [202].

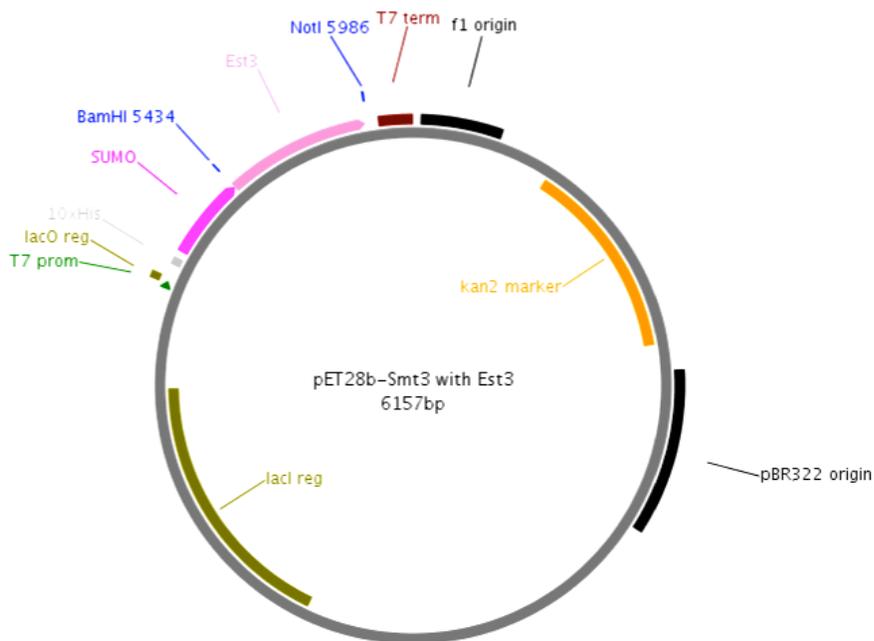
As OB-fold is a common structural motif, a lot is known about the ligand binding by these folds from numerous structures, several solved in complex with their binding partners [81]. The canonical binding face of OB-fold proteins includes the  $\beta$ -strands 1, 2 and 3 and bounded by loops L12 and L45 and sometimes L23 as well [81,202,203]. The length and sequence of the loops are highly variable and allow for accommodation of different types of ligands into the binding site. Nucleic-acid ligand binding OB-fold proteins seem to be the most abundant with members found in phage, bacteria, archaea and eukaryotes [99]. Nucleic acid binding relies on the bases being in close contact with the OB-fold protein binding site, with the phosphodiester backbone exposed to the solvent [81]. The interaction with the protein is mostly via stacking interaction between the bases and the aromatic side chains or packing with the hydrophobic side chain or polar side chain groups [81]. The variable loops L12, L45, L3 $\alpha$  and L45 significantly increase the binding surface and assist in the binding. In addition to binding a ligand at the canonical binding-site, OB-fold proteins also associate with themselves to form homo-dimers or with other proteins to form hetero-dimers and higher order oligomers. Different sites on the OB-fold surface may contribute towards this binding and formation of homo- or hetero-dimer contacts can influence or be influenced by nucleic-acid binding at the canonical site [202]. This interaction with the other proteins can be mediated by loops and helices that are either part of the core OB-fold or are found at the N- and/or C-terminal extensions.

This chapter focuses on the structure elucidation of Est3. Even with secondary and tertiary structure predictions guiding the initial insights into Est3 structure, the predictions were not of high confidence and carry a lot of caveats to make accurate prediction of some or most parts of the molecule difficult. Est3 is not a well-behaved protein in solution and required optimization of a number of variables (from protein construct to sample buffer etc.). Additionally, alternative strategies to the conventional structure determination techniques (xray and NMR) had to be explored to solve the structure of the Est3 protein as the conventional strategies were not feasible for Est3 because of low sample solubility and non-homogeneity. Final structure validations support our calculated structure. This is the first structure of any subunit of yeast telomerase subunits. The structure was used to design mutants for *in vivo* studies and also to drive *in vitro* mutational analysis for functional characterization of Est3 (Chapter 3). The knowledge of the structure will, hopefully, also guide the functional search of Est3 in other *in vivo* and *in vitro* studies.

## 2.2 Materials and Methods

### 2.2.1 Est3 expression, purification and construct optimization

Vector for Est3 expression: Fusion with His<sub>10</sub>-SUMO was achieved by introducing the *S. cerevisiae* Est3 coding sequence in the pET28b His<sub>10</sub>-Smt3 [204] vector, using 5' cgggatccatgccgaaagtaa 3' as the forward primer and 5' ttgcggccgcttataaatatttatatacaaatgggaaagtacttaacgatccg 3' as the reverse primer. pET-His<sub>10</sub>-Smt3 vector was a generous gift from Chris Lima (Sloan-Kettering). The Est3 genes were cloned between the BamHI and NotI sites in a pET-His<sub>10</sub>-Smt3 expression vector. All constructs were verified by sequencing (Figure 2.1).



**Figure 2.1 pET28b vector map with His<sub>10</sub>-SUMO-Est3 coding sequence**

Plasmid map shows the His<sub>10</sub>-SUMO-Est3 coding sequence with upstream T7 promoter sequence. Est3 sequence was introduced between BamHI and NotI restriction sites. Plasmid imparts kanamycin resistance for selection.

Construct for GB1-Est3 fusion: The GB1 coding sequence was PCR amplified from the pGEV1[205] vector, such that it was flanked by BamHI restriction site on the 5' and 3' ends. The forward primer used for this PCR reaction was 5'-d(cgggatccatgcagtacaagcttgctcttaatg)-3' and reverse primer used was 5'-d(cgggatcc cggaccgcccgggttcggttac)-3'. The amplified sequence was introduced between the SUMO and the Est3 coding sequence in the His<sub>10</sub>-SUMO-Est3 construct.

Construct for His<sub>10</sub>-SUMO-Est3<sup>ΔN</sup>: Est3<sup>ΔN</sup> construct was part of a series of constructs tried for optimization of Est3 protein stability (Figure 2.10). It proved to be relatively the best construct for stability and therefore was used in our NMR studies for structure determination. Est3<sup>ΔN</sup> was made by deleting 12 amino acids at the N-terminus of wild type Est3. It also contained a cysteine to serine mutation at residue number 142 based on a screening that showed stabilization of Est3 in the presence of this mutation. Deletion of 12 residues at the N-terminus and C142S mutation were introduced into the His<sub>10</sub>-SUMO-Est3 on the pET28b plasmid, using site-directed mutagenesis. Construct was verified by sequencing.

Protein expression: The His<sub>10</sub>-SUMO-Est3 protein, its optimized mutant His<sub>10</sub>-SUMO-Est3<sup>ΔN</sup> used in NMR studies, as well as the GB1 fused construct, were expressed in BL21(DE3) cells as follows. The cells were grown at 37°C until an O.D 600 of ~ 1.0 was reached. Cells were then cold-shocked on ice for ~1h, followed by induction of protein expression with 1.0 mM IPTG. Cells were allowed to grow post-induction for 24 hrs at 15°C and harvested by centrifugation. Growth condition

optimizations were done initially, to arrive at the temperature and duration of post-induction growth for maximum yield and solubility of SUMO-Est3 protein. Separate growth condition optimizations were done for growth in deuterated and selectively protonating media (refer to sections 2.2.5, 2.2.6 and 2.2.7).

Protein purification: The cell pellet was resuspended in lysis buffer (buffer A) containing 100 mM potassium phosphate buffer (pH 7.5), 100 mM Na<sub>2</sub>SO<sub>4</sub>, 10% glycerol, 10 mM imidazole and 3 mM βME and an EDTA-free protease inhibitor cocktail tablet (Roche). Cells were lysed by sonication and clarified by centrifugation. The clarified cell lysate was subjected to Ni<sup>2+</sup>-affinity chromatography by gravity-flow (GE Healthcare). The His<sub>10</sub>-SUMO-Est3 was eluted with buffer A containing 300 mM imidazole and concentrated in a 10,000 MWCO concentrator (Millipore) to 2 mL volume and incubated overnight at 4°C with SUMO specific protease, Ulp1, to cleave off the His<sub>10</sub>-SUMO tag [204]. Subsequently, the cleavage reaction was subjected to size-exclusion chromatography (Superdex75, GE) on an AKTA FPLC system in buffer B (100 mM potassium phosphate buffer (pH 7.5), 100 mM Na<sub>2</sub>SO<sub>4</sub>, 5% glycerol and 3 mM βME). Est3 eluted primarily as dimer and monomer fractions (ratio of the monomer to dimer was dependent on conditions of sample buffer and concentration). The eluted Est3 fractions were contaminated with His<sub>10</sub>-SUMO. The ~14 kDa His<sub>10</sub>-SUMO tag was separated from the ~20.6 kDa Est3 protein by a second round of Ni<sup>2+</sup>-affinity chromatography in buffer B containing 20 mM imidazole. Est3 eluted at >95%

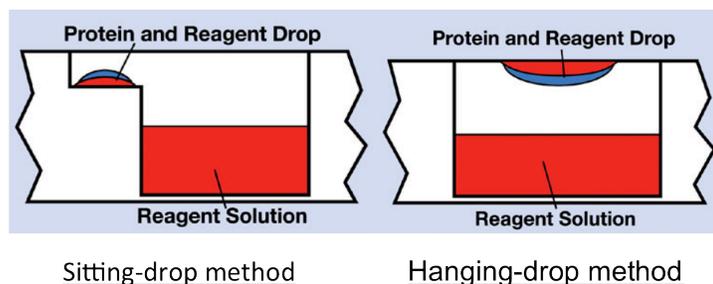
purity and was concentrated and buffer exchanged to the experimentally appropriate buffer, using a 9000 MWCO protein concentrator (Pierce). Protein concentration was estimated by measuring absorbance of the sample at 280 nm in a nanodrop 1000 spectrophotometer (Thermo Scientific). The extinction coefficient for the relevant Est3 protein (WT or mutant) was calculated using the ProtParam tool from expasy ([web.expasy.org/protparam/](http://web.expasy.org/protparam/)).

### **2.2.2 Crystallization screening**

Crystallization condition screens for Est3 were carried out hand-in-hand with the solubility screens (section 2.2.3). Conditions that helped enhance the solubility of the protein were used as storage buffer for enhanced concentration of Est3 and setting up the crystal screen with this enhanced concentration. Conversely, the conditions in the crystallization screens that gave clear protein drop over a number of days was used to guide the search for condition(s) that enhances the solubility of the protein.

A high-throughput buffer screening strategy was employed to investigate crystallization conditions for the Est3 protein. All screenings trays were set on an Art Robbins Instruments Crystal Phoenix automated drop-setter to enhance the number of conditions tested and reproducibility. This robot was used to set up 0.5  $\mu$ L volume drops in a sitting-drop technique (Figure 2.2) on Intelli-Plate 96-well plates. The premixed crystallization screens from *Hampton Research* used for condition screening were the following: Crystal Screen Lite (HR2-128), Crystal

Screen (HR2-110), Crystal Screen 2 (HR2-112), Natrix Screen (HR2-116) and PEG/Ion Screen (HR2-126).



**Figure 2.2 Hanging-drop and sitting-drop set-ups for solubility/ crystallization screens**

Protein sitting-drop set-up was used for high-throughput solubility and crystallization screenings. Manual screening was set-up using the hanging-drop set-up. (from: <http://www.genengnews.com/gen-articles/automation-of-protein-crystallography/1855/>)

The base buffer for crystallization trials (section 2.3.2) only allowed a maximum concentration for the Est3 protein to be 3 mg/mL (~145  $\mu$ M). However, the Hampton Research pre-crystallization test (PCT™) indicated this to be optimum concentration to set-up further crystallization trials. Based on how the protein drops react to the pre-formulated PCT™ reagents, the optimum concentration can be arrived at. Hampton screen 1 and 2 at 4°C, 16 °C and room temperature (RT) was set-up. No crystals were observed.

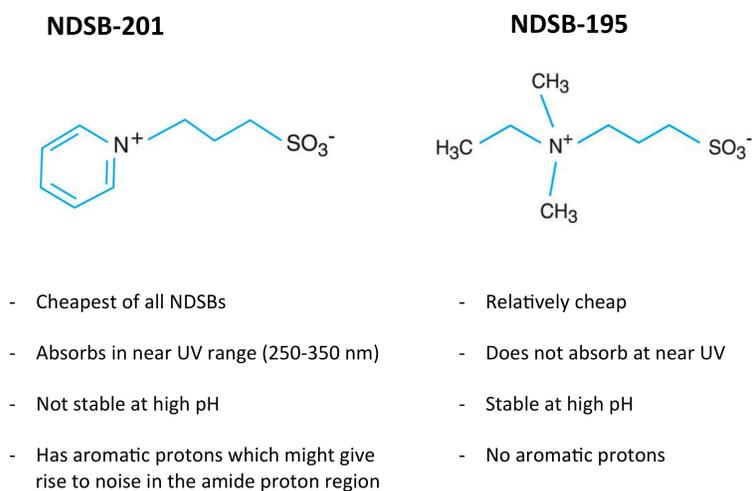
The base buffer plus 500 mM arginine allowed concentration of Est3 to be raised to 4-5 mg/mL at 4°C but protein not stable at RT. This sample-buffer was used to set-up crystallization trials at different temperatures and with Hampton screen 1 and 2 at 4°C but no crystals were observed.

The base buffer plus 400 mM to 1 M non-detergent sulfo-betaine 201 (NDSB-201) allowed Est3 protein concentration to be raised to  $\geq 10$  mg/mL. Est3 still not stable at RT with this additive (starts precipitating after 1h). Freezing concentrated Est3 with NDSB-201 led to no visible precipitation upon thawing (no glycerol added). Crystallization screen set-up with 8 mg/mL Est3 (in presence of 1 M NDSB-201) at 4°C and 16°C with Hampton Screen 1, 2, Natrix Screen and Crystal screen lite. No crystals were observed.

### **2.2.3 Solubility screening**

Untagged-Est3 was used to set-up solubility buffer screens with various buffers (100 mM of potassium phosphate (KPhos), sodium phosphate, sodium acetate, sodium Citrate, ammonium acetate, MES, HEPES, Tris) at varying pH. Out of these, potassium phosphate and Tris were chosen for further salt condition and concentration testing. Following two conditions were chosen as final base buffers for Est3: i) *for crystallization trials*: 50 mM Tris (pH 8.0), 100 mM NaCl, 3 mM  $\beta$ ME and ii) *for NMR sample buffer trials*: 50 mM KPhos (pH 7.5), 50 mM  $\text{Na}_2\text{SO}_4$  and 3 mM  $\beta$ ME. A high-throughput buffer screening strategy was employed to arrive at buffer components that enhance the solubility of the recombinantly expressed Est3 protein. All screening trays were set on an Art Robbins Instruments Crystal Phoenix automated drop-setter to enhance the number of conditions tested and reproducibility of setting 96-well trays. This robot was used to set up 0.5  $\mu\text{L}$  volume (0.25  $\mu\text{L}$  of sample + 0.25  $\mu\text{L}$  of condition) drops in a sitting-drop technique (Figure 2.3) on Intelli-Plate 96-well plates. Once the protein drops and buffers have

been dispensed by the robot, the 96-well plate was covered by a wide cellophane tape. The plate was subsequently stored at RT and the drops observed every 24 hrs for few days and scored for precipitates. A condition was assessed as enhancing solubility if the drop remained clear over the period of observation (~10 days). The identified condition(s) was then repeated by setting up a bigger hanging-drop (3  $\mu$ L volume), manually in a 24-well tray. A condition enhancing the solubility of the drop was then used to test a 50  $\mu$ L volume of protein sample that was buffer-exchanged into the identified buffer condition. This last step was necessary because it was observed that the behavior of the protein in the drop does not always scale-up to the protein behavior in the context of a bigger volume (~300  $\mu$ L) that is requisite for NMR samples.



**Figure 2.3 Comparison of NDSB-201 and NDSB-195 additives**

Non-detergent sulfobetaines (NDSBs) are useful additives for protein solubility. Characteristics of NDSB-201 and NDSB-195 are compared. NDSB-201 was used in solubility screenings, but NDSB-195 was used for NMR sample buffers.

For WT Est3, pre-formulated 96-well ready buffers from Hampton Research were used for solubility testing. These were: Detergent Screen HT™(HR2-406), Additive Screen HT™(HR2-138), Additive Screen™ (HR2-428), SaltRx HT™(HR2-136) and Lindwall Screen[206](for this the buffers/ reagents were made in Lab and dispensed by hand into 96-well buffer holding tray for use with the robot).

The additives that were identified by the screen, as enhancing solubility of the Est3 protein drops, were: Proline, 6-aminohexanoic acid, 1,6-diaminohexane, glycine, betaine hydrochloride, spermidine, sarcosine, TCEP hydrochloride, GSH (L-Glutathione reduced), D-(+)-glucose monohydrate, sucrose, xylitol, myo-Inositol, D-(+)-trehalose dihydrate, 30% glycerol, non-detergent sulfobetaines NDSBs (-195, -201, -211, -221 and -256) and ethyl acetate. The detergent screen also identified NDSBs as solubility enhancing additives for Est3. TCEP-HCl (Thermo Scientific cat# 20490) is a stable reducing agent (except in phosphate buffers at pH 7.0). It is more stable than 2-mercaptoethanol ( $\beta$ ME) that has  $t_{1/2}>100\text{h}$  at pH 6.5,  $t_{1/2}=4\text{h}$  at pH 8.5 and has high volatility and di-thiothreitol (DTT) that has  $t_{1/2}=40\text{h}$  at pH 6.5,  $t_{1/2}=1.5\text{h}$  at pH 8.5.

Given that TCEP-HCl and GSH were identified as additives enhancing Est3 protein stability in the screenings, it appeared as if Est3 might be destabilized through aggregation by inter-molecular cysteine disulfide bonds. As we were already on the edge of sensitivity, with low protein concentrations, we also needed to optimize our sample buffers for the cold-probe for NMR. High sensitivity of cryo-probes is maximally realized if the sample solution is electrically insulating (*e.g.*,

organic solvents). Conductive/ high-salt buffers add resistance to the coil as electric field of the coil interacts with the sample, contributing to noise in the spectrum [207]. The NMR buffer conditions tested subsequently were taken from i) the list of previously studied low-conductivity buffers [207], ii) low-conductivity amino acid additives arginine/glutamate [208] that have been reported to enhance solubility of proteins in NMR experiments [209] and iii) non-detergent sulfobetaines (NDSBs) that showed up in multiple screens for Est3 stabilization and have been previously reported as enhancers of protein solubility [210-212]. The NDSBs are zwitter-ionic compounds that are considered non-detergents because of short hydrophobic group that cannot aggregate to form micelles. NDSBs have been previously used in refolding of proteins found in inclusion bodies [213]. NDSB-201 is inexpensive and was already available in the lab and therefore was used to set-up manual solubility tests. However, important thing to note for NDSB-201 is that it absorbs in near UV (250 – 350 nm) region and therefore cannot be present in the protein buffer if concentration at absorbance at 280 nm ( $A_{280}$ ) needs to be recorded for protein concentration estimates. Also NDSB-201 is not stable at high pH and it contains aromatic region that might give rise to noise in the amide proton region in NMR experiments. These issues were circumvented by using NDSB-195 in the NMR samples (Figure 2.3).

The low-conductivity buffers previously identified by Kelly et al. [207] were used as guide to set-up the solubility screens for optimized Est3<sup>ΔN</sup> construct of Est3. The 4 buffers tested in this test were: BIS-TRIS propane/ HCl, Bis-Tris propane/

MOPS, HEPES and Tris/ HCl. The initial sample concentration for Est3 was ~ 6.0 mg/mL. The screen was set-up at RT in 24-well plates with reservoir volume of 700  $\mu$ L and hanging-drop volume of 3  $\mu$ L (1.5  $\mu$ L sample + 1.5  $\mu$ L reservoir buffer). All 4 buffers ( at 0.1 M) were tested in addition with following salts and additives: sodium sulfate (0.05 M, 0.1 M, 0.15 M and 0.2 M), ammonium chloride (0.05 M, 0.1 M, 0.15 M and 0.2 M), ammonium sulfate (0.05 M, 0.1 M, 0.15 M and 0.2 M), magnesium sulfate (0.05 M, 0.1 M, 0.15 M and 0.2 M), arginine glutamate (0.05 M, 0.1 M, 0.2 M and 0.4 M) and NDSB-201 (0.5 M, 1.0 M, 2.0 M). In conclusion, Bis-Tris propane/ MOPS in combination with sodium sulfate (0.2 M), arginine glutamate (0.4 M) and NDSB-201 (2.0 M) gave clear drops over 8-9 days and a combination of these salts was used to arrive at the final buffers (section 2.2.3) used in the NMR studies.

#### **2.2.4 Buffer for NMR studies of Est3 <sup>$\Delta$ N</sup>**

Low-conductivity buffering salts [207], BIS-TRIS propane and MOPS, as well as solubility enhancing additives, arginine and glutamate, and NDSB-195 were included in the final buffer. The NMR sample buffer (A) used for NMR experiments in this study was: 50 mM BIS-TRIS propane/ MOPS (pH 7.1), 50 mM Na<sub>2</sub>SO<sub>4</sub>, 100 mM Arginine, 100 mM Glutamic acid, 100 mM NDSB-195, 2 mM DTT.

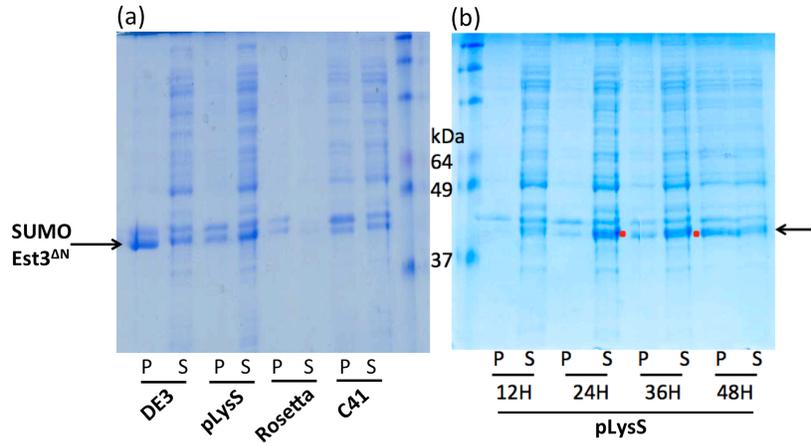
Alternative NMR buffer B: 50 mM BIS-TRIS propane/ MOPS (pH 7.1), 50 mM Na<sub>2</sub>SO<sub>4</sub>, 400 mM arginine, 400 mM glutamic acid, 2 mM DTT was arrived at by setting up arginine glutamate concentration screening using the hanging-drop method (for results of that screening refer to section 2.3.1).

## 2.2.5 Screening for expression and solubility of Est3 in deuterated media

Besides protein construct and NMR buffer optimizations, we also had to perdeuterate Est3, as our NMR experiments were still affected by low sensitivity. For perdeuterating a recombinantly expressed protein, one issue is that growth in deuterated growth medium can be really slow for the bacterial cultures. To maximize the bacterial cell growth and Est3<sup>ΔN</sup> protein expression in the perdeuterated (> 99% D<sub>2</sub>O) medium, different variables for growth were tested.

To optimize expression, different bacterial expression cell lines that were tested were: BL21(DE3), BL21(DE3) pLysS, BL21(DE3) Rosetta and BL21(DE3) C41. The pET28b-Smt3 Est3<sup>ΔN</sup> plasmid was transformed into each of these cell lines and plated on LB-kanamycin plates for DE3 and C41 and on LB-kanamycin + chloramphenicol plates for pLysS and Rosetta cells. 3 mL LB cultures were inoculated with a colony each from these plates, with appropriate antibiotics. O.D.<sub>600</sub> of these overnight cultures was allowed to go until ~0.5. 600 μL of each of this overnight culture was spun-down (at 4000 rpm at RT) and re-suspended into 3 mL each of M9/H<sub>2</sub>O (see end of this section for recipes), such that starting O.D.<sub>600</sub> for the next step was ~ 0.1. This culture was grown at 37°C for ~ 3h until O.D.<sub>600</sub> of ~0.5 was reached. 2 mL of each of this culture was spun-down and cells were resuspended into 10 mL each of M9/D<sub>2</sub>O and again allowed to grow at 37°C for ~ 5h until O.D.<sub>600</sub> of ~0.5 was reached. At this point, this 10 mL of culture was transferred to 40 mL of fresh, pre-warmed M9/D<sub>2</sub>O (in 200 mL baffled flasks each)

and allowed to grow for ~5 – 7h until O.D.<sub>600</sub> of 0.8 to 1.0 was reached for all cultures. The flasks were then kept on ice, with frequent shaking, for about an hour. Protein production was induced by 1.0 mM IPTG. Flasks were transferred to the shaker for 30h of growth at 15°C, after which the cells were harvested by centrifugation at 4000 rpm for 10 minutes and cell pellets saved. The D<sub>2</sub>O containing culture was saved in airtight container for future D<sub>2</sub>O regeneration. Cell pellets were resuspended in 15 mL each of lysis buffer (0.1 M potassium phosphate buffer pH 7.5, 0.1 M sodium sulfate, 10 mM imidazole, 10% glycerol and 3 mM βME) and sonicated with the micro-tip 3 times each for 35 second pulses. Insoluble debris (pellet fraction: p) was separated by centrifugation at 15,000 rpm for 25 minutes. The soluble supernatant (s) was decanted into separate 50 mL conical tubes and the pellet fraction brought up to same volume as the supernatant with lysis buffer. Partitioning of Est3<sup>ΔN</sup> protein into pellet and supernatant fractions was tested by mixing small and equal volumes of supernatant and pellet fraction with 2XSDS dye and run on 15% SDS PAGE gel (Figure 2.4). In conclusion, pLysS was identified as the cell-line yielding the maximum soluble double-labeled Est3 protein and this cell-line was used in subsequent double-labeled growth culture.



**Figure 2.4 Cell-line and post-induction growth duration screening for deuterated Est3<sup>AN</sup> expression**

Four *E.coli* protein expression cell-lines were screened for soluble SUMO-Est3<sup>AN</sup> expression in deuterated media. Solubility was tested by cell-sonication and separation of the insoluble (P) and soluble (S) fractions run on 15% SDS-PAGE gel (left). BL21(DE3) pLysS cells gave the maximum soluble protein. BL21(DE3) pLysS cells were tested for duration of post-induction growth and cultures were harvested at 12 h, 24 h, 36 h and 48 h post-induction growths. Harvested cultures were tested for soluble protein expression. 24 h and 36 h time-points gave equal amounts of soluble protein (gel on the right). Arrows mark the protein bands for SUMO-Est3<sup>AN</sup>.

Following optimization of the expression strain, the duration of growth was optimized. The same procedure was followed as before with 200 mL of pLysS cell culture. 50 mL each of cells were harvested for each time point: 12 h, 24 h, 36 h, 48 h. The harvested cells were resuspended in 10 mL each of lysis buffer. Pellet and supernatant fractions were mixed with SDS loading dye and run on 15% SDS-PAGE. In conclusion, 24 h and 36 h time-points gave similar levels of deuterated Est3 protein in supernatant. 24 h was chosen as the duration for post-induction growth (Figure 2.5b).

Finally, to enhance the deuterated protein yield further, another round of deuterated culture growth was done with +/- 0.4% Isogro (complex growth medium from Sigma/ Isotec). The same growth protocol was followed as before. The effect on

cell-growth was monitored by recording the O.D.<sub>600</sub> at regular intervals. The pre-induction doubling time of cell population in deuterated medium in the presence of 0.4% Isogro was found to be 1 h, compared to 2 h 20 min doubling time for cells in non-Isogro medium. 0.4% of <sup>13</sup>C, <sup>15</sup>N, D Isogro was added to subsequent deuterated media growth of Est3<sup>ΔN</sup> protein. See section 2.2.6 for final growth protocol used.

### **RECIPES:**

#### M9/ H<sub>2</sub>O recipe (50 mL):

5 ml stock of 10X M9 salts (autoclaved)

500 μL of 100 X ammonium sulfate stock

500 μL of glucose stock (20%)

50 μL of MgSO<sub>4</sub> stock (1M)

20 μL of CaCl<sub>2</sub> stock (0.2 M)

50 μL of Thiamine stock (10 mg/ mL)

50 μL of Biotin stock (10 mg/ mL)

Made up to 50 mL volume with autoclaved dH<sub>2</sub>O

Filter sterilized into 250 mL autoclaved flask.

10x M9 stock pH 7.4: 67.8g Na<sub>2</sub>HPO<sub>4</sub> (anhydrous), 30g KH<sub>2</sub>PO<sub>4</sub>, 5g NaCl.

Autoclaved this stock solution.

M9/ D<sub>2</sub>O recipe for the 500 mL growth: 250 mL each in 2.5 L flasks for maximum aeration:

<b>SALT (with final concentration)</b>	<b>AMOUNT IN 500 ML (final pH 8.2)</b>
50 mM Na <sub>2</sub> HPO <sub>4</sub> ·7H <sub>2</sub> O	6.7 gm
25 mM KH <sub>2</sub> PO <sub>4</sub>	1.7 gm
10 mM NaCl	0.3 gm
5 mM MgSO <sub>4</sub>	0.62 gm
0.2 mM CaCl <sub>2</sub> (from 0.2 M stock)	0.015 gm
Thiamine (from 10 mg/ mL stock)	500 uL (of stock)
Biotin (from 10 mg/ mL stock)	500 uL (of stock)
0.2 % Isogro	1 gm
0.15% <sup>15</sup> N (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	0.75 gm
0.3 % Deuterated <sup>13</sup> C Glucose	1.5 gm
Antibiotics: Kanamycin and Chloramphenicol	from stock made in D <sub>2</sub> O

Note: All stocks were made in 1 mL D<sub>2</sub>O each. For 0.2 M CaCl<sub>2</sub> stock (Added 0.029 gm of CaCl<sub>2</sub> in 1mL D<sub>2</sub>O). An important step is to make the final pH 8.0 to 8.2 (as the falling pH during cell growth will have an effect on final cell O.D.<sub>600</sub>)

## **2.2.6 Minimal media growth for expression of labeled Est3 for protein backbone-assignment NMR experiments**

For <sup>15</sup>N-labeled Est3 protein, the *E. coli* BL21(DE3) cells were grown in minimal M9 growth medium [3.5 g/L Na<sub>2</sub>HPO<sub>4</sub> , 3g/ L of KH<sub>2</sub>PO<sub>4</sub> , 1.5 g/L of NaCl and 1.5g/L of (<sup>15</sup>NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (Sigma/Isotec) as sole source of nitrogen] supplemented with 1X MEM vitamin solution (Invitrogen, Carlsbad CA) and 1X Metal Mix [1 μM FeCl<sub>3</sub>, 46 nM H<sub>3</sub>BO<sub>4</sub>, 102 nM CaCl<sub>2</sub>, 0.19 nM CoCl<sub>2</sub>, 0.78 nM CuSO<sub>4</sub>, 1.02 μM

MgCl<sub>2</sub>, 1 nM MnCl<sub>3</sub>, 3 pM Na<sub>2</sub>MoO<sub>4</sub>, and 1.7 nM ZnCl<sub>2</sub>] and 50 mg/ L kanamycin. 2 g/ L glucose was used as carbon source. For uniformly <sup>15</sup>N, <sup>13</sup>C-labeled Est3 2 g/ L of [<sup>13</sup>C] glucose (Sigma/Isotec) was used as the carbon source.

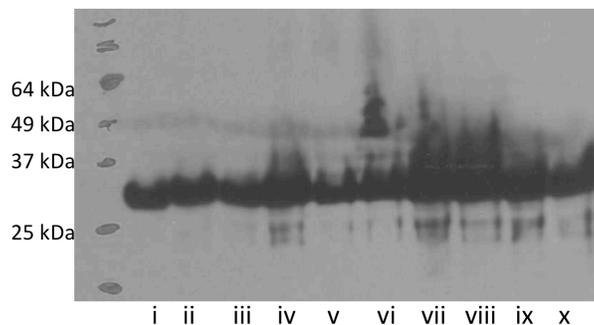
For uniformly <sup>15</sup>N, <sup>13</sup>C, <sup>2</sup>H-labeled Est3 protein, *E. coli* BL21(DE3) pLysS cells were used. Transformed cells were recovered in 500 μL of Luria Broth (LB) for 1 hr at 37°C and then transferred 250 μL each in 2 culture tubes with 6 mL LB each for overnight growth at 37°C. 5 mL of this culture was then spun-down, discarded the supernatant LB and re-suspended the cells in 50 mL M9/ H<sub>2</sub>O [214] and allowed growth at 37°C until O.D.<sub>600</sub> of 0.5 was reached. Cells were spun-down and resuspended in 50 mL M9/ D<sub>2</sub>O in 125 mL baffled flask supplemented with [<sup>2</sup>H, <sup>13</sup>C] glucose (Sigma/Isotec) as sole carbon source, (<sup>15</sup>NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> as the sole nitrogen source and 0.4% of <sup>13</sup>C, <sup>15</sup>N, D Isogro (cell growth enhancing supplement from Sigma/ Isotec). Upon reaching O.D.<sub>600</sub> of 0.5, the cells were transferred to 1 L of M9/D<sub>2</sub>O (divided between 3×2L baffled flasks for better aeration). Upon reaching O.D.<sub>600</sub> of ~ 0.9, cells were cold-shocked by placing on ice for ~ 1h. Protein expression was induced by adding 1.0 mM IPTG and then allowed to grow for 24 hrs at 15°C and harvested at O.D.<sub>600</sub> of ~ 1.7.

### **2.2.7 Screening for growth conditions for selective methylation of ILV side-chain in otherwise deuterated medium**

Perdeuterating the Est3<sup>ΔN</sup> protein helped enhance sensitivity in our NMR experiments; however, as perdeuterating replaces all non-exchangeable protons with deuterium all side-chain information is lost in the proton-detected

experiments. However, protons from amides and side-chain methyls of isoleucine, leucine and valine (ILV) are important sources of long range NOE detection. To enable us to get methyl-methyl long range NOE data, we had to optimize growth conditions with specific precursors that allow selective protonation of ILV methyls, leaving all other protons deuterated. Different small-scale tests were conducted to arrive at conditions for maximum yield and solubility of ILV methyl-protonated Est3 protein. The protocol was largely the same as previously reported[215]. I transformed His<sub>10</sub>-SUMO- Est3<sup>ΔN</sup> plasmid was transformed into the pLysS cells and then plated them onto LB-kanamycin plates. About 20 colonies from the plate were solubilized into 40 mL of LB-kanamycin+chloramphenicol and allowed to grow for ~3 h until O.D.<sub>600</sub> of ~0.6 was reached. 10 mL of this LB culture was spun-down by centrifugation at 4,000 rpm at RT and resulting pellet was resuspended in 50 mL of M9/H<sub>2</sub>O (for recipe see section 2.2.5). This 50 mL culture flask was placed in 37°C shaker and cells allowed to grow until O.D.<sub>600</sub> reached 0.6 to 0.9. The cell culture was spun-down in sterile tubes and cell pellets resuspended in ten 5 mL each of M9/D<sub>2</sub>O media (for recipe see section 2.2.5), supplemented with different growth enhancers: i) No Isogro (complex growth medium for enhancing bacterial growth and increasng protein expression); ii) 0.2 g/ L Isogro; iii) 0.2 g/ L Isogro + MEM vitamin mix (10 mL/L); iv) 0.2 g/ L Isogro + 3 g/ L glucose; v) 0.2 g/ L Isogro + ILV precursors; vi) 0.2 g/ L Isogro + 2 g/ L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; vii) 0.2 g/ L Isogro + MEM vitamin mix + 3 g/ L glucose + 2 g/ L (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> + ILV precursors; viii) 0.2 g/ L Isogro + MEM vitamin mix (from Gibco, for optimum bacterial growth) + 3 g/ L glucose + 2 g/

L  $(\text{NH}_4)_2\text{SO}_4$  + ILV precursors (in recycled  $\text{D}_2\text{O}$ ); ix) 0.5 g/ L Isogro; x) 1.0 g/ L Isogro. The ILV precursors were added to v, vii and viii culture when  $\text{O.D.}_{600}$  reached 0.7 at  $37^\circ\text{C}$  and allowed growth further until  $\text{O.D.}_{600}$  of  $\sim 0.9$  reached. The ILV precursors (alpha-keto precursor stocks) were added to the media as follows: 100 mg/L of 2-Keto-3-(methyl-d3)-butyric acid-1,2,3,4- $^{13}\text{C}_4$ , 3-d<sub>1</sub> sodium salt 99 atom %  $^{13}\text{C}$ , 98 atom % D, 98% (CP) (Isotec part # 637858) and 50 mg/L 2-Ketobutyric acid- $^{13}\text{C}_4$ ,3,3-d<sub>2</sub> sodium salt hydrate 99 atom %  $^{13}\text{C}$ , 98 atom % D, 98% (CP)(Isotec part # 607541). All the ten cultures were grown until this  $\text{O.D.}_{600}$  at  $37^\circ\text{C}$  and were subsequently placed on ice for cold-shock. Protein production was induced with 1 mM IPTG (dissolved in  $\text{D}_2\text{O}$ ) and culture tubes placed in  $15^\circ\text{C}$  shaker for post-induction growth for 24 h. 1 mL each of these cell cultures were spun-down at 4,000 rpm and cell pellets resuspended in 100  $\mu\text{L}$  each of SDS running dye. 10  $\mu\text{L}$  each of this resuspension was loaded onto a 13% SDS-PAGE gel for anti-His western blot (Figure 2.5). In conclusion, the growth conditions vii and viii seemed to express maximum protein, indicating that an increased amount of glucose (3 g/ L versus 2 g/ L for all other conditions), presence of Isogro and vitamin mix all combined to enhance the yield of the protein. This growth condition was used for the final 1.4 L scale up growth for NMR sample (see section 2.2.8).



**Figure 2.5 Western-blot for cell growth conditions screening for selective methyl protonation of Est3<sup>ΔN</sup>**

Ten different growth conditions were tested for expression of the selectively methyl protonated His<sub>10</sub>-SUMO-Est3<sup>ΔN</sup> (see text for details of the conditions). Harvested cell samples were tested on anti-His western and condition leading to maximum expression (vii and viii), were chosen for scale-up growth.

## **2.2.8 Minimal medium growth for expression of selectively methyl-protonated Est3<sup>ΔN</sup> and sample purification**

Based on the screening for growth conditions, for selective protonation of methyls in ILV side-chains (section 2.2.7), following protocol was finally used in the scale growth. A growth culture of 1.4 L was grown for selectively methyl-protonated sample for NMR experiments in section 2.2.10. Multiple pLysS colonies, containing Est3<sup>ΔN</sup> plasmid, were used to inoculate a 2x40 mL of LB-kanamycin+chloramphenicol cultures. The cultures were shaken in a 37°C shaker for ~ 3 h, until O.D.<sub>600</sub> of 0.5 was reached. The cell pellet from this was resuspended into 250 mL of M9/H<sub>2</sub>O. The culture was allowed to grow at 37°C for ~4-5 h until O.D.<sub>600</sub> of 0.7 was reached. The 250 mL culture was then spun-down and cell pellet was resuspended into two M9/D<sub>2</sub>O cultures (see recipe at the end of this section). One 900 mL M9/D<sub>2</sub>O medium from fresh D<sub>2</sub>O and one 500 mL medium with recycled D<sub>2</sub>O were made (as the screen in 2.2.7 indicated no impact on protein

expression in this recycled D<sub>2</sub>O). Media were pre-warmed at 37°C, before the cell pellet from the 250 mL M9/H<sub>2</sub>O was resuspended into these. Cultures were allowed to grow in 37°C shaker or 6 h at which point O.D.<sub>600</sub> of ~0.7 was reached. The ILV precursor ( $\alpha$ -keto precursor stocks) were added to the media as follows: 100 mg/L of 2-Keto-3-(methyl-d3)-butyric acid-1,2,3,4-<sup>13</sup>C<sub>4</sub>, 3-d<sub>1</sub> sodium salt 99 atom % <sup>13</sup>C, 98 atom % D, 98% (CP) (Isotec part # 637858) and 50 mg/ L 2-Ketobutyric acid-<sup>13</sup>C<sub>4</sub>,3,3-d<sub>2</sub> sodium salt hydrate 99 atom % <sup>13</sup>C, 98 atom % D, 98% (CP)(Isotec part # 607541). The culture was allowed to grow at 37°C until O.D.<sub>600</sub> of 1.1. The cultures were transferred to ice for ~1 h for cold-shock. Protein expression was induced with 1 mM IPTG (from 1 M stock in D<sub>2</sub>O). Post-induction growth was allowed for ~30 h at 15°C, at which point the cell were harvested by centrifugation at 5,000 rpm for 10 minutes.

**RECIPE:**

SALT	Amount (in gm) in 900 mL (final pH= 8.0)	Amount (in gm) in 500 mL (final pH= 8.0)
50 mM Na <sub>2</sub> HPO <sub>4</sub> ·7H <sub>2</sub> O	12.06	6.7
25 mM KH <sub>2</sub> PO <sub>4</sub>	3.06	1.7
10 mM NaCl	0.54	0.3
Thiamine	0.009	0.005
Biotin	0.009	0.005
0.15% <sup>15</sup> N (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	1.35	0.75
0.3 % <sup>13</sup> C, <sup>2</sup> H glucose	2.7	1.5
0.2 g/L Isogro	0.18	0.1

NOTE: The above ingredients were filtered through 0.2  $\mu$ m filter and 900  $\mu$ L and 500  $\mu$ L of the MEM metal mix (Gibco) and 9  $\mu$ L and 5  $\mu$ L of the FeCl<sub>3</sub> stock (100 mM) was added from above (because otherwise they will cause some precipitation in the

culture) and 5 mM of MgSO<sub>4</sub> (final conc.) to make growth cultures. 900 µL and 500 µL of kanamycin from the D<sub>2</sub>O stock were added for plasmid selection. The cultures were filtered through Nalgene 0.2 µm filter sterile set-up and then this total of 1400 mL culture was transferred to 6 flasks of 2 L capacity each (for maximum aeration).

### **2.2.9 Sample preparation and NMR experiments for backbone assignment\***

\* NMR experiments were run by Dr. Geoffrey Armstrong, High-field NMR facility manager, University of Colorado

Backbone atom resonance assignments are useful for obtaining information about the phi psi angles of the protein backbone [216]. This information is very useful when calculating the structures. Towards this goal relevant NMR samples were made and experimental data recorded.

For relevant labeled-protein expression see section 2.2.6. For protein purification see section 2.2.1. For backbone resonance assignment NMR experiments, <sup>2</sup>H,<sup>13</sup>C,<sup>15</sup>N-labeled and purified Est3<sup>ΔN</sup> protein was buffer exchanged into NMR sample buffer A containing 50 mM BIS-TRIS propane/ MOPS (pH 7.1), 50 mM Na<sub>2</sub>SO<sub>4</sub>, 100 mM arginine, 100 mM glutamic acid, 100 mM NDSB-195, 2 mM DTT and 6% D<sub>2</sub>O. The final sample was concentrated to ~280 µM and loaded into 5 mm Shigemi tube (Shigemi, Inc.). The sample in the Shigemi tube was kept on ice, until it was ready to be loaded on the spectrometer.

Five TROSY-type through-bond triple resonance experiments were recorded for Est3<sup>ΔN</sup> backbone assignment: HNCA, HN(CA)CB, HN(CO)CA, HN(COCA)CB

and HNCO. These experiments were collected using BioPack pulse sequences with TROSY for each of the experiments listed above. Experiments were collected at 25°C (as sensitivity at lower temperatures was low and higher temperatures than 25°C led to fast sample precipitation) at VNMRs 800-MHz spectrometer equipped with a salt-tolerant HCN cryoprobe. All NMR data were processed in NMRPipe [217] and spectra were analyzed with CcpNMR analysis software [218].

### **2.2.10 Sample preparation and NMR experiments for amide-amide and methyl-methyl NOESY\***

\*NMR experiments for structure calculation run by Geoffrey Armstrong, High-Field NMR facility Manager, CU

Besides backbone dihedral angles, structure calculation also requires information about the residues that are in close proximity ( $\sim 6$  Å or less) to each other in space. This helps to define proper folding of the protein chain. Sample preparation and NMR experiments for long distance restraints measurement for structure calculation of Est3<sup>ΔN</sup> were done as follows.

Amide-amide NOESY: Protein expression and purification are described in sections 2.2.6 and 2.2.1, respectively. NMR sample concentration was  $\sim 280$  μM (similar to the sample concentration for the backbone assignment experiments, section 2.2.9). The 3D <sup>15</sup>N NOESY-HSQC spectrum for amide <sup>1</sup>H-<sup>1</sup>H NOE restraints was collected with the <sup>2</sup>H,<sup>13</sup>C,<sup>15</sup>N-labeled and purified Est3<sup>ΔN</sup> protein in NMR sample buffer A, at a mixing time of 250 ms. 3D <sup>15</sup>N NOESY-HSQC spectra were collected on the

VNMRS 900-MHz spectrometer equipped with a salt-tolerant HCN cryoprobe. Data were collected at 25°C.

Methyl-methyl NOESY: Protein expression and purification are described in sections 2.2.8 and 2.2.1, respectively. For the methyl-methyl NOE measurement, the ILV methyl groups in the 2D  $^1\text{H}$ ,  $^{13}\text{C}$ -HMQC spectrum were first assigned using the 3D HMCMBBCA and HMCMBGBCA spectra [215] on a selectively methyl-protonated  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled and purified Est3 $\Delta^{\text{N}}$  in NMR sample buffer A. Methyl-methyl NOEs were measured on this sample using a 3D  $^1\text{H}$ -methyl,  $^1\text{H}$ -methyl,  $^{13}\text{C}$  NOESY- HMQC spectrum, collected at a long mixing time of 450-ms. The NMR data was collected on VNMRS 800-MHz spectrometer equipped with a salt-tolerant HCN cryoprobe. The experiment was conducted at 25°C.

### **2.2.11 Sample preparation and NMR experiment for Residual Dipolar Coupling (RDC) measurement\***

\* NMR experiments were run by Dr. Geoffrey Armstrong, High-field NMR facility manager, University of Colorado

RDCs provide orientational restraints for the molecule, defining relative angles/ orientations between parts of a molecule. Following steps were used for collected RDC data for Est3 $\Delta^{\text{N}}$ . For the amide  $^1\text{H}$ - $^{15}\text{N}$  RDCs, NMR data were collected using a 140  $\mu\text{M}$   $^{15}\text{N}$ -labeled and purified Est3 $\Delta^{\text{N}}$  with no Pf1 phage and the same sample in  $\sim 9.6$  mg/ mL liquid crystalline Pf1 phage [219]. For relevant labeled-protein expression see section 2.2.6. For protein purification refer to section

2.2.1. The Pf1 phage was a gift from the Dr. Arthur Pardi's Lab. 3 mL of 2.4 mg/mL Pf1 phage from Pardi Lab stock (labeled #3 08/12/02) was obtained. This was divided into 1.5 mL each and put into two TLA 100.3 ultra-centrifuge rotor tubes, for separation of the phage pellet from the storage buffer. The tubes were balanced and put into the TLA 100.3 rotor (Falke Lab, Department of Biochemistry, CU Boulder). The rotor was placed into the ultra-centrifuge and the sample was spun for 2hrs at 65,000 rpm at 4°C. The phage pellet was then re-suspended in NMR buffer A and allowed to vigorously shake at full speed, overnight, on a table-top at 4°C. The phage was again spun down for 2hrs at 65,000 rpm at 4°C. This buffer exchange procedure was repeated a total of 3 times. Finally, the phage was mixed with 350  $\mu$ L of the purified  $^{15}\text{N}$ -labeled Est3 $^{\Delta\text{N}}$  protein sample in NMR buffer A and allowed to shake overnight at 4°C to allow homogenous mixing of the phage with the protein sample. A long glass pipette was used to transfer the phage-containing sample into a Shigemi tube. The phage was aligned in the VNMRS 800 MHz spectrometer magnet. The  $^2\text{H}$  quadrupolar splitting recorded was 9.6 Hz, indicating the concentration of the Pf1 phage to be 9.6 mg/ mL [219].

$^{15}\text{N}$ -HSQC and  $^{15}\text{N}$ -TROSY HSQC spectra were recorded for both isotropic (without Pf1 phage) and aligned (with Pf1 phage) sample. The shift difference between the  $^{15}\text{N}$  chemical shifts for HSQC versus the TROSY-HSQC spectra were calculated for both samples and then divided by 2 to yield the scalar coupling constant (J) in Hz. The residual dipolar coupling value (D) was calculated by

subtracting the value for J obtained from the isotropic spectra from that obtained from the aligned spectra.

$$\text{RDC (D)} = 1/2 [\text{aligned (J+D)}_{\text{NH}} - \text{isotropic } J_{\text{NH}}]$$

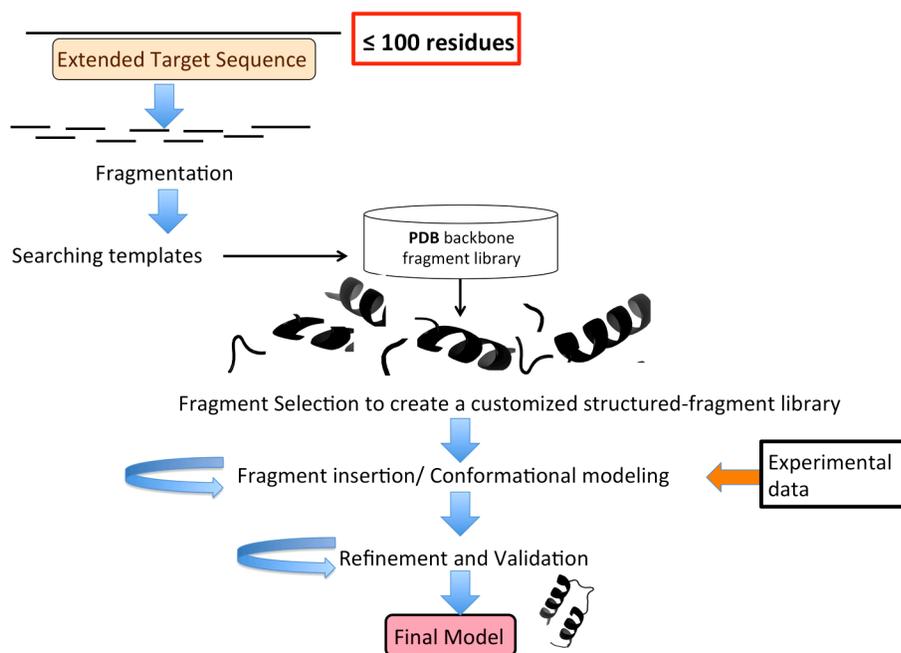
### **2.2.12 RASREC Rosetta structure calculation of Est3<sup>ΔN</sup> \***

\* Structure calculation runs set-up and run on the Janus super-computer, CU Boulder by Dr. Geoffrey Armstrong (High-field NMR facility manager, CU)

Structures were generated using a beta version of resolution-adapted structural recombination (RASREC) Rosetta[220] and CYANA-2.1 [221]. Structure validation was performed using XPLOR-NIH [222]. Rosetta calculations utilized the Janus Supercomputer at the University of Colorado. The supercomputer consists of 1368 compute nodes, each with two hexcore 2.8 GHz Intel Westmere CPUs, 24 GB RAM, and an 800 TB Lustre filesystem. The calculations were run using 528 CPUs, taking approximately 6-9 hours to generate 500 final structures (25000-50000 total targets). Multiple calculations were performed with varying experimental constraints to ensure self-consistency and convergence. CYANA and XPLOR-NIH calculations were run on a MacBook Pro 2.53 GHz Intel Core 2 Duo with 4 GB RAM. Restraint sets were generated using CcpNMR Analysis [218] software and converted to the various input formats using CcpNMR FormatConverter and in-house scripts.

## 2.2.12a Description of the Rosetta, CS-RDC-NOE Rosetta and RASREC Rosetta strategies for 3D structure calculation of proteins with minimal experimental restraint information

Rosetta is a *de novo* structure prediction server from David Baker's laboratory [223], and it has proved useful as a tool for fast and accurate structure prediction of small proteins (Figure 2.6). It is also emerging as a feasible strategy for structure prediction of those proteins for which conventional methods are not feasible and works by combining the acquired minimal experimental restraints, not sufficient for conventional structure calculations, with Rosetta structure modeling.



**Figure 2.6 Rosetta *de novo* modeling pipeline**

Step-wise Rosetta *de novo* structure calculation starts with *in situ* fragmentation of target sequence, followed by sequence homology based structured-template search. Selected structured-fragments are iteratively used in fragment insertion and conformational modeling steps to predict structured models that are finally refined and validated based on knowledge-based protein restraints. For proteins with >100 residues length, introduction of experimental restraints helps restrict the conformational space during iterative modeling.

Rosetta begins with an extended peptide chain and breaks it up into 3- and 9-residue long fragments *in situ* [224]. These short fragments are used to search the PDB for sequentially similar structured fragments on the basis of local sequence similarity to the structures in the PDB. The top hits from this search are used to create a customized library of structured fragments. Subsequently, this library of structured fragments is used in an iterative fragment insertion approach where the fragments are randomly combined using Monte Carlo simulated annealing. Backbone torsion angles from the inserted structured-fragments are used to assign a new global configuration to the protein chain. The individual conformations are evaluated using a knowledge-based scoring function. At this step, the side-chains are represented by centroids located at the center of mass of the side-chains. For the refinement step, once the peptide chain is collapsed into a compact structure, torsion angle perturbations are employed in an iterative step to optimize energy functions. Both coarse-grained (centroid-based) and fine-grained (all-atom) refinements are done that satisfy the global and local/ atomic-level energy functions respectively. Side-chain simulations can either be rapidly optimized by replacing each side-chain with lowest energy rotamer or subjected to complete combinatorial sampling by Monte Carlo simulated annealing [225]. Finally a 3D model for the target peptide chain is predicted after this refinement stage. This strategy works well for proteins up to 100 amino acids in length [220], however, as the protein size increases to >100 amino acids, the conformational space that would have to be

tested at iterative steps of fragment insertion and refinement becomes computationally too expensive. Introduction of structural restraints from experimental data can help with restricting this, otherwise unwieldy, conformational space. NMR experimental data, even with just chemical shifts for the backbone atoms can help restrict a lot of sampling by providing dihedral angle restraints for the conformational space being sampled. As the protein, size increases addition of restraints like long-distance restraints provided by NOEs and orientational restraints from RDCs becomes useful in assigning the correct folding and orientation of parts of the molecule.

In CS-RDC-NOE Rosetta, three sets of structural restraints are added to Rosetta *de novo* calculation: i) the backbone torsion angle restraints from assigned backbone atom chemical shifts, ii) orientational restraints from the residual dipolar coupling values and iii) through-space long-distance restraints from amid-amide NOEs and/ or methyl-methyl NOEs. In this strategy, the 3- and 9-residue structured-fragments, from the PDB fragment search, are scored according to their agreement with the experimental restraints. Chemical shift restraints are more important at this stage of fragment selection because they restrict the library to fragments that satisfy the correct dihedral angle restraints for constituting residues. For fragment insertion/assembly, the long-range distance restraints are more important. Short-range NOEs are used in the early stages because compaction of the primary sequence is not required. Fragment assembly starts with the 9-residue fragments. Here ~2000 to 5000 fragment insertions are attempted to put

together the right secondary structure from the fragment library already created. Subsequently, ~20,000 to ~48,000 fragment insertions are attempted. The compaction and pairing of beta-strands is rewarded and the evaluation of NOEs with increased sequence separation are cycled on and off, utilizing more long range NOEs throughout. This iteration (attempting ~20,000 to 48,000 insertions) cycles between rewarding global strand pairs only and then all-strand pairs. Periodic down weighting of the local strand pairing is to compensate for Rosetta program's tendency to overpopulate local strand pairs. The final stage of fragment insertion includes ~12,000 to 24,000 fragment insertions and optimizes the complete Rosetta energy function using 3-residue fragments. In the ranges given above, the bigger insertion attempt number is used when long-range NOEs are available. Basically, twice the number of attempts are made in the presence of NOE data. Rosetta fragment insertion/sampling has been reported to compensate for the incomplete or incorrect NMR restraints [226]. This CS-RDC-NOE Rosetta method works well for proteins up to 150 amino acids, but fragment sampling again becomes limiting for >150 amino acid proteins and the structures fail to converge even with experimental data [220]. A new approach involving resolution-adapted structural recombination (RASREC) becomes essential for accurate structure calculation in the 15-25 kDa size range[220]. The RASREC Rosetta approach allows for improved fragment sampling by recombination of the more-frequently occurring structural features at each stage of the structure calculation and this leads to enrichment of native-like features. As Est3 is a 181 amino acid protein, a combination of Rosetta

and experimental structural restraints had to be employed for an accurate structure calculation.

### **2.2.13 HetNOE experiment for assessment of dynamics\***

\* NMR experiments were run by Dr. Geoffrey Armstrong, High-field NMR facility manager, University of Colorado

A HetNOE experiment helps to assess if a protein has flexible regions and the identity of the flexible residues can be determined if the protein backbone is assigned, which was the case for Est3<sup>ΔN</sup>. For the <sup>15</sup>N-<sup>1</sup>H-heteronuclear NOE a ~290 μM <sup>15</sup>N-labeled and purified Est3<sup>ΔN</sup> sample was prepared in NMR buffer A (for protein expression see section 2.2.6 and for purification refer to section 2.2.1). A reference <sup>15</sup>N-HSQC (192 scans, duration of 52.5 hours) and an amide-proton saturated <sup>15</sup>N-<sup>1</sup>H-heteronuclear NOE HSQC (384 scans, duration of 105 hrs) were recorded at the 800 MHz spectrometer. As the signal is summed over the acquisition time, the signal for the <sup>15</sup>N-<sup>1</sup>H-heteronuclear NOE HSQC peaks was halved to account for 2X acquisition time compared to the reference experiment. The HetNOE values were calculated by taking the ratio of the peak heights of the non-saturated spectrum over the peak heights of the saturated spectrum. Replicates of the HetNOE experiment were not done, as the S/N for the acquired data was bad even after running one set of experiments for a full week. Therefore, we decided to use the analysed data qualitatively.

