Semi-rational protein engineering strategies to reduce protein aggregation within mammalian cells

---

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Sciences

University of Virginia

---

in partial fulfillment

of the requirements for the degree

Doctor of Philosophy

by

Simpson Emmanuel Gregoire

May 2014

(Defended December 19, 2013)

Acknowledgments

This dissertation is dedicated in memory of Tereston (TJ) Bertrand Jr. Although he is not here physically, he has guided me throughout graduate school spiritually. If I did not emulate his drive, passion and willingness to help others, I would not have been able to finish.

I would like to acknowledge my research advisor, Professor Inchan Kwon. Without his guidance, insight and confidence in me throughout my graduate career, I would not have been nearly as successful in completing the various projects and tasks that were asked of me. I have been blessed to have him as an advisor and hope to continue our conversations in the future.

I would also like thank Prof. Roseanne Ford, Prof. Giorgio Carta, Prof. John O'Connell and Prof. Jason Papin for serving on my dissertation committee. I truly appreciate the time, effort and advice they have provided me over the past year.

I am lucky to have such supporting lab mates, Ed Wong, Shun Zheng, Jacob Irwin and Sung In Lim. In addition to my lab mates, I would like to thank Kelly Glitzos, Arvind Rachamadugu and Yomi Famuyiwa for their assistance over the years. Finally, I would like to acknowledge my friends and family, specifically my parents, Nina and Simpson, and my siblings, Nadine and Brendon, and my best friend, Serena, for their continuing love and support during my PhD.

Abstract

Protein misfolding and aggregation are often causative to the onset of numerous neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, Huntington's disease, and Amyotrophic lateral sclerosis (ALS). Further, intracellular aggregate formation and misfolding of biopharmaceuticals in mammalian cells are often obstacles to achieving high-yield production of biopharmaceuticals. Thus, understanding protein aggregation mechanism and preventing protein aggregation is imperative from an industrial and clinical perspective.

This dissertation details two novel semi-rational protein engineering approaches to prevent protein misfolding and aggregation in mammalian cells. In particular, this work investigated the aggregation of human copper-zinc superoxide dismutase (SOD1) as a model protein. Many reports indicate that twenty percent of the familial form of amyotrophic lateral sclerosis (fALS) is caused by misfolding and aggregation of mutant SOD1. In particular, this work concentrates on two particular fALS SOD1 mutants, $SOD1^{A4V}$ and $SOD1^{G93A}$. The most common fALS-linked mutant SOD1 in the US is one containing alanine to valine mutation at 4th residue (A4V). We hypothesized that intracellular protein aggregation can be suppressed by increasing conformational stability of the protein. Using the computational software RosettaDesign, Phe20 was chosen in $SOD1^{A4V}$ as a key residue responsible for $SOD1^{A4V}$ conformational destabilization. This information was used to rationally develop a pool of candidate mutations at the Phe20 site. In order to evaluate how these mutations affect the relative aggregation propensity of the target protein inside cells, we developed a mammalian cell-based screening assay that directly correlates aggregation/misfolding propensity to mean cellular fluorescence. We performed two rounds of screening to select the most promising variants for further characterization. Three novel $SOD1^{A4V}$ variants, $SOD1^{A4V/F20G}$, $SOD1^{A4V/F20A}$ and

$SOD1^{A4V/F20A/C111S}$ were determined to have a significantly reduced aggregation propensity inside cells, demonstrating that the semi-rational protein engineering strategy to increase conformational stability by an additional mutation(s) can be used to reduce the aggregation propensity inside mammalian cells