### **2.2.14 H/D-exchange experiment for Est3**

The hydrogen-deuterium (H/D) exchange reaction can be monitored over time, to detect which amides of the protein are protected from fast-exchange to deuterium. For the H/D exchange experiment, a reference  $^{15}\text{N}$ -HSQC was collected on the  $^{15}\text{N}$ -labeled Est3 $\Delta\text{N}$  sample in NMR buffer A (for protein expression see section 2.2.6 and for purification refer to section 2.2.1). This sample was subsequently buffer exchanged to NMR buffer A made in  $\text{D}_2\text{O}$ , using an Amicon 3000 MWC0 concentrator (Millipore) and total buffer exchange and concentration took  $\sim 7$  hrs. The sample was concentrated to  $\sim 130 \mu\text{M}$  and loaded into a Shigemi tube.  $^{15}\text{N}$ -HSQCs were collected at 7 hrs, 24 hrs, 3 days, 6 days and 9 days after deuterated buffer exchange at 600 MHz at  $25^\circ\text{C}$ . HSQC spectra were overlaid and peaks assigned in CcpNMR analysis [218] software. Data was used qualitatively, by mapping the residues with protected protons onto the solved structure of Est3 $\Delta\text{N}$ .

### **2.2.15 Structure figures preparation**

All protein structure figures were prepared using the PyMOL Molecular Graphics System, Version 1.5, Schrödinger, LLC and superpositions were done using the “super” command in PyMOL or PdbStat program [227]. Electrostatic surface representations were made using the APBS plug-in in PyMOL with low and high potentials of  $-4$  and  $+4$ , respectively [228].

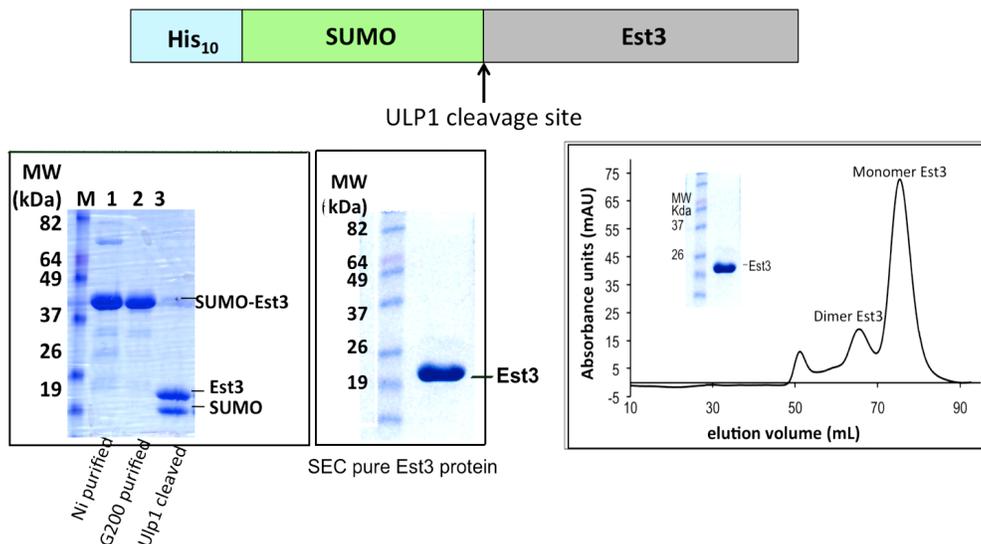
## 2.3 Results

### 2.3.1 Est3 protein sample optimization: challenges and solutions

#### 2.3.1a Protein-yield optimization for unlabeled and isotopically labeled Est3

Est3 proteins are poorly expressed and have low solubility upon purification, precluding characterization using biochemical and structural techniques. Therefore extensive optimization of the protein construct and sample conditions was required to obtain relatively stable sample for NMR studies. Initially, the *S. cerevisiae* Est3 (ScEst3) gene had been cloned into the pET21b vector for a C-terminal 6xHis tag for purification by a former member of the Wuttke Lab (Amy Gelinias). The average yield that I achieved from this Est3-6XHis protein was < 1 g/ L of growth culture. The concentration of this construct that could be achieved was ~0.4 mg/mL (20  $\mu$ M) and therefore too low for any structure determination technique. To enhance the protein yield and solubility, a number of expression vectors and bacterial growth conditions were tested. Besides Est3 from *S. cerevisiae*, Est3 genes from closely related species *S. bayanus*, *S. castellii* and *S. dairensis* were also recombinantly expressed in bacterial expression cell-lines and screened for expression and solubility. The Est3 gene was cloned into different vectors for fusion with various solubility enhancement tags (SETs) including GB1, Intein, SUMO [229-231]. Fusion with Intein in the pTXB1 vector did not give any significant enhancement of expression for Est3. Fusion of *S. cerevisiae* Est3 with N-terminal His<sub>10</sub>-SUMO

helped enhanced the expression and solubility (Figure 2.7) such that a yield of >10 mgs of soluble SUMO-Est3 was regularly obtained per liter of LB growth culture of *E. coli* BL21(DE3) cells.

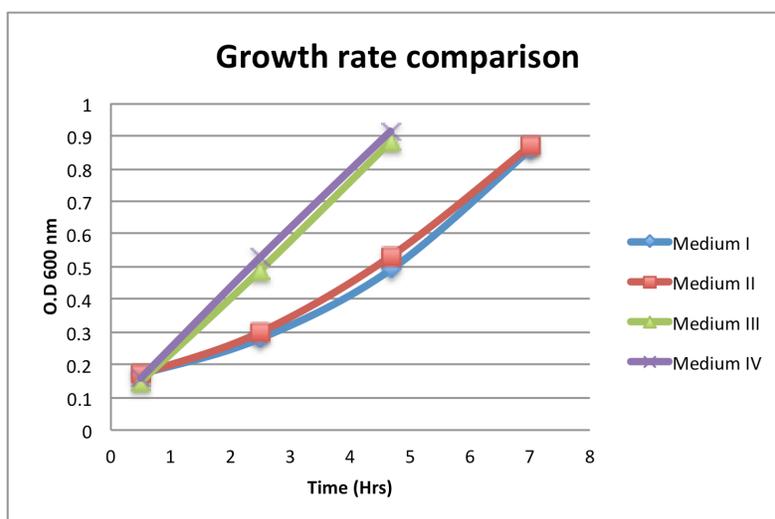


**Figure 2.7 Purification scheme for Est3 proteins**

SUMO fusion Est3 schematic is shown on top. Left most coomassie-stained SDS-PAGE shows Ni<sup>2+</sup> affinity purified SUMO-Est3 protein, which was cleaved to give Est3 and SUMO proteins. Est3 was separated from SUMO by size-exclusion chromatography (SEC), shown on the gel in the middle. Superdex G75 SEC purification profile of Est3 is shown on the right. Est3 partitions into dimer and monomer peaks on SEC.

This yield was further increased by introducing some variations to the growth conditions: higher aeration by having  $\leq 300$  mL of growth culture in a 2 L baffled flask, higher glucose (3 g/L instead of 2 g/L) in minimal media growth and using a high-cell density induction method [232] where the O.D.<sub>600</sub> of the bacterial culture was allowed to rise until 1.0-1.5 at 37°C before the cell-culture was cold-shocked and subsequently protein expression was induced by 1.0 mM IPTG followed by post-induction growth of 24-30 h at 15°C. The harvesting O.D.<sub>600</sub> of cell cultures was consistently found to be  $\geq 5.0$  for LB and M9/H<sub>2</sub>O based minimal media growth,

$\geq 3.0$  in M9/D<sub>2</sub>O based growth and  $\geq 2.0$  for ILV precursor + M9/D<sub>2</sub>O based growth. Using the Isogro growth supplement (Sigma/ Isotec) in deuterated growth media helped reach relatively high-cell densities in otherwise growth limited deuterated medium (Figure 2.8).



**Figure 2.8 Pre-induction cell-growth enhancement by presence of Isogro in minimal deuterated growth media**

Deuterated growth media, without growth supplement (Medium I and II) and with Isogro growth supplement (Medium III and IV) were compared for cell-growth at O.D.<sub>600</sub>. Media with Isogro reached harvesting O.D.<sub>600</sub> (of 0.9) 2.5 h faster than the media without Isogro.

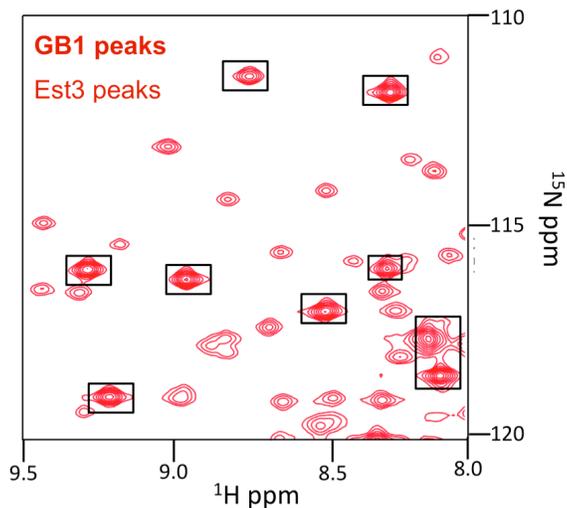
### 2.3.1b Est3 construct optimization

As the wild type Est3 is not a very soluble protein, it was not feasible to record 3D NMR experiments on the wild-type protein for structure elucidation. Strategies adopted to enhance the stability of the recombinantly expressed Est3 protein included: i) fusion with different solubility enhancement tags (SETs) [233] like GB1 and charged-peptides [233], ii) solubility tests of deletion mutants of Est3

at the N- and C-termini and iii) cysteine to serine mutants tested for reducing potential protein aggregation due to inter-molecular disulfide bonds.

GB1 is a highly soluble, small (56 amino acid) protein and its fusion with otherwise insoluble proteins has been reported to enhance their solubility for downstream NMR experiments [233]. It has a nice spectrum that can be readily identified in the context of a fusion protein. The GB1 coding sequence from pGEV1 vector [205] was introduced between the SUMO and the Est3 coding sequence in the His<sub>10</sub>-SUMO-Est3 construct. Other orientations of the GB1 were also tried: i) GB1 at the C-terminus of Est3, ii) GB1 at both the ends of the Est3 sequence in the His<sub>10</sub>-SUMO-Est3 sequence, iii) introduction of short charged-peptides [234] (4K/R, 7K/R and 10K/R) at N- and/or C-terminus Est3 in the His<sub>10</sub>-SUMO-Est3 construct and at the N- and/or C-terminus of Est3 in the His<sub>10</sub>-SUMO-GB1-Est3 construct and iv) introduction of different-sized glycine/serine linkers between GB1 and Est3 sequences. GB1-fusion at the N-terminus of Est3 drastically enhanced protein solubility and a 10 to 12 mg/mL sample of this <sup>15</sup>N,<sup>13</sup>C-labeled construct was used to collect some backbone-assignment 3D NMR experiments. The intensity differential between the GB1 and Est3 peaks was dramatic and it appeared as if the Est3 domain was behaving independently of GB1 and was aggregating, whereas the GB1 protein remained monomeric. Unfortunately, although this fusion gave a dramatic increase in solubility it still precluded Est3 assignment (Figure 2.9). Therefore, although the overall sample stability of the GB1-Est3 protein was promising, the high signal-to-noise (S/N) for the GB1 molecule compared to the really low S/N for

Est3, indicated that the two proteins were behaving independently of each other in solution and that Est3's propensity to aggregate was not alleviated by the fusion of GB1.



**Figure 2.9 HSQC spectrum of GB1-Est3 fusion protein, highlighting peak-intensity differential**

A small slice of the  $^{15}\text{N}$  HSQC spectrum of the GB1-Est3 fusion protein is shown. The spectrum shows peak-intensity difference between the two protein domains, with Est3 peaks very weak compared to those of GB1's (boxed in black). Experiment was done at 900 MHz at 25°C with a ~600  $\mu\text{M}$  sample concentration.

The charged-peptides also did not enhance the stability of Est3 protein significantly, as the solubility with the peptides was similar to the untagged protein (Figure 2.10). In summary, this strategy, although promising, was not effective towards yielding stable Est3 sample for NMR studies.

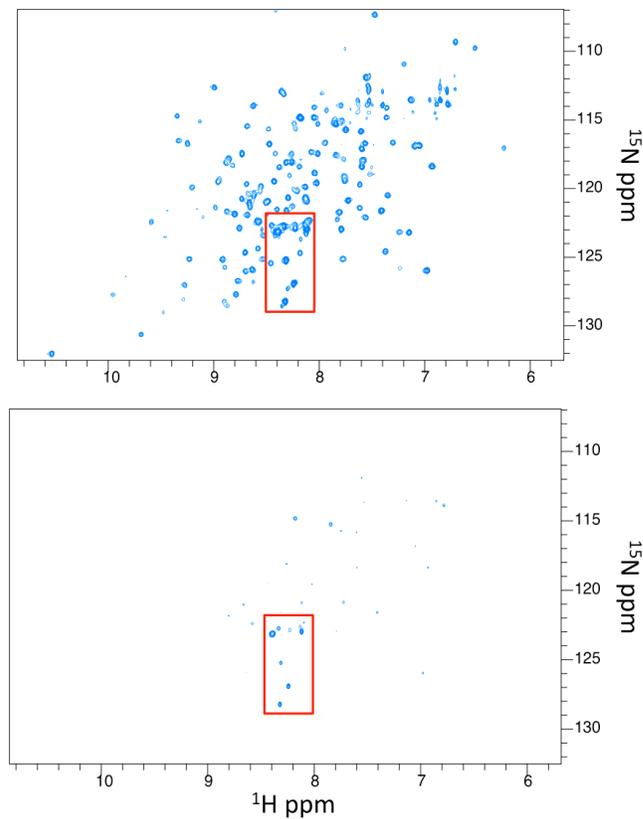
	Yield	Solubility of construct	Effect on Est3
Est3 (= WT)	+	+	Unstable protein
C142S Est3	+	++	Better than WT
GB1 Est3	+++	++	Est3 domain unstable
Est3 GB1	+	++	Est3 domain unstable
GB1 Est3 GB1	++	++	Est3 conformation changed
GB1 Est3 7 Lys	+++	++	Est3 domain unstable
GB1 C142S Est3 7 Lys	+++	+++	Better than non-C142S construct, but Est3 domain still unstable
GB1 Est3 7 Arg	++	++	Est3 domain unstable
GB1 Est3 7 Asp	+	++	Est3 domain unstable
Est3 7 Arg	++	++	Est3 domain unstable
Est3 <sup>ΔC</sup> (=Δ2)	++	+	Stability like WT
Est3 <sup>ΔN</sup> (=Δ12+C142S)	++	++	Truncated Est3, more stable than WT and usable for NMR
Est3 <sup>ΔN</sup> (=Δ54)	+ (of SUMO-Est3)	-	No solubility
Est3 <sup>ΔC</sup> (=Δ18)	+ (of SUMO-Est3)	-	No solubility

**Figure 2.10 Summary of Est3 constructs screened for protein yield and solubility**

Different solubility enhancement tags (SETs) and deletion constructs of Est3 were tried to arrive at the Est3<sup>ΔN</sup> construct that gave the best protein stability for NMR studies.

The<sup>15</sup>N-HSQC spectrum of the untagged wild-type Est3 exhibited a handful of anomalously sharp peaks, presence of which alluded to a flexible region in the protein that might contribute to its less than ideal behavior (Figure 2.11). Using N- and C-terminii deletion constructs of the Est3 protein, we were able to identify the highly intense peaks as belonging to the N-terminus of Est3. Removal of 12 residues

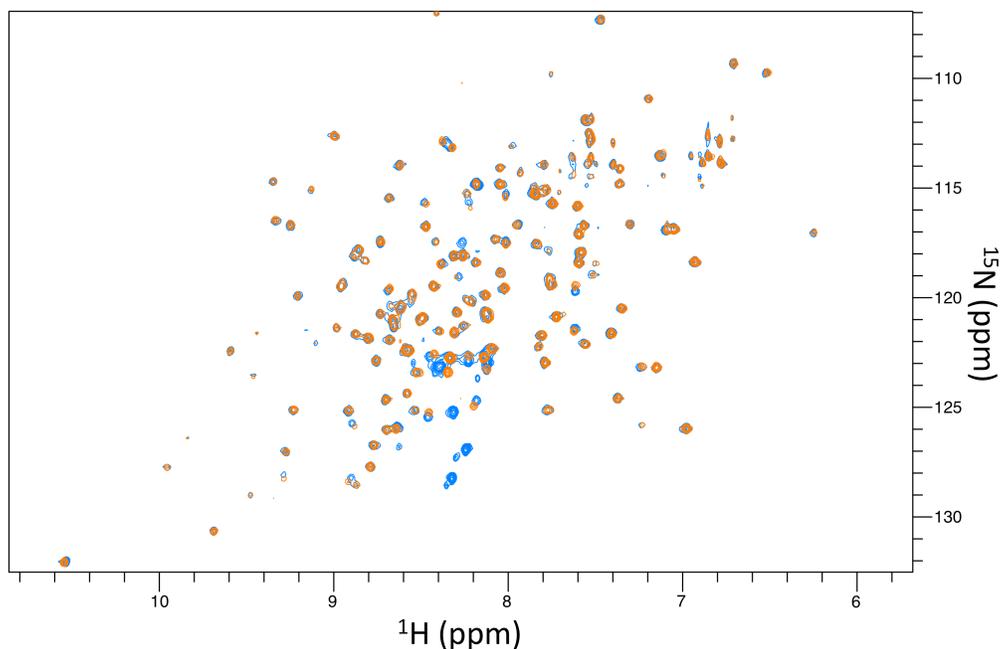
at this terminus helped enhance the stability of the protein and this construct was termed Del12\_Est3 initially. In addition, since cysteine residues on the protein surface can facilitate aggregation, a screening of cysteine to serine mutations in the non-conserved cysteines of Est3 protein was undertaken. This was done as a result of the observation that TCEP-HCl and GSH enhanced Est3 stability in the high-throughput additives screening (see section 2.3.3). Alignment of the Est3 homologs was used to identify the cysteine residues in *S. cerevisiae* Est3 (Figure 1.14). Est3 has six cysteines, two of which are highly conserved and predicted to be internalized. Four cysteines (C64, C76, C109 and C142) are not as well conserved. These 4 cysteines were modified to serine individually and in combination to identify the mutant Est3 protein that is more soluble than wild type. This strategy identified C142S mutation, as helping with solubility and this mutation was included in constructs of Est3 for structure elucidation.



**Figure 2.11**  $^{15}\text{N}$  HSQC spectrum of wild-type Est3 shows a cluster of sharp peaks

$^{15}\text{N}$  HSQC spectrum of wild-type Est3 (top-panel) contains some sharp peaks, which become apparent upon going high in contour level (bottom-panel). Presence of these peaks (boxed in red) indicated an inherently flexible region in the Est3 protein that might be contributing to sample's low stability.

Finally, Del12\_C142S\_Est3 (hereafter called Est3 $^{\Delta\text{N}}$ ) was the construct for our NMR studies. Overlay of the WT and Est3 $^{\Delta\text{N}}$   $^{15}\text{N}$ -HSQC spectra (Figure 2.12) indicates that the Est3 $^{\Delta\text{N}}$  protein retains the structural conformation of wild-type Est3.



**Figure 2.12 Superposition of  $^{15}\text{N}$  HSQC spectra of wild-type and Est3 $^{\Delta\text{N}}$  proteins**

$^{15}\text{N}$ -HSQCs spectra from wild type Est3 (blue) and Est3 $^{\Delta\text{N}}$  (orange) are compared. Removal of 12-residues from the N-terminus of the wild-type protein and inclusion of the C142S mutation, gave the Est3 $^{\Delta\text{N}}$  construct leading to the removal of the high intensity peaks as seen from this superposition. The peaks from the two spectra superpose almost completely, indicating the retention of the wild type protein conformation in the Est3 $^{\Delta\text{N}}$  construct.

### 2.3.1c Est3 storage buffer optimization

A high-throughput buffer screening strategy was employed to identify buffer conditions that enhance the stability of the Est3 protein. Low-conductivity buffering salts [207], BIS-TRIS propane and MOPS, as well as solubility enhancing additives, arginine and glutamic acid, and a non-detergent sulfobetaine NDSB-195 that emerged from the high-throughput screen were included in the final buffer. The NMR sample buffer (A) used for NMR experiments in this study was: 50 mM BIS-TRIS propane/ MOPS (pH 7.1), 50 mM  $\text{Na}_2\text{SO}_4$ , 100 mM Arginine, 100 mM Glutamic acid, 100 mM NDSB-195 and 2 mM DTT. In addition, because of stability

and dimer-monomer issues we had to work at lower concentrations ( $\leq 300 \mu\text{M}$ ) and therefore protein perdeuteration was also used as a strategy to enhance S/N [214]. Perdeuteration also helped enhance resolution of the peaks by decreasing their line-widths in the NMR spectra [235].

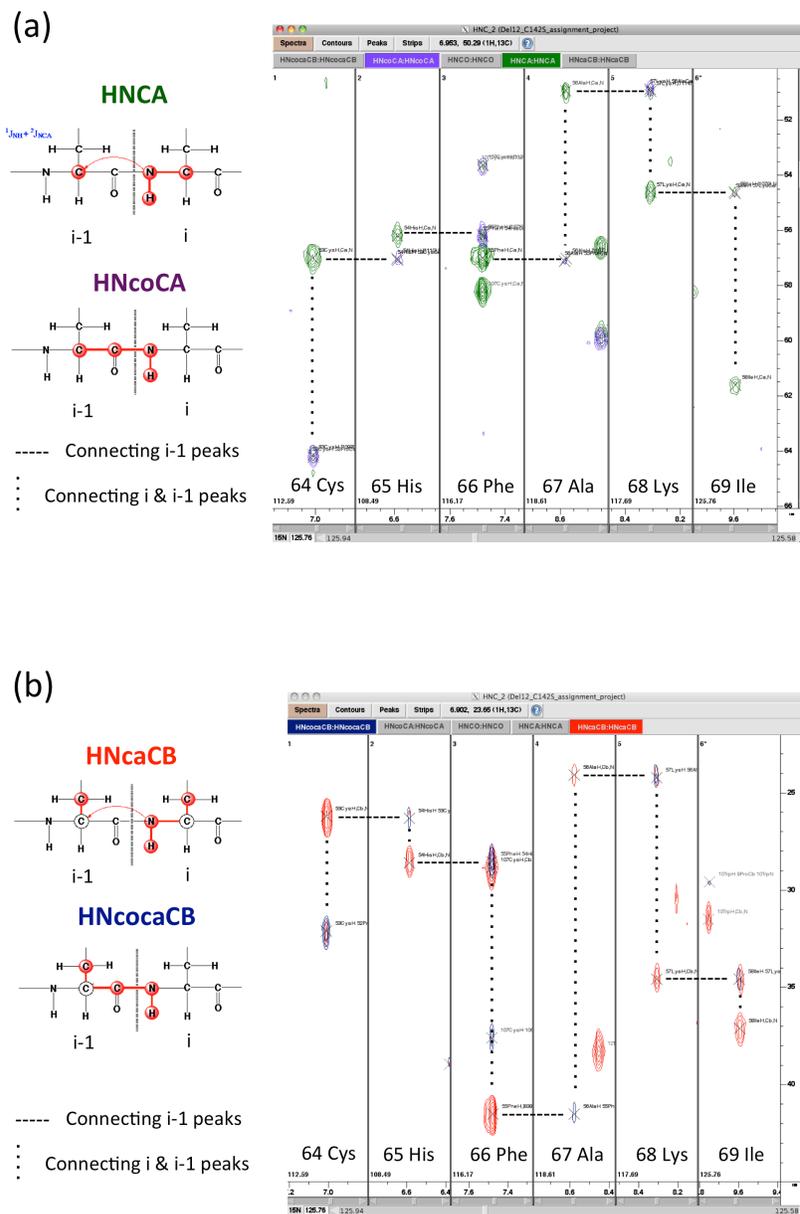
### **2.3.2 Est3 crystallization trials**

Despite numerous efforts and over 1000 crystallization conditions screened, the Est3 protein failed to crystallize. The protein drops, invariantly, either remained soluble, if protein concentration was low, or precipitated out of solution during the period of observation. The failure to get any crystals for Est3 was likely because of sample instability at even lower protein concentrations and inhomogeneity of protein solution as the wild-type Est3 and its mutants form a dimer-monomer equilibrium upon protein purification (Figure 2.7). The presence of a stable binding partner of Est3 might help alleviate protein instability, however, currently the absence of such a binding partner prevents this route of crystallization promotion for Est3. We note that we are aware that several other labs have attempted to solve the structure of Est3 using both crystallization and NMR approaches, to the best of our knowledge none of them met with success.

### 2.3.3 NMR experimental data

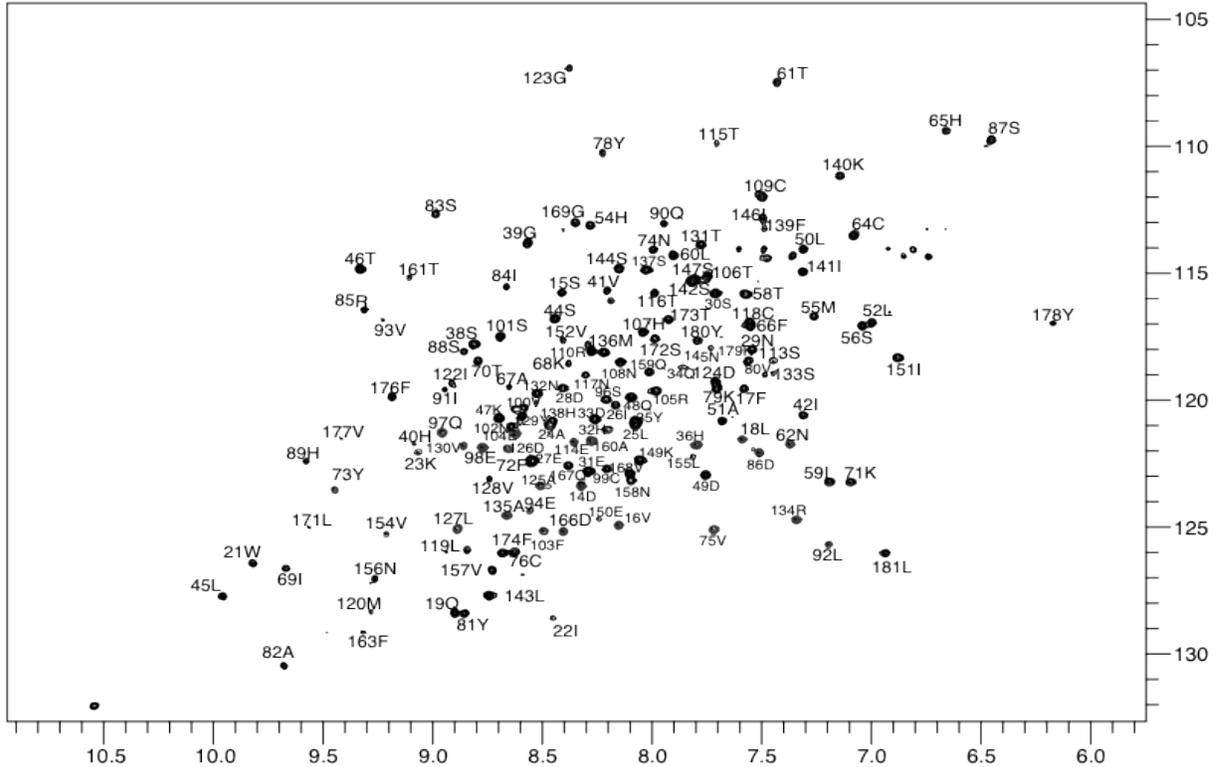
#### 2.3.3a Greater than 97% of the Est3<sup>ΔN</sup> HSQC peaks assigned

Chemical shifts define the torsion angles in the protein, because of the empirically established correlations between chemical shifts and torsion angles [236]. Automated programs, like TALOS [216], can use the chemical shifts for the backbone atoms H, N, CA, CB, CO to predict the  $\Phi$  and  $\Psi$  angles of the constituting residues. In turn, these torsion angles are important to define the structure of a protein and are used in structure calculation algorithms. Thus, to arrive at the structure of Est3<sup>ΔN</sup>, our first step was to assign, as completely as possible, the chemical shifts obtained from all the NMR experiments to their respective residue type and residue number. The CA, CB and CO sequential connectivities were provided by 5 TROSY-type through-bond triple resonance experiments: HNCA, HN(CA)CB, HN(CO)CA, HN(COCA)CB and HNCO (Figure 2.13). Manual assignment as well as PINE automated assignment (server: [pine.nmrfam.wisc.edu](http://pine.nmrfam.wisc.edu) [237]) was done for assigning 165, including prolines, of the 170 residues comprising the Est3<sup>ΔN</sup> construct, *i.e.*, a coverage of 97% was achieved. See Table A1 in Appendix for complete chemical shifts assignment. Figure 2.14 (see below) shows residue assignments of the non-proline residues on the <sup>15</sup>N-HSQC spectrum of the deuterated Est3<sup>ΔN</sup> protein. These assignments were also validated by sequential <sup>1</sup>H-<sup>1</sup>H NOEs from the 3D <sup>15</sup>N NOESY-HSQC spectrum.



**Figure 2.13  $C\alpha$  and  $C\beta$  connectivities established from corresponding three-resonance experiments**

(a) HNCA and HNcoCA spectra were used for assigning  $C\alpha$  resonances and backbone residue connectivities. (b) HNcaCB and HNcoCaCB spectra were used for assigning  $C\beta$  resonances and backbone residue connectivities. Shown for both (a) and (b) are slices from the ccpNMR analysis project for sequential connectivity from 64 Cys to 69 Ile residues of Est3.



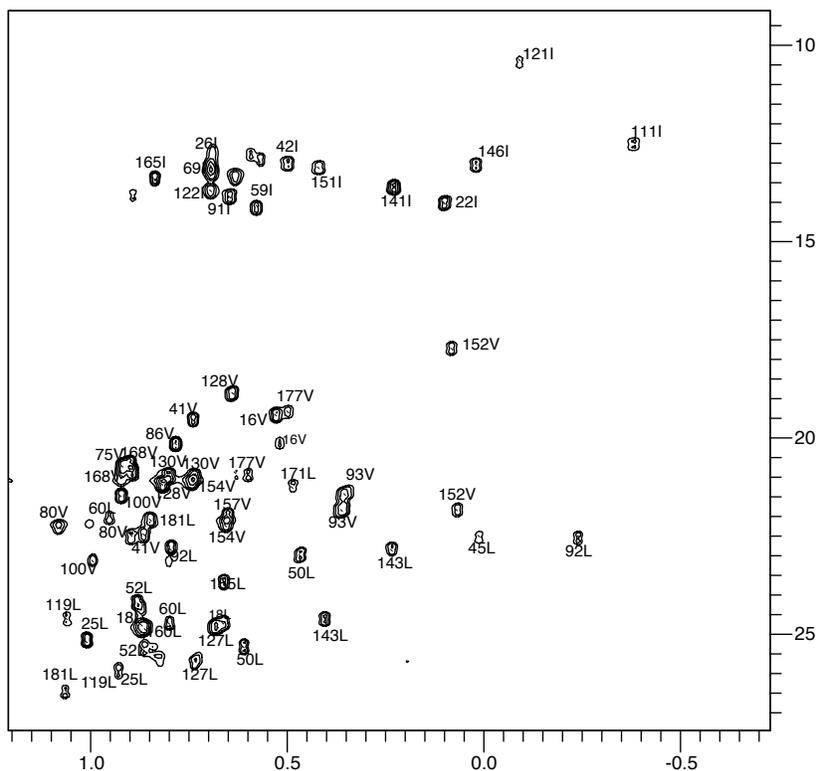
**Figure 2.14**  $^{15}\text{N}$  HSQC spectrum of Est3 $^{\Delta\text{N}}$  showing residue assignments

$^1\text{H}$ - $^{15}\text{N}$  TROSY-HSQC spectrum, showing assignments of the non-proline residues as single-letter amino acid abbreviation followed by sequence number. The data was collected on  $\sim 280 \mu\text{M}$   $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled Est3 $^{\Delta\text{N}}$  sample at 800 MHz and 25°C.

### 2.3.3b Long range and orientational structural restraints: NOEs and RDCs

Long-range structure restraints help to define proper folding of the protein chain. Protons from amides and side-chain methyls of isoleucine, leucine and valine (ILV) are important sources of long range NOE detection. A total of 219 amide-amide NOEs were obtained from the 3D  $^1\text{H}$ ,  $^1\text{H}$ ,  $^{15}\text{N}$  NOESY-HSQC experiment (refer to section 2.2.10 for details). Of these 49 were long-distance amide NOEs (all unique and greater than 3 amino acids apart). Additionally, the methyl carbon and proton

chemical shifts were assigned (Figure 2.15) for 100% of the isoleucine, leucine and valine residues (41 total), leading to 72 unique methyl-methyl NOEs (refer to section 2.2.10 for details). Table A2 (Appendix) lists all the NOEs (amide-amide and methyl-methyl) used for structure calculation.

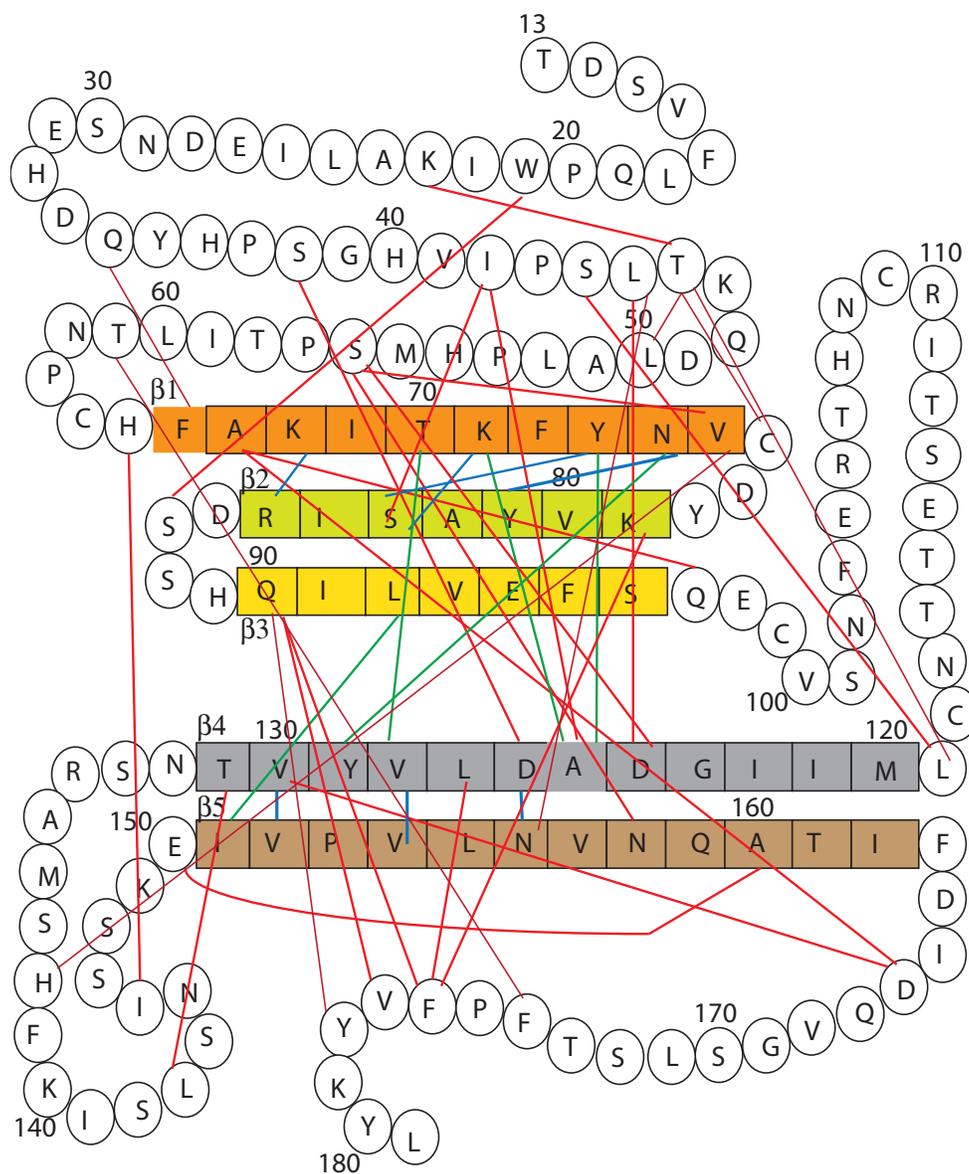


**Figure 2.15**  $^1\text{H}$ - $^{13}\text{C}$  HMQC spectrum showing methyl proton assignments

$^1\text{H}$ - $^{13}\text{C}$ -HMQC spectrum showing the assignments for methyl carbon and proton chemical shifts of the isoleucine, leucine and valine residues (41 total) in the Est3 $^{\Delta\text{N}}$  construct. The data was collected at 800 MHz on selectively ILV methyl-protonated  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ -labeled and purified Est3 $^{\Delta\text{N}}$  in NMR sample buffer A.

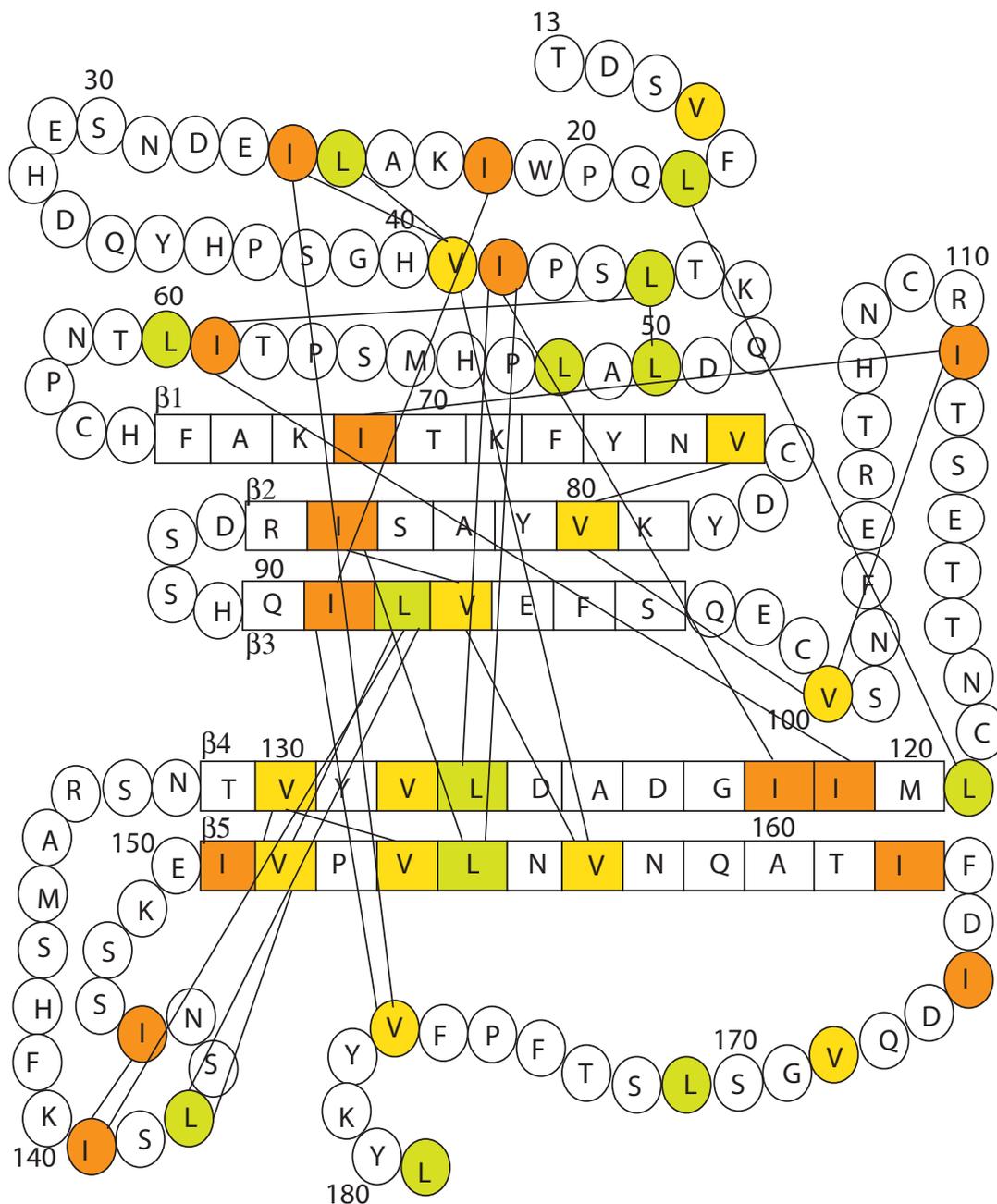
Thus, amide-amide and methyl-methyl NOEs contributed to 128 constraints defining long range NOEs for the structure calculation. Figure 2.16 and 2.17 show

schematic of the amide-amide and methyl-methyl connectivities used for structure calculation.



**Figure 2.16 Schematic of secondary structure of Est3<sup>ΔN</sup> showing long-range amide-amide NOE connectivities used for structure calculation**

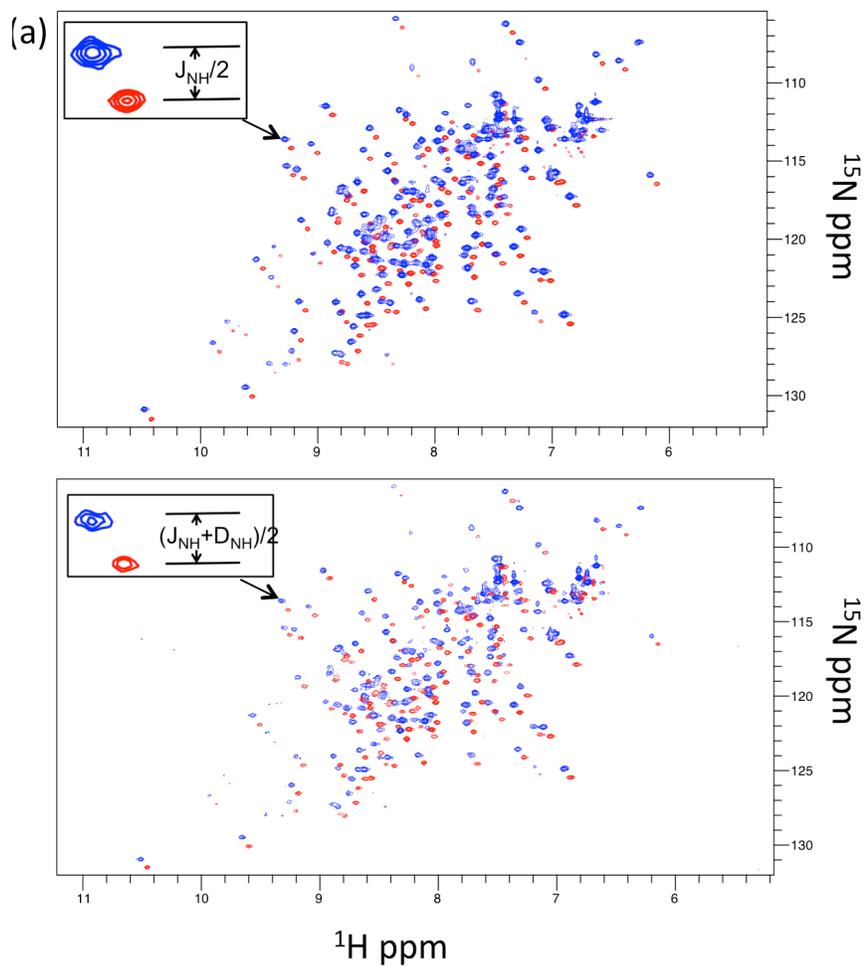
Amide-amide connections used for the Est3<sup>ΔN</sup> structure calculation are mapped on this schematic of the secondary structure of Est3, where β-strands shown as squares and helices and loops as circles. Residue numbering is labeled for every 10 residues. β-strands 1, 2 and 3 are colored in orange, green and yellow to form one β-sheet and β-strands 4 and 5 are colored in gray and brown and form another β-sheet. Intra-sheet connections are colored in blue and inter-sheet connections are colored in green. All other connectivities are shown in red.



**Figure 2.17 Schematic of secondary structure of Est3<sup>ΔN</sup> showing long-range methyl-methyl NOE connectivities used for structure calculation**

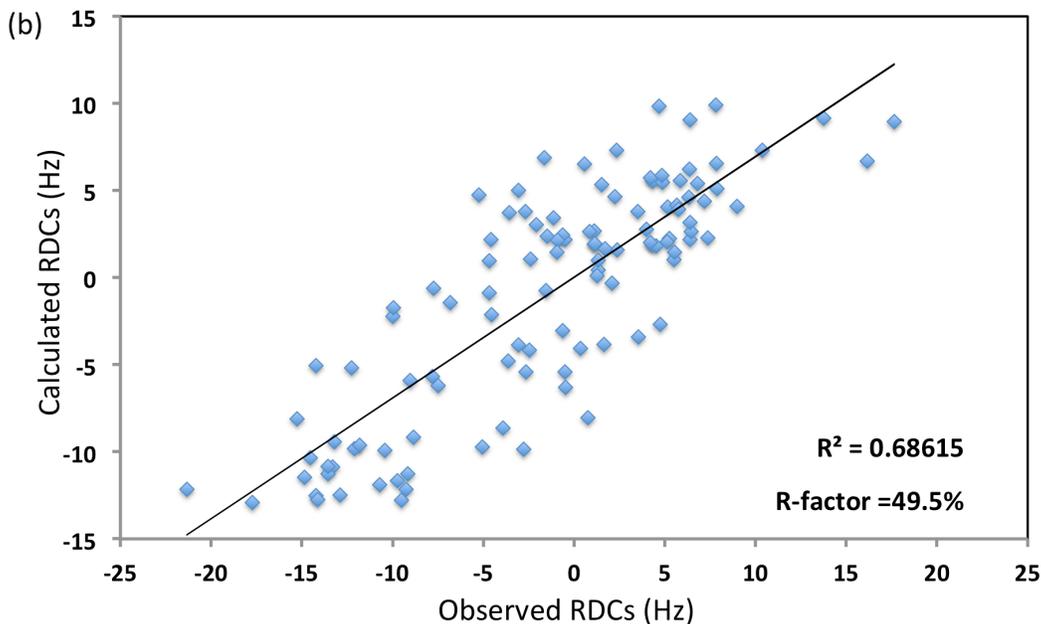
Methyl-methyl connections used for the Est3<sup>ΔN</sup> structure calculation are mapped on this schematic of the secondary structure of Est3, where  $\beta$ -strands shown as squares and helices and loops as circles. Residue numbering is labeled for every 10 residues. Selectively methyl-protonated Ile, Leu and Val residues are colored in orange, green and yellow respectively. Most of the ILV methyls show connectivities to other methyls, except for a V16 at the N-term and 162I, I65I, 171L and 181L in the C-terminal tail.

In addition, Residual Dipolar Coupling (RDC) data was collected that provides orientational information for the molecule [238].  $^{15}\text{N}$ -labeled Est3 was aligned in the presence of 9.6 mg/mL Pf1 filamentous phage.  $^{15}\text{N}$ -HSQC and -TROSY experiments on isotropic (no phage) and aligned (with phage) (Figure 2.18a) were used to derive 112 RDC measurements for use in structure calculation and the correlation plot of the back calculated RDCs (from one of the structures from the final ensemble, calculated using XPLOR-NIH) to the observed/experimental RDCs indicates that the structure is in agreement with the experimental RDC data (Figure 2.18b). See Table A3 (Appendix) for complete list of RDC restraints used in structure calculation. The cross-validation term, R-factor, of 49.5% is not better (*i.e.* lower value) likely because the rmsd over the full backbone is only 2.5 Å, thereby increasing uncertainty of N-H vector orientation in the calculated structures [239]. R-factor of 0% defines perfect agreement between observed and back-calculated RDCs from the validated structure and R-factor of 100% is the value for experimental RDCs with completely unrelated/ random structures.



**Figure 2.18a Spectra for residual dipolar coupling calculations**

Superposition of the  $^1\text{H}$ - $^{15}\text{N}$  TROSY (red) and HSQC (blue) spectrum of the unaligned (top panel) and pf1-phage aligned (bottom panel) sample of Est3<sup>AN</sup> protein. The inset shows half of the one-bond  $^{15}\text{N}$ - $^1\text{H}$  splittings  $(J_{\text{NH}} + D_{\text{NH}})/2$  measured from the difference between  $^{15}\text{N}$  frequency in the TROSY and HSQC spectra.



**Figure 2.18b Correlation plot between the back-calculated and experimental RDCs**

RDCs were back calculated from one of the structures of the ensemble from the final RASREC Rosetta calculation. Back-calculated values and R-factor were generated by XPLOR-NIH [222]. The spread of the correlation-plot is  $\sim 4$  rmsd. The coefficient of correlation and R-factor are shown on the plot. R-factor, which scales between 0 and 100%, is defined as the ratio of the root mean square deviation between observed and calculated values to the expected root mean square deviation if the vectors were randomly distributed [239].

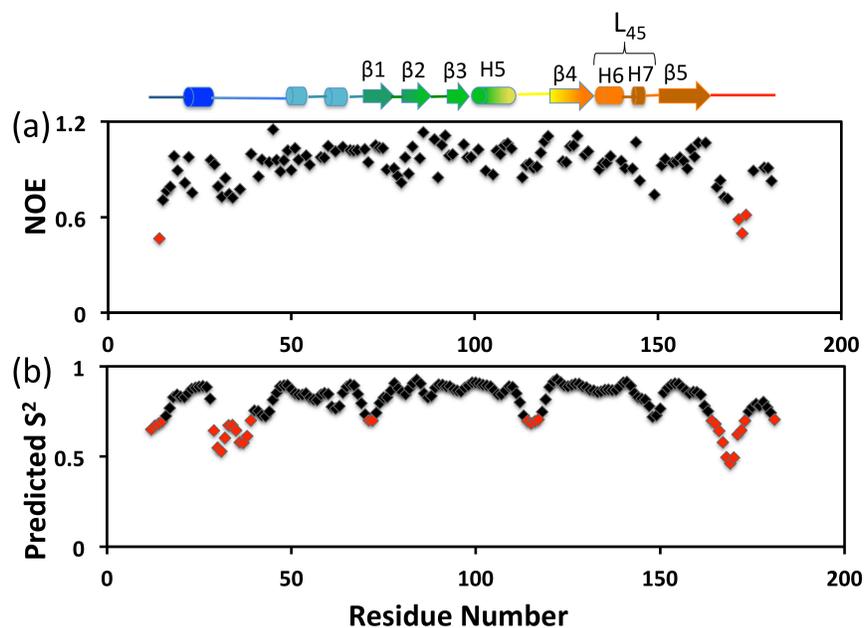
### **2.3.3c $^{15}\text{N}$ - $\{^1\text{H}\}$ -heteronuclear NOE experiment in parallel with backbone chemical shift based $S^2$ prediction helped elucidate dynamic regions of the protein**

HetNOE is a useful experiment to understand the dynamics of a protein in the fast timescales (ps-ns) that include global molecular tumbling and local bond fluctuations for internal dynamics of protein backbone [240]. A HetNOE experiment helps to assess if a protein has flexible regions and, as in the case of Est3<sup>AN</sup>, with > 97% complete backbone assignment, the identity of the flexible residues can be determined. However, the downside of this experiment is its low inherent

sensitivity [241]. The low signal-to-noise problem can be somewhat alleviated by increasing the duration of data acquisition, however, for low solubility proteins like Est3 and even its Est3<sup>ΔN</sup> mutant, this translates to difficulty in correct peak height or peak volume calculation due to poor quality of the proton saturated HetNOE spectrum compared to the reference spectrum. For HetNOE experiment on Est3<sup>ΔN</sup> mutant, the peaks that could be reasonably picked above the spectrum noise were subjected to calculation of the HetNOE values:

$$\text{HetNOE} = \frac{\text{Peak intensity in proton saturation spectrum}}{\text{Peak intensity in reference spectrum}}$$

The HetNOE values are plotted against the Est3 residue number (Figure 2.19a). HetNOE values below 0.7 are shown in red and denote the regions of higher flexibility and values above 0.7 indicate regions in the rigid core of Est3. Due to poor quality of the proton saturated HetNOE spectrum, the spread of the HetNOE values is high and prevents a fully reliable identification of the rigid core, relative to the flexible regions of the protein, except for the highly flexible N- and C-terminal regions. Despite low S/N that precluded a quantitative measurement, overall trends could be qualitatively used for validating our protein structure ensemble calculated with RASREC Rosetta. Based on the lack of sensitivity of the het NOE, we did not continue with the collection of T1 and T2 data.



**Figure 2.19** HetNOE and RCI-S<sup>2</sup> relaxation data for Est3<sup>ΔN</sup>

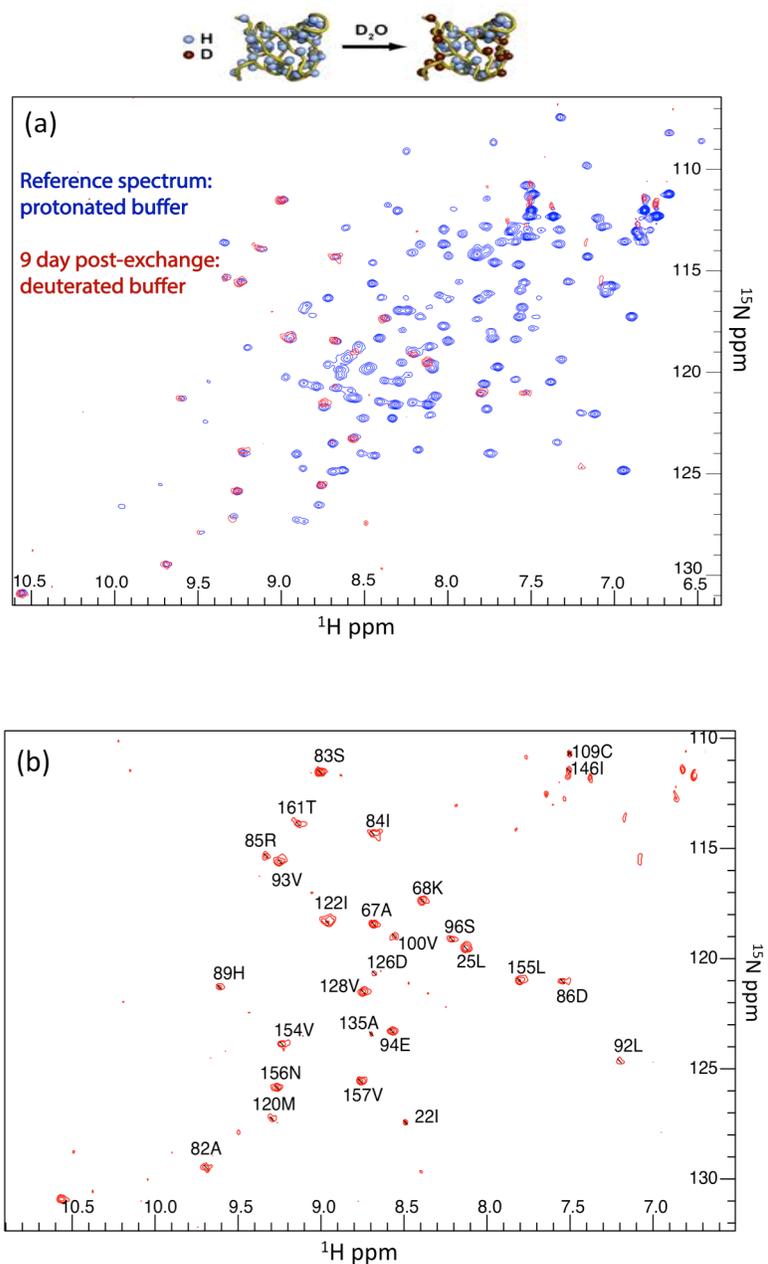
Relaxation data for Est3, to identify flexible regions of the backbone. (a) <sup>15</sup>N-<sup>1</sup>H-heteronuclear NOE and (b) predicted order parameter S<sup>2</sup> [242] along the y-axis versus Est3 residue number along the x-axis. Data for (a) was collected with a ~290 μM <sup>15</sup>N-labeled Est3<sup>ΔN</sup> at 800 MHz. Data for (b) was generated by the TALOS+ server [216] using the experimental chemical shifts of the backbone atoms generated on <sup>2</sup>H,<sup>13</sup>C,<sup>15</sup>N-labeled Est3<sup>ΔN</sup> at 800 MHz. The corresponding secondary structure elements for the residues are drawn on top of the figure.

Another method to predict protein dynamics is based on C<sub>α</sub>, C<sub>β</sub>, CO, N and H<sub>α</sub> backbone chemical shifts [242], based on the observation that random coil chemical shifts are characteristic of flexible regions of the protein. Order parameter S<sup>2</sup> is representative of protein dynamics, with values below 0.7 indicative of higher internal fluctuations and above 0.7 indicative of the less flexible or rigid core. S<sup>2</sup> values for Est3 were predicted using the C<sub>α</sub>, C<sub>β</sub>, CO, N and H<sub>α</sub> backbone chemical shifts in the TALOS+ program [216] for the Est3<sup>ΔN</sup> protein and are shown as a function of residue number in the Figure 2.19b. Regions highlighted in red are those with S<sup>2</sup> value below 0.7 and as can be seen by comparison the secondary structure

elements of Est3 structure (on top of the Figure 2.19), these flexible regions lie in the predominantly the N- and C-terminal regions of Est3 and also in some loops in the middle.