The aforementioned strategy is beneficial when the destabilizing factor can be pinpointed easily. However, this approach has the limited utility when no rational design approach to enhance conformational stability is readily available. We hypothesized that the aggregation of a target protein within mammalian cells can be reduced using a protein engineering strategy to increase kinetic stability The aggregation-prone mutant $SOD1^{G93A}$ was used as a model system. Unlike the repulsive interaction between two side chains caused by the A4V mutation in $SOD1^{A4V}$, the G93A mutation causes the repulsive interaction with the neighboring protein backbone resulting in the backbone shift and multiple side chain displacements and eventually the $SOD1^{G93A}$ aggregation. A panel of $SOD1^{G93A}$ variants containing a mutation at three residues (K36, E40, and K91) of which side chain were significantly displaced upon the G93A mutation were generated and subjected to mammalian cell-based screening. Three novel $SOD1^{G93A}$ variants ($SOD1^{G93A/K36E}$, $SOD1^{G93A/K91D}$ and $SOD1^{G93A/K91E}$) with a reduced aggregation propensity inside mammalian cells were identified. Although the conformational stability of these variants is comparable to that of $SOD1^{G93A}$, the aggregation rate of these variants is substantially lower than that of $SOD1^{G93A}$ supporting that the three variants are kinetically more stable than $SOD1^{G93A}$. From these studies, we have concluded that the semi-rational protein engineering strategies consisting of computational stability analysis, protein structure inspection, and cell-based screening of protein variant has tremendous utility for understanding and modulating protein aggregation in mammalian cells.

Table of Contents

Acknowledgements ii

Abstract iii

Table of Contents v

List of Figures viii

List of Tables xiii

Chapter 1: Introduction 1

Background and Significance 2
Protein Misfolding/Aggregation 2
Protein Stability 3
Amyotrophic Lateral Sclerosis and Human Copper/Zinc Superoxide Dismutase 4
Monitoring Protein Aggregation/Misfolding 8
Computational Design Using RosettaDesign 9
Semi-Rational Protein Engineering 10
Goals and Objectives 11

Chapter 2: A Revisited Folding Reporter for Quantitative Assay of Protein Misfolding and Aggregation in Mammalian Cells 13

Abstract 14
Introduction 15
Materials and Methods 19
Materials 19
Construction of Expression Vectors 19
Transfection of HEK293T and NSC-34 cells 19
Fluorescence Microscopy 20
Flow Cytometric Analysis of Cellular Fluorescence 20
Dot-blotting 21
Results and Discussion 21
Fluorescence Microscopic Observation of Transfected HEK293T Cells Expressing SOD1 Variants 21
Flow Cytometric Analysis of Transfected HEK293T Cells Expressing SOD1 Variants 22
Aggregate Formation of $SOD1^{A4V}$ Variants in Transfected HEK293T Cells 29
Comparison of the Cellular Fluorescence of Transfected NSC-34 and HEK293T Cells Expressing SOD1 Variants 31
Comparison of the CMV and Ubiquitin-C Promoters 36
Conclusion 40

Chapter 3: Cis-Suppression to Arrest Protein Aggregation in Mammalian Cells 42

Abstract 43
Introduction 44
Materials and Methods 46
Materials 46
Construction of Expression Vectors 46
Computational Analysis of Protein Structure 47
Transfection of HEK293T and NSC-34 Cells 48
Fluorescence Microscopy 49
Flow Cytometric Analysis of Cellular Fluorescence 49
Total Protein Extraction and Protein Fractionation of HEK293T cells 50
Western Blotting 50
Expression and Purification of SOD1 Mutants 51
Guanidine Hydrochloride-induced denaturation assay 52
Results and Discussion 53
Identification of key residues responsible for $SOD1^{A4V}$ conformational destabilization using computational protein design 53
Screening of $SOD1^{A4V}$ variants containing a mutation at the F20 site to identify mutations eliminating the $SOD1^{A4V}$ destabilization 56
Fluorescence microscopic analysis of the transfected HEK293T cells expressing the $SOD1^{A4V}$ variants 61
Identification of SOD1 triple mutants with a significantly reduced misfolding/aggregation propensity 63
Solubility assay of the novel $SOD1^{A4V}$ variants to estimate their relative aggregation propensities 70
Denaturation assay of the novel SOD1 variants to estimate their relative stabilities 72
Conclusion 74

Chapter 4: Suppressing the Protein Aggregation Expressed in Mammalian Cells by Increasing Kinetic Stability 76

Abstract 77
Introduction 78
Materials and Methods 80
Materials 80
Construction of Expression Vectors 81
Computational Analysis of Protein Structure 82
Homology Modeling and Crystal Structure Analysis 83
Transfection of HEK293T and NSC-34 Cells 83
Fluorescence Microscopy 83
Flow Cytometric Analysis of Cellular Fluorescence 84
Total Protein Extraction and Protein Fractionation of HEK293T cells 84
Western Blotting 85

Results and Discussion.....86

Determination of key factors responsible for the destabilization of $SOD1^{G93A}$ using *in silico* techniques 86
Combination of rational design and cell-based screening in order to identify to identify $SOD1^{G93A}$ variants with a reduced aggregation propensity 88
Fluorescence microscopic analysis of the promising $SOD1^{G93A}$ variants expressed in HEK293T cells.....93
Determination of the aggregation propensity of three $SOD1^{G93A}$ variants by using cell lysate fractionation.....96
Enhanced kinetic stability of the novel $SOD1^{G93A}$ variants with a reduced aggregation-propensity.....100

Conclusion 102

Chapter 5: Project Summary and Avenues for Future Work.....105

Project Summary 106
Future Work and Suggestions 108
Understanding the effect of reducing mutant SOD1 aggregation on cellular toxicity 108
Directed evolution of fALS-related SOD1 mutants to increase stability 109
Trans-suppression of SOD1.....110

References.....113
Appendix A.....125
Appendix B.....133

List of Figures

Figure 1.1: Scheme of protein folding and misfolding/aggregation 2

Figure 2.1: Fluorescence microscopy images of the transfected HEK293T cells expressing SOD1 variants fused to EGFP. (A) Images of transfected HEK293T cells expressing EGFP fusion of $SOD1^{WT}$, $SOD1^{A4V/C111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$ taken at two days post-transfection. Transfection efficiency for each SOD1 variant was determined by calculating the percentage of GFP positive cells out of total cells analyzed using flow cytometry. Values represent means ± standard deviation (n = 3). Scale bars are 250 μm. (B) Magnified fluorescence microscopic images of the transfected HEK293T cells expressing EGFP fusion of $SOD1^{WT}$ and $SOD1^{A4V}$. Arrows indicate the aggregates within the cell. Scale bars are 50 μm 23

Figure 2.2: Comparison of the cellular fluorescence intensities and expression levels of EGFP fusion of SOD1 variants. (A) Time course of fluorescence intensities of the transfected HEK293T cells expressing EGFP fusion of SOD1 variants. Fluorescence intensities of untransfected cells (day 0) and transfected cells expressing EGFP fusion of four SOD1 variants ($SOD1^{WT}$, $SOD1^{A4V/C111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$) at day 1 and 2 were determined by flow cytometry. Values and error bars represent mean cellular fluorescence and standard deviations, respectively (n = 3). Two-sided Student's t-tests were applied to the data (*: $P < 0.05$; **: $P < 0.001$). (B) Comparison of the mean cellular fluorescence of HEK293T cells expressing SOD1 variants that were transfected at two separate days. At two days post-transfection, cellular fluorescence of transfected HEK293T cells expressing EGFP fusion of SOD1 variants ($SOD1^{WT}$, $SOD1^{A4V/C111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$) was determined by flow cytometry. The data were fitted to a straight line ($R^2 = 0.99$). (a.u. = arbitrary unit) Numerical values of the mean cellular fluorescence intensities are shown in Table A1 in Appendix A (C) Dot-blot images of the transited cell lysates of EGFP fusion of $SOD1^{WT}$ (WT), $SOD1^{A4V/C111S}$ (A4V/C111S), $SOD1^{A4V}$ (A4V), and $SOD1^{A4V/C57S}$ (A4V/C57S) using anti-SOD1 polyclonal antibody. (α-tubulin; loading control).....24