### **2.3.3d H/D-exchange data identified solvent protected regions of the Est3 protein and was used in structure validation**

Exchangeable hydrogens, like amides in protein backbones, can be replaced with deuteriums by buffer exchanging the NMR buffer from H<sub>2</sub>O-based to D<sub>2</sub>O-based buffer. The exchange reaction can be monitored over time, to detect which amides of the protein are protected from exchange to deuterium. The Est3<sup>ΔN</sup> amide protons exchange to deuterium was monitored by multiple <sup>1</sup>H-<sup>15</sup>N HSQC experiments collected over a time range of 7 hrs to 9 days and compared to reference spectrum collected before the sample was exchanged into deuterium-based buffer (Figure 2.20). The amide protons that are exchanged to deuterium will cease to show signal on the <sup>1</sup>H-<sup>15</sup>N-HSQC spectra, and therefore monitoring the peaks on the HSQC will help identify the exchange protected protons and their corresponding residues in the proteins. In addition, the H/D exchange data was mapped onto the RASREC Rosetta solved structure of Est3 and used for structure validation (section 2.3.5e)



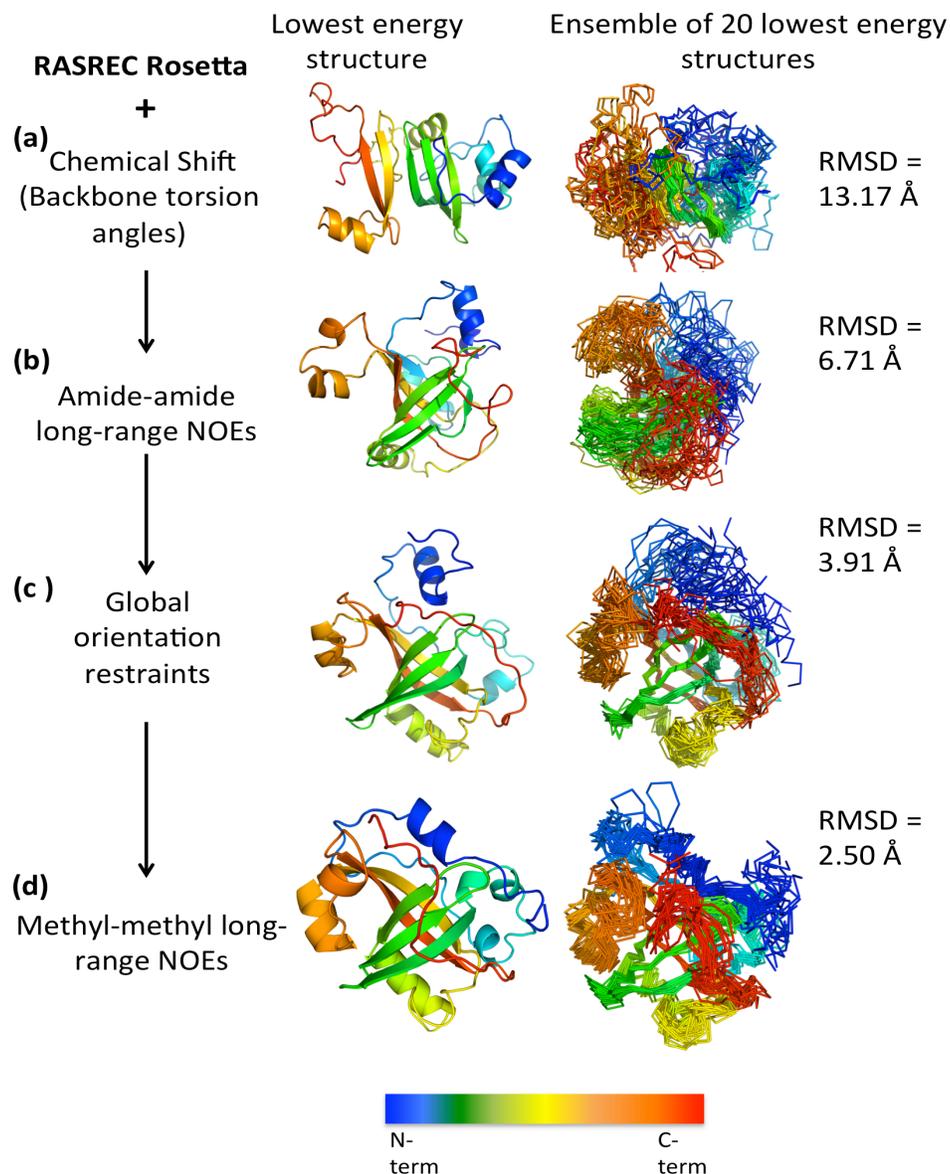
**Figure 2.20**  $^{15}\text{N}$  HSQC spectra for H/D exchange of Est3 $\Delta\text{N}$

(a) Reference (blue) and 9 day post buffer-exchange to deuterated sample buffer (red)  $^{15}\text{N}$  HSQC spectra are superimposed to show loss of most of the peak signals upon deuteration. (b) Protected peaks were identified from the residue assignments.

### 2.3.4 Relatively novel strategy, RASREC Rosetta, was used for Est3<sup>ΔN</sup> structure calculation

Conventional structure calculation methods, specifically x-ray crystallography and NMR, were not feasible for Est3<sup>ΔN</sup> structure calculation because of sample instability to precipitation, heterogeneity due to dimer-monomer equilibrium and low signal-to-noise issues. Therefore, an alternative and relatively novel strategy, RASREC Rosetta, was tried for obtaining a high-resolution structure of this protein. RASREC Rosetta is a *de novo* structure calculation approach that combines sequence homology-based structure modeling using short fragments with iterative conformational sampling using sparse experimental data [220,224,243,244]. This approach has been successfully employed previously to yield close-to-native structures of proteins [226]. Refer to section 2.2.12a for more details on the Rosetta strategy.

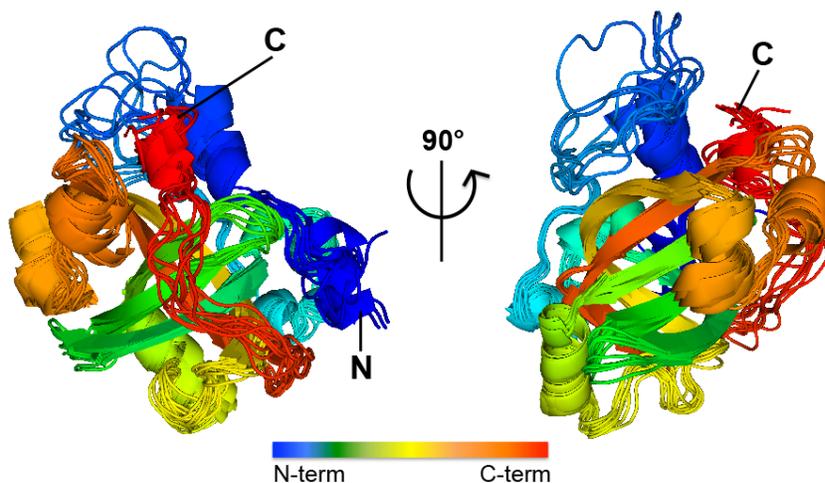
Minimal NMR experimental data, for use with RASREC Rosetta structure calculations, was collected on a deuterated Est3<sup>ΔN</sup> sample. Deuteration was necessary to enhance S/N in NMR experiments because increasing the sample concentration to >300 μM was not feasible due to sample instability. RASREC Rosetta calculation with just backbone chemical shifts, however, was not sufficient for reasonable structure outputs (Figure 2.21). Therefore, in addition to backbone chemical shifts, long-range connectivities, *i.e.*, amide-amide and methyl-methyl NOEs, were also assigned and used for structure calculation.



**Figure 2.21 Improvement of structure ensemble convergence upon step-wise addition of NMR experimental data to the RASREC Rosetta structure calculation of Est3<sup>AN</sup>**

Backbone chemical shift data (a), plus long-range amide-amide NOEs (b), plus RDCs (c), plus long-range methyl-methyl NOEs (d) were added to RASREC Rosetta for Est3<sup>AN</sup> structure calculation. Lowest energy structures from each run indicate how the overall protein folding progressed as more experimental restraints were added. Convergence of the top 20 structures (ensemble) indicate whether the experimental data in that step was sufficient to drive the calculation to a single solution. RMSDs show the quantitation of this convergence and the reducing numbers from top to bottom indicate that the best-converged structures are achieved when all the experimental restraints were included in structure calculation.

A sequential addition of minimal NMR data with RASREC Rosetta calculation of structures and comparing the level of convergence achieved for the calculated structures highlighted the importance of each set of minimal NMR data that was required for reliable Est3<sup>ΔN</sup> structure calculation. Figure 2.21 shows the step-wise improvement in structures upon addition of NMR restraints. Furthermore, RDC data was also collected that provided orientational information for the molecule [238], and was important for increasing the convergence of our calculated structure ensembles. RASREC Rosetta generated a pool of 500 structures as output of each structure calculation run and top 10 to 20 structures from each run were analyzed for convergence. The structure calculation run with the best convergence (lowest rmsds) of structures was chosen as our final structure ensemble (Figure 2.22) (Table 2.2).



**Figure 2.22 Ribbon representation of Est3<sup>ΔN</sup> structure ensemble**

Ensemble of the ten lowest energy structures of Est3<sup>ΔN</sup> generated by RASREC Rosetta calculation (see Figure 2.21 for element identification) The image on the left is rotated 90° around the vertical axis to present another view of the ensemble on the right.

---

**Table 2. 2** Structural statistics for the Est3<sup>ΔN</sup> RASREC structures

---

Number of residues	170
NOE-based distance restraints	
NOE distance restraints (violations $\geq 0.5$ Å) <sup>a</sup>	128 (24±4)
Number of restraints per residue	0.75
Other restraints	
$\phi+\psi$ dihedral-angle restraints (violations $\geq 5^\circ$ ) <sup>b</sup>	254 (26±3.7)
RDC restraints (violations $\geq 5$ Hz)	112 (7±1.6)
Average rmsd to the average structure <sup>c</sup>	
Backbone (Å)	1.3 ± 0.19
Heavy atom (Å)	1.8 ± 0.23
Ramachandran Plot summary <sup>d</sup>	
Most favored regions	84.1%
Allowed regions	15.4%
Generally allowed regions	0.5%
Disallowed regions	0.0%

---

Statistics are given for the ten lowest-energy structures.

<sup>a</sup>All NOEs, amide-amide and methyl-methyl, are long range ( $|i-j| > 3$ )

<sup>b</sup>Torsion-angle restraints were derived from TALOS+

<sup>c</sup>rmsds calculated using the iCing server.

<sup>d</sup>Analysis performed using PROCHECK

---

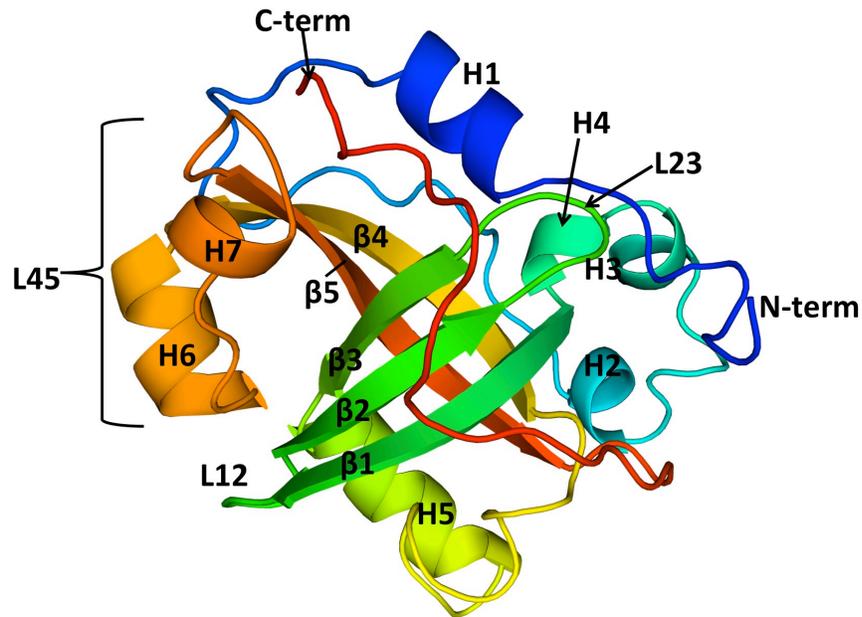
## 2.3.5 Est3 structure evaluation

### 2.3.5a Est3 is an OB-fold protein

The RASREC Rosetta structure of Est3<sup>ΔN</sup> reveals a five-stranded  $\beta$ -barrel (Figure 2.22 and Figure 2.23), which is supported by a number of long-range NOEs between neighboring antiparallel  $\beta$ -strands. This five-stranded  $\beta$ -barrel, capped by a helix between strands 3 and 4, is the canonical topology of an OB-fold [98].

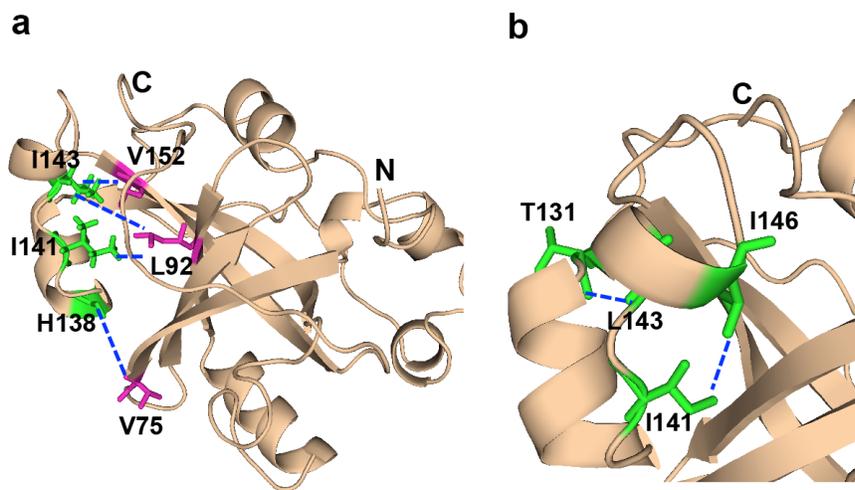
Because the deletion of 12 residues at the N-terminus enhanced stability, the first 12 residues of Est3 were excluded from the experimental data collection and therefore not present in our structures. The ensemble of converged structures (Figure 2.22) indicates that N-terminal region from Q19 to D28 sits almost parallel with the C-terminal region of the protein. Overall, the region N-terminal to the  $\beta$ 1-strand of the OB-fold, composed of the first 64 residues of the protein, makes a spiral-shaped structure that caps the top of the  $\beta$ -barrel (Figure 2.23a). The H1 helix in this N-terminal region sits very close to the short loop, L23, between strands  $\beta$ 2 and  $\beta$ 3. The N-terminal region is followed by strand  $\beta$ 1 that characteristically wraps around the axis of the barrel such that a small N-terminal portion forms a 3-stranded sheet with strands  $\beta$ 4 and  $\beta$ 5. The rest of the  $\beta$ 1 forms an antiparallel  $\beta$ -sheet with strands  $\beta$ 2 and  $\beta$ 3. L12 and L23 are short looped-turns that span five and four residues respectively. The 23-residue stretch between  $\beta$ 3 and  $\beta$ 4 forms a helix (H5) followed by a loop; both elements sit at the bottom of the  $\beta$ -barrel. H5 spans 11 residues and sits at the base of the  $\beta$ -barrel and at a  $\sim 40^\circ$  angle to the vertical axis of the  $\beta$ -barrel.  $\beta$ 4 and  $\beta$ 5 are long strands that form one of the two antiparallel sheets of the  $\beta$ -barrel. Strands  $\beta$ 4 and  $\beta$ 5 are connected by L45, an unusually long (19 residues) structural element, and is composed of the helices H6 and H7 separated by small loops. Following  $\beta$ 5, which completes the barrel, is the 19-residue C-terminal tail, which sits over the antiparallel  $\beta$ -sheet formed by  $\beta$ 1,  $\beta$ 2, and  $\beta$ 3.

(a)



**Figure 2.23a** Ribbon representation of single structure of Est3<sup>AN</sup>

The OB-fold structure of Est3<sup>AN</sup> protein is shown with secondary structure elements labeled. The protein ribbon is rainbow-colored with blue at the N-terminus and red at the C-terminus



**Figure 2. 23b** Long range NOEs define the packing of L45 against the face of  $\beta$ -barrel.

(a) Four NOE connectivities (blue dashed lines) between H atoms from three residues on the L45 (colored in green) to H atoms from three residues (colored in magenta) on the  $\beta$ -barrel are shown: I143 to V152 and L92; I141 to L92; and H138 to V75. (b) Enlarged view of L45 shows two sets of intra-L45 NOE connectivities (blue dashed lines): between T131 and L143, and between I146 and I141 (colored in green).

### ***2.5.3a1 Calculated structure ensemble is highly converged in the core OB-fold***

The ten lowest-energy structures exhibit a high degree of convergence (Figure 2.22) within the core OB-fold with an rmsd of 1.05 Å (over residues 66-163) but the loops outside of the central  $\beta$ -barrel are less converged. The loops are also dynamic regions of the protein, as indicated by the HetNOE and ordered-parameter RCI S<sup>2</sup> values (section 2.3.3c). The first turn in the N-terminal region, starting at D28, was observed to be highly flexible also contributing to lack of convergence in this region. Similarly, much of the C-tail is dynamic as indicated by the HetNOE NMR measurements and order-parameter RCI S<sup>2</sup> values (Figure 2.19). In contrast, no flexibility was observed in the long L45.

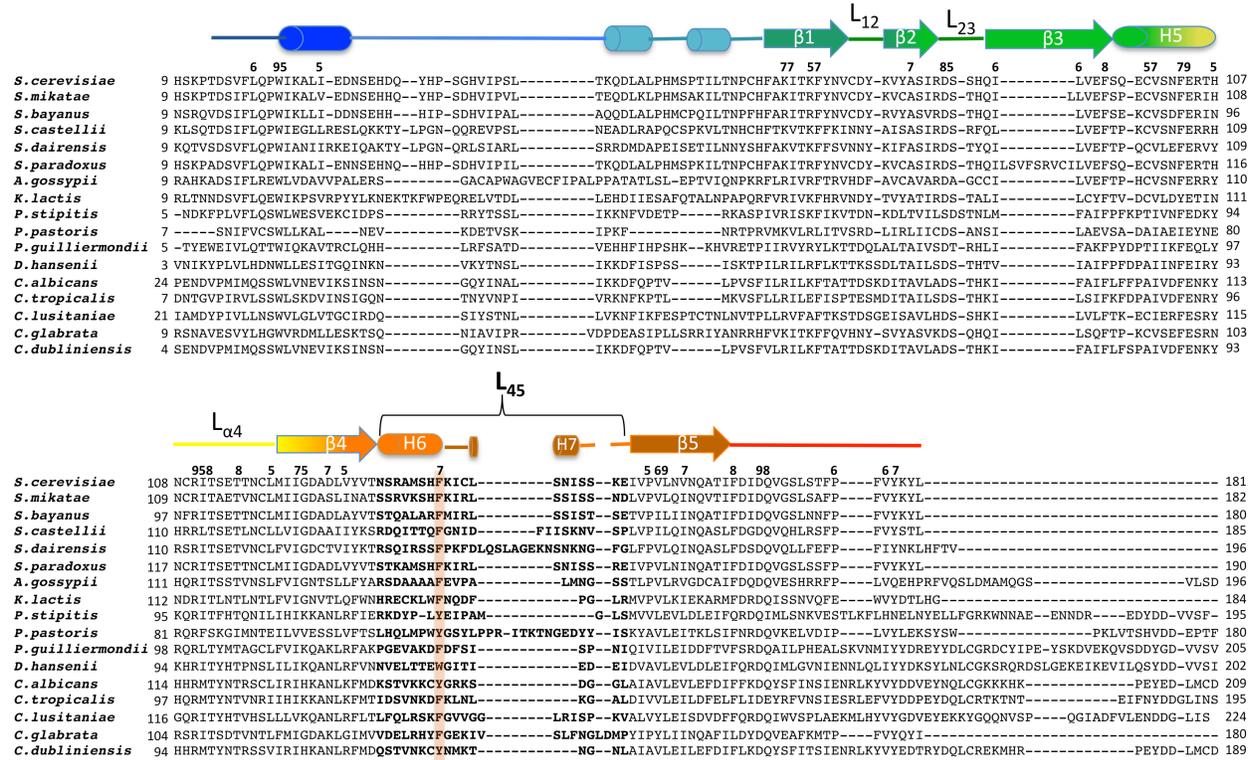
### ***2.5.3a2 Structure of the L45 is defined by several experimental restraints***

L45 is unusually long (19 residues) and structured in the Est3 structure. This is a unique structure, as the available structures of other OB folds indicate a much shorter L45 element. In our Est3<sup>ΔN</sup> structure calculation, the placement and structure of the unusually long L45 is supported by chemical shifts (Table A1) that define the helix (H6) and loop secondary structure elements, in addition to two intra-L45 long range NOEs that define the hairpin conformation of the L45 (Figure 2.23b). L45 is also highly structured as confirmed by HetNOE and RCI S<sup>2</sup> values (Figure 2.19) and it packs against the face of the barrel as defined by four long range NOEs to the  $\beta$ -sheet formed by strands  $\beta$ 1,  $\beta$ 2, and  $\beta$ 3 (Figure 2.23b). Further,

orientational restraints comprising 10 NH bond vectors (Table A3) in the L45 region validate its calculated orientation in the context of Est3.

### 2.3.5b Mapping the conserved residues on the Est3 structure

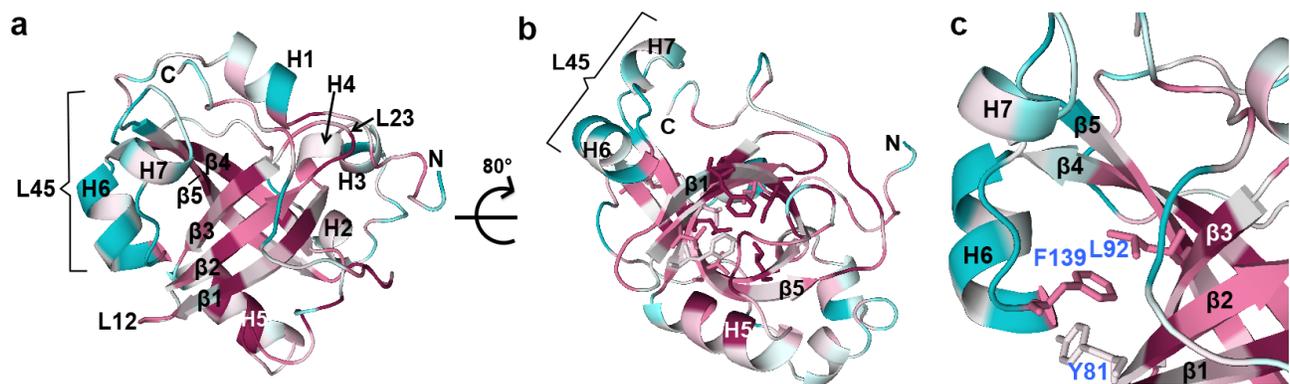
A multiple sequence alignment of Est3 proteins from 17 different yeast species (Figure 2.24) was used to identify conservation rates, which were mapped onto the structure (Figure 2.25a) using the ConSurf server [245-247].



**Figure 2.24** Sequence alignment of Est3 proteins highlighting features of the structure

Alignment of Est3 homologs from 17 yeast species was done using the PROMALS multiple sequence alignment program [188], and this alignment was used for conservation mapping by ConSurf server in Figure 2.23. Medium to highly conserved residues are indicated by numbers 5 to 9 on the bottom of the alignment, 9 being the most conserved. Secondary structure elements from RASREC Rosetta structure of Est3<sup>AN</sup> are drawn on top of the sequence alignment. The region of the loop L45 is highlighted in bold and the conserved F is highlighted in orange. All Est3 homologs have this long loop.

The N-termini (residues 1-12) are highly conserved within the Saccharomycotina clade, but only moderately conserved between all yeast Est3 sequences evaluated (data not shown). H1 helix is well conserved, however, the remainder of the N-terminal region capping the top of the OB-barrel is not (Figure 2.25a, colored in cyan). The core OB-fold region ( $\beta$ 1-5 and H5) is composed of highly conserved residues (Figure 2.25a, colored in pink). The antiparallel  $\beta$ -sheets that comprise the central  $\beta$ -barrel contain alternating and conserved hydrophobic residues with side-chains pointing inwards and this arrangement is important for the structural integrity of the barrel [81] (Figure 2.25b).



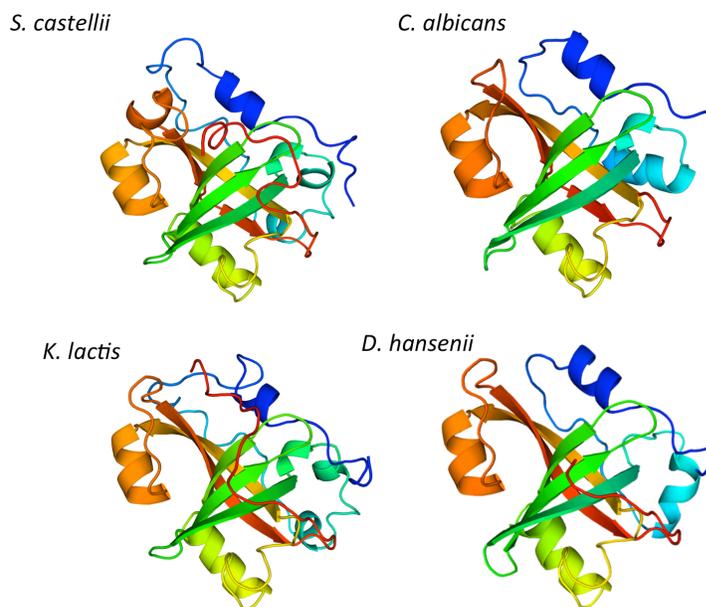
**Figure 2.25 Conserved residues mapped on the Est3<sup>AN</sup> structure**

(a) The OB-fold structure of Est3<sup>AN</sup> is presented with the ribbon color-coded in a range from maroon through turquoise indicating conserved through variable residues based on mapping by ConSurf server[247]. The secondary structure elements are labeled. (b) A view of the structure through the barrel of the OB-fold. H5 helix at the base of the OB-fold is turned towards the viewer and other secondary structure elements are labeled for orientation. The  $\beta$ -strands have conserved hydrophobic side-chains, shown as sticks that project into the core of the barrel. (c) An enlarged view of the loop L45 is shown with the conserved F139 side-chain is shown in stick. This side chain projects towards the barrel face composed of strands  $\beta$ 1,  $\beta$ 2 and  $\beta$ 3 and likely forms a hydrophobic interaction with side-chains of Y81 and L92.

Within the 19-residue C-terminus, there are three distinct regions. The first third (residues F163-V168) forms a highly conserved, negatively charged patch (Figure

2.27). The middle (G169-T173) is phylogenetically variable, and the final third is only moderately conserved.

Interestingly, the unusually long L45 is not conserved except for phenylalanine at position 139, which is a phenylalanine residue in 70% of the Est3 sequences and a tyrosine or tryptophan in others (sequence alignment: Figure 2.25c). Although the L45 sequence is variable the length does appear to be long overall, averaging around 21 residues. The sequence alignment (Figure 2.24) shows that all Est3 proteins have a long L45, ranging from 14 residues in *P. stipitis* to 28 residues in *S. dairensis*. Structures of some *S. cerevisiae* Est3 homologs were modeled using the Phyre2 server [248], with 100% confidence of prediction and indicate the presence of structured L45 loop in all the homologs (Figure 2.26).

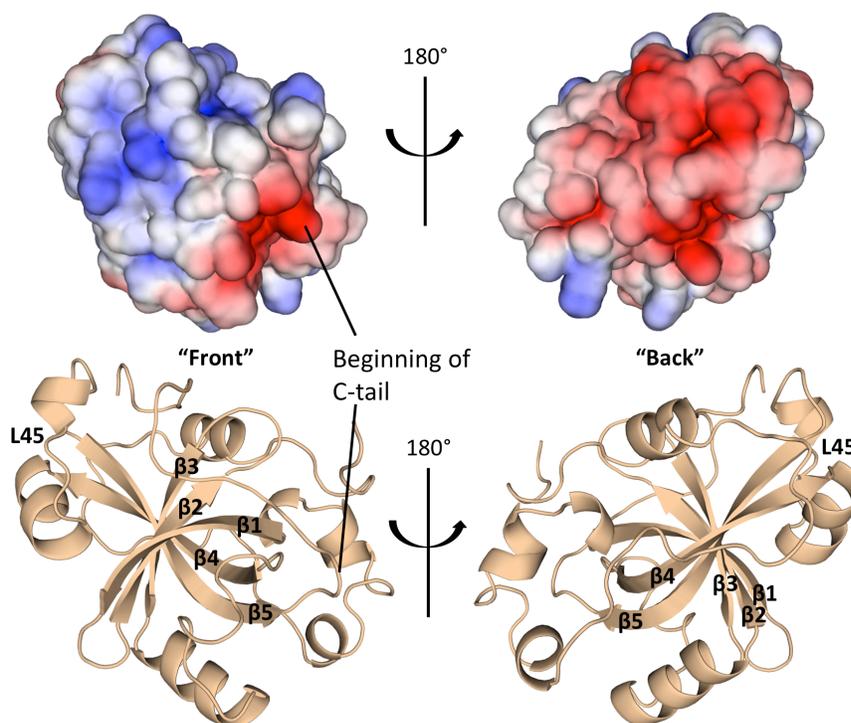


**Figure 2.26 Threaded models of some Est3 homologs based on *S. cerevisiae* Est3<sup>AN</sup> structure**

Phyre2 [248] server was used for modeling the structures of Est3 protein homologs from *S. castellii*, *C. albicans*, *K. lactis* and *D. hansenii* yeasts. Sequence identities with *S. cerevisiae* Est3 protein were 48%, 26%, 35% and 28% for *S. castellii*, *C. albicans*, *K. lactis* and *D. hansenii* Est3s respectively.

### 2.3.5c Electrostatic surface map of Est3 structure

Electrostatic charges on the protein surfaces can be important determinants of protein interactions. The electrostatic surface map of Est3<sup>ΔN</sup> structure was created using the APBS plug-in in PyMOL molecular graphics system (Figure 2.27).

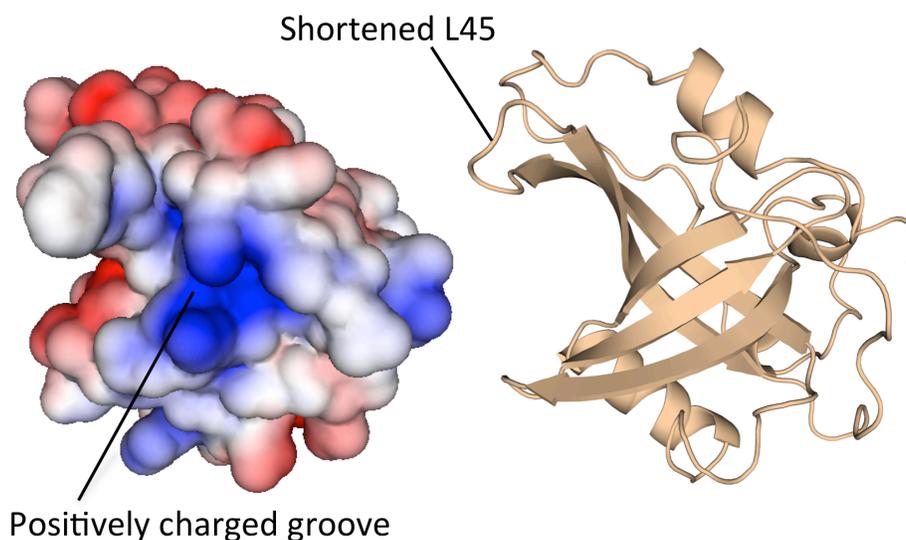


**Figure 2.27 Electrostatic surface representation of the Est3<sup>ΔN</sup> structure**

Electrostatic surface representations of two faces of the Est3<sup>ΔN</sup> structure are shown (top panel). The “Front” face (left panel) is predominantly positively charged, except the region corresponding to the beginning of the C-tail, which is negatively charged. The “Back” face (right panel) is predominantly negatively charged. Bottom panel shows the orientation of the protein in the surface representations from the top panel.

The “front” face of the molecule, represented by  $\beta$ -strands 1-3 is mostly positively charged. This face also comprises the canonical ligand-binding site (like DNA) for the OB-fold proteins and presence of the positively charged surface in Est3<sup>ΔN</sup> might

be an indicator that Est3 might interact with such a ligand, however no groove for binding is readily present on the surface of Est3. Conformational change by the moving away of L45 might lead to the opening-up of such a groove, as revealed by the electrostatic surface of the modeled Est3 structure, with L45 deleted (Figure 2.28).



**Figure 2.28 Electrostatic surface representation of the Est3 structure with loop L45 removed by modeling**

Removal of loop L45 was modeled into the structure of Est3<sup>ΔN</sup> using one-to-one threading server phyre2 [248]. The modeled structure with short loop is shown in “wheat” color on the right. The electrostatic surface representation of the structure is shown on the left, with the positively charged groove that is exposed upon removal of L45.

The region at the base of the protein has a contiguous negatively charged patch (around residues D164 and D166 at the beginning of the C-terminal tail). This surface has been implicated from mutational studies to be involved in Est3’s interaction with telomerase protein Est2 and therefore the concentration of this negatively charged patch seems to be important for protein-protein interaction for Est3. The “back” face of Est3 is constituted of  $\beta$ -strands 4 and 5 and also the N-

terminal spiral-shaped structure that caps the top of the  $\beta$ -barrel. Interestingly, almost all of this “back” face is predominantly negatively charged. However, not much is known about the involvement of this face of Est3 in its function, as no residues from this face have been ascribed to be important for Est3’s function from mutagenesis studies *in vivo*. It is likely, that this face acts the structural support to the “front” face that is involved in ligand interaction.

### 2.3.5d Structural homolog search for Est3

As structural similarity is sometimes indicative of functional similarity as well, the DALI server was used to identify proteins in the PDB database with structural similarity to Est3 [249]. The top hit was the OB fold of human TPP1 (*Hs*TPP1-OB) (see Table 2.3 for Z-scores and rmsd).

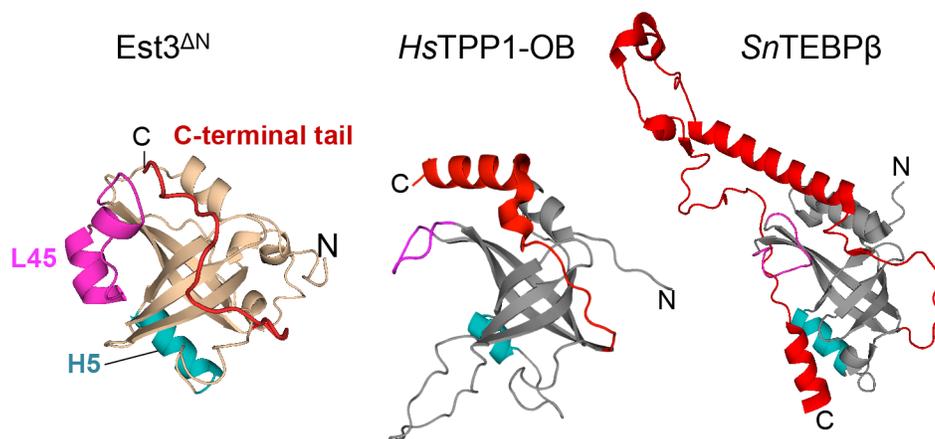
**Table 2. 3 DALI search for structurally similar proteins to Est3 in the PDB**

<b>DALI-search hit<sup>a</sup></b>	Chain	Z-score <sup>b</sup>	rmsd (Å)	Length of alignment	Number of residues	% identity
<b><i>Hs</i>TPP1-OB</b>	2i46-B	11.2	2.6	116	134	15
<i>Hs</i> lysyl-tRNA synthetase	3bjv-A	9.0	3.1	108	504	7
<i>Hs</i> RPA70	2b3g-A	8.3	2.8	106	117	9
<i>Hs</i> Gamma-Interferon-inducible protein	2oQ0-D	7.9	2.8	89	184	15
<i>Sulfolobus</i> SSB	1o7i-B	7.8	2.3	92	114	8
<b><i>Sn</i>TEBP<math>\beta</math></b>	2i0q-B	7.7	3.3	113	216	13
<b><i>Tt</i>TEB1</b>	3u58-B	7.1	2.1	85	210	9
<i>Hs</i> RPA32	1quq-A	7.0	2.8	87	121	10
<b><i>Sp</i>Stn1</b>	3K56-A	6.4	2.6	92	136	13

<sup>a</sup>Telomere-associated proteins are in bold. Proteins for which more than one structural feature matches with Est3 are in gray boxes.

<sup>b</sup>Z-score is a measure of similarity, with Z-score above 2 indicating significant structural similarity [249]

Other telomere-associated proteins were also identified as structurally related to Est3, including the *Hs*TPP1-OB homolog in the ciliate *Sterkiella nova*, *Sn*TEBP $\beta$ , *Tetrahymena thermophila* TEB1, and fission yeast Stn1. Some of the non-telomere associated proteins that showed up in the DALI search were the DNA binding subunit of human RPA70, human RPA32, human lysyl-tRNA synthetase, archaeal (*Sulfolobus*) SSB protein and human gamma-interferon-inducible protein. As expected, all the hits from the DALI-search with a Z-score cut-off at 2.0 are OB-fold proteins. However, a closer evaluation shows that only a few of these hits shared at least one feature with Est3 outside of the five-stranded  $\beta$ -barrel. These proteins were *Hs*TPP1 and *Sn*TEBP $\beta$  and to a lesser extent *Hs*RPA70 and archaeal SSB (Table 2.3). These four proteins share three specific elements with Est3. First, helix H1 is proximal to L23; second, the C-terminus crosses over the antiparallel  $\beta$ -sheet made by  $\beta$ 1-3; and third, the N- and C-terminal regions are proximal and antiparallel to each other. Additionally, *Hs*TPP1-OB and *Sn*TEBP $\beta$  share a fourth element with Est3, as helix H5 is positioned at a characteristic angle of  $\sim 40^\circ$  to the vertical axis of the  $\beta$ -barrel (Figure 2.29). Most notably, the long loop L45 was not found in any of the hits reported by the DALI search. This suggests that the long, structured L45 is a feature specific to Est3 proteins and may have a role in Est3 function.



**Figure 2.29 Structure comparisons of Est3 to its closest structural homologs *HsTPP1-OB* and *SnTEBPβ***

Comparison of Est3<sup>ΔN</sup> structure to that of the closest structural homologs in the PDB. DALI server was used to match proteins from PDB for structural similarity with Est3<sup>ΔN</sup> structure[249]. Structure of Est3<sup>ΔN</sup> is shown in the same orientation as Figure 4a. The core β-barrel and the N-terminus is “wheat” colored. Structure of *HsTPP1-OB* (PDB id: 2I46) is shown with core β-barrel and N-terminus “gray” colored. Orientation of OB-fold is the same as Est3 with β-strands 1-3 facing forward. Structure of *SnTEBPβ* (PDB id: 2I0Q) is shown with core β-barrel and N-terminus “gray” colored. Orientation of OB-fold is the same as Est3 with β-strands 1-3 facing forward. Helix at the bottom of the β-barrel, termed H5 in Est3 is colored in “teal” for all three structures. The C-terminus, colored in red for all three, sits characteristically over the face of the β-barrel formed by strands β1, β2 and β3. The L45, colored in “magenta” in all three structures, is very short in the *HsTPP1-OB* and *SnTEBPβ* structures and unusually long and structured in Est3.

### **2.3.5e Structure quality assessment and validation indicates the calculated structure of Est3<sup>ΔN</sup> is within the permitted structure quality range and is correct**

A number of structure validations were performed on the final structure ensemble to ensure that the calculated structures were not in violation of expected restraints for protein structures, were independent of any inadvertent NMR assignment error(s), and were independent of the calculation method used. Firstly, the final structure ensemble was analyzed with PROCHECK-NMR for Ramachandran violations [250] (Table 2.2). This analysis showed no residues from

the structures to be present in disallowed regions, indicating that our structures agree well with the expected residual conformational space.

Secondly, a set of internal validations were done to check if presence or absence of some NOEs led to altered final structures. It has been reported previously that RASREC Rosetta calculation accommodates for errors in the experimental data (like NOEs) [226], and therefore presence or absence of any subset of NOEs should not alter the output structure from the calculation. Hence, various subsets of the NOE constraints were fed to RASREC. These subsets of data produced structure ensembles with good agreement to that of the full calculation (Table 2.4), indicating that the Rosetta structures are well conserved between runs.

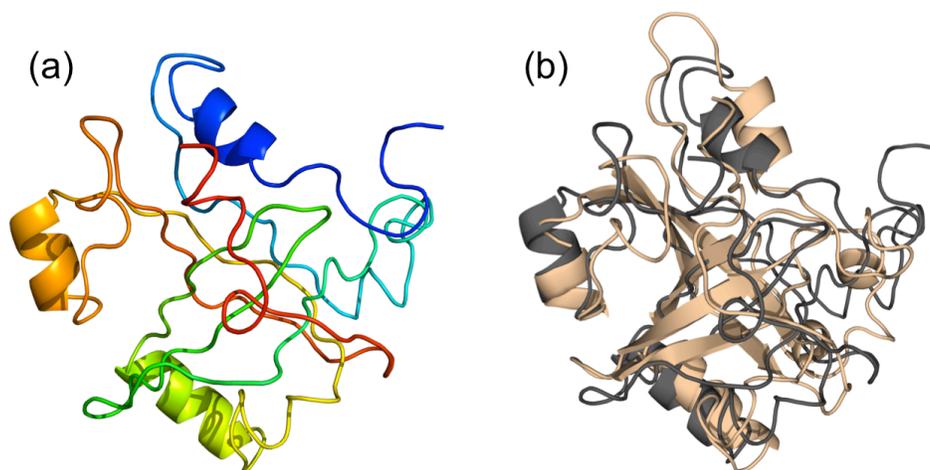
**Table 2. 4 Internal validation for convergence of RASREC Rosetta**

RASREC-Rosetta Calculation	Constraints dropped <sup>a</sup>	Number of NOE constraints	Backbone rmsd to Final Structure <sup>b</sup>
Final Structure	0%	128	
Validation I	10%	116	1.4
Validation II	10%	115	1.1
Validation III	10%	115	0.8
Validation IV	10%	115	1.0
Validation V	23%	98	0.8

<sup>a</sup>Structure calculation runs were done in presence of full-set of restraints, for the final structure, and subsequent runs of depleted restraints set

<sup>b</sup>Root-mean-square deviations (rmsd) for backbone atoms were calculated using the “super” script in PyMOL Molecular Graphics System, Version 1.5, Schrödinger, LLC

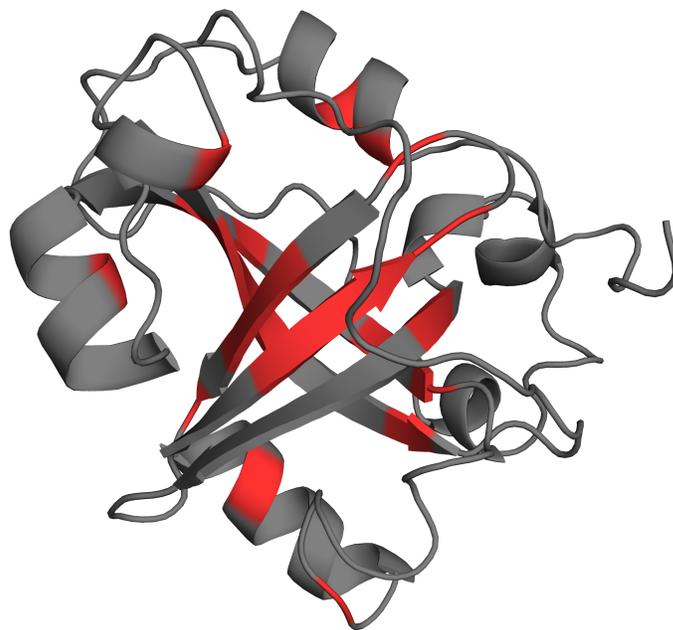
Finally, a Rosetta-independent, external validation was done by using chemical shifts and assigned NOE peaks as input for another structure calculation program, CYANA-2.1 [221]. From this input data, 100 structures were generated. Ensemble from the 20 of the lowest scoring structures was analyzed for constraint violation, and compared to the RASREC Rosetta structures. CYANA is the conventional automated NMR structure calculation program that requires a relatively complete set of chemical shift assignments from protein backbone and side-chains, in addition to long-distance restraints between backbone and side-chain atoms. Therefore, working with a limited experimental dataset for Est3<sup>ΔN</sup> structure calculation, CYANA yielded a low-resolution structure ensemble. The convergence of structures from CYANA-2.1 (Figure 2.30) was not good compared to RASREC Rosetta structure ensemble (Figure 2.22), however, the overall topology of the Est3<sup>ΔN</sup> structure was the same. The conservation of overall topology between structures calculated from these two different strategies, RASREC Rosetta and CYANA-2.1, confirmed that our observed structure was independent of the method used and thus the correct structure of Est3.



**Figure 2.30** CYANA 2.1 calculated structure of Est3<sup>AN</sup>

Est3<sup>AN</sup> structure was generated by CYANA-2.1 [251] as an external validation for the structure generated by the RAREC Rosetta using the same experimental data as for the latter strategy. (a) The CYANA structure (orientation same as Figure 2.21) shows similar topology to the RASREC Rosetta structure, however, the convergence of the ten lowest energy structures was not good (data not shown). (b) Superposition of the CYANA-2.1 structure (gray) and RASREC Rosetta structure (wheat) is 2.9 Å rmsd.

The H/D exchange experimental data was also used as validation tool (section 2.3.3d). The protected amide residues (after 9 days of H/D exchange) are mapped onto the structure of Est3<sup>AN</sup> (Figure 2.31). Predominantly, the residues in the core of the  $\beta$ -barrel, in the calculated structure, are protected from exchange. This result also acts as a validation for the calculated structure, because the  $\beta$ -barrel is expected to be more protected than the rest of the protein. Multiple of these residue side-chains are internalized, and sitting in the barrel core. Side-chains not internalized are likely protected due to hydrogen bonding. Aside from the barrel, there are some residues in the helices and three residues in the looped regions, that are also protected and these might be due to hydrogen-bonding with some other region of the protein imparting protection.

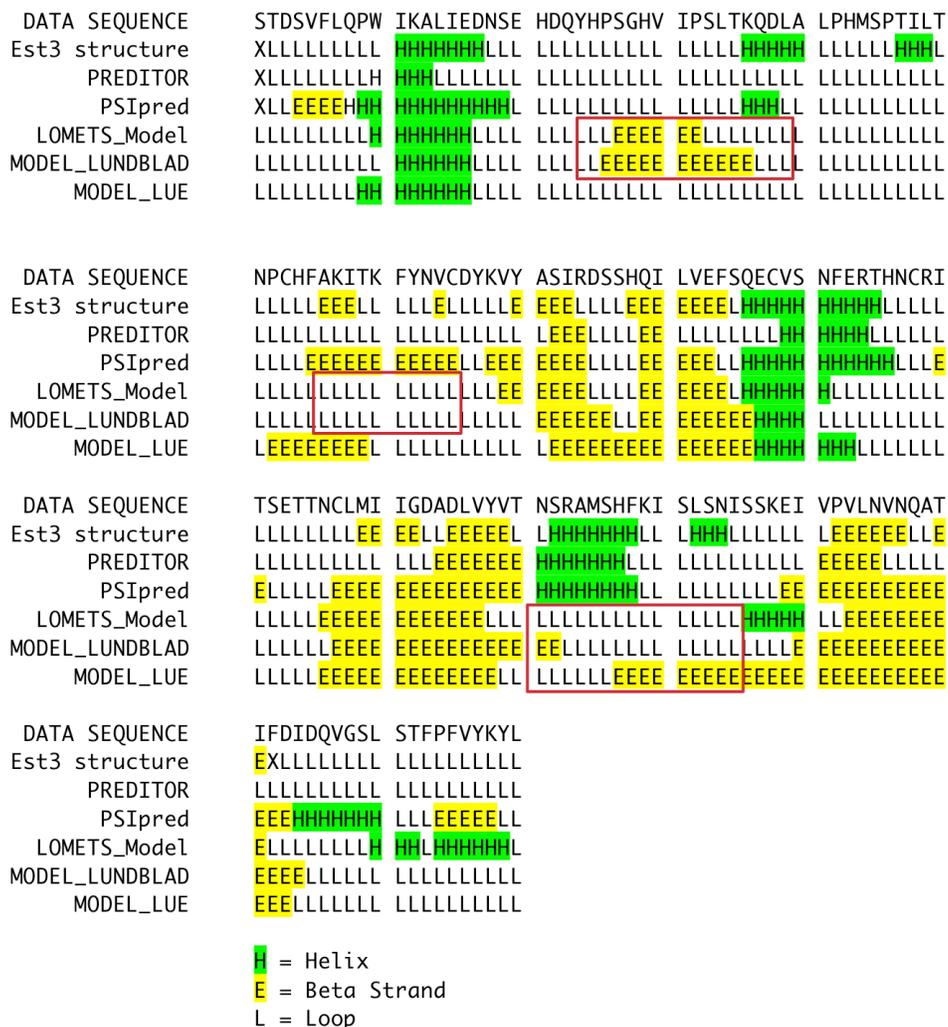


**Figure 2.31 H/D exchange protected residues mapped on the structure of Est3 $\Delta$ N**

H/D exchange protected peaks from the spectrum in Figure 2.17b are highlighted in red on the structure of Est3 $\Delta$ N, with maximum number of protected peaks mapping to the core OB-fold  $\beta$ -barrel.

### **2.3.5f Comparison of predicted secondary and tertiary elements of Est3 to the calculated Rosetta Structure reveals robustness of secondary element prediction**

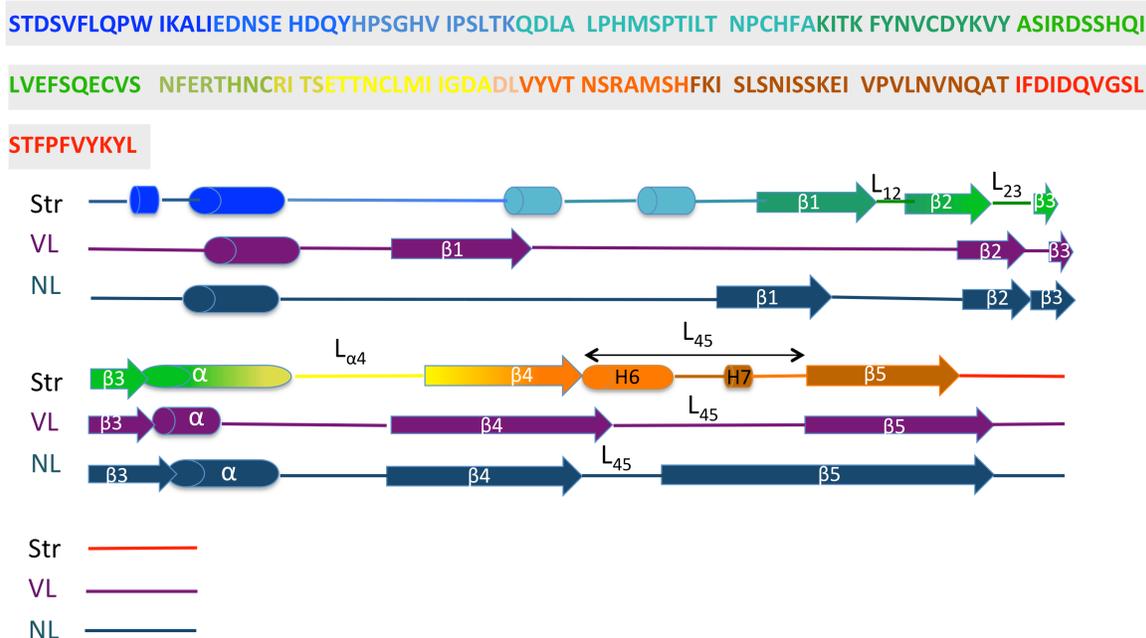
Predicted secondary and tertiary structures were compared to the calculated structure of Est3 $\Delta$ N, to assess the robustness of sequence-based structure predictions. Secondary structure elements from the calculated RASREC Rosetta structure was compared to that from PREDITOR (based on backbone chemical shift experimental data), sequence based prediction from PSIPred and from 3D models of Est3 predicted in Lundblad (VL\_model) [95] and Lue (NL\_model) [96] labs (Figure 2.32). The blocked regions in the predicted models highlight the regions that are different between the calculated and the predicted structures.



**Figure 2.32 Comparison of Est3 secondary structure predictions from various sources**

Secondary structure elements (helix, beta strand and loop) of Est3 from Est3<sup>AN</sup> structure are compared to the secondary structure predicted by PREDITOR (using backbone experimental data) [252], PSIpred (sequence based prediction) [193], LOMETS Model (sequence based 3D modeling) [199], 3D model from published from Lundblad lab [95] and 3D model published from Lue lab [96]. Some regions of the predicted elements that disagree with the calculated structure are boxed in red.

This is also shown in another schematic representation in the Figure 2.33 (see below).

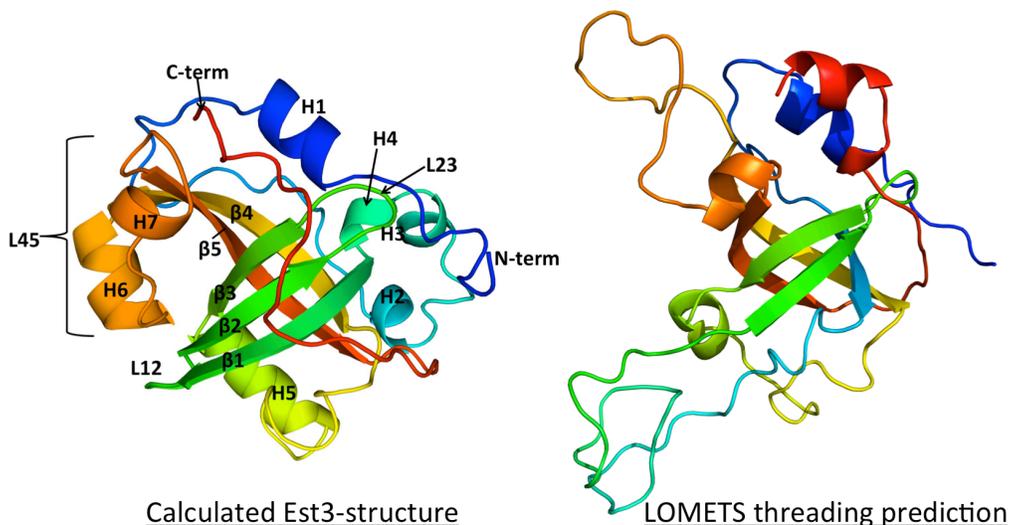


**Figure 2.33 Comparison of Est3 calculated structure with the predicted 3D models in literature**

Comparison of secondary structure from predicted 3D models from two previously published studies to the secondary structure from the RASREC-Rosetta structure calculated in this study (Str). VL in the figure indicates the predicted model from the Lundblad Lab [95] and NL is the predicted model from the Lue Lab [96].

In conclusion, it can be seen that sequence based secondary structure prediction PSIPred is robust in predicting the secondary structure elements. Predicted 3D models are conspicuously missing the loop L45, with NL\_model having a really long  $\beta$ -strand  $\beta_4$  in its stead. VL\_model has especially long loop L12, not present in the actual structure. The discrepancies in the models are likely because these are based on threading the primary sequence of the target protein (Est3) over the nearest putative structural homolog (*HsTPP1*) and threading misses new or different features from the template that the target protein adopts. This point is

again highlighted upon comparison of the calculated structure to the 3D-model predicted by the LOMETS meta-threading server [199] (Figure 2.34).



**Figure 2.34 Comparison of Est3 calculated structure to the LOMETS threaded 3D model**

RASREC Rosetta calculated structure of Est3<sup>AN</sup> (left) is compared to the sequence-based 3D modeling by LOMETS server [199] (right). Color code is blue to red (N- to C-terminus) and secondary structure elements are as labeled on the calculated structure on left. The predicted model differs in placement and prediction of strand  $\beta 1$  and folding of the helices in the L45 region. L12 also differs in being extremely long and unstructured, different from the small L12 of the calculated structure. Finally, the C-terminal tail (red) crosses over the  $\beta$ -strands 1,2 and this feature is similar to the calculated structure.

## 2.4 Discussion

### 2.4.1 Est3 structure reveals novel features previously unidentified in other OB-fold proteins

We have solved the solution NMR structure of Est3, providing the first structure of a yeast telomerase protein. Est3 is a difficult protein target for which conventional structure determination strategies of x-ray crystallography and NMR were not feasible. The structure was solved using a relatively new strategy of

combining minimal NMR experimental data (chemical shifts, RDCs and NOEs) with Rosetta *de novo* structure prediction [220,223,243]. Recent improvements in the Rosetta program due to the resolution-adapted structural recombination (RASREC) made the structure determination of this 170 amino acid Est3<sup>ΔN</sup> protein feasible, where previously the upper size-limit with experimental-data combined Rosetta was only 150-residues [220].

The core structure of Est3 was revealed to be an OB-fold with several features that distinguish it from the canonical OB-fold. The presence of the core OB-fold in Est3 had also been predicted by previous 3D-modeling predictions [95,96], however, obvious differences in the experimentally solved structure and 3D-models are readily apparent. While the secondary structures agree well, the modeling programs appear to have difficulty placing those secondary structures in a three-dimensional space. This problem is compounded when the protein of interest has very low sequence similarity (< 30 %) to previously-solved structures, as is the case for Est3 (15 % identity with the closest structural homolog in the PDB, *Hs*TPP1-OB) [200]. The modeling algorithm drives the model towards the identified top structurally similar hit(s) and in cases of low-sequence similarity between the target and template protein, the algorithm results in low accuracy. This is evidenced from the comparison of secondary structure placement from predicted models of Est3[95,96] to the experimentally calculated structure in the present work (Figure 2.32, Figure 2.33). For example, 3D-modeling predicted a long loop L12 for Est3 as the algorithms were attempting to accommodate the long L12 found

in *Hs*TPP1-OB. In addition, the predictions failed to identify the long structured L45, which is not present in *Hs*TPP1-OB. In our RASREC Rosetta strategy, however, the inclusion of even the minimal experimental-data, allowed for structure calculation independent of a sequence-similarity-based template. This allowed the identification of novel and unique features of Est3.

The high-resolution structure of Est3 allows the identification of regions on the protein surface important for interaction. Mutagenesis studies *in vivo* and biochemical studies *in vitro* have shown that Est3 interacts directly with the N-terminal domain of the yeast telomerase catalytic subunit Est2 [95,96,189,190]. The Est3 residues involved in telomerase association are all located at the base of the OB-fold and form a contiguous surface that is consistent with a protein-protein interaction site [95,96]. Interestingly, some of the residues identified for Est3-Est2 association form a highly conserved negatively charged patch (Figure 2.27), at the beginning of the C-terminal tail, which is highly flexible (Figure 2.19b). A flexible C-terminal tail is observed in several other single-stranded oligonucleotide binding proteins, where highly acidic and flexible C-terminal tails of the proteins are involved in functionally important protein-protein interactions [253-255]. Flexible tails can adopt a different conformations upon ligand binding and this allows the underlying binding surface, otherwise hidden by the tail, to become better accessible to binding partner(s) [256]. We hypothesize that the acidic C-terminal tail of Est3 directly interacts with Est2 and induces a conformational change in Est3 that moves both the C-terminal tail and the L45 away from the canonical ligand-

binding surface [257]. This “open conformation”, induced by Est2 binding, will then allow access to an as-yet-unidentified ligand, likely ssDNA, to bind. However, we cannot rule out the possibility that the L45 might be a fixture that occupies the binding site permanently.

Est3 also has an unusually long, spiral-shaped, N-terminal region. The flexibility at the very beginning of the N-terminus might be important for function of Est3 as indicated by *in vivo* mutational analysis of the K3 and L6 residues in the Lundblad lab (unpublished data). The functional relevance, if any, of the other flexible regions of the Est3 protein remains to be determined.

Structural homology can be a good indicator of some functional homology. The DALI server identified *Hs*TPP1-OB as the protein most structurally similar to Est3 from the previously solved structures in the PDB (Table 2.4 and Figure 2.30). *Hs*TPP1-OB is a central domain of the multi-domain protein *Hs*TPP1, which binds to the ssDNA binding protein hPOT1 in humans and enhances telomerase processivity *in vitro* [90,94]. *Hs*TPP1-OB is particularly involved in telomerase recruitment to telomeres, perhaps through a direct interaction with *Hs*TERT [73,94]. This reflects a common feature with yeast Est3, which is part of the yeast telomerase holoenzyme and directly interacts with *Hs*TERT homolog Est2 [73,190]. Interestingly, recent data has identified the key residues for *Hs*TPP1-OB binding to *Hs*TERT (Cech Lab, unpublished) and they lie on exactly the same surface of *Hs*TPP1-OB as the telomerase interacting residues on Est3 structure. This is another line of evidence in support of the yeast Est3 and human TPP1-OB proteins

being structural and functional homologs. The second hit from the DALI search was *SnTEBP* $\beta$ , the ciliate homolog of *HsTPP1*-OB, which is also critical in recruitment of telomerase to telomeres [48]. However, *SnTEBP* $\beta$  contains a unique C-terminus structure that enables fundamentally different end-protection and recruitment mechanisms as a heterodimer with *SnTEBP* $\alpha$  [48]. Additionally, although the OB-folds of *HsTPP1* and *SnTEBP* $\beta$  have evolved as part of proteins that perform distinct functions in telomere length maintenance, both proteins are implicated in positive and negative regulation of telomerase recruitment. We know very little about the function of Est3 in yeast beyond that it is essential for telomerase regulation *in vivo* [177]. It is possible that Est3 mirrors the roles of *HsTPP1* and *SnTEBP* $\beta$  in telomerase recruitment and/or processivity. However, it is important to note that the topology of Est3 is distinct from *HsTPP1* and *SnTEBP* $\beta$  in that it is comprised of a single OB-fold, with no other domains. Thus, it is not clear if Est3 is capable of multiple functions, like the multi-domain *HsTPP1* and *SnTEBP* $\beta$ .

#### **2.4.1a An extended Loop 45 (L45) blocks the canonical binding surface of the OB - fold in Est3**

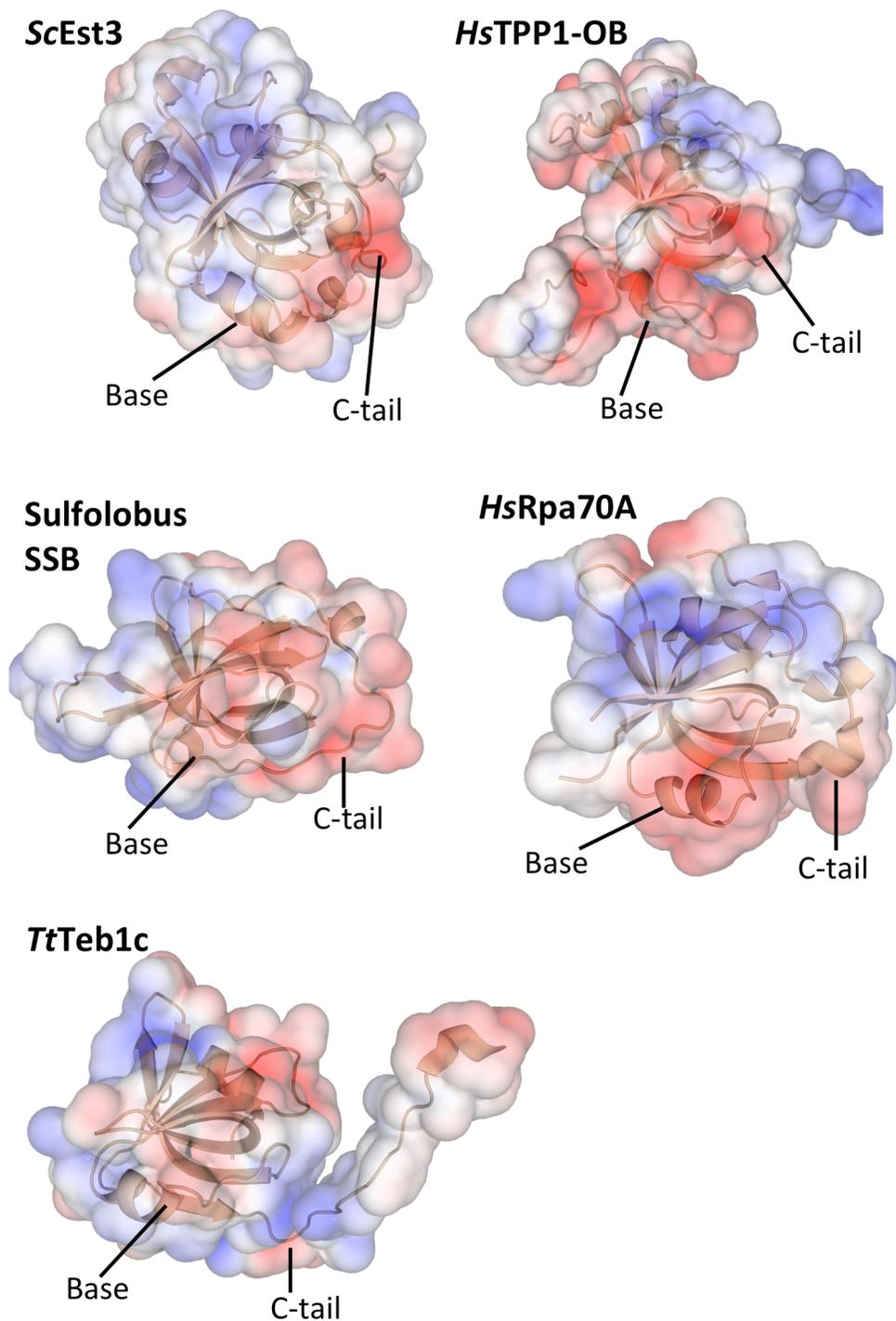
One key and unexpected feature identified from Est3 structure is the long, structured, L45. Except for F139, the sequence comprising L45 is variable (Figures 2.25 and 2.26), indicating that this loop might not be important in specific binding interactions of Est3. The canonical ligand-binding surface of OB-fold proteins is defined by L45 from above and L12 from below (orientation as in Figure 2.26) [202,253]. However, in Est3, L45 sits over and occludes the canonical ligand-binding

surface. This L45 might be a structural fixture in the Est3 proteins, acting as the “ligand” that occupies the canonical binding site, with the conserved F139 as the likely anchoring point. Consistent with that hypothesis, we have not observed flexibility in this long L45 either by  $^{15}\text{N}$ -HSQC peak intensity comparison or by HetNOE (Figure 2.19). Alternatively, L45 could be a functional switch that is displaced only in the presence of an as-yet-unidentified *in vivo* factor that does not require recognition of conserved residues on L45, and/or works through indirect displacement of L45 by conformational changes elsewhere on the protein. The displacement of the L45 might then allow better access of a ligand with the binding surface. This hypothesis is supported by the absence of nucleic acid binding activity *in vitro*, but the observation of ssDNA binding activity by Est2<sup>TEN</sup>-Est3 complex for *C. parapsilosis* Est3 by UV cross-linking [183,191].

#### **2.4.1b C-terminal of the Est3 partially sits on the canonical ligand-binding site and in conjunction with the structured L45 might be regulated to allow or block the access of Est3 ligands**

The high-resolution structure of Est3 allows the identification of regions on the protein surface important for interaction. Mutagenesis studies *in vivo* and biochemical studies *in vitro* suggest that Est3 interacts directly with the TEN domain of the yeast telomerase catalytic subunit Est2 [95,96,189,190]. The Est3 residues involved in telomerase association are all located at the bottom of the OB-fold (between strands  $\beta 3$  and  $\beta 4$ ) and form a contiguous surface that is consistent with a protein-protein interaction site [95,96]. Some of the residues identified for

Est3-Est2 association form a highly conserved negatively charged patch (Figure 2.27) on the C-terminal tail, beginning and middle portion of which is highly flexible, but the last 7-8 residues are fixed (Figure 2.19). This C-tail also sits over the  $\beta$ -sheet formed by strands  $\beta$ 1,  $\beta$ 2, and  $\beta$ 3 might be blocking ligand access to the canonical ligand-binding surface. This is similar to the observations in single-stranded oligonucleotide binding proteins (SSBs), where highly acidic and flexible C-terminal tails can adopt a different conformation upon ligand binding and this allows the underlying binding surface, otherwise hidden by the tail, to become better accessible to other binding partner(s) [253-256] (Figure 2.35). A similarly placed, acidic and flexible, C-tail is also present in the OB-folds of *Tt*Teb1 and *Hs*Rpa70A (Figure 2.35), but the functional significance of these acidic tails is as yet unknown. Interestingly, Est3 structural homolog, human TPP1-OB, has been shown to also interact with telomerase through TPP1-OB's acidic TEL (TPP1 glutamate and leucine rich) [258] patch that spans the base of its OB-fold and the beginning of the C-terminal tail, flexibility of which is not established. We hypothesize that the acidic part of the C-terminal tail of Est3 directly interacts with Est2 and induces a conformational change in Est3 that moves the beginning and middle portions of C-terminal tail, and maybe indirectly the L45, away from the canonical ligand-binding surface [257]. This "open conformation", induced by Est2 binding, will then allow access to an as-yet-unidentified ligand, likely ssDNA, to bind.



**Figure 2.35 Comparison of acidic patches on the base and C-tail of ScEst3 and homologous proteins**

ScEst3 is negatively charged at the base of the  $\beta$ -barrel and the beginning of the C-tail (labeled). Interestingly, similar pattern of the acidic patch location is found in *HsTPP1*, *Sulfolobus* SSB, *HsRpa70A* and *TtTeb1c*. All structures are oriented the same as Figure 2.27 (bottom left panel).

## Chapter 3. Structure-function analysis of Est3: *in vitro* and *in vivo* studies

### 3.1 Introduction

Est3 is the smallest subunit of the yeast telomerase holoenzyme that has been shown to be dispensable for telomerase activity *in vitro* [120]. *In vivo*, an *est3* $\Delta$  strain exhibits senescence and telomere length defect phenotypes that are identical to those observed for deletion strains of the Est2 reverse transcriptase and TLC1 RNA component as well as the Est1 regulatory subunit [57]. This indicates that Est3 protein is essential, but the precise regulatory role of Est3 remains unclear.

Several studies have employed both genetic and biochemical approaches to elucidate the role of this protein with conflicting results. One mode by which Est3 exerts its regulatory role has been proposed to be through a direct interaction with Est2. Supporting this interaction, several studies have reported co-immunoprecipitation (co-IP) of Est3 and Est2 from budding yeast extracts, with tagged endogenous proteins, and also identified mutants of Est3 that abrogate this association [95,96,177,189]. However, co-IP from cell extracts does not confirm a direct interaction of Est2 and Est3. Genetic studies have shown that Est3's association with telomerase does not depend on Est1 [183] and TLC1, the other two known subunits of the telomerase holoenzyme, therefore implicating that the likely binding partner is Est2. A domain in Est2 has been proposed for interaction with Est3, based on suppression screen of temperature sensitive mutants in the essential

N-terminal TEN domain of Est2 (Est2<sup>TEN</sup>) [189]. However, again, this suppression data is also not sufficient in establishing direct interaction between Est3 and Est2<sup>TEN</sup>. Based on ScEst3 mutants reported to affect its telomerase association [95], a direct binding of Est3 and Est2<sup>TEN</sup> domains was tested for recombinantly expressed and purified wild type and mutant proteins [190]. However, the proteins used in that study were impure; co-purification of contaminating proteins, likely chaperones, suggest that the proteins were either misfolded or otherwise unstable. Therefore, that report of a “direct” interaction between Est3 and the TEN domain is somewhat inconclusive. In addition, several observations from the affinity pull-down conflicted the observations from the *in vivo* co-IPs (see section 1.5.1 for details). Overall, the field has convincing data that Est3 directly associates with Est2, but the interaction interface remains to be convincingly identified.

In addition to Est2-Est3 protein-protein association, Est3 has been proposed to bind ssDNA and that association with Est2<sup>TEN</sup> is required to reveal Est3's DNA binding activity [191]. However, this study was performed using the putative Est3 from *C. parapsilosis*; ssDNA-binding activity has not been shown for known Est3 proteins from other species. In fact, ssDNA binding efforts by us and others [96,183] using ScEst3 and CaEst3 proteins have failed to show binding. Therefore, it remains to be seen if the canonical Est3s have the capability to bind ssDNA and if that is the mode through which Est3 exerts its function. The structure of Est3 (Chapter 2) shows that it is an OB-fold protein and therefore like many proteins in the OB-fold family, it is expected to present binding surfaces for ligand access. A

common theme for OB-fold proteins is oligonucleotide binding, although not all OB-fold proteins bind oligonucleotides. Presence of positively charged patch (Figure 2.28, 2.29) on the canonical OB-fold surface of Est3, however, suggests that Est3 might interact with DNA, however, the groove presenting that positive patch is not as extensive as reported for many ssDNA binding OB-fold proteins and therefore it remains to be seen if Est3 is a DNA binding protein. Additionally, Est3 structure (this thesis) has highlighted the presence of long structured L45 that potentially sits in the canonical binding site of OB-fold and therefore might be blocking access to a ligand, like ssDNA. The removal of L45 by a binding partner *in vivo* might be required to reveal Est3's binding face. Therefore, removal of L45 and revisiting DNA binding was one of the goals in this chapter.

Est3 has been reported by some studies to be a functional dimer [259,260]. This was mostly based on the partitioning of Est3 into monomer and dimer fractions upon purification by size exclusion chromatography and whether it is a functional dimer remains inconclusive. Some of the previous studies have used GST-tagged Est3 and the propensity of GST to dimerize might have biased the results for Est3 partitioning. Here I present my observations about untagged Est3 oligomerization state and some interpretations based on that observation.

One function of Est3 has been proposed to be activity as an enhancer of telomerase activity. This is supported by its structural and functional homology to human TPP1-OB protein one of its roles is as a processivity factor for human telomerase. Est3 might enhance baseline telomerase activity in yeasts and this was

shown to be the case for *C. albicans* and *S. castellii* telomerases, where the absence of Est3 causes a reduction in nucleotide addition activity *in vitro* [183,184]. However, this effect has yet to be conclusively observed in *S. cerevisiae* system, in which many of the primary genetics studies were performed. An attractive *in vitro* telomerase activity assay was designed by Dr. David Zappulla (in the Cech lab). He was able to reduce the bulky size of TLC1 RNA from ~1200 nt to ~500 nt (miniT) and subsequently co-express ProA-ScEst2 and miniT from the RRL system. The purified ProA-ScEst2-TLC1 complex showed telomerase activity *in vitro* as it was able to extend telomeric substrates. This is a useful tool for testing activity enhancement, if any, in the presence of other purified factors. We have collaborated with Dr. Zappulla (presently at the Johns Hopkins University), to test effect of purified ScEst3 and its mutants on telomerase activity *in vitro*.

Finally, in collaboration with Dr. Vicki Lundblad (Salk Institute), we have undertaken an integrated *in vivo* and *in vitro* analysis approach to characterize some mutants of Est3 identified from the high resolution structure of Est3 (this thesis) to test their effect on Est3's role in telomere length maintenance. This was an intensive and unbiased approach to analyze the complete surface of Est3 protein, leading to insights about previously unidentified residues that are important for Est3's role in telomere length maintenance.

## 3.2 Materials and Methods

### 3.2.1 Est3 expression and purification

His<sub>10</sub>-SUMO-Est3 proteins (wild-type *S. cerevisiae* and *S. castellii* Est3 coding sequence) were expressed in *E. coli* BL21(DE3) cells. The cells were grown in Luria Bertani broth at 37°C to an O.D.<sub>600</sub> of ~1.0 and then cold-shocked on ice for ~1h. Protein expression was induced with 1.0 mM IPTG. Cells were allowed to grow post-induction for 24 hrs at 15°C and harvested by centrifugation. Cell pellets were resuspended in lysis buffer (see end of this section for list of protein-specific lysis buffers) with an EDTA-free protease inhibitor cocktail tablet (Roche) and PMSF was added just before cell lysis by sonication followed by cell debris separation by centrifugation at 25,000xg for 20 minutes. Cells were lysed by sonication and clarified by centrifugation. The clarified cell lysate was subjected to Ni<sup>2+</sup>-affinity chromatography by gravity-flow (GE Healthcare). The His<sub>10</sub>-SUMO-Est3 proteins were subsequently eluted with buffer containing 300 mM imidazole and concentrated in a 10,000 MWCO concentrator (Millipore) to 2 mL, without buffer exchange, and incubated overnight at 4°C with Ulp1 to cleave the His<sub>10</sub>-SUMO tag [204]. Cleaved monomeric proteins were purified via size-exclusion chromatography (Superdex75, GE) on an AKTA FPLC system in buffer B (100 mM potassium phosphate buffer (pH 7.5), 100 mM Na<sub>2</sub>SO<sub>4</sub>, 5% glycerol and 3 mM βME). Contaminating ~14 kDa His<sub>10</sub>-SUMO tag was separated from the ~20.6 kDa Est3 protein by a second round of Ni<sup>2+</sup>-affinity chromatography in buffer B containing 20 mM imidazole. Est3 eluted at >95% purity and was concentrated and buffer

exchanged to the binding buffer (75 mM Tris-Cl buffer (pH 8.0), 100 mM NaCl and 2 mM DTT) by dialysis.

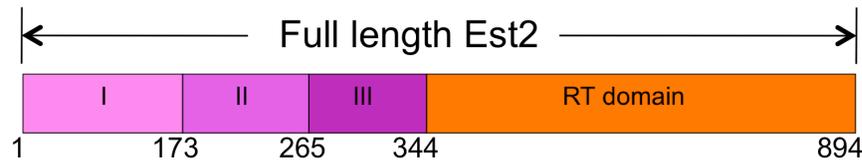
Lysis Buffers:

*S. cerevisiae* Est3: 100 mM potassium phosphate buffer (pH 7.4), 100 mM Na<sub>2</sub>SO<sub>4</sub>, 10 mM imidazole, 10% glycerol and 3 mM βME

*S. castellii* Est3: 100 mM Tris-Cl (pH 8.8), 150 mM NaCl, 10 mM imidazole, 10% glycerol and 3 mM βME

### **3.2.2 Expression and purification of Est2 N-terminal domain constructs**

To test direct binding Est3 and Est2 N-terminal domain, based on earlier reports suggesting binding, Est2 domains were identified using a multiple alignment of Est2 sequences from eight yeast species, secondary structure prediction by PROMALS [188] and the previously identified TEN domain boundary from crystal-structure of *Tetrahymena* TEN [129]. Coding sequences were cloned into both the pET28b Smt3 vector (for His<sub>10</sub>-SUMO fusion) and the pGEX-6P vector (for GST fusion). The cloned *S. cerevisiae* Est2 sequences coded for residues 1-173 for TEN domain and residues 1-344 (Figure 3.1) for whole of the N-terminal domain upstream of the RT domain in the C-terminal half (Figure 1.7). For *S. castellii*, the cloned sequence spanned from residues 1-176 for TEN domain and residues 1-368 for whole of the N-terminal domain upstream of the RT domain. Primers used for cloning the various constructs of Est2 in relevant vectors are reported in Table 3.1.



**Figure 3.1 Schematic of the *S. cerevisiae* Est2 domains**

Boundaries for the different domains of the budding yeast Est2, adapted from [261]. Region I corresponds to the TEN domain (1-173) [129] and was relatively the most soluble domain when recombinantly expressed in *E. coli*.

**Table 3. 1 Primers for cloning *S. cerevisiae* and *S. castellii* Est2 domains**

Table 3.1: Primers for cloning <i>S. cerevisiae</i> and <i>S. castellii</i> Est2 domains				
Organism	Domain	Vector	Forward primer	Reverse primer
<i>S. cerevisiae</i>	Est2 <sup>173</sup>	pET28b-Smt3	cggatccatgaaaatcttattcgagttc attcaag	cctcgagtcacgcagcagttgcggatgatg
	Est2 <sup>173</sup>	pGEX-6P	cggatccatgaaaatcttattcgagttc attcaag	cctcgagtcacgcagcagttgcggatgatg
	Est2 <sup>344</sup>	pET28b-Smt3	cggatccatgaaaatcttattcgagttc attcaagac	cctcgagtcaaatgaacaaccaccgaaaatcc tt
	Est2 <sup>344</sup>	pGEX-6P	cggatccatgaaaatcttattcgagttc attcaagac	cctcgagtcaaatgaacaaccaccgaaaatcc tt
<i>S. castellii</i>	Est2 <sup>176</sup>	pET28b-Smt3	cgggatccatgccgagagttatatttc ctcg	atttcggccgcctacaaggtgaataaac
	Est2 <sup>176</sup>	pGEX-6P	cgggatccatgaaccacttaatgact atcttc	ccctcgagtcaggtattcacagttatgtagaa ttc
	Est2 <sup>368</sup>	pET28b-Smt3	----not cloned----	----not cloned----
	Est2 <sup>368</sup>	pGEX-6P	cgggatccatgaaccacttaatgact atcttc	ccctcgagtca cgataagatttcggtatcaagattg

All primers reported in 5' to 3' direction.  
Primers were ordered through Integrated DNA technologies (idtdna.com)

All Est2 constructs were expressed in *E. coli* BL21(DE3) cells by transforming the cells with the appropriate plasmid and cells grown in Luria

Bertani broth at 37°C to an O.D.<sub>600</sub> of ~1.0 and then cold-shocked on ice for ~1h. Protein expression was induced with 1.0 mM IPTG. Cells were allowed to grow post-induction for 16-18 hrs at 15°C for SUMO vector based expression and for 6-8 hrs at 15°C for pGEX vector based expression and cells were harvested by centrifugation.

Protein purification for His<sub>10</sub>-SUMO-fused proteins was done as follows. Cell pellets for *S. cerevisiae* Est2 were resuspended in lysis buffer (see end of section for protein specific lysis buffers) with an EDTA-free protease inhibitor cocktail tablet (Roche) and PMSF was added just before cell lysis by sonication followed by cell debris separation by centrifugation at 25,000xg for 20 minutes. Cells were lysed by sonication and clarified by centrifugation. The clarified cell lysate was subjected to Ni<sup>2+</sup>-affinity chromatography by gravity-flow (GE Healthcare). The His<sub>10</sub>-SUMO-Est2 proteins were subsequently eluted with lysis buffer containing 300 mM imidazole and concentrated in a 10,000 MWCO concentrator (Millipore) and buffer exchanged to the binding buffer (75 mM Tris-Cl buffer (pH 8.0), 100 mM NaCl and 2 mM DTT) by dialysis.

Protein purification for GST-tagged proteins was done as follows. Cell pellets were resuspended in lysis buffer (see end of section for protein-specific lysis buffers). After lysis, by sonication, the soluble lysate was incubated with 1 mL of packed Glutathione Sepharose 4B (GE healthcare) beads for 2 hrs at 4°C with intermittent mixing. Beads were spun-down at 500 g for 5 minutes in the tabletop centrifuge. This was followed by four washes with 10 bed volumes of lysis buffer each. For elution, washed beads were incubated with seven bed-volumes of lysis buffer

containing 10 mM glutathione for 3 hrs at 4°C. Beads were spun-down and supernatant collected. Another two elutions were done by incubating beads with 10 mM glutathione containing lysis buffer for 2 hrs each and then collecting supernatant. Fractions from each collection were run on SDS-PAGE gel and the purest fraction chosen for binding experiment. This protein was buffer exchanged to the binding buffer (75 mM Tris-Cl buffer (pH 8.0), 100 mM NaCl and 2 mM DTT) by dialysis.

In addition to the Est2 domains, a set of small peptides were designed to include the TEN domain K111 residue that was initially proposed to be important in Est2-Est3 association (Lundblad lab, unpublished data). Different length peptides were designed around this K111 residue. The 50-residue peptide that was ultimately used for binding studies was cloned into the pET28b Smt3 plasmid between BamHI and NotI restriction sites, using the 5'-d(cgggatccggcgaattatataacaacgtactaac)-3' and 5'-d(at ttg cgg ccg ccta accgaccaa actgtggaacattttc)-3' forward and reverse primers respectively. The His<sub>10</sub>-SUMO-peptide was purified like the SUMO-fused Est2 domain in the previous paragraph.

Lysis Buffers:

SUMO-fusion proteins

*S. cerevisiae* Est2 domains: 50 mM Tris-Cl (pH 8.0), 100 mM NaCl, 10 mM imidazole, 10% glycerol and 3 mM βME

*S. castellii* Est2 domains: 100 mM Tris-Cl (pH 8.3), 150 mM NaCl, 10 mM imidazole, 10% glycerol and 3 mM  $\beta$ ME

GST-fusion proteins

*S. cerevisiae* Est2 domains: 50 mM Tris-Cl (pH 8.0), 100 mM NaCl, 10 mM imidazole, 10% glycerol and 3 mM  $\beta$ ME

*S. castellii* Est3: 100 mM Tris-Cl (pH 8.3) , 150 mM NaCl, 10 mM imidazole, 10% glycerol and 3 mM  $\beta$ ME

### **3.2.3 Construct design and cloning for a stable L45 deletion mutant of *S. cerevisiae* and *S. castellii* Est3 for DNA binding studies**

Site-directed mutagenesis was used to delete loop L45 (residues 132-150 in *S. cerevisiae* Est3 (*ScEst3*)) from the pET28b Smt3 plasmid containing *ScEst3* sequence. Forward and reverse primers for mutagenesis were 5'-d(ctgacttagtctacgtaacaatagtgcccgtttctcaatg)-3' and 5'-d(cattgagaacgggcactattgttacgtagactaagtcag)-3' respectively. Three different sized linkers were cloned-in by site-directed mutagenesis to replace the deleted loop. The linker sequences introduced were a 8-residue linker SAEVFTSS, 5-residue linker SGFGS and 4-residue linker SGGS. K71E mutation was built into the wild-type SUMO-Est3 and 4-residue linker containing SUMO-Est3 using the 5'-d(ccgtgccatttcgccc aaaattacagagttttataacgtttgcgactacaag)-3' forward and 5'-

d(cttgtagtcgcaaacgttataaaaactctgtaatgttggcgaaatggcacgg)-3' reverse primers. All plasmid constructs were confirmed by sequencing.

The plasmids were transformed into bacterial expression cell line *E. coli* BL21(DE3) and expressed and purified as follows. The cells were grown in Luria Bertani broth at 37°C to an O.D.<sub>600</sub> of ~1.0 and then cold-shocked on ice for ~1h. Protein expression was induced with 1.0 mM IPTG. Cells were allowed to grow post-induction for 24 hrs at 15°C and harvested by centrifugation. The cell pellet was resuspended in lysis buffer (buffer A) containing 100 mM potassium phosphate buffer (pH 7.5), 100 mM Na<sub>2</sub>SO<sub>4</sub>, 10% glycerol, 10 mM imidazole and 3 mM βME and an EDTA-free protease inhibitor cocktail tablet (Roche). Cells were lysed by sonication and clarified by centrifugation. The clarified cell lysate was subjected to Ni<sup>2+</sup>-affinity chromatography by gravity-flow (GE Healthcare). The His<sub>10</sub>-SUMO-Est3 was eluted with buffer A containing 300 mM imidazole and concentrated in a 10,000 MWCO concentrator (Millipore) to 2 mL volume and incubated overnight at 4°C with SUMO specific protease, Ulp1, to cleave off the His<sub>10</sub>-SUMO tag [204]. Subsequently, the cleavage reaction was subjected to size-exclusion chromatography (Superdex75, GE) on an AKTA FPLC system in buffer B (100 mM potassium phosphate buffer (pH 7.5), 100 mM Na<sub>2</sub>SO<sub>4</sub>, 5% glycerol and 3 mM βME). Est3 eluted as dimer and monomer fractions that also contained His<sub>10</sub>-SUMO. The ~14 kDa His<sub>10</sub>-SUMO tag was separated from the ~20.6 kDa Est3 protein by a second round of Ni<sup>2+</sup>-affinity chromatography in buffer B containing 20 mM imidazole. Est3 eluted at >95% purity and was concentrated and buffer exchanged to the

experimentally appropriate buffer, using a 9000 MWCO protein concentrator (Pierce). Protein concentration was estimated by measuring absorbance of the sample at 280 nm in a nano-drop spectrophotometer.

Site-directed mutagenesis was used to delete loop L45 (residues 134-154 in *S. castellii* Est3 (*ScasEst3*)) from the pET28b Smt3 plasmid containing *ScasEst3* sequence. Forward and reverse primers for mutagenesis were 5'-d(gatgctgctatcatctacaagtcattagtagcaatcttacagattaa)-3' and 5'-d(gttaatctgtaaaattggactaatgactttagatgatagcagcatc)-3' respectively. The four-residue linker SGGG was subsequently incorporated using site-directed mutagenesis to replace the deleted loop. The K73E mutation was incorporated into the wild-type SUMO-*ScasEst3* and four-residue linker containing SUMO-*ScasEst3* using the 5'-d(cactgcatctttactaaggttaactgagttttcaaaataaataactatgcatctc)-3' forward and 5'-d(gagatggcatagttattttttgaaaaactcagttaccttagtaaaatggcagtg)-3' reverse primers. All plasmid constructs were confirmed by sequencing. Protein expression and purification was done as for the *S. cerevisiae* Est3 proteins described in the previous paragraph.

### 3.2.4 Electrophoretic mobility shift assays

*S. cerevisiae* Est3 wild-type and  $\Delta L45^{SGGS}$  were expressed and purified as described in sections 3.2.1 and 3.2.3. *S. castellii* Est3 wild-type and mutants were expressed as described in sections 3.2.2 and 3.2.3. All DNA oligos were ordered from Integrated DNA Technologies (IDT) with standard desalting and resuspended in nuclease-free sterile water to form stock solutions. DNA was end-labeled with P<sup>32</sup>

for use in mobility shift assays, by setting up the labeling reaction in low-binding 1.5 mL eppendorfs as follows. Radioactive  $^{32}\text{P}$ -( $\gamma$ -ATP) vial was thawed at  $37^\circ\text{C}$  for ~30 minutes. DNA stocks were diluted to  $25\ \mu\text{M}$  concentration and the  $10\ \mu\text{L}$  labeling reaction was set-up as follows:

1  $\mu\text{L}$  DNA ( $25\ \mu\text{M}$ )

1  $\mu\text{L}$  10x Forward Reaction Buffer (Invitrogen)

1  $\mu\text{L}$  T4 Polynucleotide Kinase (Invitrogen)

6  $\mu\text{L}$  nuclease-free sterile water

1  $\mu\text{L}$   $^{32}\text{P}$ -( $\gamma$ -ATP)

Reaction was incubated  $37^\circ\text{C}$  for 30 minutes and subsequently incubated at  $95^\circ\text{C}$  for 20 minutes and then diluted to  $50\ \mu\text{L}$  by adding  $40\ \mu\text{L}$  of sterile water. The reaction tube was then transferred to ice for 10-20 minutes. Micro-spin G25 column (GE Healthcare) was used for purification of the oligo by removal of storage G25 storage buffer and subsequent application of the  $50\ \mu\text{L}$  labeling reaction to the center of the column and collection of purified oligos by centrifugation. Based on previous observation in the lab 80% recovery was assumed and therefore the labeling reaction gave a stock of  $400\ \text{nM}$  labeled DNA.

6.7% native mini-gel of 1.5 mm thickness for separating the binding reaction was prepared as follows:

1.67 mL Acrylamide (40%)

2.00 mL 5X TBE (5X TBE: 450 mM Tris-base, 450 mM boric acid, 10 mM EDTA)

0.50 mL 100% glycerol

5.83 mL sterile water

50  $\mu$ L 10% ammonium per-sulfate

30  $\mu$ L TEMED

Binding buffer (BB) for diluting protein and DNA for setting up binding reactions was 50 mM Tris pH 8.4, 75 mM KCl, 1 mg/ mL BSA, 15% glycerol and 1 mM DTT. Protein dilutions were made at 2x the final concentrations in low-binding eppendorfs. Labeled DNA was diluted to 2x the final concentration (20 pM or 40 pM). 15  $\mu$ L each of protein and DNA 2x dilutions were mixed and incubated on ice for 30 minutes. The gels were set-up in the Bio-Rad mini-gel electrophoresis apparatus and filled with 1XTBE (+5% glycerol). In the 4°C room, 25  $\mu$ L each of the binding reactions were loaded in the wells of the gel, using gel-loading tips. Thin ice packs were immersed into the running buffer and gels were run at 200 V for 20 minutes. Gels were transferred to Whatman filter paper and dried on a gel-dryer. Dried gels were exposed to the phosphorimager screen for 40-60 hours, and then scanned using a Typhoon scanner (GE Healthcare). The data were quantitated using ImageQuant 5.1 (GE Healthcare) software.

### **3.2.5 Chemical shift perturbation analysis in the $^{15}\text{N}$ HSQC spectra for protein-DNA binding**

$^{15}\text{N}$ -labeled *S. cerevisiae* Est3 bearing an N-terminal His<sub>10</sub> -SUMO tag was expressed in *E. coli* BL21(DE3) cells in minimal M9 growth medium (3.5 g/L Na<sub>2</sub>HPO<sub>4</sub>, 3g/ L of KH<sub>2</sub>PO<sub>4</sub>, 1.5 g/L of NaCl and 1.5g/L of ( $^{15}\text{NH}_4$ )<sub>2</sub>SO<sub>4</sub> (Sigma/Isotec)

as sole source of nitrogen) supplemented with 1X MEM vitamin solution (Invitrogen, Carlsbad CA) and 1X Metal Mix (1  $\mu\text{M}$   $\text{FeCl}_3$ , 46 nM  $\text{H}_3\text{BO}_4$ , 102 nM  $\text{CaCl}_2$ , 0.19 nM  $\text{CoCl}_2$ , 0.78 nM  $\text{CuSO}_4$ , 1.02  $\mu\text{M}$   $\text{MgCl}_2$ , 1 nM  $\text{MnCl}_3$ , 3 pM  $\text{Na}_2\text{MoO}_4$ , and 1.7 nM  $\text{ZnCl}_2$ ) and 50 mg/ L kanamycin. 2 g/ L glucose was used as carbon source. Cells were grown at 37°C to an O.D.<sub>600</sub> of ~1.0 and then cold-shocked on ice for ~1h. Protein expression was induced with 1.0 mM IPTG. Cells were allowed to grow post-induction for 24 hrs at 15°C and harvested by centrifugation. Following lysis, the His-tagged protein was isolated with Ni-NTA agarose beads (Qiagen) and size-exclusion chromatography to yield final a sample concentration of 150-200  $\mu\text{M}$  (for purification details, see section 3.2.1). Final NMR sample buffer was buffer A: 50 mM potassium phosphate pH 7.3, 50 mM  $\text{Na}_2\text{SO}_4$ , 250 mM arginine/ glutamate, 2 mM DTT and 7%  $\text{D}_2\text{O}$ , OR B: 14 mM Bis-tris propane/50 mM MOPS pH 7.0, 10 mM  $\text{Na}_2\text{SO}_4$ , 75 mM arginine/ glutamate, 2 mM DTT and 7%  $\text{D}_2\text{O}$ . Samples were loaded into 5 mm Shigemi tubes (Shigemi, Inc). In addition,  $^{15}\text{N}$ -labeled *S. castellii* Est3 was similarly expressed and purified.

$^1\text{H}$ - $^{15}\text{N}$  TROSY-HSQC spectra were collected on Varian Inova 800-MHz nuclear magnetic resonance (NMR) spectrometer at the W.M. Keck High Field NMR facility at the University of Colorado, Boulder, using standard Varian pulse sequences with minor modifications. Overnight spectra were obtained for identical samples of  $^{15}\text{N}$ -labeled *S. cerevisiae* Est3 in the presence and absence of 10-40-fold molar excess of a 17-mer telomeric oligomer (5'-GTGTGGGTGTGGTGTGG-3') obtained from Integrated DNA Technologies and used without further purification. Lyophilized

DNA was resuspended in smallest possible volume (10-20  $\mu\text{L}$ ) of NMR sample buffer A and then added to the protein sample. For buffer B sample, the lyophilized DNA was directly resuspended in the protein NMR sample. All spectra were collected at 25°C equipped with a salt-tolerant HCN cryoprobe. NMR data were processed in NMRPipe [217] and spectra overlaid using CcpNmr analysis software [218].

### **3.2.6 Affinity pull-down assays**

Untagged-Est3 and His- or GST-tagged Est2 proteins for binding were expressed and purified as described in sections 3.2.1 and 3.2.2. Proteins were diluted to the 2x reaction concentration of 10  $\mu\text{M}$  each. Proteins were then either diluted 1:1 with the equilibration buffer alone or 1:1 with each other to get final protein concentration of 5  $\mu\text{M}$  each. Affinity pull-down assays were done either by binding on  $\text{Ni}^{2+}$ -affinity beads or on GST sepharose beads, depending on tagged-state of the Est2 domain.

For affinity pull-down by  $\text{Ni}^{2+}$ -affinity, 20  $\mu\text{L}$  of 50% slurry of His-Select High Flow Nickel Affinity Beads (Sigma) was transferred to three 1.5 mL eppendorfs. Beads were washed twice with 500  $\mu\text{L}$  water by centrifugation at 500 g to settle the beads. Beads were then equilibrated twice with 500  $\mu\text{L}$  with equilibration buffer containing 50 mM Tris pH 8.0, 100 mM NaCl, 3 mM  $\beta\text{ME}$ , 10 mM Imidazole, 0.05% BSA and 0.05% Igepal. Finally, the beads were loaded with 100  $\mu\text{L}$  of 5  $\mu\text{M}$  HIS-tagged protein only in the first eppendorf, 100  $\mu\text{L}$  of 5  $\mu\text{M}$  non HIS-tagged protein only (as negative control) in the second eppendorf, and 100  $\mu\text{L}$  of pre-incubated 5

$\mu\text{M}$ :1  $\mu\text{M}$  HIS-tagged:non HIS tagged protein. Tubes were incubated on ice for 30 minutes, with occasional flicking. Beads were spun-down at 500 g and supernatant solution on top pipetted out and saved as flow through (FT). Beads were washed thrice with 200  $\mu\text{L}$  of equilibration buffer. This was followed with a 200  $\mu\text{L}$  50-75 mM imidazole wash step and finally elution of the His-tagged and associated non His-tagged protein (if any) with 100  $\mu\text{L}$  of 200 – 300 mM Imidazole (5 min incubation each). All saved fractions (load, FT, washes, elutions) mixed with sample loading buffer and loaded onto SDS-PAGE gel for separation and visualization of bands by either silver-staining or coomassie blue staining.

To test binding between GST-tagged and untagged proteins, affinity pull-down was done on glutathione sepharose 4B beads (GE Healthcare). Beads were treated similarly as the His-Select beads above. The protein was loaded onto the beads (with no imidazole in the equilibration buffer). After incubation of loaded proteins with beads for ~30 minutes, the beads were centrifuged at 500 g and supernatant separated by pipetting and stored as flow-through (FT). Beads were washed thrice with 200  $\mu\text{L}$  equilibration buffer and finally two elution steps were with 100  $\mu\text{L}$  of 50 mM reduced glutathione (in equilibration buffer). Visualization protocol for binding was similar to His-tagged binding reactions.

Silver staining protocol for visualization of the protein bands was as follows. All steps were done in glass round bath. SDS-PAGE gel was fixed in 50% ethanol for 15-30 minutes. Gel was then transferred to staining solution for 15-30 minutes. This was followed by three washes with 200 mL each of deionized water. Developing

solution was then added and hand-shaken until desired intensity of the stain was reached. 100-200 mL of kill solution was immediately added and shaken for 1 h on the tabletop shaker.

Recipe for silver staining solutions:

**I. Staining solution:** make in glass cylinder

100 mL deionized water

1 mL 7.56% sodium hydroxide

1.5 mL 100% ammonium hydroxide

1 mL of 4.7 M silver nitrate

Covered with Parafilm and stirred until brown solution became clear. To clear the solution, added small aliquots of ammonium hydroxide.

**II. Development Solution:**

100 mL deionized water

250  $\mu$ L 2.5% citric acid

75  $\mu$ L of 37% formaldehyde

Covered with Parafilm and stirred until solution turned clear.

**III Kill Solution:**

45% methanol + 2% acetic acid

### **3.2.7 Multi-Angle Light Scattering (MALS) study for purified Est3 protein**

Recombinant, purified Est3 was subjected to tandem size-exclusion chromatography/multi-angle light scattering (SEC-MALS) to determine its oligomerization state. Data was collected with a Wyatt Technologies SEC-MALS system outfitted with a DAWN EOS and Optilab DSP detectors. A Shodex KW802.5 sizing column equilibrated at 1 mL/min with two column volumes of buffer containing 50 mM Bis-tris propane/Cl pH 7.3, 100 mM Na<sub>2</sub>SO<sub>4</sub> and 3 mM βME that had been 0.1 μm-filtered. Two separate samples of 50 μg protein in 100 μL volume were injected for analysis: (1) pooled elution fractions from the preparative Sephadex G75 column that corresponded to an Est3 monomer and (2) pooled elution fractions from the preparative Sephadex G75 column that corresponded to an Est3 dimer. Protein peaks on the chromatogram were analyzed by using Astra 4.9 software (Wyatt Technologies).

### **3.2.8 *In vitro* telomerase assay\***

\* Assays performed by Rachel Niederer in the laboratory of Dr. David Zappulla at the Johns Hopkins University.

Yeast *in vitro* telomerase activity assay protocol is described previously [167]. In short, yeast telomerase activity was reconstituted *in vitro* from Mini-T RNA (500-nt shortened but active form of the ~1200 nt wild-type TLC1 RNA) and ProA-tagged Est2 coding sequence containing plasmids were introduced into RRL transcription and translation system. Purified ProA-Est2-Mini-T complex on IgG-Sephadex 6

beads was incubated with telomerase assay buffer containing 40 mM Tris-HCl pH 7.5, 50 mM NaCl, 5% glycerol, 0.5 mM spermidine, 0.5 mM DTT, 0.9  $\mu$ M [ $\alpha$ - $^{32}$ P]dGTP, 100  $\mu$ M TTP, 3.5 MgCl<sub>2</sub>, and 2.5  $\mu$ M telomeric oligonucleotide substrates at 30°C for 20 minutes. Telomerase extension products were separated on 7 M urea denaturing gel by electrophoresis and analyzed and quantitated by Phosphorimaging.

Est3 proteins and its mutants were expressed and purified as discussed in section 3.2.1 and sent to Dr. Zappulla's lab for inclusion in the telomerase activity assay reagents to test effect of these Est3 proteins on telomerase activity.

### **3.2.9 *In vivo* analysis of est3- mutants\***

\* *In vivo* work performed by John Lubin in the laboratory of Dr. Victoria Lundblad (Salk Institute)

To quickly screen for Est3 mutants that confer EST phenotype on the yeast strain, a synthetic lethal screen was performed in an *EST3 yku80- $\Delta$*  strain. *yku80- $\Delta$*  background was used because this strain is highly sensitive to defects in telomerase, leading to observable phenotypes of growth defect or cell lethality [95]. The presence of wild type copy of *EST3* ensures selection of only functional mutants that, when overexpressed, exhibit growth defects, *i.e.*, an overexpression dominant negative (ODN) phenotype. High copy (2  $\mu$ ) plasmids expressing wild-type and mutant alleles of *EST3*, under control of ADH promoter, were transformed into a *yku80- $\Delta$  YKU80 URA3* strain and examined for growth on media that selects for

loss of *YKU80* plasmid. We examined growth of *est3* mutants at both 30 °C and 32 °C and identified mutants that survive at the permissive temperature (30 °C) but die at non-permissive temperature (32 °C). The mutants selected for growth defects in these assays were further tested for telomere length, in an assay as previously described [95]. The protocol for these *in vivo* experiments is also described in detail in section 4.1.2.

### **3.3 Results**

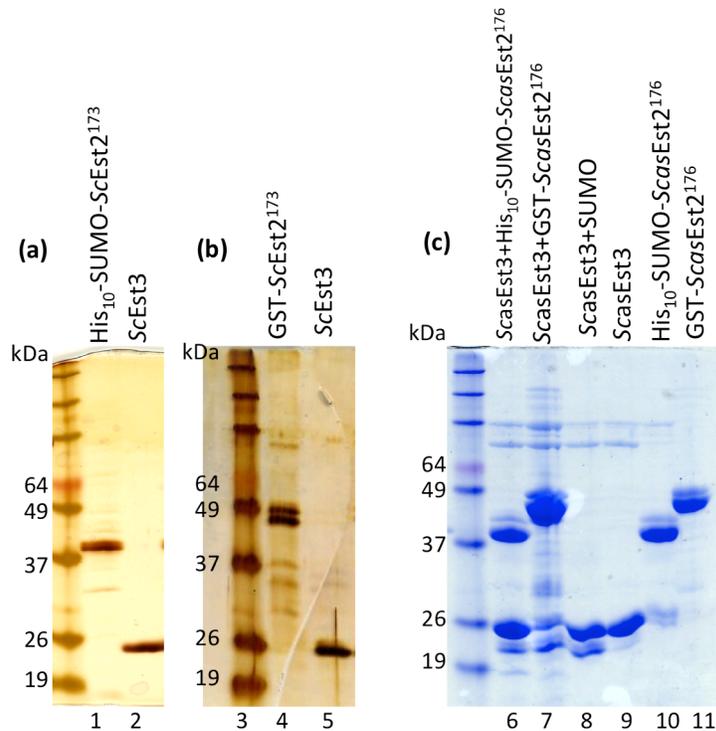
#### **3.3.1 *In vitro* studies**

##### **3.3.1a Unambiguous binding of Est3 to N-terminal domain of Est2 was not observed**

The N-terminal TEN domain of Est2 (Figure 1.7, 1.8) has been proposed to be involved in direct interaction with Est3 [189]. Biochemical data in support of this model was published last year [190]. However, several caveats of this study, such as protein purity, protein stability, and no characterization of the proper folding of point mutants, prevented the study from being conclusive. In addition to this biochemical study, a recent report describes direct interactions between the putative Est3 and Est2<sup>TEN</sup> proteins of *C. parapsilosis* and *L. elongisporus* [191]. However, these interactions were only observed when the long N- and C-terminal extensions in the Est3 proteins were intact, directly contradicting the result published on ScEst3. However, these N- and C-terminal extensions are not

phylogenetically conserved in other known Est3 protein (Figure 2.24 and 2.25) and indeed these Est3s are essential proteins in their respective organisms, raising the concern that they have a completely distinct activity (Figure 2.24 and 2.25). Therefore, if a direct interaction between Est3 and Est2<sup>TEN</sup> is present, the mechanism of interaction may differ from that in *C. parapsilosis* and *L. elongisporus*. To probe the direct interaction between these two proteins, we recombinantly expressed and purified them as described in the methods section. The purified proteins were used to assess direct association by affinity pull-down assays.

Several constructs of *S. cerevisiae* and *S. castellii* Est2 were tested for expression and solubility in *E. coli*. Only *ScEst2*<sup>173</sup> and *ScasEst2*<sup>176</sup>, corresponding to the respective TEN domains, showed somewhat stable protein expression (Figure 3.2). However, as shown in the gel, purity of the proteins was low with several contaminating protein bands visible in the lanes. The GST-tagged *ScEst2* was especially unstable and precipitated out of solution. Comparatively, the *S. castellii* GST-tagged Est2<sup>TEN</sup> was purer, with fewer contaminants. Both Est2<sup>TEN</sup> domains were used for binding with Est3 proteins.

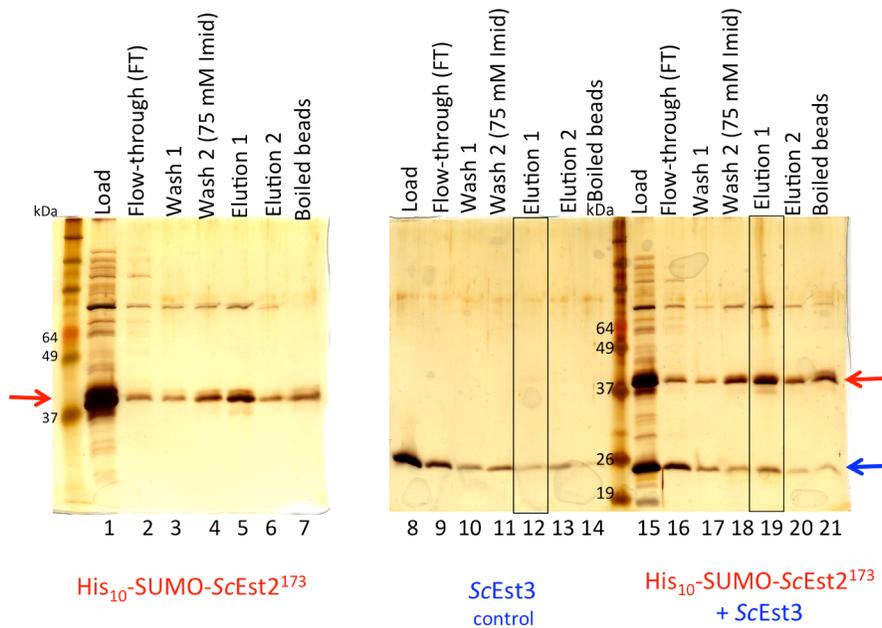


**Figure 3.2 Purified *S. cerevisiae* and *S. castellii* Est3 and Est2 proteins used for affinity pull-down studies**

(a) and (b) *S. cerevisiae* proteins visualized on silver-stained gel. (c) *S. castellii* proteins visualized on coomassie-stained gel. Purity of the Est2 proteins was low, especially GST-ScEst2<sup>173</sup> (lane 4) co-purified with several protein contaminants.

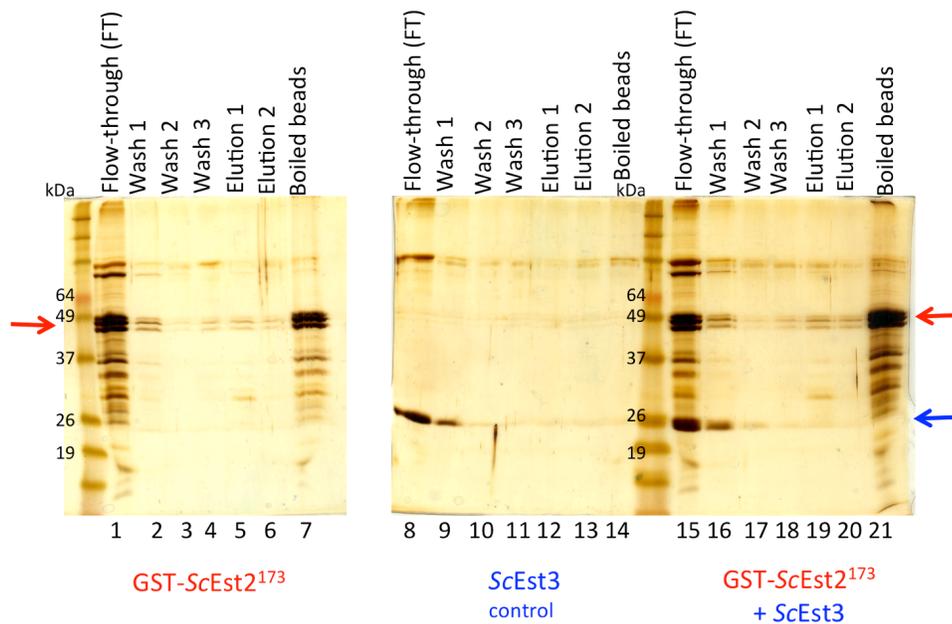
Binding of *S. cerevisiae* Est3 and Est2 proteins, using His<sub>10</sub>-SUMO-tagged ScEst2<sup>173</sup> as bait, was not observed (Figure 3.3). The first elution fraction showed slightly increased level of ScEst3, compared to the elution 1- lane in the negative control pull-down. However, the effect of impurities (which are likely chaperone proteins) in the bait protein cannot be ruled out. In addition, using slightly higher salt in the equilibration buffer, *i.e.*, 300 mM NaCl instead of 100 mM NaCl, removed this slight ScEst3 pull-down (data not shown). Also, even with 100 mM NaCl salt, the slight Est3 pull-down was not consistent between different trials. This problem was exacerbated by the tendency of the ScEst3 protein to non-specifically bind and

also to precipitate on the Ni<sup>2+</sup> affinity beads and even though this issue was slightly improved in the presence of 0.05% Igepal detergent, complete non-precipitation could not be achieved. Similarly, using GST-tagged ScEst2<sup>173</sup> as bait, no pull-down of ScEst3 was observed (Figure 3.4), but this could be because the GST-tagged ScEst2<sup>173</sup> protein was very unstable and aggregated readily. Therefore, to address the question whether Est3 and Est2<sup>TEN</sup> domains directly interact, we turned to the recombinantly expressed and purified *S. castellii* proteins.



**Figure 3.3 Affinity pull-down assay with SUMO-ScEst2<sup>173</sup> and ScEst3**

As observed from the “elution 1” in the control pull-down, ScEst3 non-specifically associates with Ni<sup>2+</sup> affinity beads because of presence Est3 band in all the lanes in the silver-stained gel. Despite somewhat increased Est3 in the “elution 1” fraction in the presence of His-tagged Est2 domain, the non-specific binding and presence of other protein impurities makes interpretation difficult. The slight retention of the “prey” protein ScEst3 was not reproducible. “Bait” proteins are highlighted in red and “prey” protein in blue.

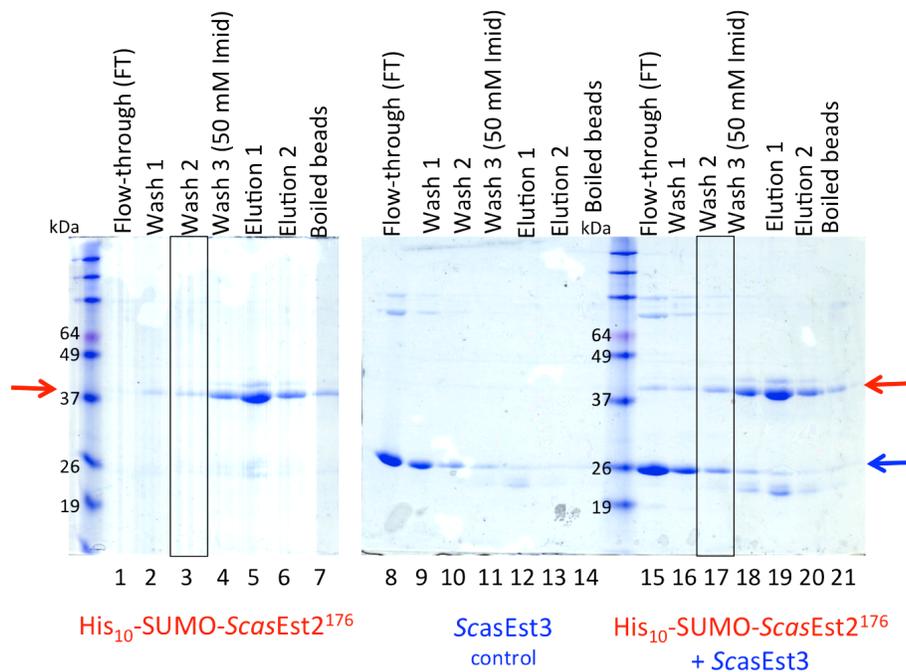


**Figure 3.4 Affinity pull-down assay with GST-ScEst2<sup>173</sup> and ScEst3**

As seen from the FT fractions on this silver-stained gel, the GST-tagged ScEst2<sup>173</sup> domain either does not get retained on the GST beads or precipitates on the beads. Thus, purified GST-tagged ScEst2<sup>173</sup> was highly unstable and not useful for binding tests. “Bait” proteins are highlighted in red and “prey” protein in blue.