Figure 2.3: Mean cellular fluorescence intensities of transfected HEK293T cells co-expressing split GFP fragments fused to $SOD1^{WT}$ or $SOD1^{A4V}$. In order to achieve split GFP complementation, both pCMV-mGFP_1-10 and pCMV-mGFP_Cterm_S11 plasmids encoding large and small GFP fragment genes, respectively, which were originally developed by Waldo group, and obtained from Theranostech, Inc. (Albuquerque, NM) (Cabantous et al. 2005b; Cabantous and Waldo 2006). PCR-amplified $SOD1^{WT}$ and $SOD1^{A4V}$ genes were inserted between *Nhe*I and *BamH*I restriction sites of pCMV-mGFP_Cterm_S11 to generate pCMV_$SOD1^{WT}$-mGFP_S11 and pCMV-$SOD1^{A4V}$-mGFP_S11, respectively. HEK293T cells were co-transfected with pCMV-mGFP_1-10 and either pCMV_$SOD1^{WT}$-mGFP_S11 or pCMV-$SOD1^{A4V}$-mGFP_S11 plasmid. Mean cellular fluorescence intensities of the transfected HEK293T cells were determined by flow cytometry at two days post-transfection. Values indicate the mean cellular fluorescence intensities and standard deviations (n = 3) 30

Figure 2.4: Aggregate formation of $SOD1^{A4V/C111S}$ and $SOD1^{A4V/C57S}$ in transfected HEK293T cells. Fluorescence microscopy images of the transfected HEK293T cells expressing EGFP

fusion of $SOD1^{A4VC111S}$ and $SOD1^{A4V/C57S}$ taken at two days post-transfection. White arrows indicate intracellular aggregates 32

Figure 2.5: The percentage of transfected HEK293T cells exhibiting SOD1 aggregates. The number of cells exhibiting intracellular SOD1 aggregates was determined by analyzing fluorescence microscopy images of the transfected cells expressing EGFP fusion of four SOD1 variants ($SOD1^{WT}$, $SOD1^{A4V/C111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S)}$ at day two post-transfection. Values and error bars represent mean and standard deviations, respectively (n = 3). Two-sided Student's t-tests were applied to the data (*: $P < 0.05$). Numerical values of the mean cellular fluorescence intensities are shown in Table A2 in Appendix A.....33

Figure 2.6: Comparison of the orders of fluorescence intensities of transfected HEK293T and NSC-34 cells expressing EGFP fusion of SOD1 variants. (A) Fluorescence intensities of the transfected NSC-34 cells expressing EGFP fusion of SOD1 variants ($SOD1^{WT}$, $SOD1^{A4VC111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$) at two days post-transfection were determined by flow cytometry. Values and error bars represent mean cellular fluorescence and standard deviations, respectively (n = 3). (a.u. = arbitrary unit) Two-sided Student's t-tests were applied to the data (*: $P < 0.05$; **: $P < 0.01$). Numerical values of the mean cellular fluorescence intensities are shown in Table A1 in Appendix A. (B) Comparison of the cellular fluorescence of the transfected HEK293T at day two and NSC-34 cells at day three post-transfection. The cellular fluorescence of transfected HEK293T and NSC-34 cells expressing EGFP fusion of SOD1 variants ($SOD1^{WT}$, $SOD1^{A4VC111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$) was determined by flow cytometry. The data were fitted to a straight line ($R^2 = 0.98$). $FL(SOD1^{X})/FL(SOD1^{WT})$ indicates the ratio of fluorescence intensity of cells expressing a SOD1 variant ($SOD1^{X}$) over the fluorescence intensity of cells expressing $SOD1^{WT}$ 35

Figure 2.7: Comparison of CMV and Ubc promoter activities for the fluorescence intensities of transfected HEK293T cells expressing EGFP fusion of SOD1 variants. Cellular fluorescence of HEK293Tcells expressing EGFP fusion of $SOD1^{WT}$ and $SOD1^{A4V}$ under the control of either CMV or Ubc promoter was determined by flow cytometry. Values and error bars represent mean cellular fluorescence and standard deviations, respectively (n = 3). (a.u. = arbitrary unit) Numerical values of the mean cellular fluorescence intensities are shown in Table A1 in Appendix A.....38