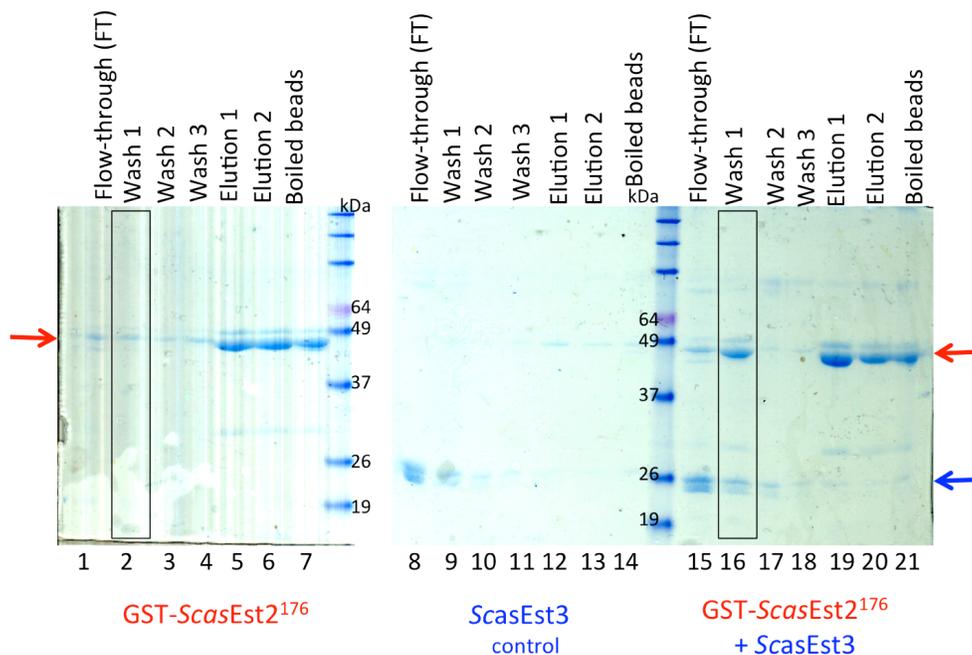
Binding of *S. castellii* Est3 and Est2 proteins, was first tested by using His<sub>10</sub>-SUMO-tagged *ScasEst2*<sup>176</sup> as bait (Figure 3.5). Again, no increased pull-down of Est3 was observed in the elution fractions in the presence of His<sub>10</sub>-SUMO-tagged *ScasEst2*<sup>176</sup>. However, binding of the His-tagged *ScasEst2*<sup>176</sup> to the affinity-beads was affected as the tagged protein was increasingly present in the flow-through and wash fractions when *ScasEst3* was present. This result was recapitulated when GST-tagged *ScasEst2*<sup>176</sup> was used as bait, to capture proteins on glutathione sepharose beads, instead on the Ni<sup>2+</sup>-affinity beads (Figure 3.6). As both the GST

and His<sub>10</sub>-SUMO tags were on the N-terminal ends of the Est2 sequence, observation of reduced binding to affinity beads might indicate that binding with Est3 disrupts the respective affinity-binding sites of the fusion-tags. To prove that this is true, either untagged Est2 with His- or GST-tagged Est3 or C-terminally tagged Est2 would have to be used for binding. Untagged Est2<sup>TEN</sup> was not stable and a suitable C-terminal tagging strategy has not yet been identified. However, as this result was not observed with the *S. cerevisiae* proteins, conclusive direct binding cannot be deduced.



**Figure 3.5 Affinity pull-down assay with SUMO-*ScasEst2*<sup>176</sup> and *ScasEst3***

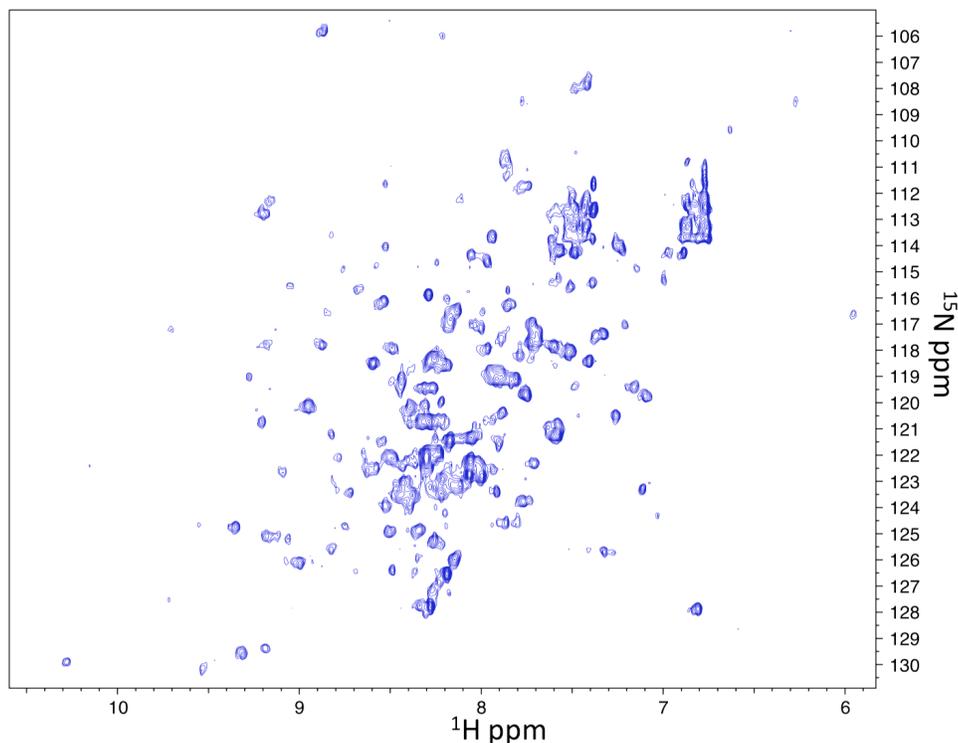
This coomassie-stained gel indicates no increased retention of *ScasEst3* on the Ni<sup>2+</sup> affinity beads in the presence of the bait protein (compare “elution 1” lane in control and test). However, presence of *ScasEst3* prevented complete retention of His-tagged SUMO Est2 on the beads (compare lanes 10 and 17). “Bait” proteins are highlighted in red and “prey” protein in blue.



**Figure 3.6 Affinity pull-down assay with GST-*ScasEst2*<sup>176</sup> and *ScasEst3***

This coomassie-stained gel indicates no increased retention of *ScasEst3* on the GST-affinity beads in the presence of the bait protein (compare lanes 12 and 19). However, presence of *ScasEst3* prevented complete retention of GST-tagged Est2 on the beads (compare lanes 9 and 16). “Bait” proteins are highlighted in red and “prey” protein in blue.

Additionally, *ScasEst3* protein forms soluble aggregates as observed from the visually soluble NMR sample that yields <sup>1</sup>H-<sup>15</sup>N HSQC spectrum of a mostly aggregated protein (Figure 3.7). Thus, the likelihood that *ScasEst3* protein used for binding studies is partially unfolded introduces another variable, making interpretation of small differences between negative control and test conditions difficult.

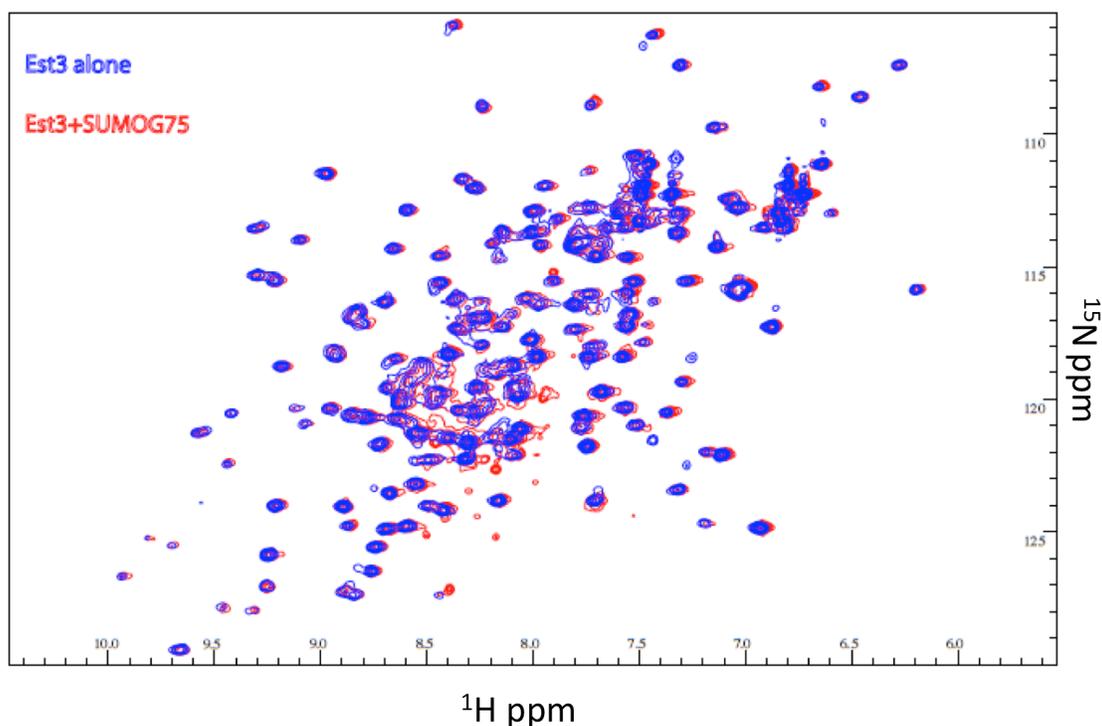


**Figure 3.7**  $^{15}\text{N}$  TROSY HSQC spectrum of *ScasEst3*

Differential peak-intensities and clumping of peaks in the HSQC spectrum are indicative of *ScasEst3* protein being aggregated in solution. The experiment was conducted at 25°C at the 900 MHz spectrometer. NMR sample was in buffer (pH 7.1) containing 50 mM Bis-tris propane, MOPS, 50 mM  $\text{Na}_2\text{SO}_4$ , 100 mM arginine/ glutamate, 150 mM NDSB-195, 2 mM DTT and 6%  $\text{D}_2\text{O}$ . Aggregated protein spectrum was recapitulated with the potassium phosphate buffer containing no additives (data not shown).

In addition to affinity pull-down assays, binding of Est3 and Est2<sup>TEN</sup> was also attempted by designing small peptides from the Est2<sup>TEN</sup> region around the putative Est3 interacting residue K111 (Lundblad lab, unpublished data). One such peptide was 50 residues long: GELYNNVLTFGYKIARNEDVNNSLFCHSANVNVTL**L**KGAAWKMFHSLVG (with K111 underlined and in bold). The peptide was expressed and purified as a fusion with His<sub>10</sub>-SUMO tag and added to the  $^{15}\text{N}$ -labeled *ScEst3* NMR sample. Chemical

shift perturbation analysis of the superimposed spectra of Est3 alone and Est3 in the presence of the peptide showed no significant chemical shift perturbations indicating lack of Est3 binding to Est2 peptide (Figure 3.8). This study, however, does not rule out the probability that Est3 and Est2<sup>TEN</sup> bind. This is because the peptide might be unable to recapitulate a binding interaction that requires the intact and folded protein domains. In addition, the putative interacting residue K111 might not define the key region of Est2<sup>TEN</sup> binding with Est3.

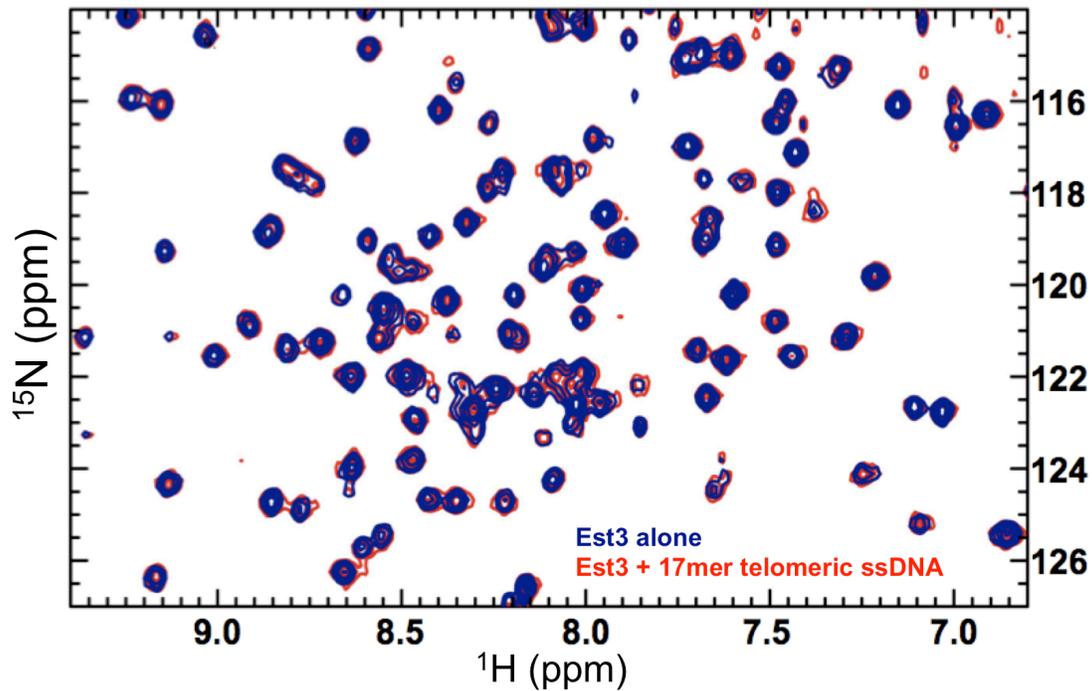


**Figure 3.8 Chemical shift perturbation of <sup>15</sup>N-labeled Est3 in the presence of small peptide of Est2**

<sup>15</sup>N HSQC spectrum from Est3 (blue) alone and Est3 mixed with a 50-residue peptide of Est2<sup>TEN</sup> domain (red). No significant shifts in the peaks were observed, indicating no direct binding in our test.

### 3.3.1b Purified Est3 does not bind to telomeric ssDNA *in vitro*

Est3 has been reported to enhance telomerase activity *in vitro* in *C. albicans* [184] and *S. castellii* [183] systems. This indicates that Est3 might potentially interact with telomeric DNA and stabilize the telomerase on substrate ssDNA overhang, as it is being extended. In previous studies published by us and others, purified Est3 has not been observed to bind DNA [96,183]. In our study direct binding was assessed by addition of ~40-fold molar excess of Tel17 telomeric ssDNA (5'-d(GTGTGGGTGTGGTGTGG)-3') to the <sup>15</sup>N-labeled *ScEst3* sample (in buffer A containing 50 mM potassium phosphate pH 7.3, 50 mM Na<sub>2</sub>SO<sub>4</sub>, 250 mM arginine/ glutamate, 2 mM DTT and 7% D<sub>2</sub>O) to assess chemical shift perturbation in the <sup>1</sup>H-<sup>15</sup>N TROSY HSQC experiment (Figure 3.10). We observed no chemical shift perturbation, indicating that Est3 does not directly interact with telomeric ssDNA [183].

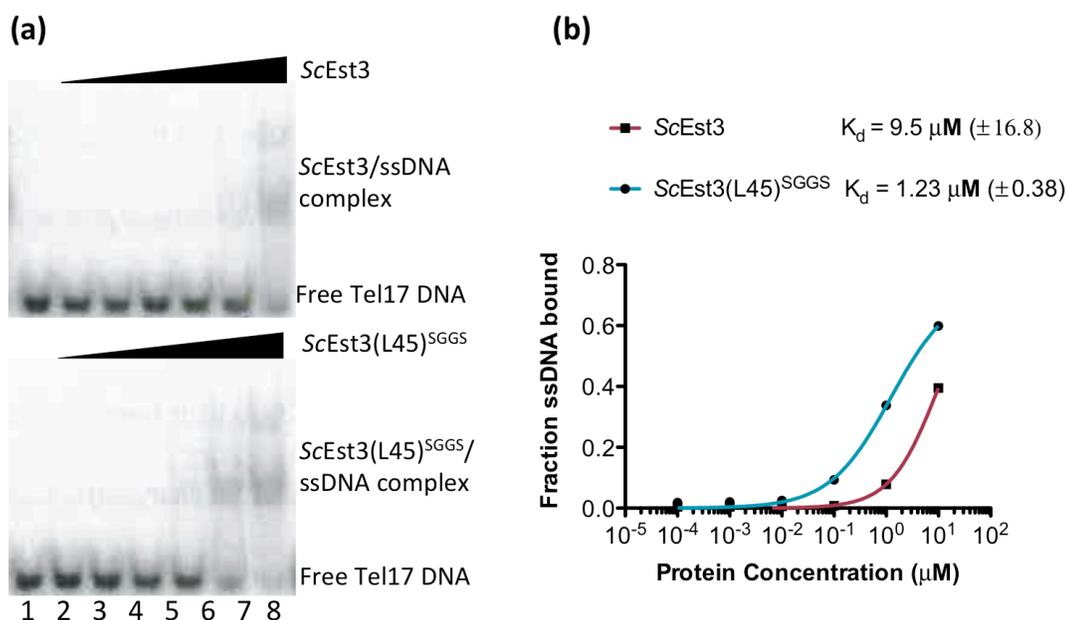


**Figure 3.10 No Est3-ssDNA binding is detected from minimal NMR chemical shifts analysis**

HSQC spectrum with (red) and without (blue) DNA sample of the  $^{15}\text{N}$ -labeled Est3. No binding was observed by HSQC as of peak positions in the test sample matching the peaks in Est3 alone. (from [183])

In a recent study of a putative Est3 homolog from *C. parapsilosis* (*Cp*), ssDNA binding by Est3 was observed via ultraviolet-mediated cross-linking [191]. However, this observation was only in the presence of *CpEst2*<sup>TEN</sup> domain. Using our new high-resolution structure of *ScEst3* (Chapter 2), we compared the canonical OB fold ligand binding surface on *ScEst3* to that of other known ssDNA-binding OB-folds, including *ScCdc13*, *HsPot1*, and *SpPot1* [203,262,263]. The long loop L45 that is visible in the multiple sequence alignment of Est3 proteins was found to be occupying the canonical binding surface of *ScEst3* OB-fold (section 2.4.1a). This

suggested that ssDNA binding activity of Est3 might be masked by the presence of L45 in the binding site. In a hypothetical model, L45 might move away from the canonical binding site in the presence of an *in vivo* factor (Est2), thus revealing the ssDNA binding activity of Est3. To directly test the hypothesis that removal of L45 opens the Est3 binding surface to ligands, we replaced loop L45 with short linker sequence, Ser-Gly-Gly-Ser (L45<sup>SGGS</sup>; section 3.2.3) and evaluated protein stability, structure, and ssDNA binding activity. The L45<sup>SGGS</sup> mutant protein could be partially purified. However, the protein precipitated out of solution upon Ulp1-mediated cleavage of the His<sub>10</sub>-SUMO fusion; even without Ulp1, the His<sub>10</sub>-SUMO-Est3(L45)<sup>SGGS</sup> spontaneously cleaved into His<sub>10</sub>-SUMO and Est3 fractions. Therefore, it was not feasible to prepare even a low-concentration (100  $\mu$ M) <sup>15</sup>N-labeled Est3(L45)<sup>SGGS</sup> NMR sample, to revisit chemical shift perturbation in the presence of telomeric ssDNA. As a preliminary test, binding activity of *S. cerevisiae* wild type Est3 and crudely pure Est3(L45)<sup>SGGS</sup> to Tel17 telomeric DNA was assessed by electrophoretic mobility shift assay (EMSA) (Figure 3.11). The  $K_{D,app}$ s were in  $\mu$ M range, but not reproducible likely because of protein purity problems. In addition, *Scas*Est3(L45)<sup>SGGS</sup> mutant was not soluble, and therefore was not evaluated for binding activity. Thus, the hypothesis that the long L45 removal aids in DNA binding-site access could not be assessed properly because of poorly behaved Est3 protein, upon L45 deletion.

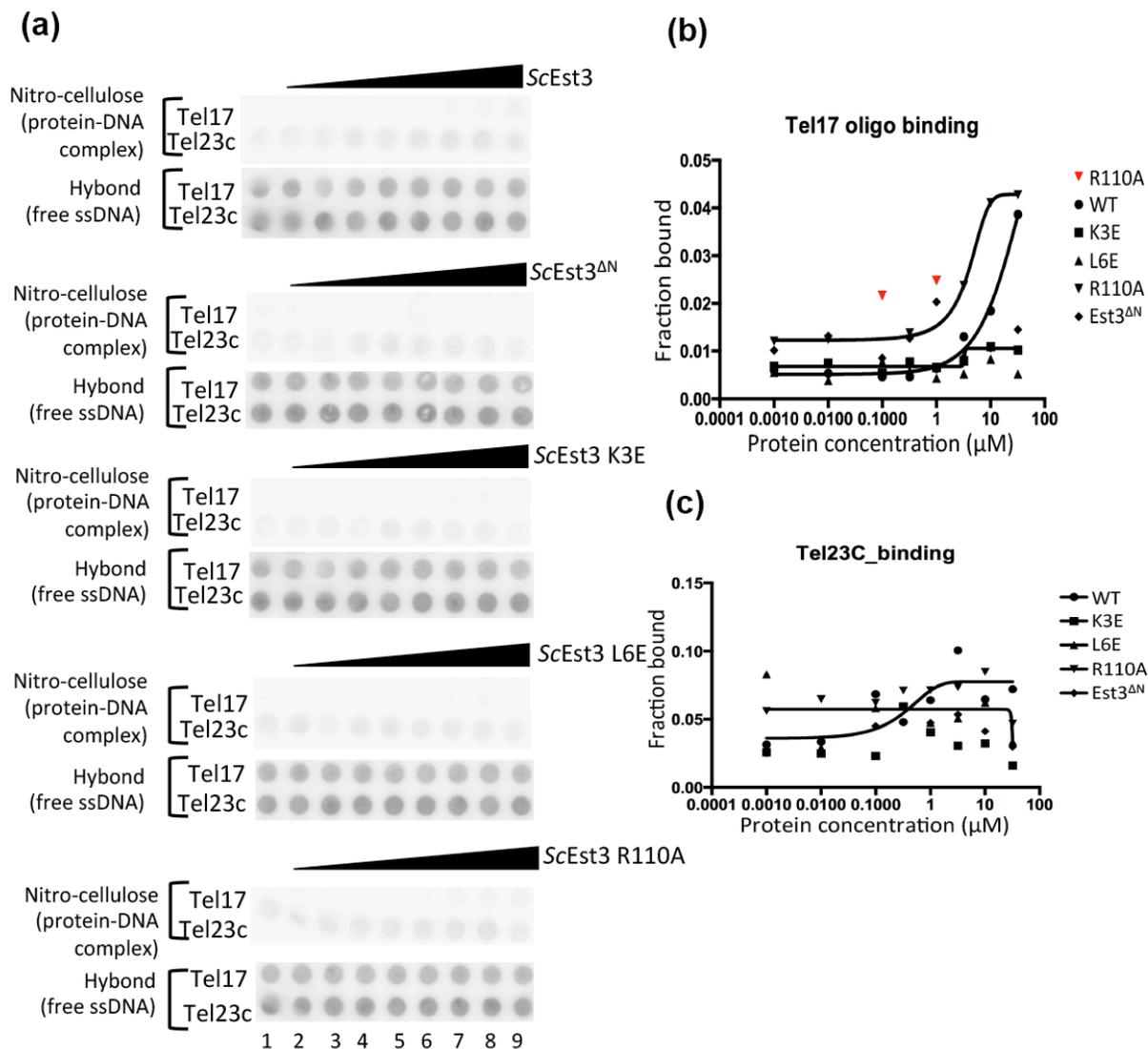


**Figure 3.11** Electrophoretic mobility shift assay to test direct binding of ScEst3 and telomeric Tel17 DNA

(a) Gel-shift assay showing slower migration of wild type and ScEst3 L45 mutant bound Tel17 ssDNA. Lanes 1-8 represent binding with increasing concentrations of the protein (0, 10<sup>-4</sup>, 10<sup>-3</sup>, 10<sup>-2</sup>, 10<sup>-1</sup>, 1.0 and 10 μM). (b) Quantitation of the gel in (a).

Finally, in an alternative and faster protein-DNA binding assay, called the filter-binding assay, we tested DNA binding to wild type ScEst3 and its mutants (ScEst3<sup>ΔN</sup>, ScEst3 K3E, ScEst3 L6E and ScEst3 R110A) (Figure 3.12). ScEst3<sup>ΔN</sup> was tested because it is our construct for structure elucidation (Chapter 2). K3E and L6E mutants (Lundblad lab, unpublished data) and R110A mutant [95] have been shown to be important for Est3's telomere-length maintenance function. Thus, we wanted to test if these residues of Est3 are involved in DNA binding activity. Binding was tested to the telomeric Tel17 ssDNA (5'-d(GTGTGGGTGTGGTGTGG)-3') and a telomere complementary sequence Tel23c ssDNA 5'-

d(CACACCCACACACCCACACACCC)-3'. No significant retention of Tel17 ssDNA signal on the nitro-cellulose membrane (DNA-bound protein complex) was observed for either the wild type Est3 or its mutants, as quantitated in the Figure 3.12(b). Some retention of the Tel23c ssDNA signal was observed on the nitro-cellulose membrane; however, quantitation indicated that the DNA binding was non-specific (Figure 3.12(c)). In conclusion, we have used several strategies to detect a DNA binding activity for the *ScEst3* protein. No activity was detected for the purified proteins. Thus, *ScEst3* does not interact with telomeric ssDNA by itself. This does not rule out the possibility that *in vivo* the presence of a binding factor (maybe Est2) promotes binding of Est3 to telomeric DNA.



**Figure 3.12 Filter binding assay to test direct binding of ScEst3 and its mutants to telomeric DNA**

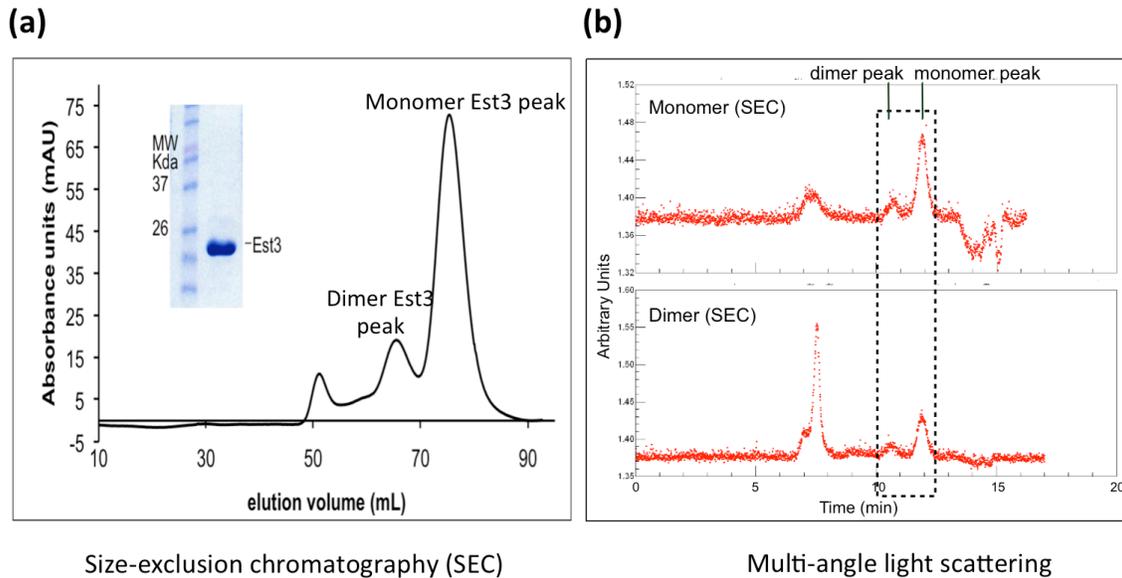
(a) Wild type ScEst3 and its mutants (*ScEst3<sup>ΔN</sup>*, *ScEst3 K3E*, *ScEst3 L6E* and *ScEst3 R110A*) were tested for binding to telomeric Tel17 DNA and a telomere complementary Tel23c DNA oligos. Protein-bound DNA was captured on nitro-cellulose membrane (top half of each protein panel). Unbound DNA was captured and quantitated from hybond membrane (bottom half of each protein panel). Protein concentrations from wells 1 to 9 were 0.0, 0.001, 0.01, 0.1, 0.32, 1.0, 3.2, 10 and 32  $\mu\text{M}$ . (b) and (c) are quantitations of Est3 proteins binding to Tel17 and Tel23c DNA oligos, respectively. Red inverted triangles in (b) indicate the outliers in R110A.

### **3.3.1c Est3 protein solution is comprised of an equilibrium between dimer and monomer forms**

It has been proposed that dimerization is important for Est3's function [259] and that *ScEst3* residues D49, Q34 and H40 are important for this dimerization. Mutational analysis of these residues resulted in reduced telomere length *in vivo*. However, as it was not established whether these mutants were simply disrupted in the core structure of Est3, the *in vivo* effects cannot be taken as conclusive (refer to Chapter 4 for importance of establishing intactness of mutant structure for functional analysis).

Purified Est3 partitions into monomer and dimer fractions on the preparative Sephadex G75 size-exclusion chromatography (SEC) column (Figure 3.12a). It is not clear if Est3 is a functional dimer *in vivo*, despite some correlative studies between mutations that change the volume of the dimer peak on SEC column and effect telomere length [259]. To test if monomer and dimer fractions of Est3 from SEC can be isolated, the protein fractions under the monomer peak were pooled and concentrated and similarly the fractions under the dimer peak were pooled and concentrated. Upon injection of these two separated and pooled fractions at 0.5 mg/mL concentration onto another protein separation column for multi-angle light scattering analysis, both monomer and dimer injections re-partitioned into dimer and monomer peaks (Figure 3.12b). The size of the monomer and dimer peaks was estimated to be ~18 kDa and ~35 kDa respectively, corresponding approximately to the ~20 kDa size prediction for Est3 monomer using the ProtParam tool from

expasy: web.expasy.org/protparam. The presence of dimer and monomer peaks from both the pooled monomer and pooled dimer injections indicates that the Est3 protein exists in dimer-monomer equilibrium, leading to sample heterogeneity.



**Figure 3.13 ScEst3 purification on size-exclusion chromatography and multi-angle light scattering**

(a) Purification profile of Est3 upon elution from SEC Sephadex-G75 column. (b) Pooled dimer peak was injected on the MALS column and found to partition again into dimer and monomer peaks. Same result was observed for the SEC pooled monomer peak. Back calculation of protein size from the MALS column indicated monomer peak size of ~18 kDa and dimer peak size of ~35 kDa.

### 3.3.1d *S. cerevisiae* Est3, does not enhance yeast telomerase activity *in vitro*\*

\* Telomerase assays performed by Rachel Niederer, laboratory of Dr. David Zappulla, Johns Hopkins University. I purified wild type and mutant Est3 proteins to test in these telomerase assays.



### **3.3.2 Structure based *in vivo* analysis of Est3 mutants led to identification of residues important for Est3's function\***

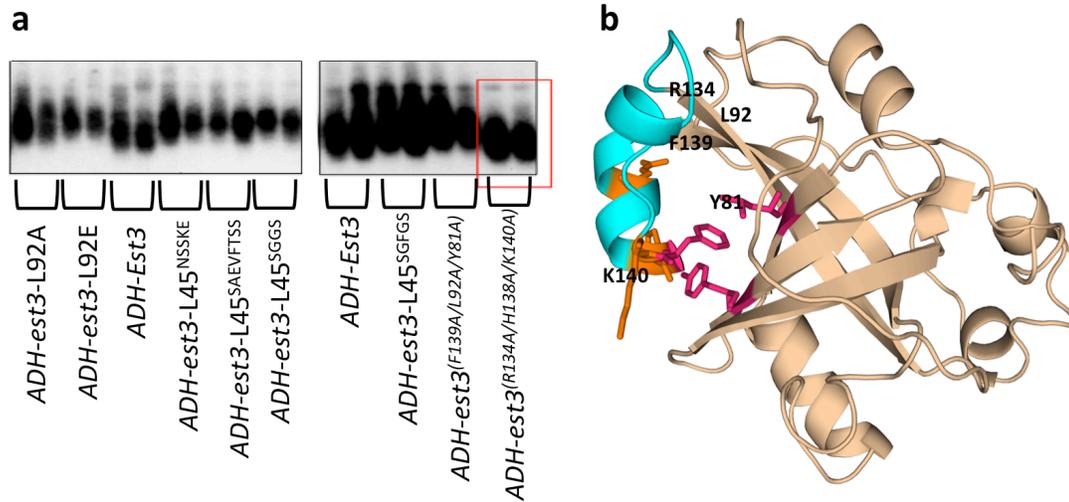
\* All *in vivo* work performed by John Lubin in the laboratory of Dr. Victoria Lundblad, Salk Institute. The constructs were designed in collaboration between myself, Dr. Wuttke and Dr. Lundblad. Cloning and genetic experiments were done by John Lubin.

#### **3.3.2a Mutational analysis of L45**

The high-resolution structure of Est3 shows a long loop L45 that sits over and occludes the canonical binding-site of the OB-fold motif. Based on this unusual L45 and the published report of a latent ssDNA binding activity [191], it was hypothesized that ssDNA binding of Est3 might be hidden because of presence of L45 in the binding-site and that removal of this L45 in the presence of another binding protein (likely Est2) might reveal the ssDNA binding activity of Est3. To test this hypothesis *in vivo*, F139 residue (highly conserved and likely anchoring the L45 via hydrophobic interaction with L92 on the surface of the  $\beta$ -barrel) was mutated to alanine in combination with L92A mutation (F139A+L92A). Secondly, a triple mutant (R134A+H138A+K140A) of solvent-exposed residues on the surface of L45 was also tested. In addition, the loop L45 was shortened, such that all residues, except five, were deleted to leave a shortened loop ( $\Delta$ R134-S147). Also, the entire loop was replaced with different sized linkers 8-residue SAEVFTSS, 5-residue SGFGS and 4-residue SGGs. All of these mutants were tested for overexpression

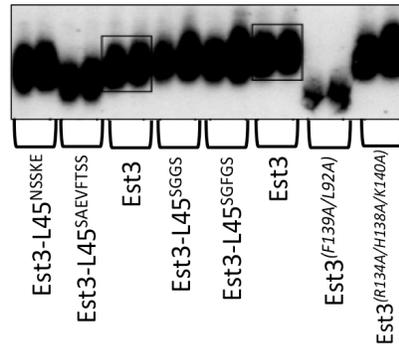
dominant-negative (Figure 3.15) and loss-of-function (Figure 3.16) phenotypes, by assessing telomere-length in both cases to the telomere length from wild-type Est3 expressing yeast cells. Except the triple mutant (R134A+H138A+K140A), none of the other mutants showed a dominant-negative phenotype. The absence of dominant-negative phenotype was found to be due to these mutants carrying a wild-type phenotype for telomere-length, as observed from the loss-of-function test, and not because of protein destabilization, which would lead to some degree of EST phenotype (*i.e.*, telomere-length shortening) in the loss-of-function assay. Wild type telomere lengths were seen for all L45 shortened or deleted mutants in the loss-of-function assay (Figure 3.16), except the 8-residue linker containing Est3 mutant that showed slightly reduced telomere length. Triple mutant (R134A+H138A+K140A) that showed slight dominant-negative effect also showed no reduction in telomere length in the loss-of-function assay. The obviously shortened telomere-length in loss-of-function assay for F139A+L92A double mutant indicates that this mutant Est3 protein is likely destabilized because this mutant does not show a dominant-negative effect in the ODN assay. Thus, in conclusion, the removal of loop L45 did not manifest as an effect on the telomere-length in these *in vivo* assays. This indicates that L45 is not required for Est3's role in telomere-length maintenance and that presence or absence of L45 does not affect the access of a putative binding partner to the OB-fold surface or that the canonical binding-site in Est3 simply does not bind to any interacting partner. These results, however,

do not rule out the possibility that L45 might be playing some other role in Est3 that is not readily apparent from just testing for telomere-length upon mutagenesis.



**Figure 3.15 Mutants in L45 region of Est3 do not show dominant-negative phenotypes**

(a) None of the mutants in the long L45 region of Est3 showed significant effect in telomere-length maintenance. This result was recapitulated by the loss-of-function mutants in these residues. The region of interest is boxed in red. (b) Loop L45 is colored in cyan. Other mutants used in this study are mapped on the structure.



**Figure 3.16 Mutants in L45 region of Est3 do not show dominant-negative phenotype**

None of the mutants of Est3 showed significant effect on telomere-length maintenance. No mutants, except the F139A, H138, and K140 triple mutant, showed somewhat decreased telomere length. Region of interest for wild-type Est3 is boxed in black.

### **3.3.2b Conservation-independent, complete surface mutagenesis of Est3 identifies functionally important residues**

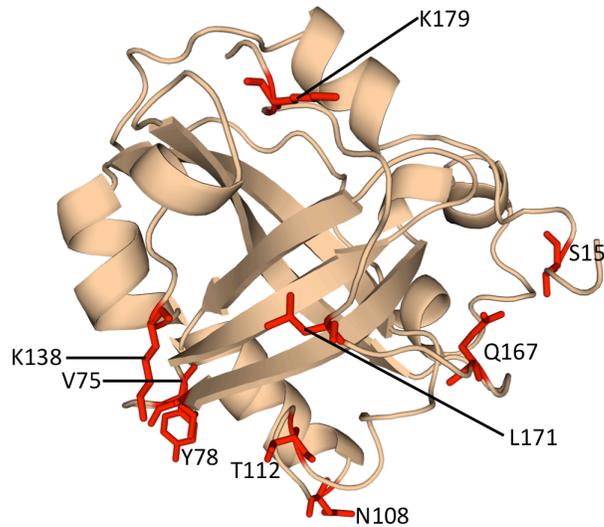
Since Est3 is an important regulatory element in yeast telomere length homeostasis, we wanted to assess the complete surface of this protein for residues involved in this function. Previous studies had mutated some of the conserved residues in Est3 and tested the effect of these mutations on telomere length *in vivo* [95,96]. The presence of a structure for Est3 allowed us to better identify all the residues with surface exposed side-chains and assess these residues for their importance in Est3's telomere length function as well as establish the spatial inter-relationship of the residues known to be involved in function. Towards that goal we identified 111 residues as being surface exposed and in an unbiased screen mutated all of them to either alanine or a charge-swap residue and tested for *in vivo* telomere length defects. Telomere length defects were assessed using a synthetic lethal screen (Table 3.2). The screen identified 10, previously unidentified, residues that when mutated led to an ODN phenotype. These residues were S15, V75, Y78, N108, T112, K140, Q167, L171, S172 and K179 (Figure 3.17). Interestingly, only three out the 10 residues were highly conserved and the rest were only moderately conserved or variable, thus highlighting the importance of an unbiased mutational analysis approach towards identification of functionally relevant protein surfaces.

Table 3. 2 Overexpression dominant negative (ODN) phenotypes of point-mutants in solvent exposed residues of Est3

#	Residue number	Mutation	ODN phenotype	#	Residue number	Mutation	ODN phenotype
1	D14	D14R	no	52	S87	S87E	no
2	<b>S15</b>	<b>S15E</b>	<b>moderate</b>	53	S88	S88E	no
3	V16	V16E	no	54	H89	H89E	no
4	F17	F17R, F19E	no, no	55	Q90	Q90R	no
5	Q19	Q19R	no	56	E94	E94R	no
6	P20	P20E	no	57	S96	S96E, S96I	no, no
7	K23	K23E	no	58	Q97	Q97R	no
8	A24	A24R	no	59	E98	E98R	no
9	L25	L25E	no	60	S101	S101R	no
10	E27	E27R	no	61	N102	N102R	no
11	D28	D28R	no	62	R105	R105E	no
12	N29	N29R	no	63	T106	T106R	no
13	E31	E31R	no	64	H107	H107R	no
14	H32	H32R, H32E	no, no	65	<b>N108</b>	<b>N108E</b>	<b>yes</b>
15	D33	D33R	no	66	C109	C109R	no
16	Q34	Q34R	no	67	<b>T112</b>	<b>T112A, T112E</b>	<b>YES, YES very strong</b>
17	Y35	Y35R, Y35E	no, no	68	S113	S113R	no
18	H36	H36E		69	T116	T116E	<b>moderate to slight</b>
19	P37	P37E	no	70	D124	D124R	no
20	S38	S38R	no	71	D126	D126R	no
21	G39	G39R	no	72	V128	V128E	no
22	H40	H40R, H40I	no, no	73	T131	T131R	no
23	V41	V41E	no	74	N132	N132R	no
24	P43	P43R	no	75	S133	S133E	no
25	S44	S44R	no	76	R134	R134E	no
26	T46	T46R	no	77	M136	M136E	no
27	K47	K47E	no	78	S137	S137E	no
28	Q48	Q48R	no	79	H138	H138R, H138I	no, no
29	D49	D49R	no	80	<b>K140</b>	<b>K140E</b>	<b>very slight</b>
30	L52	L52E	no	81	C142	C142E	no
31	H54	H54R, H54I	no, no	82	S144	S144R	no
32	M55	M55E	no	83	N145	N145R	no
33	S56	S56R	no	84	I146	I146E	no
34	P57	P57R	no	85	S148	S148E	no
35	T58	T58R	no	86	K149	K149E	no
36	L60	L60E	no	87	E150	E150R	no
37	T61	T61R	no	88	I151	I151E	no
38	N62	N62R	no	89	N156	N156R	no
39	P63	P63R	no	90	N158	N158R	no
40	C64	C64R	no	91	Q159	Q159R	no
41	T70	T70E	no	92	T161	R161R	no
42	Y73	Y73R, Y73E	no, no	93	F163	F163R, F163E	no, no
43	N74	N74R	no	94	<b>Q167</b>	<b>Q167R</b>	<b>moderate</b>
44	<b>V75</b>	<b>V75E</b>	<b>YES, very strong</b>	95	<b>L171</b>	<b>L171E</b>	<b>YES very strong</b>
45	C76	C76R	no	96	<b>S172</b>	<b>S172E</b>	<b>slight</b>
46	D77	D77K	no	97	T173	T173E	no
47	<b>Y78</b>	<b>Y78A, Y78F</b>	<b>yes, yes, yes</b>	98	P175	P175E	no
48	K79	K79E	no	99	F176	F176A, F176I	no, no, no
49	Y81	Y81R, Y81E	no, no	100	<b>K179</b>	<b>K179E</b>	<b>slight</b>
50	S83	S83E	no	101	Y180	Y180R, Y180I	no, no
51	R85	R85E	no	102	L181	L181E	no

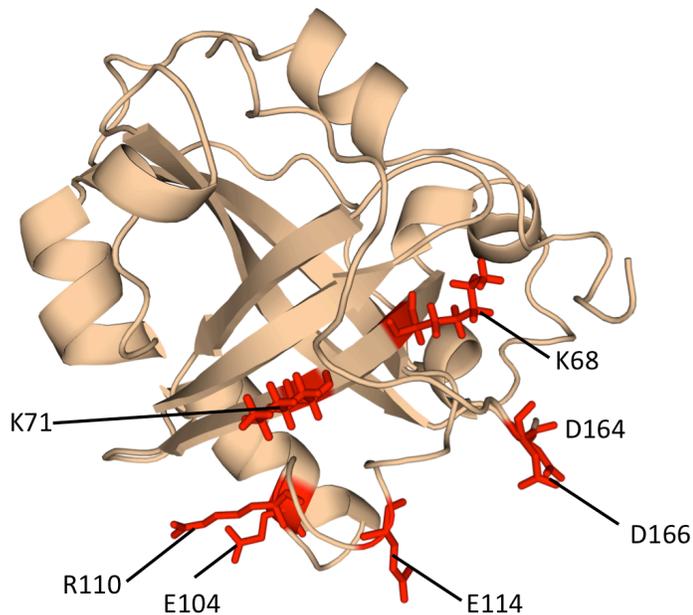
To more directly assess the involvement of the 10 newly identified residues in telomere length control, telomere lengths of the strains carrying these Est3 mutants were directly assessed. Telomere length analysis was done both in the *est3Δ* background (loss-of-function assay) and in the presence of the wild-type *EST3* copy (overexpression dominant negative assay) (data not shown). Mutations to both alanine and a charge-swap residue were made due to the fact that sometimes a charge-swap mutation is better able to disrupt the functional interaction of a residue leading to revelation of its importance for that function.

Mapping the newly identified functional residues (from unbiased whole surface mutagenesis) (Figure 3.17) and the previously identified residues (based on highly conserved residue mutational analysis) (Figure 3.18) onto Est3 structure identifies the region of Est3 functionally involved in telomere length maintenance function of Est3. This region is defined by  $\beta$ -sheet formed by strands  $\beta$ 1,  $\beta$ 2 and  $\beta$ 3 and at the bottom of the  $\beta$ -barrel by the helix H5 and loop preceding  $\beta$ 4 (orientation as in Figure 2.26). Surprisingly, even in the presence of several surface exposed residues, many of them highly conserved, the residues outside of this region are not important for telomere length maintenance function of Est3 as assessed by the synthetic lethal screen for telomere replication defects. This might be indicative of an alternative role, as yet unidentified, for the Est3 protein.



**Figure 3.17 Est3 residues important for telomere-length maintenance, picked from the extensive point-mutant screening of the entire Est3 surface**

Residues newly identified from the screening, as involved in Est3's telomere-length maintenance function, are highlighted in red and labeled.



**Figure 3.18 Est3 residues important for telomere-length maintenance, picked from conserved residues of Est3**

Residues identified previously [95] as involved in Est3's telomere-length maintenance function are mapped onto the Est3 structure and highlighted in red and labeled.

## 3.4 Discussion

### 3.4.1 Direct binding of ScEst3 with the TEN domain of Est2 was not established *in vitro*, but the existence of a common surface for telomerase association points to functional homology between Est3 and human TPP1-OB

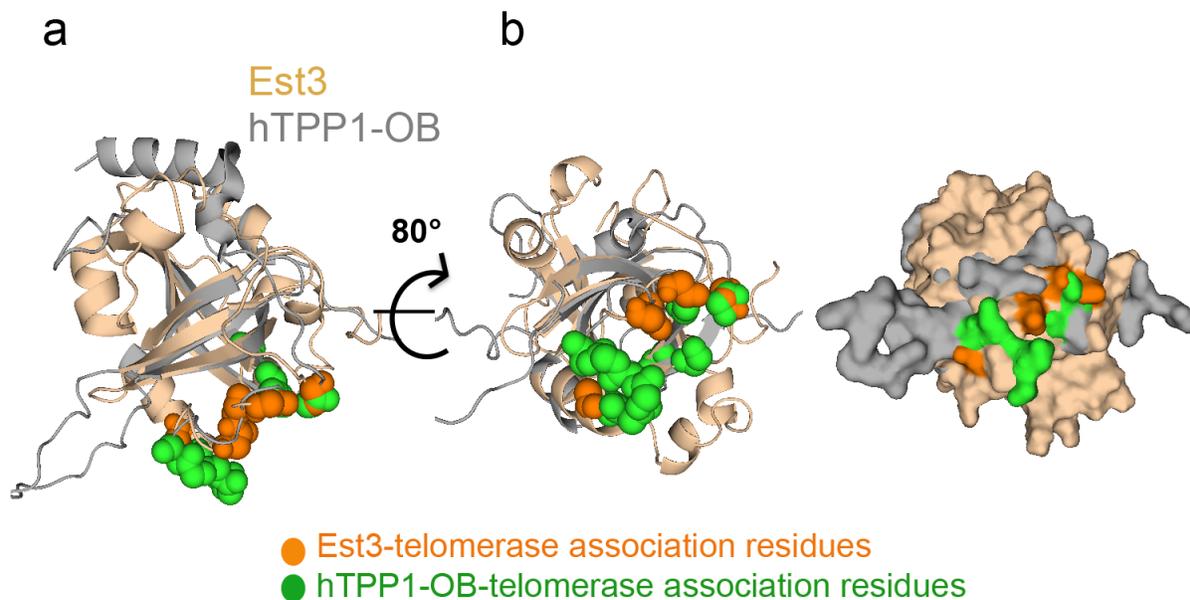
In a structure-function analysis approach, we have tested the binding of Est3 to its proposed binding partners, Est2<sup>TEN</sup> and telomeric ssDNA, by various methods. The binding to Est2<sup>TEN</sup> still remains inconclusive because of purified proteins being unstable and impure. Absence of pull-down might be indicative of weak or no-association of the Est2<sup>TEN</sup> domain with Est3, or that better conditions for binding are required or that the yeast full-length Est2 protein is required for binding tests if Est2<sup>TEN</sup> is better functional/ active in the context of the full-length protein.

In conclusion, the Est2<sup>TEN</sup> domain did not unambiguously and reproducibly bind the Est3 protein in our studies. The caveats associated with our observations are also applicable to the direct binding result reported from the Friedman lab [190], with poor protein quality affecting the interpretation of the data. Therefore, the direct binding of Est2<sup>TEN</sup> domain to Est3 is still an open-ended question. Furthermore, it remains to be seen whether the Est3-binding interface on Est2 is outside the TEN domain.

Our experimental structural model of *S. cerevisiae* Est3 (Chapter 2) shows that the closest structural homolog is the OB-fold of human TPP1 (*Hs*TPP1-OB) [90].

Genetic and biochemical studies of *Hs*TPP1-OB can be used as a starting place to further understand the potentially physiologically relevant surfaces on Est3. A recent study probed the interaction between *Hs*TPP1-OB and the telomerase core components *Hs*TERT (the reverse transcriptase) and *Hs*TR (the RNA) [92]. Several point mutations were identified in *Hs*TPP1-OB that reduced the recruitment of telomerase to telomeres and were therefore implicated in direct association between *Hs*TPP1-OB and *Hs*TERT [92,93,258]. A few of the key residues on the *Hs*TERT surface that are involved in *Hs*TPP1-OB association were shown to map to the disparate regions of TEN and CTE domains of *Hs*TERT (see Figure 1.7 for *Hs*TERT domain map). Support for a model in which *Hs*TPP1-OB and Est3 bind their respective reverse transcriptases, *Hs*TERT and Est2, similarly, can be seen when residues identified to be involved in *Hs*TERT association are mapped on the *Hs*TPP1-OB structure (Figure 3.9). On both *Hs*TPP1-OB and Est3, the telomerase-interaction mutation sites form a cluster that indicates a binding interface. Interestingly, the surface to which the previously identified telomerase association residues of Est3 map, precisely overlays with the binding interface on *Hs*TPP1-OB, when the two structures are superimposed [95] (Figure 3.9). This suggests a functional homology between the two proteins, in addition to the established structural homology. It stands to reason, therefore, that absence or presence of *Hs*TPP1-OB association with the *Hs*TEN domain would be indicative of absence or presence of direct association between yeast Est3 and Est2<sup>TEN</sup> and vice versa. However, the direct association can only be conclusively proven upon identification

of key residues on the yeast Est2 surface as well that affect co-immunoprecipitation of Est3 with telomerase and if those mutants complement the mutants on the telomerase-association region of Est3. Also important to note is that although the telomerase interacting surfaces and their electrostatic natures are roughly the same, the nature of the corresponding residues in Est3 and TPP1-OB are not conserved (see Supplemental Table 3 for residue comparison). This likely also suggests a complimentary divergence of the corresponding binding surfaces of the telomerase from both yeast and humans.



**Figure 3.9 Mapping residues of *ScEst3* and *HsTPP1-OB* that are potentially involved in binding to telomerase**

Est3 and *HsTPP1-OB* share a common telomerase-association surface. Est3<sup>ΔN</sup> (tan) and *HsTPP1-OB* (gray) are superimposed. (a) Previously identified residues in Est3 that mediate binding to telomerase (E104, E114, T115, N117 and D166) [95] are depicted as orange space-fill, while recently identified residues in *HsTPP1-OB* that mediate binding to telomerase (D166, E168 and K170)[92], (D166-F172, L183 and E215)[93] and (E168, E169, E171, R180, L183, L212 and E215)[258] are shown as green space-fill. (b) 80° rotation around a horizontal axis shows that the telomerase interacting residues cluster at the base of the β-barrels. A surface representation demonstrates that the interacting residues form a protein-protein interaction surface.

### 3.4.2 Isolated ScEst3 does not bind telomeric ssDNA or enhance telomerase activity *in vitro*

Est3 has been reported to enhance telomerase activity of *C. albicans* [184] and *S. castellii* [183] telomerases *in vitro*. One mode of this activity enhancement can be through a potential interaction with telomeric ssDNA, by stabilizing the telomerase complex on the extending substrate. Therefore, in *S. cerevisiae* Est3, we wanted to test if purified ScEst3 enhances telomerase activity *in vitro* and if a direct interaction with telomeric ssDNA oligo can be detected. However, in the *in vitro* telomerase activity assay, we have been unable to detect a reproducible enhancement of activity in the presence of ScEst3. This might be because either ScEst3 is not a telomerase activity enhancer, unlike CaEst3 and ScasEst3 proteins or the conditions for the activity assay are not optimum for ScEst3's activity. At the same time our results also do not rule out the possibility of ScEst3 being a telomerase activity enhancer but that the *in vitro* assay available does not provide a complete supplement of *in vivo* partners required for ScEst3 activity. For example Est3 might require the presence of another holoenzyme subunit, Est1, for its activity. However, until late full-length Est1 has been difficult to express/ purify. Wuttke lab has recently been able to successfully express and purify Est1 and this might aid future efforts in testing if presence of Est1 is required for Est3's effect as an enhancer of telomerase activity.

In addition to absence of any detectable telomerase activity by isolated ScEst3, we were unable to detect any ssDNA binding activity using different

binding strategies (NMR  $^{15}\text{N}$  HSQC, EMSA and filter-binding). This again might indicate the requirement for additional factors for Est3's DNA binding function. Est3 might not interact with telomeric DNA and a different binding partner, as yet unidentified, might be involved in Est3's function.<sup>12</sup>

### **3.4.3 Long and structured L45 is a unique feature, not involved in Est3's telomere length regulation function**

One key and unexpected feature of Est3 is the extensive structure in L45. Except for F139, the sequence of L45 is variable indicating that this region may not be important in specific binding interactions of Est3. This is in contrast to a shorter L45 that forms the canonical ligand-binding cleft of OB-fold proteins along with L12 and assists in specific ligand binding [202,253]. In Est3, however, L45 sits over and occludes the canonical ligand-binding surface. The extensions in L45 may be a structural fixture in the Est3 proteins, acting as the "ligand" that occupies the canonical binding site, with the conserved F139 as the likely anchoring point. Consistent with this hypothesis, there is no evidence of flexibility in this long L45 either by  $^{15}\text{N}$ -HSQC peak intensity comparison or by heteronuclear NOE (Chapter 2). This ligand-like role of L45 could be an evolutionary feature that led to OB-fold of Est3 to lose ligand-binding function in the presence of some evolutionary advantage.

Alternatively, L45 could be a functional switch, displaced only in the presence of an as-yet-unidentified *in vivo* factor that does not require recognition of

conserved residues on L45. This factor may also work through indirect displacement of L45 by binding elsewhere on the protein. The displacement of L45 would then allow better access of a ligand to the binding surface. This hypothesis is supported by the absence of nucleic acid binding activity *in vitro* for purified Est3 proteins, but the report of cross-linked ssDNA binding activity when Est3 from *C. parapsilosis* is in complex with the essential N-terminal TEN domain of Est2 [183,191]. In other words, association with telomerase may promote L45 displacement and better access of ligand (ssDNA) to the canonical binding cleft of Est3. By binding to both telomeric ssDNA and telomerase, Est3 may assist in completing the proper assembly of holoenzyme onto the telomere, where recruitment has been initiated by Est1 [179]. This mutual dependence of Est1 and Est3 has been reported in *C. albicans* [184]. In addition to assembly, telomeric DNA interaction by Est3 might be important for its role in enhancement of telomerase activity as has been reported in *S. castellii* [183] and *C. albicans* [184] systems. Thus, L45 might be important in regulating Est3's telomere-length maintenance activities by controlling access of telomeric DNA to the binding surface of Est3. However, this is not the case, based on the *in vivo* wild type phenotype of the L45 deletion mutants, indicating that L45 is not involved in Est3's telomere length regulation function.

In conclusion, our studies on L45 indicate that either there is no requirement for ligand binding to the canonical ligand binding face of the OB-fold in Est3 or that *in vivo* either by some post-translational modification or by association with some factor the L45 is constitutively flexible, allowing unhindered ligand access. In other

words, L45 might either be an evolutionarily acquired structural plug, which blocks all ligand access, or it is a constitutively flexible structural artifact with no functional importance. Alternatively, L45 may be involved in some other function of Est3, independent of telomere-length regulation.

### **3.4.4 Region between $\beta 3$ and $\beta 4$ is the hub for telomere length regulation function of Est3**

Most of the previously and newly discovered telomere-length regulating residues of Est3 lie on the bottom of the  $\beta$ -barrel, highlighting this region's importance for Est3's telomere length regulation function; mostly via direct interaction with telomerase [95]. Some of the functionally important residues are also located in the unstructured N-terminal region (not included in our structure). This can be because of the likely proximity of this region to the telomerase-associating region of Est3 and hence a regulatory function of the N-terminus towards telomerase association is implied. In an unbiased surface mutagenesis screen we also identified some of the less conserved residues along  $\beta 1$  and C-tail as being functionally important for telomere length regulation. This highlights the importance of an unbiased surface mutagenesis strategy to screen for functionally relevant residues. Interestingly, the residues located on the  $\beta$ -sheet formed by strands  $\beta 4$  and  $\beta 5$  are not involved in telomere length regulation function, as indicated by the *in vivo* mutagenesis screen. Several of these residues are highly conserved and surface exposed, indicative of a thus far undiscovered function of

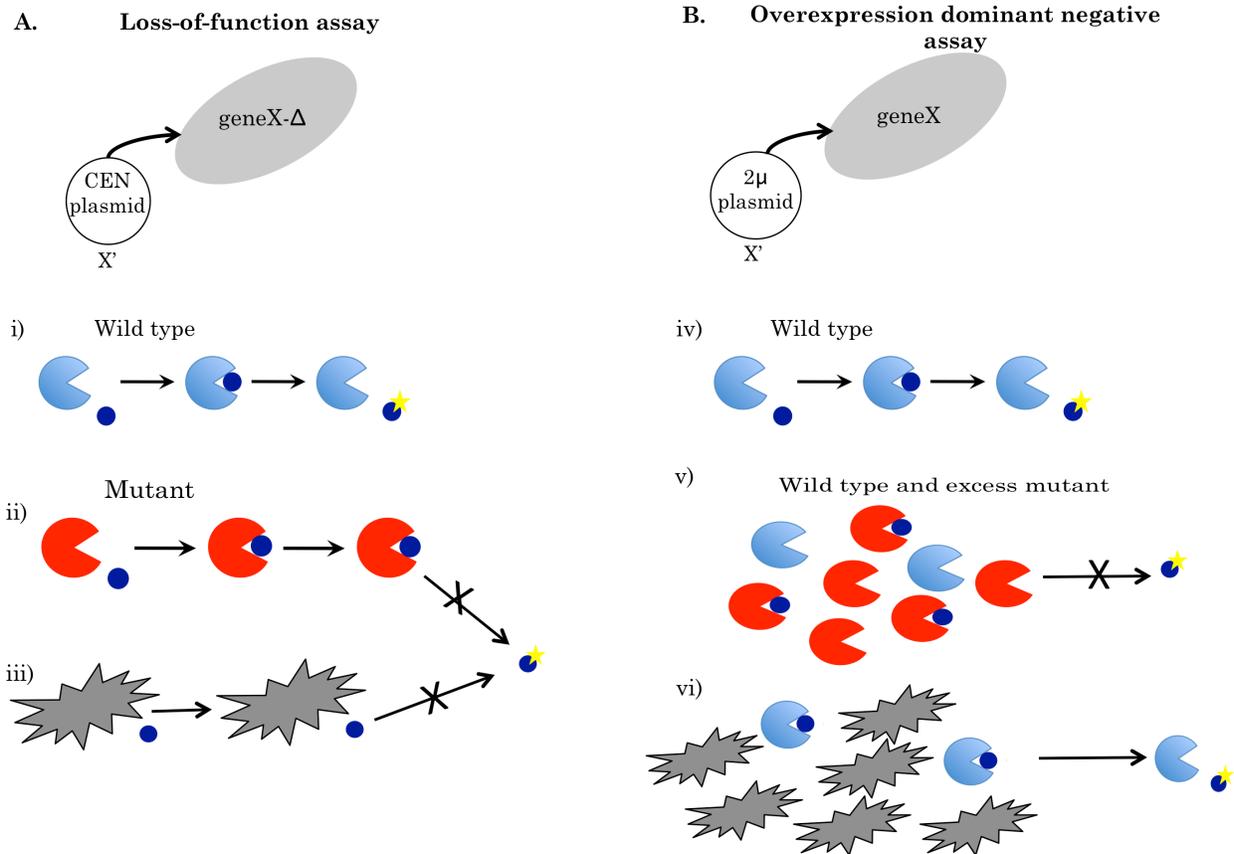
Est3 that requires these conserved residues for function. As an initial test of whether the Est3 protein might perform an additional role in genome maintenance that employs this surface, we tested whether defects in residues located on this surface might confer enhanced sensitivity to DNA damaging agents (HU, MMS and UV) (data not shown), however, none of the mutations did so. Future efforts will focus on screening for other functions as well, to define Est3's interaction network.