Figure 2.8: Mean cellular fluorescence intensities of transfected HEK293T cells co-expressing firefly luciferase (Luc) and either $SOD1^{WT}$-EGFP or $SOD1^{A4V}$-EGFP. pAAV-Luc plasmid encodes Luc gene under the control of CMV promoter (Jang et al. 2010). HEK293T cells were co-transfected with pAAV-Luc and either pEGFP-N3-$SOD1^{WT}$ or pEGFP-N3-$SOD1^{A4V}$ plasmid. Mean cellular fluorescence intensities of the transfected HEK293T cells were determined by flow cytometry at two days post-transfection. Values indicate the mean cellular fluorescence intensities and standard deviations (n = 3).....39

Figure 3.1: The crystal structure of $SOD1^{A4V}$ (PDB ID: 1U70). The residues near Val4 in one subunit (cyan color) are shown in green color. The other subunit is shown in gray spheres. The structure is generated using PyMol graphic software (Schrodinger 2010)..... 54

Figure 3.2: The mean cellular fluorescence of the transfected HEK293T (A) and NSC-34 (B) cells expressing EGFP fusion of SOD1 variants. (C) The SOD1 bands of the transfected HEK293T cells expressing the SOD1 variants. Values and error bars represent mean cellular fluorescence and standard deviations, respectively (n = 3). In order to determine whether the mean cellular fluorescence of the SOD1 double mutants is greater than that of $SOD1^{A4V}$, two-sided Student's t-tests were applied to the data (* $p < 0.05$; ** $p < 0.01$).....58

Figure 3.3: The adjusted mean cellular fluorescence of the transfected cells expressing the EGFP fusion of $SOD1^{WT}$, $SOD1^{A4V}$, and $SOD1^{A4V}$ variants containing a second mutation. The values were calculated by dividing the mean cellular fluorescence of each sample by the relative expression level to that of $SOD1^{WT}$. Values = mean ± standard deviation (n = 3). In order to determine whether the mean cellular fluorescence of the SOD1 double mutants is greater than that of $SOD1^{A4V}$, two-sided Student's t-tests were applied to the data (* $p < 0.05$; ** $p < 0.01$) 60

Figure 3.4: The fluorescence microscopic images of the transfected HEK293T cells expressing EGFP fusion of SOD1 variants. The images of transfected HEK293T cells expressing EGFP fusion of $SOD1^{WT}$, $SOD1^{A4V}$, and $SOD1^{A4V}$ double mutants were taken at two days post-transfection.....62

Figure 3.5: The percentage of the transfected HEK293T cells exhibiting SOD1 aggregates. The number of cells exhibiting intracellular SOD1 aggregates was determined by analyzing fluorescence microscopy images of the transfected cells expressing EGFP fusion of five SOD1 variants ($SOD1^{WT}$, $SOD1^{A4V}$, $SOD1^{A4V/C111S}$, $SOD1^{A4V/F20A}$, and $SOD1^{A4V/F20G}$) at two days post-transfection.....64

Figure 3.6: The mean cellular fluorescence of the transfected (A) HEK293T cells and (B) NSC-34 expressing EGFP fusion of the SOD1 variants ($SOD1^{WT}$, $SOD1^{A4V}$, $SOD1^{A4V/F20A/C111S}$, and $SOD1^{A4V/F20G/C111S}$). (C) The SOD1 bands of the transfected HEK293T cells expressing the SOD1 variants.Values and error bars represent mean cellular fluorescence and standard deviations, respectively (n = 3). In order to determine whether the mean cellular fluorescence of the SOD1 triple mutants is greater than that of $SOD1^{A4V}$, two-sided Student's t-tests were applied to the data (* $p < 0.05$; ** $p < 0.01$).....65