## **Chapter 4. Correlation of the *in vivo* over-expression dominant-negative phenotype with *in vitro* stability measurements: a new strategy for *in vivo* protein characterization**

### **4.1 Introduction**

#### **4.1.1 Loss-of-function assays fail to distinguish separation-of-function alleles from alleles that encode disrupted protein structure**

Separation-of-function mutants identify those residues on the protein that are important for protein's function. In this instance separation-of-function are missense mutations that alter function without impinging on the structural integrity of the protein. To identify separation-of-function mutants scientists generally design a loss-of-function assay that aims to assess the disruption of an activity of the protein when important residues involved in that process are mutated. A major caveat associated with the loss-of-function assay is that the mutants that are disrupted in the function being tested as well as the mutants that are structurally unfolded exhibit this loss of function (Figure 4.1A). Consequently, one ends up with a host of mutants, majority of which are no longer structurally viable.



**Figure 4.1** Cartoon representations of a loss-of-function and an overexpression dominant-negative case

**A. In a loss-of-function assay**, yeast strain (grey oval) deleted for the gene of interest is transformed with a CEN plasmid carrying either the wild type or the mutant copy of that gene i) If wild type is transformed, protein (blue pacman) interacts with a ligand (small dark-blue circle) and modifies it. ii) The mutant of this protein (red pacman) retains the structure of the wild type and binds the ligand, but is unable to modify it. (iii) Structurally unfolded mutant (grey with jagged edges) is unable to bind the substrate, again leading to no modified product.

**B. In the overexpression dominant-negative assay**, yeast strain (grey oval) expressing wild type copy of the gene of interest is transformed with high copy  $2\mu$  plasmid carrying either the wild type or the mutant copy of that gene iv) is similar to case i). v) The mutant allele (red pacman) is overexpressed in the presence of the wild-type allele and is able to bind the ligands away from the wild type. As the mutant allele is in excess, wild-type does not get to bind the ligand, leading to loss of ligand-modification function vi) Overexpressed but structurally unfolded mutant (grey with jagged edges) is unable to bind the substrate, allowing the wild type copies to bind and give modified substrate. (adapted from [264])

An assay testing for overexpression dominant negative phenotype of the mutants is more adept at identifying true separation-of-function mutants. The

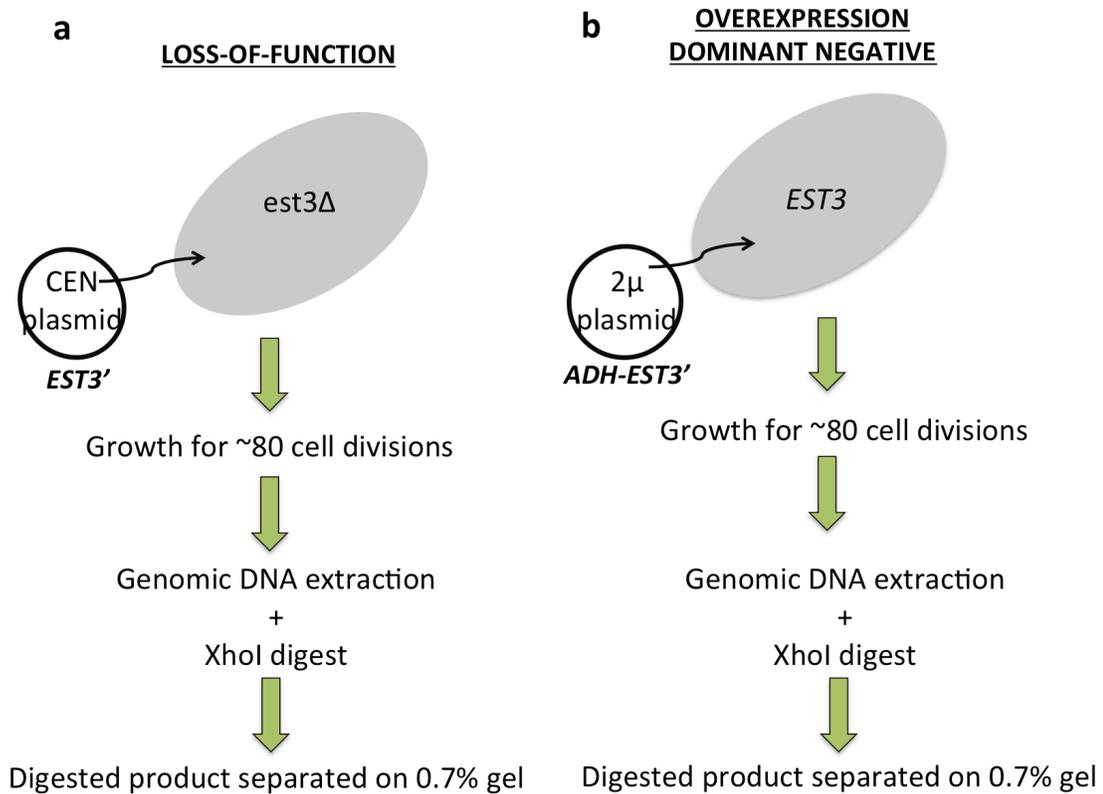
premise for this assay is that in a heterozygous mutant that expresses both a dominant negative and a wild-type copy of the protein, the protein containing the dominant negative mutation will show an inhibitory phenotype even in the presence of the wild-type protein [265]. This implies loss of some, but not all functions of the protein, thus enabling identification of functionally important residues or regions of the protein. The creation of inhibitory variants of proteins is possible because proteins can have multiple functional sites such as sites for oligomerization, substrate-binding, catalysis etc. and these sites can be mutated independent of each other (Figure 4.1) [264]. Overexpression of the dominant-negative genes can be used to saturate a cell with mutant forms of a particular protein, causing all functional effectors of that protein to be bound into inactive complexes [266] or in case of multimeric proteins contribute inactive monomers (Figure 4.1B v). Since the dominant negative protein is able to bind with the effector molecule or the wild type monomer to make inactive multimer, the assumption is that the dominant negative mutant proteins retain an intact and functional subset of the wild-type protein domains [264]. This is in contrast to when structurally unfolded proteins are overexpressed in presence of wild type. The unfolded mutants are unable to bind the substrate, which is available for wild type copies to interact with and give the modified end product (Figure 4.1B vi). Thus, using an overexpression dominant negative assay one should be able to cull out just the separation-of-function mutants from the full set of mutants being tested.

We have used *EST3* as our gene of interest, to identify loss-of-function and overexpression dominant negative mutants in collaboration with the Lundblad lab (Salk Institute). The assumption that the dominant negative mutants are structurally viable and the ones that are not dominant negative are structurally unfolded was tested by structural analysis using circular dichroism (CD).

#### **4.1.2 Est3 as a model protein for identification of loss-of-function mutants that also exhibit overexpression dominant negative phenotype *in vivo*\***

\* The assays described in this section are from an earlier publication [95] from Dr. Victoria Lundblad's lab. A similar strategy was used by John Lubin (Lundblad lab) to generate new set of loss-of function and dominant-negative mutants in Est3. Mutants were analyzed by both Lundblad and Wuttke labs. I performed the *in vitro* studies for structural stability.

Yeast strains expressing point mutants of *EST3* were generated and loss-of-function effect, if any, of these mutations was assessed as reduced telomere length and growth defects in the *est3-Δ* background [95] (Figure 4.2) (Lundblad lab, Salk Institute). This is because of the observation that *est3-Δ* yeast strain exhibits an EST (ever shorter telomeres) and cell senescence phenotype *in vivo* [57]. Residues for mutational analysis were selected on the basis of conservation and predicted surface-exposure. Mutations were to alanine and/ or a charge-swap mutation to enhance the effect of the mutation.

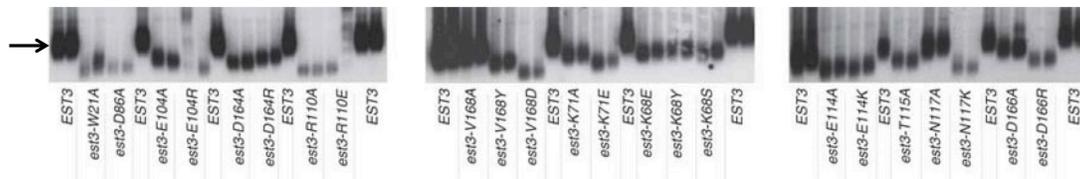


**Figure 4.2 Simplified scheme of the loss-of-function and overexpression dominant negative experiments**

*EST3'* denotes the mutant form of the *EST3* gene being transferred on the plasmide. *ADH* is the strong promoter for overexpression of the protein. Details of the experiments are explained in section 4.1.2

Loss-of-function was assessed in *est3Δ* background to prevent compensatory effect by any wild-type *Est3* copies in the cell. For this assay, a freshly derived *est3-Δ* strain was transformed with single-copy CEN plasmids expressing mutant or wild-type *Est3* (untagged and under the control of native *EST3* promoter) and cells were propagated for ~80 cell-divisions. Genomic DNA was extracted from these cells and subjected to cleavage by *XhoI* restriction enzyme and separated on 0.7% agarose gels. A southern-blot of this gel was used to test for telomere-length by probing with

telomere-specific oligo d(GT/CA) probes (Figure 4.3). As a result, several point-mutants of Est3 were identified that led to severe to moderate telomere-length defects compared to wild-type Est3. These mutants were W21A, D86A, E104A/R, D164A/R, R110A/E, V168Y/D, K71A/E, K68E/Y/S, E114A/K, T115A, N117A/K and D166A/R. However, to distinguish which of these mutants manifest loss-of-function phenotype due to severely defective protein stability versus overall well-folded protein where the mutation has disrupted a biochemical function, the mutants were also tested for dominant-negative effect. A mutant that is dominant-negative would suggest the protein is folded and able to out-compete the endogenous wild-type protein whereas a mis-folded protein would not be able to disrupt the activity of the endogenous wild type protein.

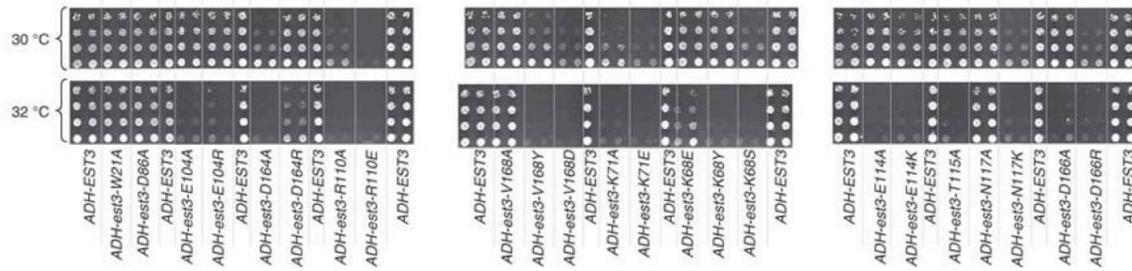


**Figure 4.3 Telomere-length analysis of the Est3 mutants in a loss-of-function assay**

Single copy plasmids of wild type and mutant alleles were transformed into *est3Δ* strain background. Cells were propagated for ~80 cell-divisions and subsequently tested for telomere-length after separation of XhoI-digested genomic DNA on 0.7% agarose gel and probing the southern-blot with telomeric d(GT/CA) probes. The black arrow marks region of observation on the blot (from[95])

The Lundblad lab has reported using two assays to identify dominant-negative phenotypes [95] (Figure 4.2). One of the assays tests for synthetic lethality in an *EST3 yku80-Δ* strain. *yku80-Δ* background was used because this strain is highly sensitive to defects in telomerase, leading to observable phenotypes of growth defect

or cell lethality [267]. The *yku80-Δ* strains are already defective in telomere maintenance, therefore introduction of a mutant allele defective in telomere length maintenance will affect the cells drastically, whereas if the mutant allele is not defective, then no additional destabilization of telomere length will happen. Since the expectation is growth defect, the study examined growth of *est3* mutants at both 30 °C and 32 °C because growth defects in *yku80-Δ* background are enhanced at even slightly higher temperatures than 30 °C. This conditional lethality allows for selection of colonies that survive at the permissive temperature but otherwise die at non-permissive temperature. High copy (2 μ) plasmids expressing wild-type and mutant alleles of *EST3*, under control of ADH promoter, were transformed into a *yku80-Δ YKU80 URA3* strain and examined for growth by plating five-fold serial dilutions of equivalent numbers of cells on media that selects for loss of *YKU80* plasmid. Mutation of highly conserved W21A and D86A were predicted to be internalized and important for structure core stabilization [95] and, accordingly, perturbation of either of these two residues failed to confer a dominant negative phenotype (Figure 4.4). Conversely, the mutations in predicted surface-exposed residues, *i.e.*, whose side chains are not involved in internal contacts that are important for structural stability, conferred modest to severe synthetic lethality in the *yku80-Δ* strain as seen from effect on growth at the non-permissive temperature of 32 °C. Thus, K68Y, K68S, K71A, K71E, E104A, E104R, R110A, R110E, E114A, E114K, T115A, N117K, D164A, D166A, D166R, V168Y and V168D mutations in Est3 were identified as dominant negative mutations from this assay.

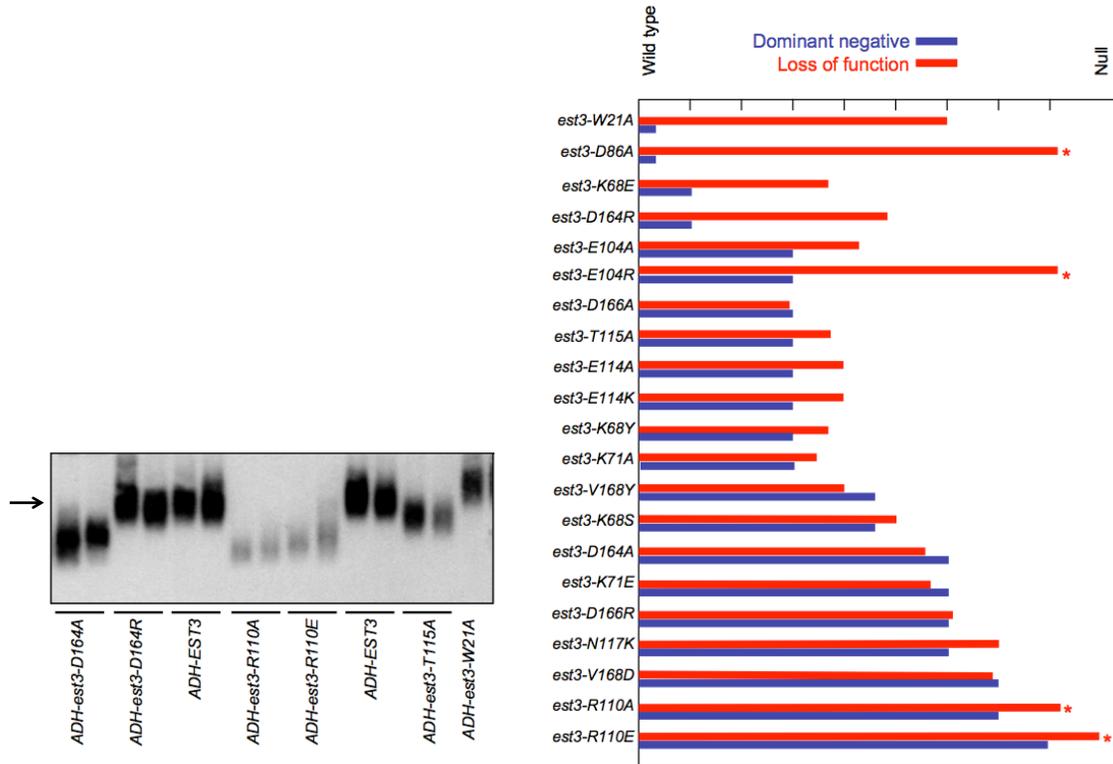


**Figure 4.4 Synthetic lethality test for mutants in Est3 protein**

Est3 mutant alleles were overexpressed in a *yku80*  $\Delta$  strain expressing wild-type Est3 protein. Equivalent number of cells were plated on selective media plates and tested for growth at 30°C and 32°C because of the observation that *yku80*  $\Delta$  strain are sensitized to defects in telomerase machinery and even modest increase in temperature enhances the growth defects. (from [95])

A second assay to identify dominant negative mutants was to examine telomere length in the *EST3 YKU80* strain. An *EST3* wild type strain was transformed with high copy (2  $\mu$ ) plasmids expressing either wild type or mutant alleles of *EST3* and examined for telomere length after  $\sim 75$  generations of propagation (Figure 4.5). When wild type was overexpressed, telomere length reduced by  $\leq 50$  bp. Overexpression of Est3-W21A mutant had no effect on telomere length, suggesting that this mutant is non-functional. This non-functionality is likely because this protein is unfolded. Conversely, mutations in predicted surface-exposed side-chain residues like E104R, D166A/R, D164A/R, R110A and T115A show increased loss of telomere length relative to that observed with the overexpressed wild-type control (Figure 4.4). These mutations, in addition to K68Y, K68S, K71A, K71E, R110E, E114A, E114K, N117K, V168Y and V168D, were thus dominant negative and the mutant proteins were likely structurally intact but disrupted in some functional region of the Est3 protein. The study showed a good

correlation between the dominant negative mutants identified by both the assays, synthetic lethality and telomere length.



**Figure 4.5 Dominant-negative effects of Est3 mutants on telomere-length**

A subset of Est3 mutants tested for dominant-negativeness is assessed for telomere length in this southern blot. Arrow indicates the region of interest on the gel, pointing to wild-type length. For the overexpression dominant-negative, protein were expressed under a strong, ADH, promoter, unlike loss-of-function assay where native *EST3 promoter* was used. Panel on the right correlated dominant-negative and loss-of-function effects. Mostly, the mutants that are dominant-negative also show equivalent loss-of-function effect, except for E104R. (from [95])

### 4.1.3 Telomerase-association versus separation-of-function mutants

Key residues in Est3 protein that exhibit a dominant negative phenotype, were further tested by our collaborators for their ability to disrupt association with telomerase [95]. Since TLC1 RNA is a key core component of telomerase, the ability

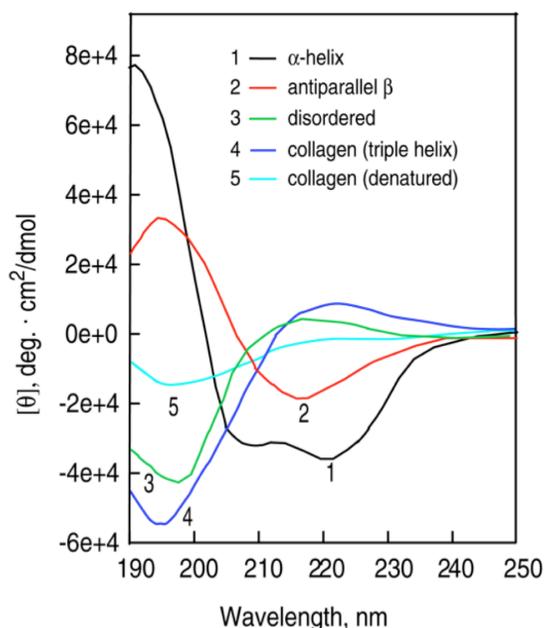
of Flag tagged wild type and mutant Est3 proteins to co-immunoprecipitate TLC1 was tested. Northern blot analysis of anti-Flag immunoprecipitates was done using probes to detect TLC1 RNA and the unrelated U1 RNA. As the Est2 protein was untagged, probing for RNA confers the advantage of not generating an anti-body for the protein component. Many of the strong dominant negative mutants like K71E, R110E and D164A were still able to efficiently pull-down TLC1, whereas some others like E104A/R, E114A/K, T115A and N117K were defective in TLC1 association. Based on this test, the identified residues were separated into 2 categories: ones that are involved in association of Est3 to telomerase and ones that are involved in some other, as yet un-identified, function of Est3. Thus, a separation of function was proposed. Since all of the mutants identified as disrupting the telomere maintenance function of Est3 were picked on the basis of their overexpression dominant negative phenotype, we wanted to pick a strategy for relatively quick assessment of structure of these mutant proteins. CD is one such technique for structure assessment.

#### **4.1.4 CD as a tool for protein structure assessment**

To address the question of whether mutants identified as over-expression dominant negative, in the *in vivo* screen, are structurally intact, we decided to use CD as the tool for assessing secondary structure.

CD arises from the differential absorption of the left and the right circularly polarized light [268,269]. CD signals arise only when absorption of the polarized light occurs by chromophores that are either naturally associated with the molecule

under study or are artificially attached to it. In proteins, chromophores that absorb are the peptide bond (below 240 nm), aromatic amino acid side chains (absorb from 260 nm to 320 nm) and disulfide bonds (around 260 nm) [269,270]. A chromophore will be optically active (chiral) if it is intrinsically chiral because of its asymmetric structure, is covalently linked to a chiral center in the molecule or is placed in an asymmetric environment because of the tertiary structure adopted by the molecule [269]. The different types of regular secondary structures like  $\alpha$ -helix,  $\beta$ -sheet and random coil give rise to characteristic CD spectra in the far UV (180 to 250 nm) range (Figure 4.5). The spectra in the near UV (260 to 320 nm) can also be used to obtain a fingerprint of the tertiary structure of a protein [269]. Thus, CD spectra can be used to observe structural retention or differences between native and recombinant proteins [270] or between wild-type and mutant protein versions [271]. Loss of CD signals upon thermal or chemical denaturation are also used as tools to assess structural stability of the wild-type and mutant proteins [271,272].



**Figure 4.6 Representative far-UV circular-dichroism spectra for secondary structure elements of proteins**

Spectra of proteins with completely  $\alpha$ -helical (black),  $\beta$ -sheet (red) and random coil (green) content and placental collagen in its triple helix (blue) and denatured (cyan) forms. (from [268]).

Instrumentation: The CD light source is a high-pressure Xenon lamp. A monochromator and linear polarizer is placed in its path, to select a particular wavelength for light that passes through. This linearly polarized light is then turned into left- or right-circularly polarized light using a photo-elastic modulator (PEM) alternately and made to pass through the sample. A detector is placed on the other side of the sample, to record intensity of the transmitted light and the intensity difference between the left- and right-circularly polarized light reported as the CD signal. If the sample is not chiral, there is no intensity differential between of the left- and right-circularly polarized light. A chiral sample will absorb the two lights differentially and a difference is reported either in ellipticity or absorbance units.

In collaboration with Dr. Victoria Lundblad's lab (Salk Institute), we have undertaken to determine whether the Est3 mutants that show dominant negative phenotype are also structurally stable. The behavior of dominant-negative mutants is premised on the fact that they aren't unfolded proteins, although many mutations would be expected to radically destabilize proteins. To the best of our knowledge, no one has ever tested this with in vivo and in vitro characterized proteins. We have tested several Est3 mutants that show both dominant-negative behavior and just loss-of-function phenotype, for intact structure using circular dichroism (CD) analysis. The results from this study, on Est3, can be used as a general proof-of-concept for mutational analysis of other proteins. In the absence of structure, the direct correlation of overexpression dominant-negative phenotype to intact protein structural core allows identification of functionally relevant residues. In light of the newly solved high-resolution structure of Est3, the correlation of dominant-negative phenotype leading to identification of solvent exposed residues was further confirmed by mapping the identified residues on the structure of Est3.

## **4.2 Materials and Methods**

### **4.2.1 Generating Est3 mutants for the CD stability study**

The *S. cerevisiae* Est3 gene was cloned into the pET-His-Smt3 vector (a generous gift from Chris Lima, Sloan-Kettering) between the BamHI and NotI sites using 5' -d(cgggatccatgccgaaagtaa)- 3' forward primer and 5' -

d(ttgcgccgcttataaaatatttatatacaaatgggaaagtacttaacgatccg)- 3' reverse primer [204] as described in Chapter 2. Mutants of SUMO-Est3 were generated by doing Site-directed mutagenesis in His<sub>10</sub>-Smt3 plasmid containing the wild-type Est3 sequence. The forward and reverse primers for the mutagenesis were designed using the Quikchange Primer Design program ([www.genomics.agilent.com](http://www.genomics.agilent.com)). The formula used for calculating the melting temperature of the primers was:

$$T_m = 81.5 + 0.41(\%GC) - 675/N - \%mismatch$$

where:

N = primer length

% GC = percentage of GC content in the primer

Designed primers for generating mutants are listed in Table 4.1. Primers were ordered from Integrated DNA Technologies. The Quikchange PCR reaction was essentially that of the Quikchange manual with minor modifications. The reaction set-up was as following: 5 µl of 10× reaction buffer, 1 µl (5–50 ng) of dsDNA template, 2.5 µl (125 ng) of oligonucleotide forward primer, 2.5 µl (125 ng) of oligonucleotide reverse primer, 4 µl of dNTP mix (2.5 mM), 1 µl of dUTPase, 1 µl of *Pfu* DNA polymerase (2.5 U/µl) and ddH<sub>2</sub>O to a final volume of 50 µl.

All mutations were confirmed by sequencing.

**Table 4. 1 Primers for point-mutants of Est3 for CD study**

**Table 4.1 Primers used for generating Est3 mutants**

Est3 mutant	Forward Primer	Reverse Primer
K3E	ggtggatccatgccggaagtaattctggagtctc	gagactocagaattacttccggcatggatccaccaatc
L6E	gatccatccgaaagtaattgaggagtctcattcaaagcc	tgaatgagactcctcaattacttccggcatggatccacca
W21A	cagactcagttttctacaaccagcgataaaaggcattaattgaagacaac	gttgtctcaattaatgcctttatcgctggttagaaaaactgagtctg
I22A	gactcagttttctacaaccatgggcaaaggcattaattgaagacaactc	gagttgtctcaattaatgcctttgcccattggttagaaaaactgagtc
I22E	gactcagttttctacaaccatgggagaaggcattaattgaagacaactcgg	ccgagttgtctcaattaatgccttctccatgggttagaaaaactgagtc
D86A	ggtatacgcacgataagagctcctcacaccaaactctg	caagtattggtgtgaggaagcctctatcgatcgtatacc
K68A	ttaaccaatccgtgccatttcccgcaattacaaaatttataacgtttgc	gcaaacgttataaaatttgaattcgccgaaatggcacggattggttaa
K68E	ttaaccaatccgtgccatttcccgagattacaaaatttataacgtttgcg	cgcaaacgttataaaatttgaatctcgccgaaatggcacggattggttaa
K71E	ccgtgccatttcccaaaattacagagtttataacgtttgcgactacaag	cttgtagtcgcaaacgttataaaactctgtaatttggcgaatggcacgg
E104R	ccaagagtggtatctaaatttagaaggactcataaattgcaggatc	gategtcgaattatgagtccttctaaaattagatacacactcttggg
R110A	tgaaggactcataaattgcgcgatcacatctgagacgacc	ggtcgtctcagatgtgatcgcgaattatgagtcctttca
V157A	gtgcccttctcaatgcaaaaccaggccacgata	tatcgtggcctggtttgcattgagaacggggcac
V157E	gtgcccttctcaatgcaaaaccaggccacgata	tatcgtggcctggtttcattgagaacggggcac
D164A	gtaaacaggccacgatattgccattgatcaagtcggatcg	cgatccgacttgatcaatggcaaatatcgtggcctggtttac
D164R	gtaaacaggccacgatattcgtattgatcaagtcggatcg	cgatccgacttgatcaatgcaaatatcgtggcctggtttac
D166A	gtaaacaggccacgatattgatattgccaagtcggatcg	acgatccgacttgggcaatatcaaatatcgtggcctggtttac
D166R	gtaaacaggccacgatattgatattcgtcaagtcggatcg	acgatccgacttgacgaatatcaaatatcgtggcctggtttac
V168E	ccacgatattgatattgatcaaggggatcgttaagtactttccatttg	caaatgggaaagtacttaacgatccctcttgatcaatatcaaatatcgtgg
F139A	caaattctcgacaaatgtctcagccaaaatttgctcaagcaacttcc	gaaatgttcttaggcaaaatttggcgtgagacattgctcgagaatttg

All listed oligos are in 5' to 3' direction

## 4.2.2 Expression and solubility test of Est3 and its point mutants

A set of mutant Est3 proteins, selected from the large set tested *in vivo* in the dominant-negative and loss-of-function assays, was recombinantly expressed and purified for *in vitro* stability measurements.

Expression: His<sub>10</sub>-SUMO-Est3 proteins were expressed in *E. coli* BL21(DE3) cells.

The cells were grown in Luria Bertani broth at 37°C to an O.D.<sub>600</sub> of ~1.0 and then cold-shocked on ice for ~1h. Protein expression was induced with 1.0 mM IPTG.

Cells were allowed to grow post-induction for 24 hrs at 15°C and harvested by centrifugation.

Solubility test: Cell pellets were resuspended in lysis buffer (buffer A) containing 100 mM potassium phosphate buffer (pH 7.5), 100 mM Na<sub>2</sub>SO<sub>4</sub>, 10% glycerol, 10 mM imidazole and 3 mM βME with an EDTA-free protease inhibitor cocktail tablet (Roche). PMSF was added just before cell lysis by sonication, followed by cell debris separation by centrifugation at 25,000xg for 20 minutes. The soluble cell lysate (supernatant) was decanted into a fresh 50 mL conical tube. The pelleted cell debris was resuspended with lysis buffer and brought to same volume as the supernatant. The pellet and supernatant fractions were mixed with SDS running dye. 5 – 7 μl of this mix was loaded onto a 13 % acrylamide SDS-PAGE gel and the proteins were separated by running the gel for ~1.0-1.5 hrs at 210 volts. Running buffer for separating proteins on the SDS-PAGE gel was: 15 g/ L Tris base, 72 g/ L glycine, 5 g/ L sodium dodecyl sulfate (SDS) and sterile water (to make up the 1 L volume).

Western blot analysis against the His tag was used to assess solubility. Equal volume aliquots of the pellet and supernatant fractions were mixed with SDS sample buffer and proteins separated on a 13% SDS-PAGE gel. The gel was transferred to nitrocellulose at 100V for ~1 h. The membrane was blocked for 1 h with blocking buffer (1xTBS, 0.05% Tween-20 and 1% BSA) followed by incubation with HisProbe-HRP (Pierce) in the blocking solution for 1 h at room temperature. The membrane was washed twice with 1xTBS-0.05%Tween-20 followed by two washes with 1xTBS for 15 minutes each. Protein bands were detected by chemiluminescence. The solubility of the proteins was determined by comparing the His<sub>10</sub>-SUMO-Est3 bands from the pellet and supernatant fractions.

### 4.2.3 Purification of Est3 and its point mutants

Soluble Est3 mutants were purified to homogeneity. Cell pellets were resuspended in buffer A containing 100 mM potassium phosphate buffer (pH 7.5), 100 mM Na<sub>2</sub>SO<sub>4</sub>, 10% glycerol, 10 mM imidazole and 3 mM βME with an EDTA-free protease inhibitor cocktail tablet (Roche) and PMSF added just before cell lysis. Cells were lysed by sonication and clarified by centrifugation at 25,000xg for 20 minutes. Cells were lysed by sonication and clarified by centrifugation. The clarified cell lysate was subjected to Ni<sup>2+</sup>-affinity chromatography by gravity-flow (GE Healthcare). The His<sub>10</sub>-SUMO-Est3 proteins were subsequently eluted with buffer A containing 300 mM imidazole and concentrated in a 10,000 MWCO concentrator (Millipore) to 2 mL, without buffer exchange, and incubated overnight at 4°C with Ulp1 to cleave the His<sub>10</sub>-SUMO tag [204]. Cleaved monomeric proteins were purified via size-exclusion chromatography (Superdex75, GE) on an AKTA FPLC system in buffer B (100 mM potassium phosphate buffer (pH 7.5), 100 mM Na<sub>2</sub>SO<sub>4</sub>, 5% glycerol and 3 mM βME). Contaminating ~14 kDa His<sub>10</sub>-SUMO tag was separated from the ~20.6 kDa Est3 protein by a second round of Ni<sup>2+</sup>-affinity chromatography in buffer B containing 20 mM imidazole. Est3 eluted at >95% purity and was concentrated and buffer exchanged to the CD buffer (50 mM potassium phosphate buffer (pH 7.4)) using a 9000 MWCO protein concentrator (Pierce).

#### 4.2.4 CD melting curves: acquisition and analysis

All CD experiments were carried out in a 0.05 cm cuvette using a ChirascanPlus spectropolarimeter (Applied Photophysics) equipped with a Quantum Mechanics Peltier temperature controller. Buffer baselines were acquired using the flow-through buffer from the protein concentrator for each mutant. Thermal denaturation spectra of ~10  $\mu$ M Est3 proteins were collected from 180 to 300 nm in 1 nm increments, over a melting range of 5°C to 75°C with steps of 10°C and equilibration times of 3 minutes per step. Additional data points, with same wavelength range and equilibration time, were collected between 30°C to 50°C with steps of 2°C for better estimation of the melting transition. Data acquisition and initial processing like baseline subtraction and curve smoothing was performed using ProData Viewer (Applied Photophysics). Further processing was done in Microsoft Excel. The data was recorded in millidegree (*mdeg*) units and the equation used to convert *mdeg* to Mean Residue Ellipticity MRE (a unit more comparable between different samples and concentrations):

$$\text{MRE} = mdeg / (10 \times C_r \times l)$$

where:

mdeg = raw CD data

$C_r$  = mean residue molar concentration

$l$  = path length in cm

and,

$$Cr = (n \times 1000 \times c_g) / M_r$$

where:

n = number of residues in the protein

$c_g$  = protein concentration in g/ mL

$M_r$  = molecular weight of species

The units of MRE are  $\text{deg.cm}^2.\text{dmol}^{-1}.\text{residue}^{-1}$

(<http://dichroweb.cryst.bbk.ac.uk/html/userguide.shtml>)

Integration of the 30°C to 50°C (2°C step size) data into the 5°C to 75°C (10°C step size) data was done and melting curves at 208 nm and 222 nm were fitted with the Boltzmann sigmoid equation using the Prism GraphPad software and mid-point of the curves calculated from the equation.

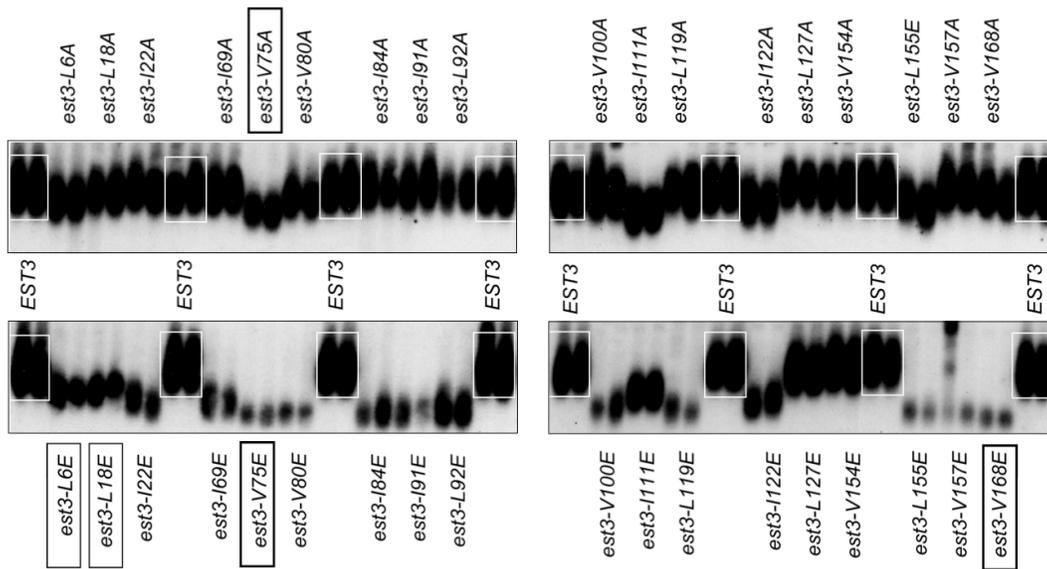
## 4.3 Results

### 4.3.1 Point-mutants of Est3, tested for loss-of-function and over-expression dominant negative effect, reveal importance of several residues in Est3's role in telomere-length regulation \*

The Lundblad Lab have reported, in previously published data [95], their results for test of loss-of-function and dominant negative phenotype for a number of Est3 mutants (see section 4.1.2 for details).

In a continuation of this loss-of-function study, mutations in several hydrophobic residues of Est3 were tested for telomere-length defects upon change to

either alanine or to the charged glutamate (Figure 4.7) (John Lubin, Lundblad lab, unpublished). 18 isoleucine, leucine and valine residues mutated and tested for telomere-length effect were L6, L18, I22, I69, V75, V80, I84, I91, L92, V100, I111, L119, I122, L127, V154, L155, V157 and V168. Compared to the wild-type length, most hydrophobic residues to alanine mutations did not show severe telomere-length reduction, except for V75A, I111A, I122A and L155A and a moderate reduction for L6A. On the other hand, mutations at these sites to a charged glutamate resulted in telomere-length reduction for all 18 hydrophobic residues, with severely short telomere-lengths for most and moderately short for L6E and L18E.



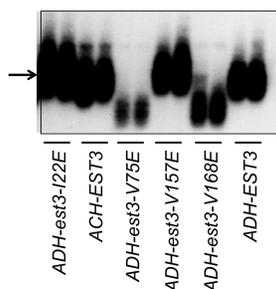
**Figure 4.7 Length defect in loss-of-function mutations in hydrophobic residues of Est3**

Eighteen hydrophobic (Ile, Leu and Val) residues of Est3 were mutated to alanine (top panel) and glutamate (bottom panel). The length defects in the loss-of-function assay were assessed by comparison to telomere length in wild-type transformed strain. Telomere-length in the strain containing wild-type plasmid is highlighted with white box. Mutants boxed in black are the ones that showed over-expression dominant-negative phenotype, in addition to the loss-of-function (reduced length) shown here. Mutants that are not boxed in black showed loss-of-function (reduced length of telomeres) but not overexpression-dominant-negative phenotype (from John Lubin, Lundblad lab, unpublished data).

To test how many of these hydrophobic-to-charged residue mutants are loss-of-function mutants because of overall intact structure of the protein, mutants were tested for overexpression dominant-negative (ODN) effect as well. Results of the loss-of-function and ODN assays are summarized in Table 4.2. A dominant-negative effect on telomere-length is also shown for a subset of these 18 mutants in the southern-blot gel in Figure 4.8. Out of all the 18 mutants tested, only L6E, L18E, V75E and V168E were dominant-negative, suggesting that these residues are surface exposed and conceivably involved in some interaction with a factor (maybe Est2) that is important for mediating Est3's role in telomere-length maintenance. Point mutations of these residues to a charged residue resulted in moderate to severe telomere-length defects. Mutations in any other hydrophobic residue exhibited no dominant-negative effect leading to the hypothesis that they are likely present in the protein core and involved in contacts important for structural stability of Est3. To determine whether this correlation between dominant-negative phenotypes and structural stability holds true, we have undertaken a study to assess the stability of a subset of Est3 mutant proteins by recombinantly expressing them and testing the purified proteins for solubility and secondary structure stability *in vitro*.

**Table 4. 2** Result of loss-of-function and dominant-negative assays for the hydrophobic to –alanine and –charged residue (Glu/E) mutations (from John Lubin, Lundblad Lab, unpublished data)

Mutant	→ A loss of function	→ E loss of function	→ E of ODN	Level of conservation
L6	slightly shorter	med short	moderate	moderate
L18	wild type	med short	moderate	high
I22	wild type	<b>short</b>	no	high
I26	wild type	wild type	no	high
I69	wild type	<b>short</b>	no	high
V75	med short	<b>NULL</b>	<b>SEVERE</b>	high
V80	slightly shorter	<b>NULL</b>	slight	high
I84	wild type	<b>short</b>	no	high
I91	wild type	<b>short</b>	no	high
L92	wild type	<b>short</b>	no	high
V100	wild type	<b>short</b>	slight	high
I111	med short	med short	no	high
L119	wild type	<b>NULL</b>	no	high
I122	slightly shorter	med short	no	INVARIANT
L127	wild type	wild type	no	high
V154	wild type	wild type	no	high
L155	slightly shorter	<b>NULL</b>	no	INVARIANT
V157	wild type	<b>NULL</b>	no	high
V168	wild type	<b>NULL</b>	<b>SEVERE</b>	high



**Figure 4.8** Telomere-length defects in a subset of overexpression dominant-negative mutants in hydrophobic residues of Est3

A subset of the mutants tested in the dominant-negative assay are shown here for telomere-length comparison between strains with wild-type Est3 and its mutant alleles in *EST3* background. For the overexpression dominant-negative, protein were expressed under a strong, ADH, promoter, unlike loss-of-function assay where native *EST3 promoter* was used (from John Lubin, Lundblad lab, unpublished data).

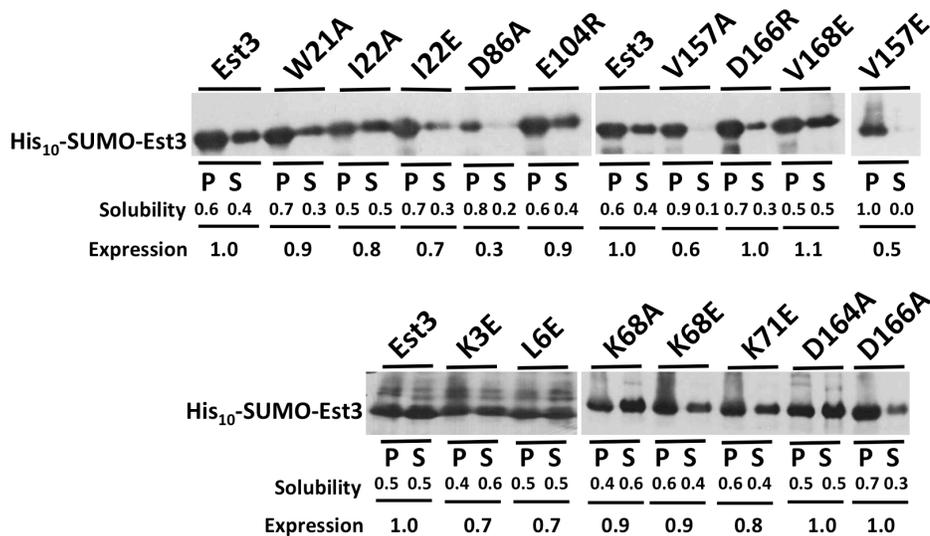
### 4.3.2 Point mutations in *S. cerevisiae* Est3 affect expression and solubility of the recombinant proteins

Dominant negative mutations show an inhibitory phenotype when both the dominant negative and wild type copy of the protein are present in the cell. To be functionally active, the dominant negative mutants must retain an intact and functional subset of the wild-type protein domains, which presumably requires that the overall structure remain intact [264]. To test whether this assumption holds true, we tested several key mutations from the dominant negative screen of *S. cerevisiae* Est3 for structure stability of the recombinantly expressed Est3. Structure stability was assessed based on expression and solubility of the His<sub>10</sub>-SUMO tagged Est3 mutants and secondary structure analysis by CD of the purified and untagged soluble mutants of Est3.

Five representative subsets of the mutations identified in the *in vivo* dominant-negative study were chosen to evaluate the relationship between *in vivo* phenotype and protein stability: i) mutations in the putative structural core of Est3, W21A and D86A; ii) mutants affecting Est3's association with telomerase, E104R, D166A and D166R; iii) mutants affecting an unidentified function of Est3, K68A, K68E, K71E, R110A, D164A, D164R and V168E; iv) mutations in conserved hydrophobic residues L6E, I22A, I22E, V157A, V157E and V168E; and v) mutants in N-terminal, functionally relevant, region of Est3, K3E and L6E.

Wild type and mutant Est3 proteins were initially tested for expression and solubility by evaluating the partitioning of recombinant *E. coli* expressed protein

into insoluble versus soluble fractions after cell lysis by Western blot analysis (Figure 4.9). The wild-type SUMO-Est3 protein expressed well and partitioned almost equally between the insoluble and soluble fractions. Expression levels similar to wild type are observed for most of the mutant proteins (summarized in Table 4.3 and Table 4.4). V168E exhibited slightly increased expression relative to wild-type, whereas somewhat reduced expression was observed for K3E, L6E, I22E and V157A/E mutants. D86A expression was most markedly reduced; exhibiting expression levels only 30% of the wild type.



**Figure 4.9 Western-blot of recombinantly expressed His<sub>10</sub>-SUMO-Est3 wild type and its mutants**

Expression and solubility of the wild-type His<sub>10</sub>-SUMO-Est3 (labeled Est3) and its point mutants were assessed by Anti-His Western blot. Solubility was assessed by quantitating the partitioning of the His<sub>10</sub>-SUMO-Est3 proteins into pellet (P) and supernatant (S) fractions after the bacterial cell-lysis. Expression was quantitated relative to the wild type. Image quantitation performed using the ImageJ software

**Table 4. 3 Comparison of point-mutants of *S. cerevisiae* Est3, recombinantly expressed from *E.coli* cells**

Type of Est3 point-mutant	Mutant name	Expression <sup>a</sup>	Solubility <sup>b</sup>	Mid-point of 208 nm mel (°C) <sup>c</sup>
Wild-type	WT	+++	+++	39
N-terminal	K3E	+++	+++	Alternate conformation after 45°C
	L6E	+++	+++	39
Structure-destabilizing	W21A	+++	++	-
	D86A	+	-	-
Telomere-association effect	E104R	+++	+++	35
	D166A	+++	++	37
	D166R	+++	++	36
Separation-of-function (non-telomere association)	K68A	+++	+++	33
	K68E	+++	++	Unfolded
	K71E	+++	+++	36
	R110A	--	--	37
	D164A	+++	+++	39
	D164R	--	--	37
Hydrophobic-to-charged residue	I22A	+++	+++	-
	I22E	++	+	-
	V157A	++	-	-
	V157E	++	-	-
	V168E	++++	+++	40

<sup>a</sup>Expression level for WT SUMO-Est3 is represented by +++, slightly reduced expression by ++ and severely reduced expression by +. Increased expression is represented by +++++. Expression levels were based on the quantitation of the western blot in Figure 1.

<sup>b</sup>The levels of solubility (division of protein into inclusion body versus soluble supernatant upon cell-lysis) represented by similar notation as for expression<sup>a</sup>. - indicates “not soluble” and -- indicates not assessed by western-blot.

<sup>c</sup>Curve-fitting of the melting-curve at 208 nm was done with Boltzmann sigmoid equation using the Prism GraphPad software and mid-point of the curves calculated from the equation. “-” indicates no data acquired as these proteins could not be purified.

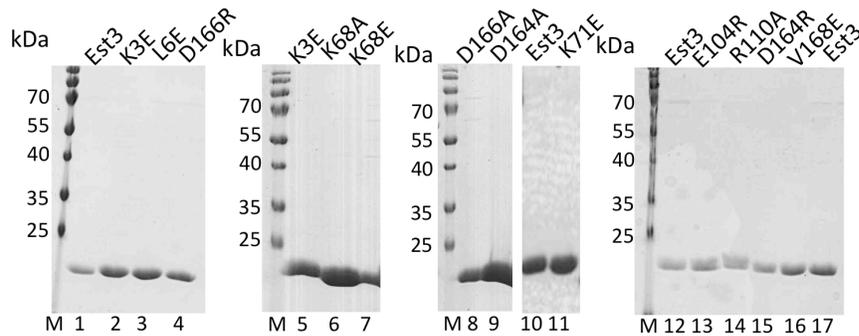
Overall, the solubility of the proteins roughly paralleled their expression levels in that poorly expressed proteins also were more prone to partition to the insoluble state. Of the mutants that expressed at similar levels as wild type, most also partitioned similar to the wild-type Est3. The reduced expression of D86A and V157A/E mutants correlated with the isolation of very little, if any, soluble protein. The expression and solubility result also correlates well with the *in vivo* dominant

negative screen as the mutants most affected in expression and solubility, D86A and V157A/E, were also not dominant negative, whereas other mutants, except for W21A and I22A/E, were also dominant negative to variable degrees. Interestingly, the mutations in hydrophobic residues I22 and V157 led to reduced expression and/or drastically reduced solubility and this effect was exacerbated for hydrophobic-to-charged mutations of I22E and V157E. This suggests that these hydrophobic residue side-chains are likely internalized and important for structural integrity of the protein. This is not true for all the conserved hydrophobic substitutions, L6 and V168 hydrophobic substitutions, as their mutations to charged residues did not impact either the expression or solubility of the protein. This suggests that these residues are not involved in structural stability but in some other function of the protein, consistent with the observation that these mutants show moderate and strong dominant negative effect, respectively, *in vivo*.

Further purification of soluble proteins by Ni<sup>2+</sup>-affinity and size-exclusion chromatography, as described in 4.2.3, resulted in some mutants forming soluble aggregates. For example, the mutants W21A and I22A eluted in the void of the size-exclusion column and I22E precipitated entirely out of solution after Ni<sup>2+</sup>-affinity purification. As these mutants exhibited no dominant negative phenotype *in vivo*, loss of protein due to aggregation reinforces the correlation between dominant negative phenotype and structural stability. The mutant Est3 proteins that were successfully isolated as monomers with >95% purity were K3E, L6E, K68A, K68E, K71E, E104R, R110A, D164A, D164R, D166A, D166R and V168E.

### 4.3.3 Secondary structure analysis of Est3 mutants by CD correlates well with their *in vivo* dominant negative phenotypes

The secondary structure of the recombinantly expressed and purified wild-type Est3 and soluble point-mutants K3E, L6E, K68A, K68E, K71E, E104R, R110A, D164A, D164R, D166A, D166R and V168E was evaluated by circular dichroism (CD) at 5°C. The purity of the proteins used for CD analysis was assessed by separation on SDS-PAGE gel (Figure 4.10).

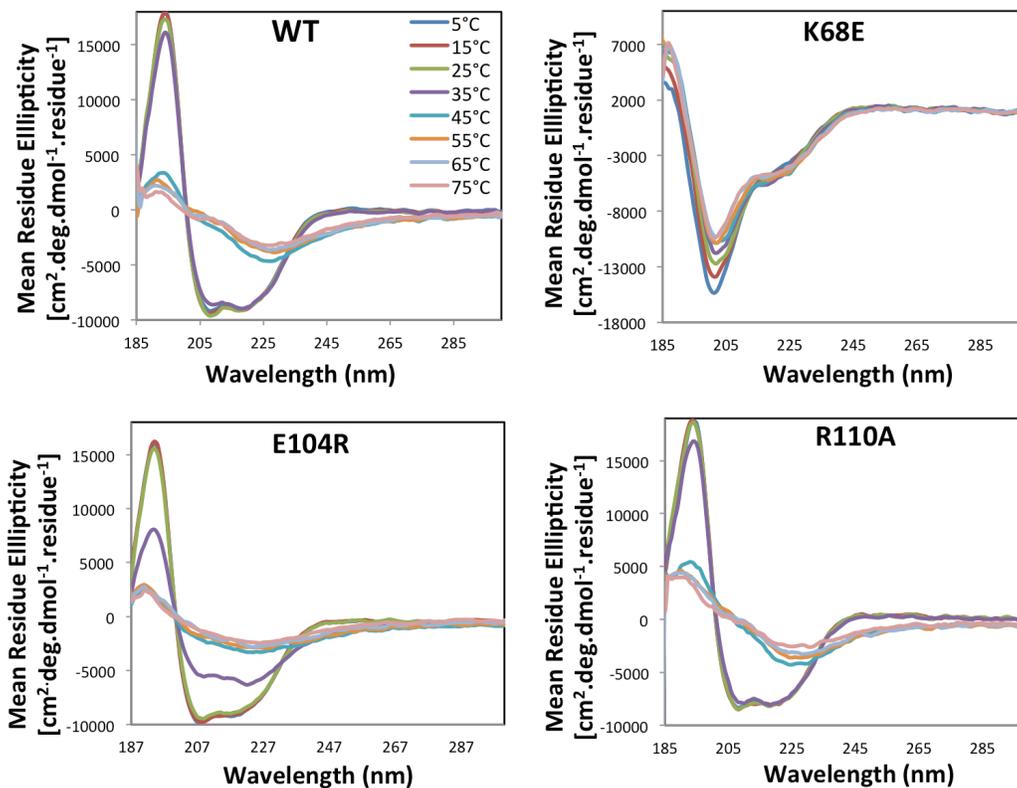


**Figure 4.10 Purity of the Est3 wild type and mutants used for CD analysis**

Coomassie-stained gels of the *S. cerevisiae* WT Est3 and its mutants expressed from *E. coli* BL21(DE3) cells and purified to >95% purity. These purified proteins were used for the secondary structure analysis by circular dichroism.

The spectrum of wild-type Est3 is representative of a mixed  $\alpha/\beta$  structure, with double minima at 208 nm and 222 nm and a global maximum at 195 nm (Figure 4.11). For the wild-type protein, this spectrum remains constant with respect to temperature until 35 °C, after which increased temperature unfolds the protein. This thermal transition is complete by 55°C and no further change in secondary structure was observed above this temperature. Figure 4.11 compares the

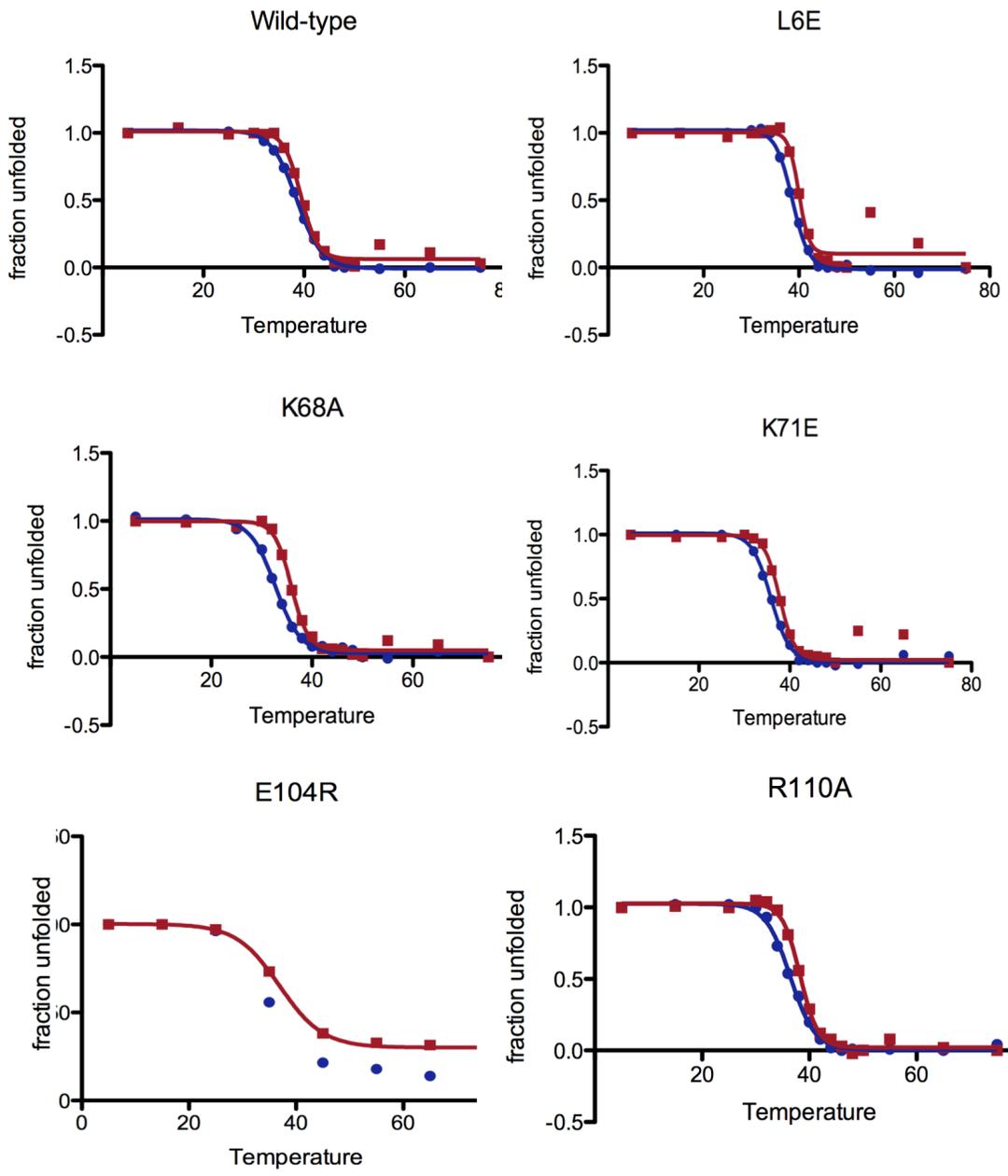
representative spectra of three mutants of Est3 with the wild-type spectrum and these mutants exhibit a range of denaturation behaviors.

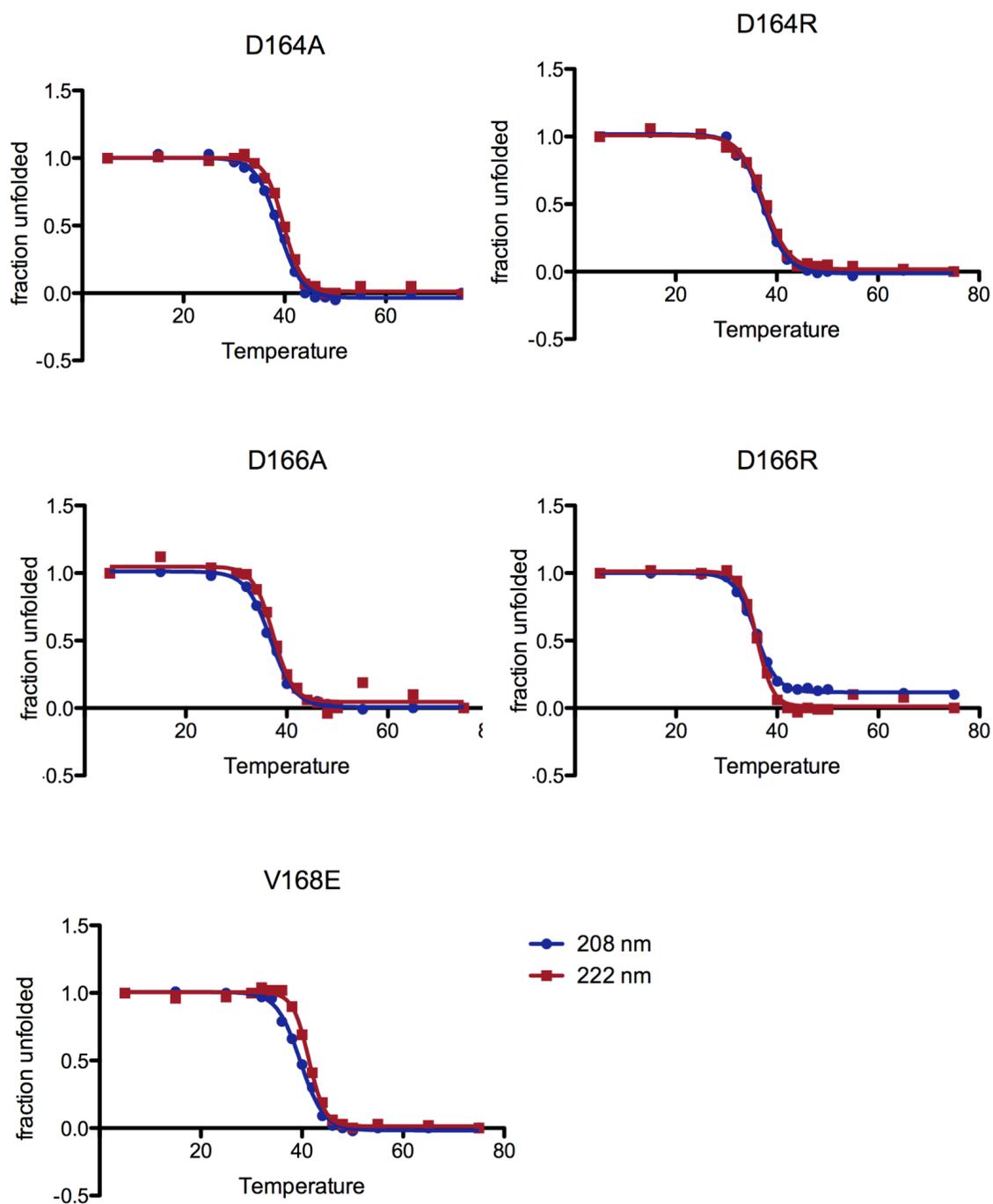


**Figure 4.11 Secondary structure analysis from the circular dichroism spectra of wild-type Est3 (WT) and its point mutants as a function of temperature**

Far-UV CD-spectra were collected at protein concentrations of  $\sim 10 \mu\text{M}$  in 50 mM potassium phosphate buffer (pH 7.4) from 5 °C to 75 °C with step-size of 10°C. Color-code for the temperatures at which the spectra were collected is shown on the plot-area of the WT. The WT spectrum resembles that of a mixed  $\alpha/\beta$  structure, with double minima at 208 and 222 nm and a global maximum at 195 nm. Spectrum of K68E resembles that of an unfolded protein. E104R and R110A spectra are similar to WT, with E104R showing reduced CD signal at 35°C unlike WT and R110A.

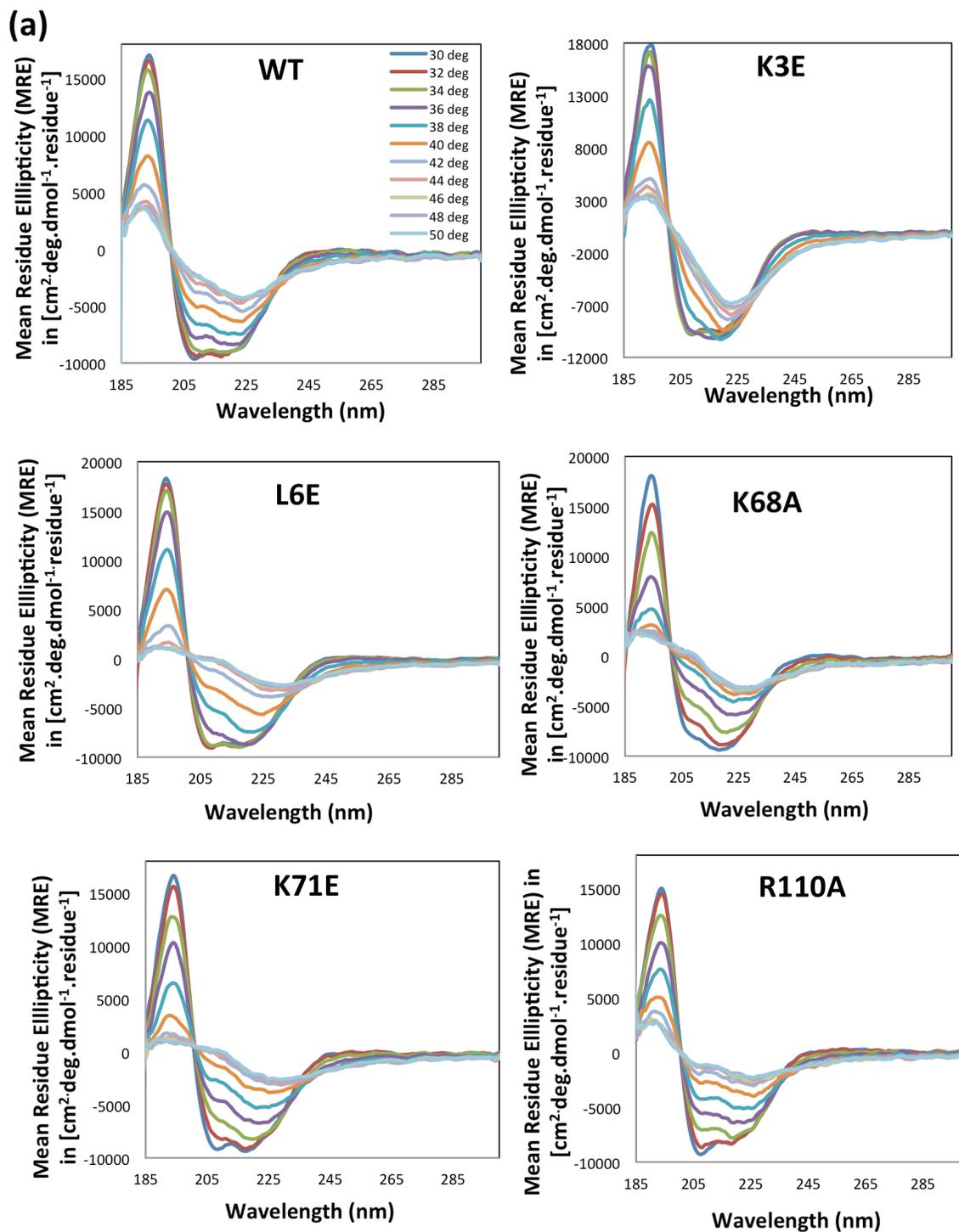
The mid-points of the melting curves at 208 nm and 222 nm estimated (Figure 4.12) for the wild type Est3 and its point mutants integrated data from 30 °C to 50 °C (2 °C step size) data (Figure 4.13) into the 5 °C to 75 °C (10°C step size) data (Figure 4.14) are reported in Table 4.3 and Table 4.4.



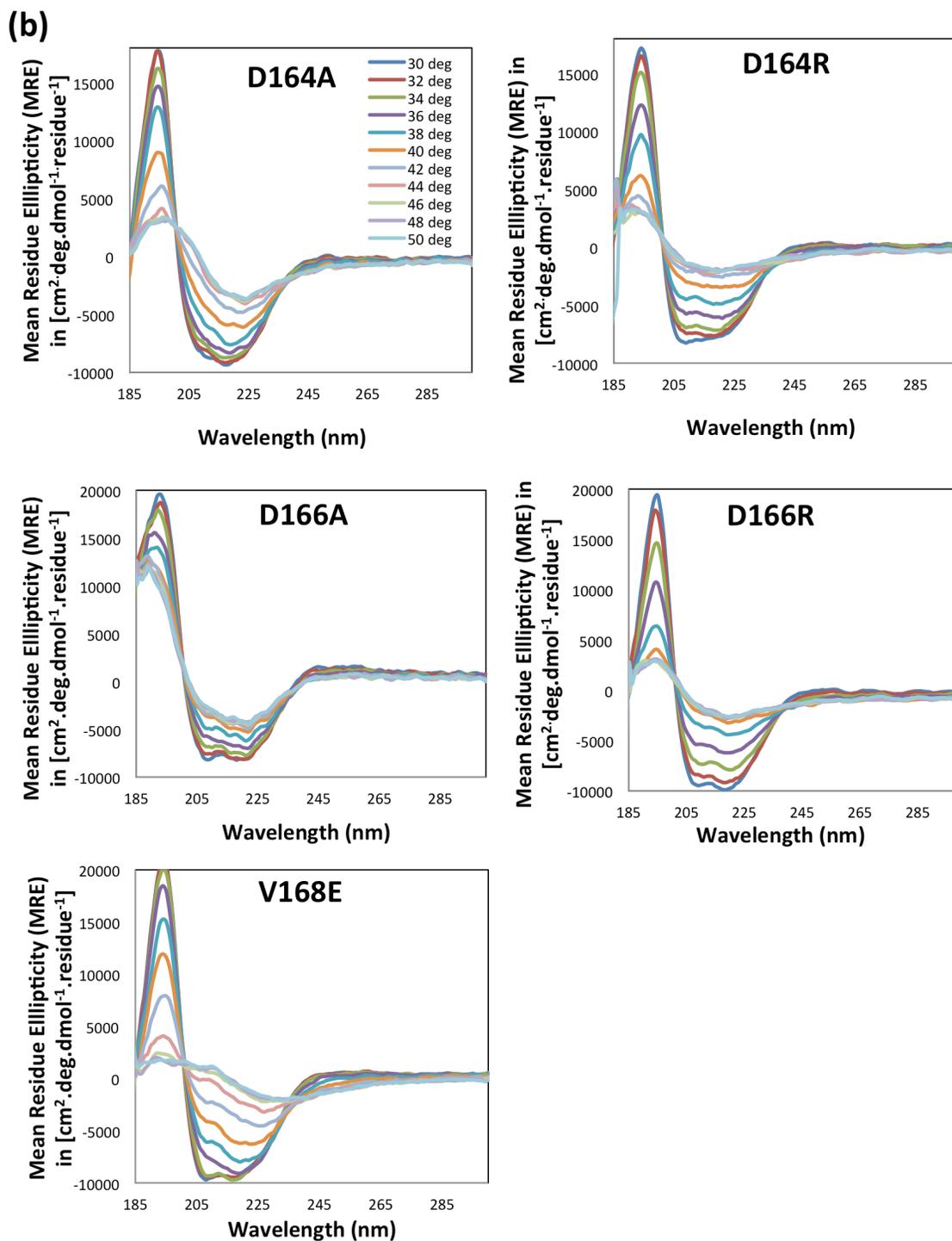


**Figure 4.12 Melting curve at 208nm and 222nm for wild type and mutant S3 protein**

Curve-fitting of the melting-curve at 208 nm and 222 nm was done with Boltzmann sigmoid equation using the Prism GraphPad software and mid-point of the curves calculated from the equation.

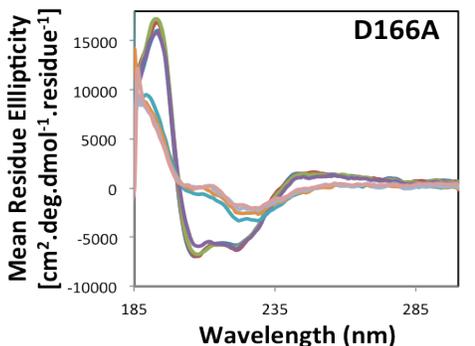
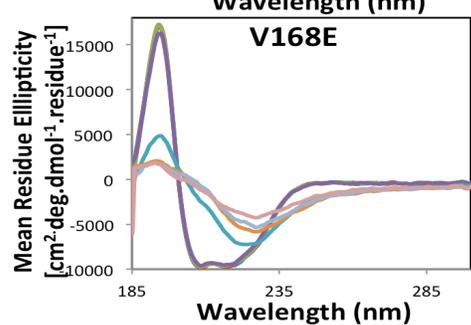
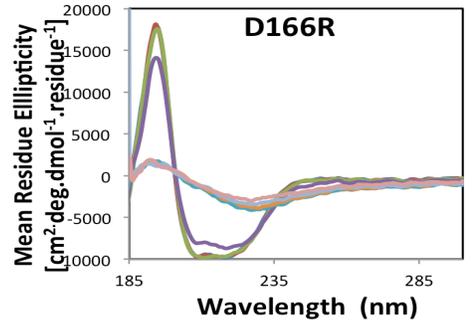
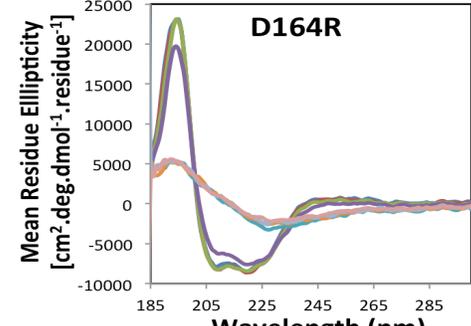
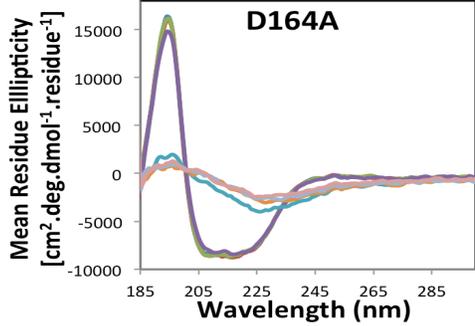
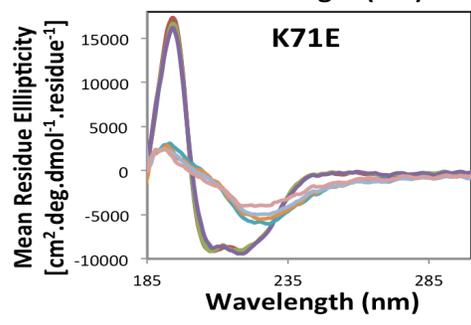
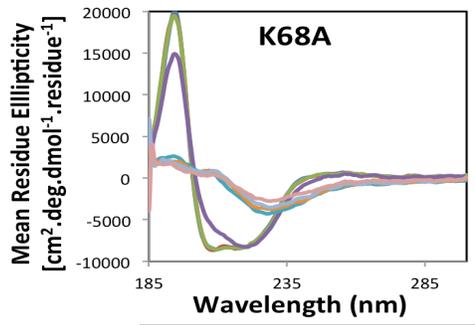
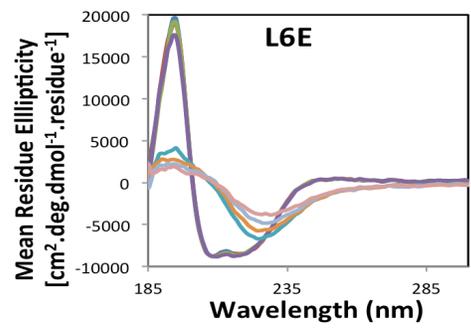
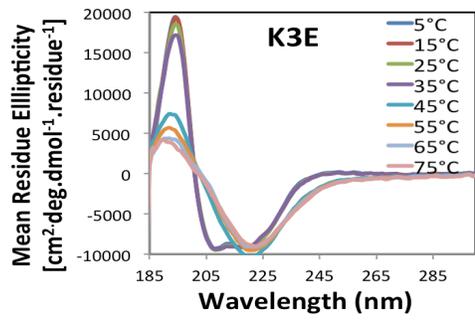


continued....



**Figure 4.13 Circular Dichroism spectra of Est3 mutants recorded from 30°C to 50°C**

Est3 wild type and its mutant were subjected to CD spectral analysis, recording far UV CD spectra from 30°C to 50°C with a step-size of 2°C each. The complete set is shown in (a) and (b).



**Figure 4.14 Secondary structure analysis from the circular dichroism spectra of point mutants of Est3 as a function of temperature**

Far-UV CD-spectra were collected at protein concentrations of ~10  $\mu$ M in 50 mM potassium phosphate buffer (pH 7.4) from 5°C to 75°C with step-size of 10°C. Color-code for the temperatures at which the spectra were collected is shown on the plot-area of K3E mutant.

The spectrum of R110A mutant shows wild type like secondary structure content and similar thermal stability and this correlates well with the strong dominant negative effect of this mutant *in vivo*. Conversely, the charge-swap K68E mutant has a native structure that is primarily denatured into a different state with a single minimum around 200 nm (Figure 4.11) whereas mutation of this residue to alanine, K68A, gives a protein with an unfolded CD spectrum similar to wild-type but overall reduced in thermal stability. Again, this correlates well with the weak and very weak dominant negative effects observed for K68A and K68E mutants respectively *in vivo*. A third case is the E104R mutant that shows a moderate dominant negative effect *in vivo* and the CD spectrum suggests reduced thermal stability compared to the wild-type because of loss of secondary structure after 25°C (Figure 4.11). Similar impact of reduced stability is seen in the D164R mutant, which has a very weak dominant negative effect (Table 4.3, Table 4.4 and Figure 4.14). Surprisingly, despite being a strong dominant negative mutant, D166R was also found to be slightly less stable than wild type with mid-point of melting curve for D166R at 36°C, 3°C lower than the wild type mid-point (Table 4.3 and Figure 4.13). Est3 mutants K71E, D164A (Figure 4.14) and D166A have similar thermal stability as the wild type and this correlated with the strong to moderate dominant negative effect observed for these mutants *in vivo*. Interestingly, as observed for

D164 and D166 positions, the less conservative change to alanine resulted in wild type like structural stability, whereas a charge-swap mutation to arginine produced mutants with reduced stability. Hydrophobic-to-charged residue mutants L6E and V168E had spectra and thermal stability similar to wild type (Table 4.3 and Figure 4.11) confirming their involvement in some other function of Est3 than structural stability. K3E, charge-swap mutation in a N-terminal residue, produced an atypical denaturation curve (Figure 4.14), where the signal at ~222 nm did not decrease after 35°C (n=2), indicating a retention of  $\beta$ -sheet content or some other effect that is not readily apparent.

## 4.4 Discussion

### 4.4.1 Mutant proteins with dominant-negative effect *in vivo* are structurally stable

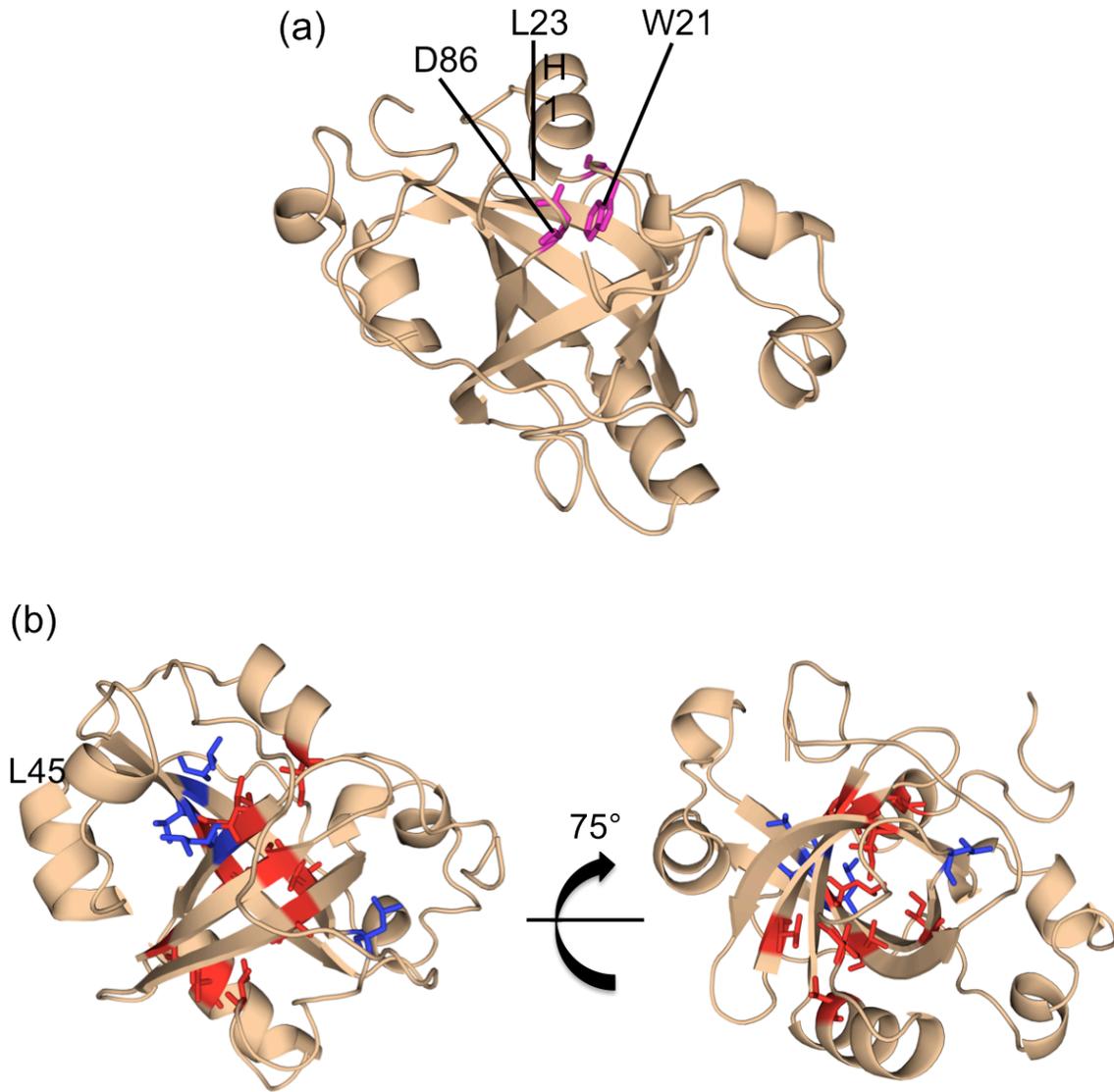
This correlation study of *in vivo* over-expression dominant-negative phenotype with *in vitro* stability measurements supports the two-step *in vivo* mutant selection strategy from the Lundblad lab. The two-step strategy proposes, firstly, the screening of mutants in a dominant-negative assay to assess both the effect of the mutation and also to choose those mutations that show the effect despite overall retention of structure and not because of the mutation rendering the protein misfolded. This is really important, removing a caveat that plagues *in vivo* mutational analysis. The extent of effect by the mutation can then be assessed in the second

step of the strategy by a loss-of-function assay. Remarkable, all of the Est3 mutants that were not dominant-negative (W21A, D86A, I22A/E and V157A/E) were also found to be insoluble in our direct assessment of protein solubility. The opposite also held true, *i.e.*, all mutants that were dominant-negative could be purified *in vitro* and gave circular-dichroism spectra same or somewhat similar to the wild-type purified Est3. Although it could be purified, K68E gave a CD-spectrum of unfolded protein and this correlates with the very weak dominant-negative effect observed for this mutation. K3E mutant was the exception because of its atypical CD melting spectra and this behavior could not be predicted from just the observation of its dominant-negative effect. In conclusion, the test for dominant-negative phenotypes is a viable and important step to parse out those mutants from the *in vivo* mutagenesis screens that render the protein unstable and hence those residues are not determinants of protein function.

#### **4.4.2 Predictions from the *in vivo* dominant-negative and *in vitro* protein stability correlation study hold true in the context of the Est3 structure**

Although the residues chosen for Est3 mutagenesis screens were based on sequence alignment and rough model building, the conclusions from the dominant-negative assay for the location (surface or internalized) of tested residues holds true in the context of the Est3 protein structure reported in Chapter 2 of this thesis. With the high-resolution structure of Est3 in hand, the location of the residues tested in the mutagenesis screens could be mapped on the protein structure. W21

and D86 were predicted to be important for structural stability and as can be seen from their location on the structure, they are likely involved in holding the loop L23 and Helix H1 close together (Figure 4.15a). A long-range amide-amide NOE (section 2.3.3b) is also observed between W21 and S87, indicating that indeed L23 and H1 are in proximity. Therefore, the fact that disruption of either W21 or D86 destabilizes the Est3 protein is completely consistent with the structure. Hydrophobic residues (Table 4.2) that were predicted to be structurally important, because of a lack of dominant-negative phenotype upon mutation, mostly map to the  $\beta$ -barrel of the OB-fold and interestingly the side-chains of most project into the core of the barrel (Figure 4.15b). Those hydrophobic side-chains not projecting into the barrel are involved in interaction with other structural elements of Est3, like loop L45 etc. In fact, all residues that showed no dominant-negative effect upon mutation map to internal sites.

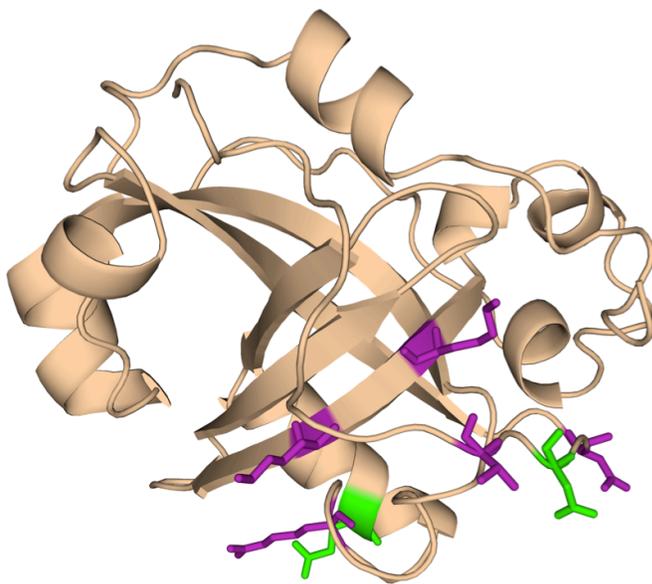


**Figure 4.15 Mapping of residues from Est3 mutagenesis screen that do not show dominant-negative effect**

(a) W21 and D86 residues are highlighted in magenta and loop L23 and helix H1 are labeled. (b) Hydrophobic residues with side-chains projecting into the core of the  $\beta$ -barrel are colored in red and hydrophobic residues that point towards other elements of the structure and not the barrel core are colored in blue. The structure on the left is rotated by  $75^\circ$  to show view of the  $\beta$ -barrel core (right panel) with side-chains projecting inwards.

Finally, the residues that were predicted to be functionally important, by exhibiting a dominant-negative phenotype, map onto the surface of Est3 with side-chains exposed to allow interaction with other *in vivo* factors (Figure 4.16). Thus, mapping

of the residues to the Est3 structure verified the conclusion that dominant-negative assay can correctly predict the location of the residues on the protein structure (whether exposed or internalized) and identify functionally important residues, distinguishable from those residues important for structural integrity. This is a pretty unique combination of *in vivo* and *in vitro* tools.



**Figure 4.16 Mapping of Est3 residues, identified from dominant-negative and loss-of-function assay, to be important for Est3's function**

Mutants that were dominant-negative were identified in residues that are exposed to the surface and involved in the Est3's function. Side-chains colored in green belong to residues important in telomerase-association of Est3 (E104 and D166). Residues in purple indicate residues involved in some other function of Est3 than telomerase association (K68, K71, R110, D164 and V168).

**Table 4. 4 Summary of the *in vivo* and *in vitro* data for protein stability analysis of Est3 mutants**

Mutant Name	Location Structure	<i>in vivo</i>				<i>in vitro</i>			
		Loss-of-function effect in estΔ background	ODN effect in EST3 background	Expression western-Blot	pellet vs sup western-Blot	partitioning into SEC void size exclusion chromatography	CD melting curve circular dichroism	208 nm (°C) melting curve	222 nm (°C) melting curve
WT				+++	0.6/0.4	negligible	regular 2° structure	39	39
N-term									
K3E	exposed	short telomeres	yes	+++	60/40	similar to WT	alternate conformation	39	40
L6E	exposed	medium short telomeres	moderate	+++	60/40	bit more in void than WT	regular 2° structure		
<b>Negative control</b>									
W21A	internalized	extremely short telomeres	No	++	0.7/0.3	All in void	-	-	-
D86A	internalized	extremely short telomeres	No	-	0.8/0.2	No protein	-	-	-
<b>Telomerase-association</b>									
E104R	exposed	extremely short telomeres	moderate	+++	0.6/0.4	similar to WT	regular 2° structure	35	37
D166A	exposed	medium short telomeres	moderate	+++	NR	similar to WT	regular 2° structure	37	37
D166R	exposed	short telomeres	strong	+++	0.7/0.3	similar to WT	regular 2° structure	36	36
<b>Separation-of-function</b>									
K68A	exposed	NR (not reported)	NR (not reported)	+++	0.4/0.6	similar to WT	regular 2° structure	33	36
K68E	exposed	medium short telomeres	very weak	+++	0.6/0.4	similar to WT	unfolded	unfolded	unfolded
K71E	exposed	short telomeres	strong	+++	0.6/0.4	similar to WT	regular 2° structure	36	38
R110A	exposed	extremely short telomeres	strong	Not recorded	NR	similar to WT	regular 2° structure	37	38
D164A	exposed	medium short telomeres	strong	+++	0.5/0.5	similar to WT	regular 2° structure	39	40
D164R	exposed	medium short telomeres	very weak	+++	0.7/0.3	similar to WT	regular 2° structure	37	38
<b>Hydrophobic to charged residues</b>									
I22A	internalized	wild type	No	+++	0.5/0.5	All in void	-	-	-
I22E	internalized	short telomeres	No	++	0.7/0.3	No protein	-	-	-
V157A	internalized	wild type	No	++	0.9/0.1	No protein	-	-	-
V157E	internalized	extremely short telomeres	No	++	1/0	No protein	-	-	-
V168E	exposed	extremely short telomeres	Severe	++++	0.5/0.5	similar to WT	regular 2° structure	40	43

## References:

- [1] H.J. Muller, The remaking of chromosomes, *Collecting Net*, 8 (1938) 181–198.
- [2] E. Gilson, E. Ségal-Bendirdjian, The telomere story or the triumph of an open-minded research, *Biochimie*, 92 (2010) 321–326.
- [3] E.H. Blackburn, J.G. Gall, A tandemly repeated sequence at the termini of the extrachromosomal ribosomal RNA genes in *Tetrahymena*, *J. Mol. Biol.*, 120 (1978) 33–53.
- [4] D. Cassidy-Hanley, Y. Bisharyan, V. Fridman, J. Gerber, C. Lin, E. Orias, J.D. Orias, H. Ryder, L. Vong, E.P. Hamilton, Genome-wide characterization of *Tetrahymena thermophila* chromosome breakage sites II Physical and genetic mapping, *Genetics*, 170 (2005) 1623–1631.
- [5] R.S. Coyne, D.L. Chalker, M.C. Yao, Genome downsizing during ciliate development: nuclear division of labor through chromosome restructuring, *Annu. Rev. Genet.*, 30 (1996) 557–578.
- [6] M. Yao, S. Duhaucourt, Genome-wide rearrangements of DNA in ciliates, *Mobile DNA II*, 2002.
- [7] D.R. Corey, Telomeres and telomerase: from discovery to clinical trials, *Chem. Biol.*, 16 (2009) 1219–1223.
- [8] J.W. Szostak, E.H. Blackburn, Cloning yeast telomeres on linear plasmid vectors, *Cell*, 29 (1982) 245–255.
- [9] V.A. Zakian, Structure and function of telomeres, *Annu. Rev. Genet.*, 23 (1989) 579–604.
- [10] E.H. Blackburn, Structure and function of telomeres, *Nature*, 350 (1991) 569–573.
- [11] N. Hug, J. Lingner, Telomere length homeostasis, *Chromosoma*, 115 (2006) 413–425.
- [12] A.M. Olovnikov, A theory of marginotomy: the incomplete copying of template margin in enzymic synthesis of polynucleotides and biological significance of the phenomenon, *J Theor Biol*, 41 (1973) 181–190.
- [13] J. Lin, E. Epel, E. Blackburn, Telomeres and lifestyle factors: roles in cellular aging, *Mutat. Res*, 730 (2012) 85–89.
- [14] T. de Lange, How telomeres solve the end-protection problem, *Science*, 326 (2009) 948–952.
- [15] K. Sakabe, K. Sugimoto, A. Sugino, R. Okazaki, Mechanism of DNA chain growth - possible discontinuity and unusual secondary structure of newly synthesized chains, *Proc. Natl. Acad. Sci. USA*, 59 (1968) 598.
- [16] T. Ogawa, T. Okazaki, Discontinuous DNA replication, *Annu. Rev. Biochem.*, 49 (1980) 421–457.
- [17] T.T. Chow, Y. Zhao, S.S. Mak, J.W. Shay, W.E. Wright, Early and late steps in telomere overhang processing in normal human cells: the position of the final RNA primer drives telomere shortening, *Genes Dev*, 26 (2012) 1167–1178.

- [18] P.P. Wu, M.M. van Overbeek, S.S. Rooney, T.T. de Lange, Apollo contributes to G overhang maintenance and protects leading-end telomeres, *Mol. Biol. Cell*, 39 (2010) 606–617.
- [19] P. Wu, H. Takai, T. de Lange, Telomeric 3' overhangs derive from resection by Exo1 and Apollo and fill-in by POT1b-associated CST, *Cell*, 150 (2012) 39–52.
- [20] L. Hayflick, The limited *in vitro* lifetime of human diploid cell strains, *Exp. Cell Res*, 37 (1965) 614–636.
- [21] J.D. Watson, Origin of concatemeric T7 DNA, *Nat. New Biol*, 239 (1972) 197–201.
- [22] R.J. Wellinger, A.J. Wolf, V.A. Zakian, *Saccharomyces* telomeres acquire single-strand TG1-3 tails late in S phase, *Cell*, 72 (1993) 51–60.
- [23] A.A. Neumann, R.R. Reddel, Telomere maintenance and cancer -- look, no telomerase, *Nat. Rev. Cancer*, 2 (2002) 879–884.
- [24] A. Bianchi, D. Shore, How telomerase reaches its end: mechanism of telomerase regulation by the telomeric complex, *Mol. Cell*, 31 (2008) 153–165.
- [25] M.T. Hemann, C.W. Greider, Wild-derived inbred mouse strains have short telomeres, *Nucleic Acids Res*, 28 (2000) 4474–4478.
- [26] M.J. McEachern, E.H. Blackburn, A conserved sequence motif within the exceptionally diverse telomeric sequences of budding yeasts, *Proc. Natl. Acad. Sci. USA*, 91 (1994) 3453–3457.
- [27] W. Palm, How shelterin protects mammalian telomeres, *Annu. Rev. Genet*, (2008).
- [28] M. Larrivé, C. LeBel, R.J. Wellinger, The generation of proper constitutive G-tails on yeast telomeres is dependent on the MRX complex, *Genes Dev*, 18 (2004) 1391–1396.
- [29] J. Prescott, E.H. Blackburn, Functionally interacting telomerase RNAs in the yeast telomerase complex, *Genes Dev*, 11 (1997) 11.
- [30] J. Meyne, R.L. Ratliff, R.K. Moyzis, Conservation of the human telomere sequence (TTAGGG)<sub>n</sub> among vertebrates, *Proc. Natl. Acad. Sci. USA*, 86 (1989) 7049–7053.
- [31] O. Samassekou, M. Gadji, R. Drouin, J. Yan, Sizing the ends: normal length of human telomeres, *Ann. Anat*, 192 (2010) 284–291.
- [32] N.K. Jacob, A.R. Stout, C.M. Price, Modulation of telomere length dynamics by the subtelomeric region of *Tetrahymena* telomeres, *Mol. Biol. Cell*, 15 (2004) 3719–3728.
- [33] W. Chai, Q. Du, J.W. Shay, W.E. Wright, Human telomeres have different overhang sizes at leading versus lagging strands, *Mol. Cell*, 21 (2006) 427–435.
- [34] J. Shampay, J.W. Szostak, E.H. Blackburn, DNA sequences of telomeres maintained in yeast, *Nature*, 310 (1984) 154–157.
- [35] M. Cohn, E.H. Blackburn, Telomerase in yeast, *Science*, 269 (1995) 396–400.

- [36] M. Cohn, M.J. McEachern, E.H. Blackburn, Telomeric sequence diversity within the genus *Saccharomyces*, *Curr. Genet*, 33 (1998) 83–91.
- [37] M.J. McEachern, J.B. Hicks, Unusually large telomeric repeats in the yeast *Candida albicans*, *Mol. Cell. Biol*, 13 (1993) 551–560.
- [38] A.W. Murray, N.P. Schultes, J.W. Szostak, Chromosome length controls mitotic chromosome segregation in yeast, *Cell*, 45 (1986) 529–536.
- [39] M.M. Dahlen, T.T. Olsson, G.G. Kanter-Smoler, A.A. Ramne, P.P. Sunnerhagen, Regulation of telomere length by checkpoint genes in *Schizosaccharomyces pombe*, *Mol. Biol. Cell*, 9 (1998) 611–621.
- [40] J.D. Griffith, L. Comeau, S. Rosenfield, R.M. Stansel, A. Bianchi, H. Moss, T. de Lange, Mammalian telomeres end in a large duplex loop, *Cell*, 97 (1999) 503–514.
- [41] L. Tomaska, J. Nosek, J. Kramara, J.D. Griffith, Telomeric circles: universal players in telomere maintenance? *Nat. Struct. Mol. Biol*, 16 (2009) 1010–1015.
- [42] L. Tomaska, A.M. Makhov, J.D. Griffith, J. Nosek, t-Loops in yeast mitochondria, *Mitochondrion*, 1 (2002) 455–459.
- [43] T. de Lange, T-loops and the origin of telomeres, *Nat. Rev. Mol. Cell Biol*, 5 (2004) 7.
- [44] T. de Lange, Shelterin: the protein complex that shapes and safeguards human telomeres, *Genes Dev*, 19 (2005) 2100–2110.
- [45] T. de Lange, Cell biology Telomere capping-one strand fits all, *Science*, 292 (2001) 1075–1076.
- [46] D.E.D. Gottschling, V.A.V. Zakian, Telomere proteins: specific recognition and protection of the natural termini of *Oxytricha* macronuclear DNA, *Cell*, 47 (1986) 195–205.
- [47] M.P. Horvath, V.L. Schweiker, J.M. Bevilacqua, J.A. Ruggles, S.C. Schultz, Crystal structure of the *Oxytricha nova* telomere end binding protein complexed with single strand DNA, *Cell*, 95 (1998) 963–974.
- [48] K. Paeschke, S. Juranek, T. Simonsson, A. Hempel, D. Rhodes, H.J. Lipps, Telomerase recruitment by the telomere end binding protein-beta facilitates G-quadruplex DNA unfolding in ciliates, *Nat. Struct. Mol. Biol*, 15 (2008) 598–604.
- [49] J. Berman, C.Y. Tachibana, B.K. Tye, Identification of a telomere-binding activity from yeast, *Proc. Natl. Acad. Sci. USA*, 83 (1986) 3713–3717.
- [50] M.N. Conrad, J.H. Wright, A.J. Wolf, V.A. Zakian, RAP1 protein interacts with yeast telomeres *in vivo*: overproduction alters telomere structure and decreases chromosome stability, *Cell*, 63 (1990) 739–750.
- [51] L.E. Hang, X. Liu, I. Cheung, Y. Yang, X. Zhao, SUMOylation regulates telomere length homeostasis by targeting Cdc13, *Nat. Struct. Mol. Biol*, 18 (2011) 920–926.
- [52] D. Shore, A. Bianchi, Telomere length regulation: coupling DNA end processing to feedback regulation of telomerase, *Embo J*, 28 (2009) 2309–2322.

- [53] C.M. Price, K.A. Boltz, M.F. Chaiken, J.A. Stewart, M.A. Beilstein, D.E. Shippen, Evolution of CST function in telomere maintenance, *Cell Cycle*, 9 (2010) 3157–3165.
- [54] B. Garvik, M. Carson, L. Hartwell, Single-stranded DNA arising at telomeres in *cdc13* mutants may constitute a specific signal for the RAD9 checkpoint, *Mol. Cell. Biol.*, 15 (1995) 6128–6138.
- [55] N. Grandin, C. Damon, M. Charbonneau, Ten1 functions in telomere end protection and length regulation in association with Stn1 and Cdc13, *Embo J.*, 20 (2001) 1173–1183.
- [56] N. Grandin, S.I. Reed, M. Charbonneau, Stn1, a new *Saccharomyces cerevisiae* protein, is implicated in telomere size regulation in association with Cdc13, *Genes Dev.*, 11 (1997) 512–527.
- [57] T.S. Lendvay, D.K. Morris, J. Sah, B. Balasubramanian, V. Lundblad, Senescence mutants of *Saccharomyces cerevisiae* with a defect in telomere replication identify three additional EST genes, *Genetics*, 144 (1996) 1399–1412.
- [58] A. Chandra, T.R. Hughes, C.I. Nugent, V. Lundblad, Cdc13 both positively and negatively regulates telomere replication, *Genes Dev.*, 15 (2001) 404–414.
- [59] E. Pennock, K. Buckley, V. Lundblad, Cdc13 delivers separate complexes to the telomere for end protection and replication, *Cell*, 104 (2001) 387–396.
- [60] H. Qi, V.A. Zakian, The *Saccharomyces* telomere-binding protein Cdc13p interacts with both the catalytic subunit of DNA polymerase  $\alpha$  and the telomerase-associated est1 protein, *Genes Dev.*, 14 (2000) 1777–1788.
- [61] J. Sun, Y. Yang, K. Wan, N. Mao, T.-Y. Yu, Y.-C. Lin, D.C. Dezwaan, B.C. Freeman, J.-J. Lin, N.F. Lue, M. Lei, Structural bases of dimerization of yeast telomere protein Cdc13 and its interaction with the catalytic subunit of DNA polymerase  $\alpha$ , *Cell Res.*, 21 (2010) 258–274.
- [62] C.-L. Hsu, Y.-S. Chen, S.-Y. Tsai, P.-J. Tu, M.-J. Wang, J.-J. Lin, Interaction of *Saccharomyces* Cdc13p with Pol1p, Imp4p, Sir4p and Zds2p is involved in telomere replication, telomere maintenance and cell growth control, *Nucleic Acids Res.*, 32 (2004) 511–521.
- [63] A.D. Gelinas, M. Paschini, F.E. Reyes, A. Héroux, R.T. Batey, V. Lundblad, D.S. Wuttke, Telomere capping proteins are structurally related to RPA with an additional telomere-specific domain, *Proc. Natl. Acad. Sci. USA*, 106 (2009) 19298–19303.
- [64] D. Jain, J.P. Cooper, Telomeric strategies: means to an end, *Annu. Rev. Genet.*, 44 (2010) 243–269.
- [65] D. Shore, K. Nasmyth, Purification and cloning of a DNA binding protein from yeast that binds to both silencer and activator elements, *Cell*, 51 (1987) 721–732.
- [66] A.R. Buchman, N.F. Lue, R.D. Kornberg, Connections between transcriptional activators, silencers, and telomeres as revealed by functional analysis of a yeast DNA-binding protein, *Mol. Cell. Biol.*, 8 (1988)

- 5086–5099.
- [67] B. Li, S. Oestreich, T. de Lange, Identification of Human Rap1, *Cell*, 101 (2000) 471–483.
  - [68] D.L. Levy, E.H. Blackburn, Counting of Rif1p and Rif2p on *Saccharomyces cerevisiae* telomeres regulates telomere length, (2004).
  - [69] E.A. Feeser, C. Wolberger, Structural and functional studies of the Rap1 C-terminus reveal novel separation-of-function mutants, *J. Mol. Biol.*, 380 (2008) 520–531.
  - [70] D. Loayza, T. de Lange, POT1 as a terminal transducer of TRF1 telomere length control, *Nature*, 423 (2003) 1013–1018.
  - [71] D.D. Liu, A.A. Safari, M.S.M. O'Connor, D.W.D. Chan, A.A. Laegeler, J.J. Qin, Z.Z. Songyang, PTOP interacts with POT1 and regulates its localization to telomeres, *Nat. Cell. Biol.*, 6 (2004) 673–680.
  - [72] B.A. Moser, T.M. Nakamura, Protection and replication of telomeres in fission yeast, *Biochem. Cell Biol.*, 87 (2009) 747–758.
  - [73] E. Abreu, E. Aritonovska, P. Reichenbach, G. Cristofari, B. Culp, R.M. Terns, J. Lingner, M.P. Terns, TIN2-tethered TPP1 recruits human telomerase to telomeres *in vivo*, *Mol. Cell. Biol.*, 30 (2010) 2971–2982.
  - [74] M.S. O'Connor, A. Safari, H. Xin, D. Liu, Z. Songyang, A critical role for TPP1 and TIN2 interaction in high-order telomeric complex assembly, *Proc. Natl. Acad. Sci. USA*, 103 (2006) 11874–11879.
  - [75] J.P. Cooper, E.R. Nimmo, R.C. Allshire, T.R. Cech, Regulation of telomere length and function by a Myb-domain protein in fission yeast, *Nature*, 385 (1997) 744–747.
  - [76] P. Konig, R. Giraldo, L. Chapman, D. Rhodes, The crystal structure of the DNA-binding domain of yeast RAP1 in complex with telomeric DNA, *Cell*, 85 (1996) 125–136.
  - [77] R. Court, L. Chapman, L. Fairall, D. Rhodes, How the human telomeric proteins TRF1 and TRF2 recognize telomeric DNA: a view from high-resolution crystal structures, *EMBO Rep*, 6 (2005) 39–45.
  - [78] D. Broccoli, S. Smolikov, L. Chong, T. de Lange, Human telomeres contain two distinct Myb-related proteins, TRF1 and TRF2, *Nat. Genet.*, 17 (1997) 231–235.
  - [79] A. Bianchi, S. Smith, L. Chong, P. Elias, T. de Lange, TRF1 is a dimer and bends telomeric DNA, *Embo J*, 16 (1997) 1785–1794.
  - [80] L. Fairall, L. Chapman, H. Moss, T. de Lange, D. Rhodes, Structure of the TRFH dimerization domain of the human telomeric proteins TRF1 and TRF2, *Mol. Cell*, 8 (2001) 351–361.
  - [81] D.L. Theobald, R.M. Mitton-Fry, D.S. Wuttke, Nucleic acid recognition by OB-fold proteins, *Annu. Rev. Biophys. Biomol. Struct.*, 32 (2003) 115–133.
  - [82] E. Gilson, M. Roberge, R. Giraldo, D. Rhodes, S.M. Gasser, Distortion of the DNA double helix by RAP1 at silencers and multiple telomeric binding sites, *J. Mol. Biol.*, 231 (1993) 293–310.
  - [83] A. Smogorzewska, T. de Lange, Regulation of telomerase by telomeric

- proteins, *Annu. Rev. Biochem.*, 73 (2004) 177–208.
- [84] S. Matsuoka, B.A. Ballif, S. Smolikov, E.R. McDonald, K.E. Hurov, J. Luo, C.E. Bakalarski, Z. Zhao, N. Solimini, Y. Lerenthal, Y. Shiloh, S.P. Gygi, S.J. Elledge, ATM and ATR substrate analysis reveals extensive protein networks responsive to DNA damage, *Science*, 316 (2007) 1160–1166.
- [85] R.E. Verdun, J. Karlseder, Replication and protection of telomeres, *Nature*, 447 (2007) 924–931.
- [86] A. Sfeir, T. de Lange, Removal of shelterin reveals the telomere end-protection problem, *Science*, 336 (2012) 593–597.
- [87] B.R. Houghtaling, L. Cuttonaro, W. Chang, S. Smith, A dynamic molecular link between the telomere length regulator TRF1 and the chromosome end protector TRF2, *Curr. Biol.*, 14 (2004) 1621–1631.
- [88] J.Z.-S. Ye, D. Hockemeyer, A.N. Krutchinsky, D. Loayza, S.M. Hooper, B.T. Chait, T. de Lange, POT1-interacting protein PIP1: a telomere length regulator that recruits POT1 to the TIN2/TRF1 complex, *Genes Dev.*, 18 (2004) 1649–1654.
- [89] A.J. Zaug, E.R. Podell, J. Nandakumar, T.R. Cech, Functional interaction between telomere protein TPP1 and telomerase, *Genes Dev.*, 24 (2010) 613–622.
- [90] M.L. Feng Wang, The POT1-TPP1 telomere complex is a telomerase processivity factor, *Nature*, 445 (2007) 506–510.
- [91] C.M. Latrick, T.R. Cech, POT1-TPP1 enhances telomerase processivity by slowing primer dissociation and aiding translocation, *Embo J.*, 29 (2010) 924–933.
- [92] F.L. Zhong, L.F.Z. Batista, A. Freund, M.F. Pech, A.S. Venteicher, S.E. Artandi, TPP1 OB-fold domain controls telomere maintenance by recruiting telomerase to chromosome ends, *Cell*, 150 (2012) 481–494.
- [93] A.N. Sexton, D.T. Youmans, K. Collins, Specificity requirements for human telomere protein interaction with telomerase holoenzyme, *J. Biol. Chem.*, 287 (2012) 34455–34464.
- [94] H. Xin, D. Liu, M. Wan, A. Safari, H. Kim, W. Sun, M.S. O'Connor, Z. Songyang, TPP1 is a homologue of ciliate TEBP-beta and interacts with POT1 to recruit telomerase, *Nature*, 445 (2007) 559–562.
- [95] J. Lee, E. Mandell, T. Tucey, D. Morris, V. Lundblad, The Est3 protein associates with yeast telomerase through an OB-fold domain, *Nat. Struct. Mol. Biol.*, 15 (2008) 990–997.
- [96] E.Y. Yu, F. Wang, M. Lei, N.F. Lue, A proposed OB-fold with a protein-interaction surface in *Candida albicans* telomerase protein Est3, *Nat. Struct. Mol. Biol.*, 15 (2008) 985–989.
- [97] M. Wan, J. Qin, Z. Songyang, D. Liu, OB fold-containing protein 1 (OBFC1), a human homolog of yeast Stn1, associates with TPP1 and is implicated in telomere length regulation, *J. Biol. Chem.*, 284 (2009) 26725–26731.
- [98] A.G. Murzin, OB(oligonucleotide/oligosaccharide binding)-fold: common

- structural and functional solution for non-homologous sequences, *Embo J*, 12 (1993) 861–867.
- [99] V. Arcus, OB-fold domains: a snapshot of the evolution of sequence, structure and function, *Curr. Opin. Struct. Biol*, 12 (2002) 794–801.
- [100] M.P. Longhese, D. Bonetti, N. Manfrini, M. Clerici, Mechanisms and regulation of DNA end resection, *Embo J*, 29 (2010) 2864–2874.
- [101] C. Ducray, J.P. Pommier, L. Martins, F.D. Boussin, L. Sabatier, Telomere dynamics, end-to-end fusions and telomerase activation during the human fibroblast immortalization process, *Oncogene*, 18 (1999) 4211–4223.
- [102] P. Abdallah, P. Luciano, K.W. Runge, M. Lisby, V. Géli, E. Gilson, M.T. Teixeira, A two-step model for senescence triggered by a single critically short telomere, *Nat. Cell. Biol*, 11 (2009) 988–993.
- [103] M.T. Hemann, M.A. Strong, L.-Y. Hao, C.W. Greider, The shortest telomere, not average telomere length, is critical for cell viability and chromosome stability, *Cell*, 107 (2001) 67–77.
- [104] K.E. Huffman, S.D. Levene, V.M. Tesmer, J.W. Shay, W.E. Wright, Telomere shortening is proportional to the size of the G-rich telomeric 3'-overhang, *J. Biol. Chem*, 275 (2000) 19719–19722.
- [105] V.V. Lundblad, E.H.E. Blackburn, An alternative pathway for yeast telomere maintenance rescues est1- senescence, *Cell*, 73 (1993) 347–360.
- [106] D. Lydall, Hiding at the ends of yeast chromosomes: telomeres, nucleases and checkpoint pathways, *J. Cell. Sci*, 116 (2003) 4057–4065.
- [107] A.W. Murray, T.E. Claus, J.W. Szostak, Characterization of two telomeric DNA processing reactions in *Saccharomyces cerevisiae*, *Mol. Cell. Biol*, 8 (1988) 4642–4650.
- [108] S. Marcand, A protein-counting mechanism for telomere length regulation in yeast, *Science*, 275 (1997) 986–990.
- [109] M.T. Teixeira, M. Arneric, P. Sperisen, J. Lingner, Telomere length homeostasis is achieved via a switch between telomerase- extendible and - nonextendible states, *Cell*, 117 (2004) 323–335.
- [110] B. Van Steensel, Control of telomere length by the human telomeric protein TRF1, *Nature*, (1997).
- [111] A. Smogorzewska, B. van Steensel, A. Bianchi, S. Oelmann, M.R. Schaefer, G. Schnapp, T. de Lange, Control of human telomere length by TRF1 and TRF2, (2000).
- [112] G. Kyrion, K.A. Boakye, A.J. Lustig, C-terminal truncation of RAP1 results in the deregulation of telomere size, stability, and function in *Saccharomyces cerevisiae*, *Mol. Cell. Biol*, 12 (1992) 5159–5173.
- [113] J.L. Milica Arnerić, Tel1 kinase and subtelomere-bound Tbf1 mediate preferential elongation of short telomeres by telomerase in yeast, *EMBO Rep*, 8 (2007) 1080.
- [114] C.W. Greider, E.H. Blackburn, Identification of a specific telomere terminal transferase activity in *Tetrahymena* extracts, *Cell*, 43 (1985) 405–413.
- [115] C.W. Greider, E.H. Blackburn, The telomere terminal transferase of

- Tetrahymena* is a ribonucleoprotein enzyme with two kinds of primer specificity, *Cell*, 51 (1987) 887–898.
- [116] N.W. Kim, M.A. Piatyszek, K.R. Prowse, C.B. Harley, M.D. West, P.L. Ho, G.M. Coviello, W.E. Wright, S.L. Weinrich, J.W. Shay, Specific association of human telomerase activity with immortal cells and cancer, *Science*, 266 (1994) 2011–2015.
- [117] A.G. Bodnar, M. Ouellette, M. Frolkis, S.E. Holt, C.P. Chiu, G.B. Morin, C.B. Harley, J.W. Shay, S. Lichtsteiner, W.E. Wright, Extension of life-span by introduction of telomerase into normal human cells, *Science*, 279 (1998) 349–352.
- [118] N.F. Lue, D. Bosoy, T.J. Moriarty, C. Autexier, B. Altman, S. Leng, Telomerase can act as a template- and RNA-independent terminal transferase, *Proc. Natl. Acad. Sci. USA*, 102 (2005) 9778–9783.
- [119] S.L. Weinrich, R. Pruzan, L. Ma, M. Ouellette, V.M. Tesmer, S.E. Holt, A.G. Bodnar, S. Lichtsteiner, N.W. Kim, J.B. Trager, R.D. Taylor, R. Carlos, W.H. Andrews, W.E. Wright, J.W. Shay, C.B. Harley, G.B. Morin, Reconstitution of human telomerase with the template RNA component hTR and the catalytic protein subunit hTRT, *Nat. Genet.*, 17 (1997) 498–502.
- [120] J. Lingner, T.R. Cech, T.R. Hughes, V. Lundblad, Three ever shorter telomere (EST) genes are dispensable for *in vitro* yeast telomerase activity, *Proc. Natl. Acad. Sci. USA*, 94 (1997) 11190–11195.
- [121] E. Arnold, A. Jacobo-Molina, R.G. Nanni, R.L. Williams, X. Lu, J. Ding, A.D. Clark, A. Zhang, A.L. Ferris, P. Clark, Structure of HIV-1 reverse transcriptase/DNA complex at 7 Å resolution showing active site locations, *Nature*, 357 (1992) 85–89.
- [122] J. Lingner, T.R. Hughes, A. Shevchenko, M. Mann, V. Lundblad, T.R. Cech, Reverse transcriptase motifs in the catalytic subunit of telomerase, *Science*, 276 (1997) 561–567.
- [123] C. Autexier, N.F. Lue, The structure and function of telomerase reverse transcriptase, *Annu. Rev. Biochem.*, 75 (2006) 493–517.
- [124] M. Belfort, M.J. Curcio, N.F. Lue, Telomerase and retrotransposons: reverse transcriptases that shaped genomes, in: *Proc. Natl. Acad. Sci. U.S.A.*, 2011: pp. 20304–20310.
- [125] K.A. Lewis, D.S. Wuttke, Telomerase and telomere-associated proteins: structural insights into mechanism and evolution, *Structure*, 20 (2012) 28–39.
- [126] A.J. Gillis, A.P. Schuller, E. Skordalakes, Structure of the *Tribolium castaneum* telomerase catalytic subunit TERT, *Nature*, 455 (2008) 633–637.
- [127] M. Mitchell, A. Gillis, M. Futahashi, H. Fujiwara, E. Skordalakes, Structural basis for telomerase catalytic subunit TERT binding to RNA template and telomeric DNA, *Nat. Struct. Mol. Biol.*, 17 (2010) 513–518.
- [128] S. Rouda, E. Skordalakes, Structure of the RNA-binding domain of

- telomerase: implications for RNA recognition and binding, *Structure*, 15 (2007) 1403–1412.
- [129] S.A. Jacobs, E.R. Podell, T.R. Cech, Crystal structure of the essential N-terminal domain of telomerase reverse transcriptase, *Nature*, 13 (2006) 218–225.
- [130] A.J. Zaug, E.R. Podell, T.R. Cech, Mutation in TERT separates processivity from anchor-site function, *Nat. Struct. Mol. Biol.*, 15 (2008) 870–872.
- [131] G. Gavory, M. Farrow, S. Balasubramanian, Minimum length requirement of the alignment domain of human telomerase RNA to sustain catalytic activity *in vitro*, *Nucleic Acids Res.*, 30 (2002) 4470–4480.
- [132] M. Xie, A. Mosig, X. Qi, Y. Li, P.F. Stadler, J.J.L. Chen, Structure and function of the smallest vertebrate telomerase RNA from teleost fish, *J. Biol. Chem.*, 283 (2007) 2049–2059.
- [133] S. Gunisova, E. Elboher, J. Nosek, V. Gorkovoy, Y. Brown, J.-F. Lucier, N. Laterreur, R.J. Wellinger, Y. Tzfati, L. Tomaska, Identification and comparative analysis of telomerase RNAs from *Candida* species reveal conservation of functional elements, *Rna*, 15 (2009) 546–559.
- [134] K. Chakrabarti, M. Pearson, L. Grate, T. Sterne-Weiler, J. Deans, J.P. Donohue, M. Ares, Structural RNAs of known and unknown function identified in malaria parasites by comparative genomics and RNA analysis, *Rna*, 13 (2007) 1923–1939.
- [135] E.H. Blackburn, K. Collins, Telomerase: an RNP enzyme synthesizes DNA, *Cold Spring Harb. Perspect. Biol.*, 3 (2011).
- [136] Q. Zhang, N.-K. Kim, J. Feigon, Architecture of human telomerase RNA, *Proc. Natl. Acad. Sci. USA*, 108 (2011) 20325–20332.
- [137] E.Y. Yu, Telomeres and telomerase in *Candida albicans*, *Mycoses*, 55 (2012) e48–59.
- [138] P.-M. Dehé, O. Rog, M.G. Ferreira, J. Greenwood, J.P. Cooper, Taz1 enforces cell-cycle regulation of telomere synthesis, *Mol. Cell*, 46 (2012) 797–808.
- [139] W.E.W. Wright, V.M.V. Tesmer, M.L.M. Liao, J.W.J. Shay, Normal human telomeres are not late replicating, *Exp. Cell Res.*, 251 (1999) 492–499.
- [140] Y. Zhao, A.J. Sfeir, Y. Zou, C.M. Buseman, T.T. Chow, J.W. Shay, W.E. Wright, Telomere extension occurs at most chromosome ends and is uncoupled from fill-in in human cancer cells, *Cell*, 138 (2009) 463–475.
- [141] K.J. Buchkovich, C.W. Greider, Telomerase regulation during entry into the cell cycle in normal human T cells, *Mol. Biol. Cell*, 7 (1996) 1443–1454.
- [142] D.A. Skvortzov, M.P. Rubzova, M.E. Zvereva, F.L. Kiselev, O.A. Donzova, The regulation of telomerase in oncogenesis, *Acta Naturae*, 1 (2009) 51–67.
- [143] C.K. Lai, M.C. Miller, K. Collins, Roles for RNA in telomerase nucleotide and repeat addition processivity, *Mol. Cell*, 11 (2003) 1673–1683.
- [144] L.K. White, W.E. Wright, J.W. Shay, Telomerase inhibitors, *Trends Biotechnol.*, 19 (2001) 114–120.
- [145] J.K. Alder, J.D. Cogan, A.F. Brown, C.J. Anderson, W.E. Lawson, P.M.

- Lansdorp, J.A. Phillips, J.E. Loyd, J.J.-L. Chen, M. Armanios, Ancestral mutation in telomerase causes defects in repeat addition processivity and manifests as familial pulmonary fibrosis, *PLoS Genet*, 7 (2011) e1001352–e1001352.
- [146] A.S. Venteicher, E.B. Abreu, Z. Meng, K.E. McCann, R.M. Terns, T.D. Veenstra, M.P. Terns, S.E. Artandi, A human telomerase holoenzyme protein required for Cajal body localization and telomere synthesis, *Science*, 323 (2009) 644–648.
- [147] Y. Mochizuki, J. He, S. Kulkarni, M. Bessler, P.J. Mason, Mouse dyskerin mutations affect accumulation of telomerase RNA and small nucleolar RNA, telomerase activity, and ribosomal RNA processing, *Proc. Natl. Acad. Sci. USA*, 101 (2004) 10756–10761.
- [148] S.B. Cohen, M.E. Graham, G.O. Lovrecz, N. Bache, P.J. Robinson, R.R. Reddel, Protein composition of catalytically active human telomerase from immortal cells, *Science*, 315 (2007) 1850–1853.
- [149] T. Kiss, E. Fayet-Lebaron, B.E. Jády, Box H/ACA small ribonucleoproteins, *Mol. Cell*, 37 (2010) 597–606.
- [150] D. Fu, K. Collins, Purification of human telomerase complexes identifies factors involved in telomerase biogenesis and telomere length regulation, *Mol. Cell*, 28 (2007) 773–785.
- [151] V. Pogacić, F. Dragon, W. Filipowicz, Human H/ACA small nucleolar RNPs and telomerase share evolutionarily conserved proteins NHP2 and NOP10, *Mol. Cell. Biol*, 20 (2000) 9028–9040.
- [152] F. Dragon, V. Pogacić, W. Filipowicz, *In vitro* assembly of human H/ACA small nucleolar RNPs reveals unique features of U17 and telomerase RNAs, *Mol. Cell. Biol*, 20 (2000) 3037–3048.
- [153] B.E. Snow, N. Erdmann, J. Cruickshank, H. Goldman, R.M. Gill, M.O. Robinson, L. Harrington, Functional conservation of the telomerase protein Est1p in humans, *Curr. Biol*, 13 (2003) 698–704.
- [154] P. Reichenbach, M. Höss, C.M. Azzalin, M. Nabholz, P. Bucher, J. Lingner, A human homolog of yeast Est1 associates with telomerase and uncaps chromosome ends when overexpressed, *Curr. Biol*, 13 (2003) 568–574.
- [155] A.G. Seto, A.J. Livengood, Y. Tzfati, E.H. Blackburn, T.R. Cech, A bulged stem tethers Est1p to telomerase RNA in budding yeast, *Genes Dev*, 16 (2002) 2800–2812.
- [156] D.C.F. Sealey, A.D. Kostic, C. LeBel, F. Pryde, L. Harrington, The TPR-containing domain within Est1 homologs exhibits species-specific roles in telomerase interaction and telomere length homeostasis, *BMC Mol. Biol*, 12 (2011) 45.
- [157] J.L. Stern, K.G. Zyner, H.A. Pickett, S.B. Cohen, T.M. Bryan, Telomerase recruitment requires both TCAB1 and Cajal bodies independently, *Mol. Cell. Biol*, 32 (2012) 2384–2395.
- [158] R.A. Marciniak, F.B. Johnson, L. Guarente, Dyskeratosis congenita, telomeres and human ageing, *Trends Genet*, 16 (2000) 193–195.

- [159] I. Dokal, Dyskeratosis congenita, Hematology Am. Soc. Hematol. Educ. Program, 2011 (2011) 480–486.
- [160] T.J. Vulliamy, S.W. Knight, N.S. Heiss, O.P. Smith, A. Poustka, I. Dokal, P.J. Mason, Dyskeratosis congenita caused by a 3' deletion: germline and somatic mosaicism in a female carrier, Blood, 94 (1999) 1254–1260.
- [161] N.S. Heiss, A. Poustka, S.W. Knight, S. Aradhya, D.L. Nelson, R.A. Lewis, T. Esposito, A. Ciccodicola, M. D'Urso, A. Smahi, S. Heuertz, A. Munnich, P. Vabres, H. Woffendin, S. Kenwrick, Mutation analysis of the DKC1 gene in incontinentia pigmenti, J. Med. Genet, 36 (1999) 860–862.
- [162] H. Yamaguchi, R.T. Calado, H. Ly, S. Kajigaya, G.M. Baerlocher, S.J. Chanock, P.M. Lansdorp, N.S. Young, Mutations in TERT, the gene for telomerase reverse transcriptase, in aplastic anemia, N. Engl. J. Med, 352 (2005) 1413–1424.
- [163] C.K. Garcia, W.E. Wright, J.W. Shay, Human diseases of telomerase dysfunction: insights into tissue aging, Nucleic Acids Res, 35 (2007) 7406–7416.
- [164] Y. Zhang, N. Kaminski, Biomarkers in idiopathic pulmonary fibrosis, Curr. Opin. Pulm. Med, 18 (2012) 441–446.
- [165] D. Hartmann, U. Srivastava, M. Thaler, K.N. Kleinhans, G. N'kontchou, A. Scheffold, K. Bauer, R.F. Kratzer, N. Kloos, S.-F. Katz, Z. Song, Y. Begus-Nahrman, A. Kleger, G. von Figura, P. Strnad, A. Lechel, C. Günes, A. Potthoff, K. Deterding, H. Wedemeyer, et al., Telomerase gene mutations are associated with cirrhosis formation, Hepatology, 53 (2011) 1608–1617.
- [166] V. Lundblad, J.W. Szostak, A mutant with a defect in telomere elongation leads to senescence in yeast, Cell, 57 (1989) 633–643.
- [167] D.C. Zappulla, K. Goodrich, T.R. Cech, A miniature yeast telomerase RNA functions in vivo and reconstitutes activity *in vitro*, Nat. Struct. Mol. Biol, 12 (2005) 1072–1077.
- [168] M.S. Singer, D.E. Gottschling, TLC1: template RNA component of *Saccharomyces cerevisiae* telomerase, Science, 266 (1994) 404–409.
- [169] M. Pedram, C.N. Sprung, Q. Gao, A.W.I. Lo, G.E. Reynolds, J.P. Murnane, Telomere position effect and silencing of transgenes near telomeres in the mouse, Mol. Cell. Biol, 26 (2006) 1865–1878.
- [170] D.C. Zappulla, T.R. Cech, Yeast telomerase RNA: a flexible scaffold for protein subunits, Proc. Natl. Acad. Sci. USA, 101 (2004) 10024–10029.
- [171] M. Zuker, Mfold web server for nucleic acid folding and hybridization prediction, Nucleic Acids Res, 31 (2003) 3406–3415.
- [172] I.L. Hofacker, M. Fekete, P.F. Stadler, Secondary structure prediction for aligned RNA sequences, J. Mol. Biol, 319 (2002) 1059–1066.
- [173] A.S. Chappell, V. Lundblad, Structural elements required for association of the *Saccharomyces cerevisiae* telomerase RNA with the Est2 reverse transcriptase, Mol. Cell. Biol, 24 (2004) 7720–7736.
- [174] C.I. Nugent, V. Lundblad, The telomerase reverse transcriptase: components and regulation, (1998).

- [175] C.T. Tuzon, Y. Wu, A. Chan, V.A. Zakian, The *Saccharomyces cerevisiae* telomerase subunit Est3 binds telomeres in a cell cycle- and Est1-dependent manner and interacts directly with Est1 *in vitro*, *PLoS Genet*, 7 (2011) e1002060.
- [176] A.D. Mozdy, T.R. Cech, Low abundance of telomerase in yeast: Implications for telomerase haploinsufficiency, *Rna*, 12 (2006) 1721–1737.
- [177] T.R. Hughes, S.K. Evans, R.G. Weilbaecher, V. Lundblad, The Est3 protein is a subunit of yeast telomerase, *Curr. Biol*, 10 (2000) 809–812.
- [178] J. Zhou, K. Hidaka, B. Futcher, The Est1 subunit of yeast telomerase binds the Tlc1 telomerase RNA, *Mol. Cell. Biol*, 20 (2000) 1947–1955.
- [179] S.K. Evans, V. Lundblad, Est1 and Cdc13 as comediators of telomerase access, *Science*, 286 (1999) 117–120.
- [180] A.K.P. Taggart, S.-C. Teng, V.A. Zakian, Est1p as a cell cycle-regulated activator of telomere-bound telomerase, *Science*, 297 (2002) 1023–1026.
- [181] V. Virta-Pearlman, D.K. Morris, V. Lundblad, Est1 has the properties of a single-stranded telomere end-binding protein, *Genes Dev*, 10 (1996) 3094–3104.
- [182] J.L. Osterhage, J.M. Talley, K.L. Friedman, Proteasome-dependent degradation of Est1p regulates the cell cycle-restricted assembly of telomerase in *Saccharomyces cerevisiae*, *Nat. Struct. Mol. Biol*, 13 (2006) 720–728.
- [183] J. Lee, E.K. Mandell, T. Rao, D.S. Wuttke, V. Lundblad, Investigating the role of the Est3 protein in yeast telomere replication, *Nucleic Acids Res*, 38 (2010) 2279–2290.
- [184] M. Hsu, E.Y. Yu, S.M. Singh, N.F. Lue, Mutual dependence of *Candida albicans* Est1p and Est3p in telomerase assembly and activation, *Eukaryot. Cell*, 6 (2007) 1330–1338.
- [185] D.K. Morris, V. Lundblad, Programmed translational frameshifting in a gene required for yeast telomere replication, *Curr. Biol*, 7 (1997) 969–976.
- [186] P.J. Farabaugh, E. Kramer, H. Vallabhaneni, A. Raman, Evolution of +1 programmed frameshifting signals and frameshift-regulating tRNAs in the order *Saccharomycetales*, *J. Mol. Evol*, 63 (2006) 545–561.
- [187] D. Taliaferro, P.J. Farabaugh, An mRNA sequence derived from the yeast EST3 gene stimulates programmed +1 translational frameshifting, *Rna*, 13 (2007) 606–613.
- [188] J.J. Pei, N.V.N. Grishin, PROMALS: towards accurate multiple sequence alignments of distantly related proteins, *Bioinformatics*, 23 (2007) 802–808.
- [189] K.L. Friedman, J.J. Heit, D.M. Long, T.R. Cech, N-terminal domain of yeast telomerase reverse transcriptase: recruitment of Est3p to the telomerase complex, *Mol. Biol. Cell*, 14 (2003) 1–13.
- [190] J.M. Talley, D.C. Dezwaan, L.D. Maness, B.C. Freeman, K.L. Friedman, Stimulation of yeast telomerase activity by the Ever Shorter Telomeres 3 (Est3) subunit is dependent on direct interaction with the catalytic protein Est2, *J. Biol. Chem*, 286 (2011) 26431–26439.

- [191] W.-F. Yen, L. Chico, M. Lei, N.F. Lue, Telomerase regulatory subunit Est3 in two *Candida* species physically interacts with the TEN domain of TERT and telomeric DNA, *Proc. Natl. Acad. Sci. USA*, 108 (2011) 20370–20375.
- [192] J. Prescott, E.H. Blackburn, Telomerase RNA mutations in *Saccharomyces cerevisiae* alter telomerase action and reveal nonprocessivity in vivo and *in vitro*, *Genes Dev*, 11 (1997) 528–540.
- [193] L.J. McGuffin, K. Bryson, D.T. Jones, The PSIPRED protein structure prediction server, *Bioinformatics*, 16 (2000) 404–405.
- [194] S.B. Hedges, The origin and evolution of model organisms, *Nat. Rev. Genet*, 3 (2002) 838–849.
- [195] S.M. Singh, O. Steinberg-Neifach, I.S. Mian, N.F. Lue, Analysis of telomerase in *Candida albicans*: potential role in telomere end protection, *Eukaryot. Cell*, 1 (2002) 967–977.
- [196] S.M. Singh, N.F. Lue, Ever shorter telomere 1 (EST1)-dependent reverse transcription by *Candida* telomerase in vitro: evidence in support of an activating function, *Proc. Natl. Acad. Sci. USA*, 100 (2003) 5718–5723.
- [197] J.J. Söding, A.A. Biegert, A.N.A. Lupas, The HHpred interactive server for protein homology detection and structure prediction, *Nucleic Acids Res*, 33 (2005) W244–W248.
- [198] A. Roy, A. Kucukural, Y. Zhang, I-TASSER: a unified platform for automated protein structure and function prediction, *Nat. Protoc*, 5 (2010) 725–738.
- [199] S. Wu, Y. Zhang, LOMETS: a local meta-threading-server for protein structure prediction, *Nucleic Acids Res*, 35 (2007) 3375–3382.
- [200] V. Mariani, F. Kiefer, T. Schmidt, J. Haas, T. Schwede, Assessment of template based protein structure predictions in CASP9, *Proteins*, 79 (2011) 37–58.
- [201] D.E. Kim, D. Chivian, D. Baker, Protein structure prediction and analysis using the Robetta server, *Nucleic Acids Res*, 32 (2004) W526–31.
- [202] M.P. Horvath, Structural anatomy of telomere OB proteins, *Crit. Rev. Biochem. Mol. Biol*, 46 (2011) 409–435.
- [203] R.M. Mitton-Fry, E.M. Anderson, D.L. Theobald, L.W. Glustrom, D.S. Wuttke, Structural basis for telomeric single-stranded DNA recognition by yeast Cdc13, *J. Mol. Biol*, 338 (2004) 241–255.
- [204] E. Mossessova, C.D. Lima, Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast, *Mol. Cell*, 5 (2000) 865–876.
- [205] J.R. Huth, C.A. Bewley, B.M. Jackson, A.G. Hinnebusch, G.M. Clore, A.M. Gronenborn, Design of an expression system for detecting folded protein domains and mapping macromolecular interactions by NMR, *Protein Sci*, 6 (1997) 2359–2364.
- [206] G. Lindwall, M.-F. Chau, S.R. Gardner, L.A. Kohlstaedt, A sparse matrix approach to the solubilization of overexpressed proteins, *Protein Eng*, 13 (2000) 67–71.

- [207] A.E. Kelly, H.D. Ou, R. Withers, V. Dötsch, Low-conductivity buffers for high-sensitivity NMR measurements, *J. Am. Chem. Soc.*, 124 (2002) 12013–12019.
- [208] K. Shiraki, M. Kudou, S. Fujiwara, T. Imanaka, M. Takagi, Biophysical effect of amino acids on the prevention of protein aggregation, *J. Biochem.*, 132 (2002) 591–595.
- [209] G.M. Hautbergue, A.P. Golovanov, Increasing the sensitivity of cryoprobe protein NMR experiments by using the sole low-conductivity arginine glutamate salt, *J. Magn. Reson.*, 191 (2008) 335–339.
- [210] M. Goldberg, N. Expert-Bezancon, L. Vuillard, T. Rabilloud, Non-detergent sulphobetaines: a new class of molecules that facilitate *in vitro* protein renaturation, *Fold. Des.*, 1 (1995) 21–27.
- [211] H.M. Pereira, A. Cleasby, S.D.J. Pena S, G.R. Franco G, R.C. Garratt, Cloning, expression and preliminary crystallographic studies of the potential drug target purine nucleoside phosphorylase from *Schistosoma mansoni*, *Acta Crystallogr. D Biol. Crystallogr.*, 59 (2003) 1096–1099.
- [212] L. Xiang, T. Ishii, K. Hosoda, A. Kamiya, M. Enomoto, N. Nameki, Y. Inoue, K. Kubota, T. Kohno, K. Wakamatsu, Interaction of anti-aggregation agent dimethylethylammonium propane sulfonate with acidic fibroblast growth factor, *J. Magn. Reson.*, 194 (2008) 147–151.
- [213] M.S. Willis, J.K. Hogan, P. Prabhakar, X. Liu, K. Tsai, Y. Wei, T. Fox, Investigation of protein refolding using a fractional factorial screen: a study of reagent effects and interactions, *Protein Sci.*, 14 (2005) 1818–1826.
- [214] K.H. Gardner, L.E. Kay, Multidimensional 2H-based NMR methods for resonance assignment, structure determination, and the study of protein dynamics, in: N.R. Krishna, L.J. Berliner (Eds.), *Modern Techniques in Protein NMR*, Plenum Publishing Corporation, 1998: pp. 27–74.
- [215] V. Tugarinov, L.E. Kay, Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods, *J. Am. Chem. Soc.*, 125 (2003) 13868–13878.
- [216] Y. Shen, F. Delaglio, G. Cornilescu, A. Bax, TALOS+: a hybrid method for predicting protein backbone torsion angles from NMR chemical shifts, *J. Biomol. NMR.*, 44 (2009) 213–223.
- [217] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR.*, 6 (1995) 277–293.
- [218] W.F. Vranken, W. Boucher, T.J. Stevens, R.H. Fogh, A. Pajon, M. Llinas, E.L. Ulrich, J.L. Markley, J. Ionides, E.D. Laue, The CCPN data model for NMR spectroscopy: development of a software pipeline, *Proteins*, 59 (2005) 687–696.
- [219] M.R. Hansen, L. Mueller, A. Pardi, Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions, *Nat. Struct. Biol.*, 5 (1998) 1065–1074.
- [220] O.F. Lange, D. Baker, Resolution-adapted recombination of structural

- features significantly improves sampling in restraint-guided structure calculation, *Proteins*, 80 (2012) 884–895.
- [221] P. Güntert, C. Mumenthaler, K. Wüthrich, Torsion angle dynamics for NMR structure calculation with the new program DYANA, *J. Mol. Biol.*, 273 (1997) 283–298.
- [222] C.D.C. Schwieters, J.J.J. Kuszewski, N.N. Tjandra, G.M.G. Clore, The Xplor-NIH NMR molecular structure determination package, *J. Magn. Reson.*, 160 (2003) 65–73.
- [223] R. Das, D. Baker, Macromolecular modeling with rosetta, *Annu. Rev. Biochem.*, 77 (2008) 363–382.
- [224] C.A. Rohl, Protein structure estimation from minimal restraints using Rosetta, *Methods Enzymol.*, 394 (2005) 244–260.
- [225] C.A. Rohl, C.E.M. Strauss, K.M.S. Misura, D. Baker, Protein structure prediction using Rosetta, *Methods Enzymol.*, 383 (2004) 66–93.
- [226] L.R. Warner, K. Varga, O.F. Lange, S.L. Baker, D. Baker, M.C. Sousa, A. Pardi, Structure of the BamC two-domain protein obtained by Rosetta with a limited NMR data set, *J. Mol. Biol.*, 411 (2011) 83–95.
- [227] A. Bhattacharya, R. Tejero, G.T. Montelione, Evaluating protein structures determined by structural genomics consortia, *Proteins*, 66 (2007) 778–795.
- [228] N.A. Baker, D. Sept, S. Joseph, M.J. Holst, J.A. McCammon, Electrostatics of nanosystems: application to microtubules and the ribosome, *Proc. Natl. Acad. Sci. USA*, 98 (2001) 10037–10041.
- [229] D. Esposito, D.K. Chatterjee, Enhancement of soluble protein expression through the use of fusion tags, *Curr. Opin. Biotechnol.*, 17 (2006) 353–358.
- [230] S. Züger, H. Iwai, Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies, *Nat. Biotechnol.*, 23 (2005) 736–740.
- [231] J.G. Marblestone, S.C. Edavettal, Y. Lim, P. Lim, X. Zuo, T.R. Butt, Comparison of SUMO fusion technology with traditional gene fusion systems: enhanced expression and solubility with SUMO, *Protein Sci.*, 15 (2006) 182–189.
- [232] A. Sivashanmugam, V. Murray, C. Cui, Y. Zhang, J. Wang, Q. Li, Practical protocols for production of very high yields of recombinant proteins using *Escherichia coli*, *Protein Sci.*, 18 (2009) 936–948.
- [233] P. Zhou, G. Wagner, Overcoming the solubility limit with solubility-enhancement tags: successful applications in biomolecular NMR studies, *J. Biomol. NMR*, 46 (2010) 23–31.
- [234] A. Kato, K. Maki, T. Ebina, K. Kuwajima, K. Soda, Y. Kuroda, Mutational analysis of protein solubility enhancement using short peptide tags, *Biopolymers*, 85 (2007) 12–18.
- [235] K.H. Gardner, L.E. Kay, The use of  $^2\text{H}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$  multidimensional NMR to study the structure and dynamics of proteins, *Annu. Rev. Biophys. Biomol. Struct.*, 27 (1998) 357–406.
- [236] G. Cornilescu, F. Delaglio, A. Bax, Protein backbone angle restraints from

- searching a database for chemical shift and sequence homology, *J. Biomol. NMR*, 13 (1999) 289–302.
- [237] A. Bahrami, A.H. Assadi, J.L. Markley, H.R. Eghbalnia, Probabilistic interaction network of evidence algorithm and its application to complete labeling of peak lists from protein NMR spectroscopy, *PLoS Comput. Biol*, 5 (2009) e1000307.
- [238] K. Chen, N. Tjandra, The use of residual dipolar coupling in studying proteins by NMR, *Top. Curr. Chem*, 326 (2012) 47–67.
- [239] G.M. Clore, D.S. Garrett, R-factor, Free R, and complete cross-validation for dipolar coupling refinement of NMR structures, *J. Am. Chem. Soc*, 121 (1999) 9008–9012.
- [240] N. Tjandra, S.E. Feller, R.W. Pastor, A. Bax, Rotational diffusion anisotropy of human ubiquitin from <sup>15</sup>N NMR relaxation, *J. Am. Chem. Soc*, 117 (1995) 12562–12566.
- [241] C. Renner, M. Schleicher, L. Moroder, T.A. Holak, Practical aspects of the 2D <sup>15</sup>N-[<sup>1</sup>H]-NOE experiment, *J. Biomol. NMR*, 23 (2002) 23–33.
- [242] M.V. Berjanskii, D.S. Wishart, A simple method to predict protein flexibility using secondary chemical shifts, *J. Am. Chem. Soc*, 127 (2005) 14970–14971.
- [243] S. Raman, O.F. Lange, P. Rossi, M. Tyka, X. Wang, J. Aramini, G. Liu, T.A. Ramelot, A. Eletsky, T. Szyperski, M.A. Kennedy, J. Prestegard, G.T. Montelione, D. Baker, NMR structure determination for larger proteins using backbone-only data, *Science*, 327 (2010) 1014–1018.
- [244] Y. Shen, R. Vernon, D. Baker, A. Bax, *De novo* protein structure generation from incomplete chemical shift assignments, *J. Biomol. NMR*, 43 (2009) 63–78.
- [245] H. Ashkenazy, E. Erez, E. Martz, T. Pupko, N. Ben-Tal, ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids, *Nucleic Acids Res*, 38 (2010) W529–533.
- [246] M. Landau, I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko, N. Ben-Tal, ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures, *Nucleic Acids Res*, 33 (2005) W299–302.
- [247] F. Glaser, T. Pupko, I. Paz, R.E. Bell, D. Bechor-Shental, E. Martz, N. Ben-Tal, ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information, *Bioinformatics*, 19 (2003) 163–164.
- [248] L.A. Kelley, M.J.E. Sternberg, Protein structure prediction on the Web: a case study using the Phyre server, *Nat. Protoc*, 4 (2009) 363–371.
- [249] L. Holm, P. Rosenström, Dali server: conservation mapping in 3D, *Nucleic Acids Res*, 38 (2010) W545–549.
- [250] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemical quality of protein structures, *J. Appl. Crystallogr*, 26 (1993) 283–291.
- [251] P. Güntert, Automated NMR structure calculation with CYANA, *Methods Mol. Biol*, 278 (2004) 353–378.

- [252] M.V. Berjanskii, S. Neal, D.S. Wishart, PREDITOR: a web server for predicting protein torsion angle restraints, *Nucleic Acids Res*, 34 (2006) W63–9.
- [253] I.D. Kerr, R.I.M. Wadsworth, L. Cubeddu, W. Blankenfeldt, J.H. Naismith, M.F. White, Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of *Sulfolobus* SSB protein, *Embo J*, 22 (2003) 2561–2570.
- [254] T.M. Lohman, M.E. Ferrari, *Escherichia coli* single-stranded DNA-binding protein: multiple DNA-binding modes and cooperativities, *Annu. Rev. Biochem*, 63 (1994) 527–570.
- [255] Y. Shamoo, A.M. Friedman, M.R. Parsons, W.H. Konigsberg, T.A. Steitz, Crystal structure of a replication fork single-stranded DNA binding protein (T4 gp32) complexed to DNA, *Nature*, 376 (1995) 362–366.
- [256] B. Marintcheva, A. Marintchev, G. Wagner, C.C. Richardson, Acidic C-terminal tail of the ssDNA-binding protein of bacteriophage T7 and ssDNA compete for the same binding surface, *Proc. Natl. Acad. Sci. USA*, 105 (2008) 1855–1860.
- [257] N.F. Lue, Z. Li, Modeling and structure function analysis of the putative anchor site of yeast telomerase, *Nucleic Acids Res*, 35 (2007) 5213–5222.
- [258] J. Nandakumar, C.F. Bell, I. Weidenfeld, A.J. Zaug, L.A. Leinwand, T.R. Cech, The TEL patch of telomere protein TPP1 mediates telomerase recruitment and processivity, *Nature*, (2012) In press.
- [259] C.-P. Yang, Y.-B. Chen, F.-L. Meng, J.-Q. Zhou, *Saccharomyces cerevisiae* Est3p dimerizes in vitro and dimerization contributes to efficient telomere replication *in vivo*, *Nucleic Acids Res*, 34 (2006) 407–416.
- [260] A.G. Malyavko, N.A. Logvina, M.E. Zvereva, O.A. Dontsova, *In vitro* dimerization of telomerase protein Est3p is stimulated by magnesium cations, *Dokl. Biochem. Biophys*, 433 (2010) 152–154.
- [261] K.L. Friedman, T.R. Cech, Essential functions of amino-terminal domains in the yeast telomerase catalytic subunit revealed by selection for viable mutants, *Genes Dev*, 13 (1999) 2863–2874.
- [262] M. Lei, E.R. Podell, P. Baumann, T.R. Cech, DNA self-recognition in the structure of Pot1 bound to telomeric single-stranded DNA, *Nature*, 426 (2003) 198–203.
- [263] M. Lei, E.R. Podell, T.R. Cech, Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection, *Nat. Struct. Mol. Biol*, 11 (2004) 1223–1229.
- [264] I. Herskowitz, Functional inactivation of genes by dominant negative mutations, *Nature*, 329 (1987) 219–222.
- [265] A.O. Wilkie, The molecular basis of genetic dominance, *J. Med. Genet*, 31 (1994) 89–98.
- [266] S. Primrose, R. Twyman, B. Old, Principles of gene manipulation, 2004.
- [267] S.K. Evans, V. Lundblad, The Est1 subunit of *Saccharomyces cerevisiae* telomerase makes multiple contributions to telomere length maintenance,

- Genetics, 162 (2002) 1101–1115.
- [268] N.J. Greenfield, Using circular dichroism spectra to estimate protein secondary structure, *Nat. Protoc*, 1 (2006) 2876–2890.
- [269] S.M. Kelly, T.J. Jess, N.C. Price, How to study proteins by circular dichroism, *Biochim. Biophys. Acta*, 1751 (2005) 119–139.
- [270] S.M. Kelly, N.C. Price, The use of circular dichroism in the investigation of protein structure and function, *Curr. Protein Pept. Sci*, 1 (2000) 349–384.
- [271] G. Dantas, B. Kuhlman, D. Callender, M. Wong, D. Baker, A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins, *J. Mol. Biol*, 332 (2003) 449–460.
- [272] C.M. Johnson, A.R. Fersht, Protein stability as a function of denaturant concentration: the thermal stability of barnase in the presence of urea, *Biochemistry*, 34 (1995) 6795–6804.

# APPENDIX

## I. Chemical shift assignments for Est3<sup>ΔN</sup>

**Table A 1 <sup>1</sup>HN <sup>15</sup>N <sup>13</sup>CO <sup>13</sup>Ca and <sup>13</sup>Cb protein backbone resonances assignments for the Est3<sup>ΔN</sup> protein**

Residue number	Residue name	<sup>1</sup> HN (ppm)	<sup>15</sup> N (ppm)	<sup>13</sup> Ca (ppm)	<sup>13</sup> Cb (ppm)	<sup>13</sup> CO (ppm)
13	THR	0.00	----	61.47	69.20	174.30
14	ASP	8.32	123.23	54.00	41.60	176.24
15	SER	8.40	115.46	58.02	63.08	175.99
16	VAL	8.14	124.54	62.75	30.86	176.10
17	PHE	7.57	119.24	58.93	38.13	176.09
18	LEU	7.59	121.29	54.05	41.53	173.87
19	GLN	8.89	128.12	53.13	30.30	----
20	PRO	----	----	62.46	29.69	176.97
21	TRP	9.81	126.04	58.24	31.46	176.62
22	ILE	8.44	128.28	66.63	37.45	176.78
23	LYS	9.06	121.66	60.47	31.06	177.81
24	ALA	8.45	120.45	54.08	17.22	179.29
25	LEU	8.07	120.48	58.10	41.64	180.27
26	ILE	8.16	119.87	63.55	35.79	179.62
27	GLU	8.54	122.07	59.14	28.46	179.10
28	ASP	8.40	119.23	55.66	40.40	176.99
29	ASN	7.54	117.54	53.69	40.07	174.28
30	SER	7.70	115.44	58.80	63.84	173.94
31	GLU	8.29	122.46	55.63	29.62	176.23
32	HIS	8.21	120.95	56.36	29.47	174.78
33	ASP	8.26	120.47	54.28	40.01	175.22
34	GLN	7.85	118.52	55.39	28.59	175.27
35	TYR	8.07	120.83	56.75	36.71	174.88
36	HIS	7.79	121.43	53.28	29.22	----
37	PRO	----	----	2.95	31.73	177.02
38	SER	8.81	117.55	57.49	63.79	174.28
39	GLY	8.56	113.55	44.46	0.00	173.69
40	HIS	9.07	121.53	55.80	26.89	175.79
41	VAL	8.20	115.42	62.42	31.20	176.25
42	ILE	7.31	120.21	58.66	39.58	----

43	PRO	----	----	2.45	32.49	177.08
44	SER	8.44	116.53	57.64	63.03	173.40
45	LEU	9.95	127.44	52.53	42.74	178.10
46	THR	9.33	114.57	60.44	71.16	175.22
47	LYS	8.69	120.40	59.34	30.96	179.43
48	GLN	8.08	119.49	58.36	27.36	178.69
49	ASP	7.74	122.54	57.52	41.15	175.76
50	LEU	7.31	113.80	56.06	41.16	178.24
51	ALA	7.68	120.42	53.72	18.19	178.66
52	LEU	6.99	116.64	51.31	42.10	----
53	PRO	----	----	65.47	31.76	177.59
54	HIS	8.27	112.75	57.70	28.72	175.19
55	MET	7.25	116.33	53.19	32.24	175.52
56	SER	7.04	116.68	54.27	63.23	----
57	PRO	----	----	64.33	31.04	178.71
58	THR	7.56	115.44	65.41	68.21	175.81
59	ILE	7.18	122.82	64.23	37.28	176.67
60	LEU	7.89	113.85	56.59	40.73	179.29
61	THR	7.43	107.27	61.52	69.92	173.99
62	ASN	7.37	121.38	49.98	39.88	----
63	PRO	----	----	64.13	32.28	176.87
64	CYS	7.07	113.21	56.94	26.24	170.20
65	HIS	6.65	108.99	56.14	28.56	174.30
66	PHE	7.55	116.82	56.99	41.54	174.31
67	ALA	8.65	119.24	50.84	24.14	175.50
68	LYS	8.37	118.34	54.56	34.58	176.01
69	ILE	9.66	126.32	61.58	37.16	176.71
70	THR	8.79	118.15	60.88	68.64	175.08
71	LYS	7.09	122.88	56.46	34.14	173.36
72	PHE	8.53	122.14	57.98	39.36	175.07
73	TYR	9.44	123.21	60.46	39.63	176.29
74	ASN	7.99	113.79	52.67	39.80	172.65
75	VAL	7.72	124.91	61.42	33.19	175.18
76	CYS	8.68	125.70	57.69	27.99	----
77	ASP	----	----	----	----	----
78	TYR	8.21	109.73	60.08	35.63	173.59
79	LYS	7.69	119.18	55.24	33.52	175.20
80	VAL	7.48	118.70	61.09	34.05	172.95
81	TYR	8.85	128.06	57.84	37.59	175.95
82	ALA	9.67	130.23	51.70	24.00	174.29
83	SER	8.98	112.45	56.66	65.40	175.06

84	ILE	8.66	115.29	57.82	41.31	173.57
85	ARG	9.31	116.14	52.72	33.19	174.59
86	ASP	7.51	121.62	51.67	38.97	175.80
87	SER	6.45	109.49	59.78	62.82	174.34
88	SER	8.85	117.71	60.63	66.39	172.32
89	HIS	9.57	121.99	55.53	36.62	173.25
90	GLN	7.94	112.82	53.06	32.98	175.64
91	ILE	8.94	119.35	61.02	41.08	172.31
92	LEU	7.19	125.41	55.08	40.64	175.85
93	VAL	9.22	116.67	58.64	35.84	173.01
94	GLU	8.55	124.25	53.42	32.13	----
95	PHE	----	----	56.71	38.77	175.47
96	SER	8.20	119.65	57.23	65.25	173.02
97	GLN	8.95	120.89	58.80	26.86	179.42
98	GLU	8.76	121.47	59.06	28.65	177.40
99	CYS	8.19	122.36	61.68	27.00	178.67
100	VAL	8.58	120.07	66.50	31.30	178.19
101	SER	8.69	117.16	61.18	62.11	177.93
102	ASN	8.64	120.93	56.11	37.46	178.05
103	PHE	8.50	124.91	61.73	38.98	177.89
104	GLU	8.61	120.93	59.07	28.32	180.04
105	ARG	7.98	119.20	58.18	29.30	178.10
106	THR	7.74	114.73	65.16	68.96	175.32
107	HIS	8.03	116.91	56.64	29.83	174.77
108	ASN	8.14	118.20	53.86	36.62	173.33
109	CYS	7.49	111.64	55.23	29.52	171.93
110	ARG	8.27	117.68	54.91	30.25	179.02
111	ILE	9.55	124.68	63.11	37.53	----
112	THR	6.48	106.98	60.03	67.95	175.16
113	SER	7.45	118.28	58.47	62.73	173.81
114	GLU	8.35	121.27	56.91	27.46	----
115	THR	7.69	109.18	61.91	69.07	177.03
116	THR	7.98	115.34	63.83	68.13	175.02
117	ASN	8.30	118.56	53.55	37.55	174.41
118	CYS	7.55	116.61	58.21	28.87	174.43
119	LEU	8.84	125.61	54.87	42.39	174.61
120	MET	9.27	128.08	53.29	----	174.38
121	ILE	8.53	120.09	55.54	36.33	176.48
122	ILE	8.91	119.13	57.79	41.61	174.90
123	GLY	8.37	106.65	43.70	----	170.57
124	ASP	7.71	119.02	53.99	39.16	174.48

125	ALA	8.50	123.00	49.26	23.47	175.80
126	ASP	8.66	121.64	51.98	43.38	174.13
127	LEU	8.89	124.87	53.94	43.76	175.01
128	VAL	8.74	122.82	58.88	34.06	173.24
129	TYR	8.58	120.37	57.27	38.93	176.28
130	VAL	8.85	121.47	59.32	34.19	174.14
131	THR	7.77	113.65	59.75	70.86	175.86
132	ASN	8.52	119.41	56.64	38.27	----
133	SER	----	----	61.29	----	176.76
134	ARG	7.33	124.19	57.03	29.05	178.39
135	ALA	8.66	124.16	55.24	16.06	180.06
136	MET	8.21	117.77	58.21	31.49	178.84
137	SER	8.02	114.61	60.59	62.24	176.20
138	HIS	8.46	120.79	58.30	29.59	174.53
139	PHE	7.47	114.05	58.72	38.58	173.54
140	LYS	7.14	110.73	56.11	27.89	175.95
141	ILE	7.31	114.72	58.80	41.59	173.51
142	SER	7.81	115.03	55.68	62.91	176.44
143	LEU	8.74	127.41	56.18	40.19	178.50
144	SER	8.14	114.48	60.24	61.69	----
145	ASN	7.73	117.30	54.27	38.28	175.94
146	ILE	7.49	112.48	60.84	39.12	174.16
147	SER	7.79	114.96	56.77	64.08	----
148	SER	----	----	58.19	62.84	174.29
149	LYS	8.05	122.10	54.61	33.71	----
150	GLY	8.61	120.07	56.70	29.04	174.16
151	ILE	6.87	118.01	59.09	39.77	175.07
152	VAL	8.40	117.28	56.34	33.34	----
153	PRO	----	----	60.89	31.11	175.09
154	VAL	9.21	124.95	59.70	35.36	174.32
155	LEU	7.82	122.27	52.54	43.02	174.86
156	ASN	9.26	126.66	51.38	38.59	174.96
157	VAL	8.73	126.55	62.74	33.90	174.99
158	ASN	8.09	122.93	53.15	40.96	174.15
159	GLN	8.01	118.64	54.96	33.26	174.15
160	ALA	8.28	121.36	51.34	20.75	176.62
161	THR	9.10	115.03	59.99	69.77	----
162	ILE	0.00	----	62.03	39.36	175.08
163	PHE	9.31	128.73	59.06	39.54	----
164	ASP	----	----	----	----	----
165	ILE	----	----	60.71	38.55	175.30

166	ASP	8.40	124.97	53.99	41.11	175.58
167	GLN	8.38	122.32	55.01	28.83	175.70
168	VAL	8.10	122.55	62.32	31.98	176.59
169	GLY	8.35	112.86	44.51	----	----
170	SER	----	----	----	----	----
171	LEU	----	----	54.60	40.89	177.34
172	SER	7.99	117.42	57.86	63.12	172.76
173	THR	7.92	116.66	61.13	69.26	173.57
174	PHE	8.62	125.74	54.74	38.36	----
175	PRO	----	----	61.55	31.61	178.23
176	PHE	9.18	119.60	59.63	39.49	178.36
177	VAL	9.41	121.28	62.67	31.44	174.92
178	TYR	6.17	116.60	56.58	35.76	176.84
179	LYS	7.55	118.11	57.54	31.02	176.38
180	TYR	7.79	117.35	57.41	38.02	174.48
181	LEU	6.93	125.62	55.39	42.10	----

“----” indicates resonance assignments not assigned for these atoms

## II. NOEs used for structure calculation

**Table A 2 All NOESs (amide-amide and methyl-methyl) for Est3<sup>AN</sup> structure calculation**

	NOE residue number	from Atom type	NOE residue number	to Atom type	Estimated distance (Å)
1	82	HN	93	HN	3.64
2	82	HN	83	HN	4.07
3	119	HN	163	HN	4.07
4	67	HN	120	HN	3.79
5	73	HN	81	HN	3.55
6	74	HN	81	HN	3.3
7	75	HN	81	HN	4.01
8	21	HN	22	HN	3.67
9	45	HN	124	HN	3.95
10	143	HN	144	HN	3.39
11	143	HN	145	HN	3.73
12	126	HN	156	HN	2.99
13	96	HN	157	HN	3.36
14	21	HN	22	HN	3.52

15	180	HN	181	HN	2.74
16	179	HN	181	HN	3.33
17	119	HN	163	HN	3.7
18	119	HN	166	HN	3.86
19	76	HN	78	HN	3.42
20	76	HN	79	HN	2.93
21	128	HN	154	HN	3.27
22	119	HN	166	HN	3.92
23	41	HN	127	HN	3.64
24	103	HN	105	HN	3.42
25	103	HN	106	HN	3.95
26	75	HN	76	HN	3.67
27	74	HN	75	HN	4.76
28	134	HN	135	HN	3.33
29	134	HN	136	HN	3.79
30	16	HN	17	HN	3.3
31	111	HN	112	HN	3.7
32	135	HN	136	HN	3.05
33	94	HN	155	HN	4.01
34	131	HN	135	HN	4.11
35	134	HN	135	HN	3.11
36	73	HN	81	HN	3.52
37	72	HN	73	HN	4.01
38	73	HN	74	HN	2.96
39	41	HN	125	HN	3.27
40	124	HN	125	HN	3.21
41	42	HN	125	HN	3.08
42	128	HN	154	HN	3.21
43	167	HN	168	HN	3.17
44	46	HN	49	HN	3.55
45	48	HN	49	HN	3.33
46	49	HN	50	HN	3.11
47	49	HN	52	HN	3.79
48	59	HN	60	HN	2.96
49	58	HN	59	HN	2.83
50	71	HN	72	HN	4.01
51	70	HN	71	HN	2.83
52	99	HN	101	HN	2.99
53	30	HN	31	HN	3.27
54	167	HN	168	HN	3.52
55	88	HN	89	HN	3.3

56	86	HN	89	HN	3.7
57	87	HN	89	HN	3.76
58	93	HN	155	HN	3.92
59	26	HN	27	HN	2.8
60	27	HN	29	HN	3.24
61	71	HN	72	HN	3.86
62	86	HN	89	HN	3.21
63	86	HN	88	HN	3.21
64	86	HN	87	HN	3.55
65	23	HN	24	HN	3.55
66	98	HN	99	HN	3.02
67	126	HN	156	HN	3.27
68	41	HN	126	HN	3.76
69	42	HN	126	HN	4.01
70	130	HN	131	HN	3.73
71	130	HN	152	HN	2.83
72	159	HN	160	HN	3.55
73	35	HN	36	HN	3.27
74	60	HN	62	HN	3.36
75	90	HN	177	HN	3.55
76	177	HN	178	HN	3.39
77	97	HN	99	HN	3.3
78	104	HN	105	HN	2.8
79	104	HN	106	HN	3.24
80	114	HN	115	HN	3.67
81	24	HN	25	HN	2.83
82	138	HN	139	HN	3.39
83	138	HN	140	HN	3.42
84	35	HN	36	HN	2.93
85	23	HN	24	HN	3.21
86	50	HN	51	HN	2.8
87	51	HN	52	HN	2.77
88	42	HN	125	HN	3.08
89	41	HN	42	HN	2.77
90	42	HN	124	HN	3.7
91	23	HN	25	HN	3.79
92	24	HN	25	HN	2.8
93	26	HN	27	HN	2.83
94	121	HN	163	HN	4.54
95	121	HN	161	HN	3.55
96	26	HN	29	HN	3.79

97	48	HN	49	HN	3.27
98	48	HN	50	HN	3.58
99	76	HN	79	HN	2.83
100	78	HN	79	HN	2.71
101	16	HN	17	HN	3.36
102	26	HN	28	HN	3.21
103	28	HN	29	HN	2.9
104	84	HN	91	HN	3.3
105	67	HN	68	HN	3.7
106	96	HN	97	HN	3.86
107	67	HN	120	HN	3.42
108	104	HN	105	HN	2.8
109	105	HN	106	HN	2.96
110	45	HN	124	HN	3.67
111	117	HN	118	HN	3.92
112	123	HN	159	HN	3.08
113	124	HN	159	HN	3.52
114	179	HN	180	HN	2.93
115	179	HN	181	HN	3.24
116	178	HN	179	HN	3.24
117	70	HN	71	HN	2.9
118	68	HN	85	HN	3.17
119	106	HN	108	HN	3.73
120	108	HN	109	HN	3.21
121	150	HN	151	HN	3.61
122	151	HN	152	HN	3.89
123	177	HN	179	HN	3.92
124	28	HN	29	HN	2.9
125	88	HN	89	HN	3.27
126	86	HN	88	HN	3.64
127	87	HN	88	HN	3.3
128	179	HN	180	HN	2.96
129	180	HN	181	HN	2.74
130	130	HN	152	HN	3.02
131	99	HN	101	HN	3.48
132	106	HN	107	HN	3.17
133	107	HN	109	HN	3.36
134	178	HN	180	HN	3.79
135	64	HN	66	HN	3.24
136	65	HN	66	HN	2.83
137	54	HN	56	HN	3.24

138	51	HN	52	HN	2.71
139	55	HN	56	HN	2.62
140	50	HN	52	HN	3.14
141	177	HN	178	HN	3.36
142	178	HN	179	HN	3.3
143	82	HN	93	HN	3.76
144	44	HN	124	HN	3.61
145	118	HN	119	HN	4.2
146	117	HN	118	HN	3.36
147	54	HN	55	HN	3.14
148	55	HN	56	HN	2.71
149	68	HN	85	HN	3.17
150	58	HN	60	HN	3.24
151	58	HN	59	HN	2.83
152	41	HN	125	HN	3.42
153	41	HN	42	HN	2.93
154	115	HN	116	HN	3.55
155	30	HN	31	HN	3.21
156	84	HN	91	HN	3.21
157	40	HN	41	HN	3.79
158	121	HN	161	HN	3.64
159	46	HN	49	HN	3.3
160	46	HN	47	HN	3.86
161	46	HN	48	HN	4.14
162	46	HN	50	HN	3.58
163	135	HN	137	HN	3.45
164	137	HN	138	HN	3.17
165	137	HN	139	HN	3.36
166	137	HN	140	HN	3.55
167	104	HN	106	HN	3.45
168	106	HN	107	HN	2.77
169	146	HN	147	HN	2.77
170	60	HN	61	HN	2.9
171	59	HN	60	HN	2.93
172	73	HN	74	HN	2.86
173	74	HN	81	HN	3.48
174	138	HN	139	HN	3.39
175	137	HN	139	HN	3.79
176	54	HN	55	HN	3.08
177	54	HN	56	HN	3.24
178	90	HN	177	HN	3.33

179	90	HN	176	HN	3.76
180	90	HN	178	HN	3.98
181	71	HN	83	HN	3.3
182	108	HN	109	HN	2.77
183	137	HN	140	HN	3.89
184	139	HN	140	HN	2.86
185	138	HN	140	HN	3.42
186	76	HN	78	HN	4.04
187	78	HN	79	HN	3.24
188	114	HN	115	HN	3.86
189	115	HN	116	HN	3.89
190	87	HN	89	HN	3.14
191	87	HN	88	HN	2.83
192	86	HN	87	HN	3.7
193	65	HN	66	HN	3.05
194	64	HN	65	HN	3.02
195	60	HN	61	HN	3.48
196	123	HN	159	HN	3.11
197	123	HN	124	HN	3.83
198	111	HN	112	HN	3.89
199	16	HN	18	HN	4.17
200	47	HN	48	HN	3.89
201	47	HN	49	HN	3.83
202	50	HN	52	HN	3.11
203	50	HN	51	HN	2.65
204	48	HN	50	HN	3.55
205	46	HN	50	HN	3.92
206	64	HN	66	HN	3.3
207	64	HN	65	HN	2.86
208	74	HN	75	HN	3.7
209	21	HN	87	HN	4.76
210	93	HN	94	HN	4.17
211	131	HN	134	HN	3.79
212	138	HN	141	HN	3.48
213	136	HN	141	HN	3.76
214	137	HN	141	HN	3.76
215	141	HN	143	HN	3.92
216	143	HN	144	HN	3.58
217	146	HN	147	HN	2.9
218	41	HN	127	HN	3.73
219	94	HN	155	HN	4.04

220	69	HD1*	111	HD1*	5.28
221	100	HG1*	111	HD1*	5.62
222	100	HG2*	111	HD1*	5.4
223	59	HD1*	121	HD1*	5.09
224	146	HD1*	171	HD2*	5.87
225	141	HD1*	146	HD1*	4.53
226	22	HD1*	84	HD1*	4.97
227	141	HD1*	171	HD2*	5.18
228	141	HD1*	146	HD1*	4.56
229	92	HD1*	141	HD1*	5.52
230	42	HD1*	155	HD2*	5.59
231	42	HD1*	122	HD1*	4.6
232	42	HD1*	155	HD1*	4.9
233	45	HD1*	59	HD1*	5.46
234	59	HD1*	121	HD1*	5.06
235	45	HD2*	59	HD1*	5.52
236	45	HD2*	59	HD1*	5.4
237	42	HD1*	155	HD2*	5.9
238	22	HD1*	84	HD1*	5.15
239	84	HD1*	93	HG2*	4.97
240	84	HD1*	155	HD2*	5.34
241	22	HD1*	91	HD1*	5.74
242	91	HD1*	177	HG2*	5.15
243	42	HD1*	122	HD1*	4.69
244	69	HD1*	111	HD1*	5.56
245	92	HD1*	141	HD1*	5.31
246	92	HD1*	152	HG1*	6.24
247	130	HG1*	152	HG2*	5.83
248	143	HD2*	152	HG2*	6.05
249	143	HD1*	152	HG2*	5.74
250	92	HD1*	152	HG2*	6.58
251	143	HD2*	152	HG1*	5.9
252	143	HD1*	152	HG1*	6.17
253	143	HD1*	152	HG1*	5.43
254	93	HG1*	157	HG2*	4.72
255	143	HD2*	152	HG1*	5.55
256	92	HD1*	143	HD2*	7.38
257	45	HD1*	50	HD1*	5.83
258	141	HD1*	171	HD2*	5.12
259	146	HD1*	171	HD2*	5.62
260	26	HD1*	177	HG2*	5.43

261	93	HG2*	157	HG1*	5.34
262	130	HG1*	154	HG1*	5.09
263	130	HG2*	154	HG1*	5.43
264	42	HD1*	155	HD1*	5.62
265	75	HG2*	80	HG2*	5.34
266	130	HG2*	154	HG1*	5.18
267	80	HG2*	100	HG2*	5.18
268	75	HG1*	80	HG1*	5.86
269	100	HG2*	111	HD1*	5.4
270	80	HG2*	100	HG2*	5.09
271	18	HD2*	119	HD1*	5.77
272	18	HD1*	119	HD1*	5.68
273	25	HD1*	41	HG2*	5.49
274	25	HD2*	41	HG2*	5.58
275	41	HG1*	127	HD1*	5.9
276	18	HD1*	119	HD1*	5.8
277	18	HD2*	119	HD1*	6.02
278	100	HG1*	111	HD1*	6.27
279	130	HG1*	154	HG1*	4.84
280	130	HG1*	152	HG2*	5.68
281	26	HD1*	177	HG1*	5.4
282	75	HG2*	80	HG1*	6.27
283	25	HD1*	41	HG2*	5.65
284	25	HD2*	41	HG2*	5.52
285	41	HG2*	42	HD1*	5.9
286	41	HG2*	155	HD1*	5.09
287	91	HD1*	177	HG2*	5.62
288	26	HD1*	177	HG2*	5.09
289	45	HD1*	59	HD1*	6.18
290	41	HG1*	127	HD2*	5.74
291	41	HG2*	155	HD1*	5.21
292	75	HG1*	80	HG1*	5.65

---

### III. Orientational restraints (RDC data) for structure calculation

Table A 3 RDC restraints for Est3<sup>AN</sup>

#	Residue number	bond vector	RDC value
1	12	HN	-2.69
2	13	HN	-10.418
3	14	HN	-9.524
4	15	HN	0.572
5	16	HN	-9.177
6	17	HN	-3.923
7	18	HN	0.348
8	19	HN	-13.294
9	20	HN	3.554
10	21	HN	4.292
11	22	HN	4.697
12	23	HN	2.256
13	24	HN	4.347
14	25	HN	-2.096
15	26	HN	-5.249
16	27	HN	-1.547
17	28	HN	5.246
18	29	HN	1.331
19	30	HN	1.127
20	31	HN	6.395
21	32	HN	1.118
22	33	HN	-0.496
23	34	HN	7.82
24	35	HN	6.385
25	36	HN	-4.577
26	37	HN	-4.679
27	38	HN	1.518
28	39	HN	6.445
29	40	HN	-0.491
30	41	HN	-9.245
31	42	HN	5.492
32	43	HN	13.763
33	44	HN	-9.044

34	45	HN	3.985
35	46	HN	5.147
36	47	HN	-7.503
37	48	HN	-14.208
38	49	HN	-12.247
39	50	HN	5.141
40	51	HN	5.522
41	52	HN	4.872
42	53	HN	4.527
43	54	HN	-1.484
44	55	HN	7.835
45	56	HN	3.52
46	57	HN	-9.995
47	58	HN	-14.529
48	59	HN	-9.731
49	60	HN	5.651
50	61	HN	17.65
51	62	HN	4.215
52	63	HN	-2.786
53	64	HN	-12.902
54	65	HN	-1.131
55	66	HN	1.335
56	67	HN	-15.272
57	68	HN	-14.234
58	69	HN	-13.563
59	70	HN	-17.733
60	71	HN	-21.343
61	72	HN	-5.065
62	73	HN	-10.713
63	74	HN	-14.836
64	75	HN	-8.844
65	76	HN	4.749
66	77	HN	-3.639
67	78	HN	-0.627
68	79	HN	16.158
69	80	HN	4.208
70	81	HN	-2.64
71	82	HN	-2.446
72	83	HN	-2.402
73	84	HN	-9.953
74	85	HN	1.637

75	86	HN	0.779
76	87	HN	6.801
77	88	HN	-0.627
78	89	HN	7.384
79	90	HN	2.363
80	91	HN	0.87
81	92	HN	2.078
82	93	HN	7.867
83	94	HN	10.379
84	95	HN	-12.094
85	96	HN	-7.813
86	97	HN	-13.21
87	98	HN	-13.552
88	99	HN	-7.742
89	100	HN	-4.663
90	101	HN	6.346
91	102	HN	-0.907
92	103	HN	6.314
93	104	HN	-14.143
94	105	HN	5.868
95	106	HN	8.997
96	107	HN	6.382
97	108	HN	2.347
98	109	HN	-3.076
99	110	HN	5.774
100	111	HN	1.723
101	112	HN	-4.537
102	113	HN	1.276
103	114	HN	-3.061
104	115	HN	-0.455
105	116	HN	-1.636
106	117	HN	4.838
107	118	HN	7.169
108	119	HN	-0.957
109	120	HN	-6.827
110	121	HN	-11.812
111	122	HN	-3.562
112	123	HN	1.143

---