Figure 3.7: The adjusted mean cellular fluorescence of the transfected cells expressing the EGFP fusion of $SOD1^{WT}$, $SOD1^{A4V}$, $SOD1^{A4V/F20A/C111S}$, and $SOD1^{A4V/F20G/C111S}$. The values were calculated by dividing the mean cellular fluorescence of each sample by the relative expression level to that of $SOD1^{WT}$. Values = mean ± standard deviation (n = 3). In order to determine whether the mean cellular fluorescence of the SOD1 triple mutants is greater than that of $SOD1^{A4V}$, two-sided Student's t-tests were applied to the data (* $p < 0.05$; NS: not significant). The mean cellular fluorescence of the $SOD1^{A4V/F20A/C111S}$ is not significantly different from that of $SOD1^{WT}$ 67

Figure 3.8: The fluorescence microscopic images of the transfected HEK293T cells expressing EGFP fusion of SOD1 variants. The images of the transfected HEK293T cells expressing EGFP

fusion of the $SOD1^{WT}$, $SOD1^{A4V}$ and $SOD1^{A4V}$ triple mutants were taken at two days post-transfection.....68

Figure 3.9: The percentage of the transfected HEK293T cells exhibiting SOD1 aggregates for the SOD1 variants ($SOD1^{WT}$, $SOD1^{A4V}$, $SOD1^{A4V/F20G}$, and $SOD1^{A4V/F20A/C111S}$) 69

Figure 3.10: (A) The western-blot bands of the insoluble and soluble fractions of the transfected HEK293T cells expressing the SOD1 variants. (B) The relative aggregation propensity of SOD1 proteins at 48 hrs. NC: untransfected cells. Values = mean ± standard deviation (n = 3). In order to determine whether the SOD1 variant aggregation propensity is smaller than that of $SOD1^{A4V}$, two-sided Student's t-tests were applied to the data (*; $p < 0.05$).....71

Figure 3.11: The denaturation curves of the $SOD1^{WT}$, $SOD1^{A4V}$, $SOD1^{A4V/F20G}$, $SOD1^{A4V/F20A}$, and $SOD1^{A4V/F20A/C111S}$ 73

Figure 4.1: Structural analysis of $SOD1^{WT}$ (PDB ID: 2C9V) and $SOD1^{G93A}$ (PDB ID:2WKO). (**A**) Ala93 (Blue) mutation causes shift in backbone of $SOD1^{G93A}$ in comparison to $SOD1^{WT}$ ($SOD1^{WT}$ backbone in yellow; $SOD1^{G93A}$ backbone in green) (**B**) Ala93 (blue) shifts the amino acid back bone of Leu38 (gray) (**C**) The crystal structure (PDB ID: 2C9V) of the holo-form of $SOD1^{WT}$. Residues of interest within 5 angstroms were highlighted as specified. (**D**) The crystal structure (PDB ID: 2WKO) of the holo-form of $SOD1^{G93A}$. Residues of interest within 5 angstroms were highlighted as specified 89

Figure 4.2 Residues of $SOD1^{G93A}$ substantially displaced compared to those of $SOD1^{WT}$. Residues K36 (*purple*) and G93 (*red*) (**A**); E40 (*orange*) and G93 (*red*) (**B**); and K91 (*yellow*) and G93 (*red*) (**C**) within crystal structure of the holo-form of $SOD1^{WT}$ (PDB ID: 2C9V). Residues K36 (*purple*) and A93 (*blue*) (**D**); E40 (*orange*) and A93 (*blue*) (**E**); and K91 (*yellow*) and A93 (*blue*) (**F**) within crystal structure of the holo-form of $SOD1^{G93A}$ (PDB ID: 2WKO)....90

Figure 4.3: The mean cellular fluorescence of the transfected HEK293T (**A**) and NSC-34 (**B**) cells expressing EGFP fusion of SOD1 variants. (**C**) The SOD1 bands of the transfected HEK293T cells expressing the SOD1 variants. Values and error bars represent mean cellular fluorescence and standard deviations, respectively (n=3). In order to determine whether the mean cellular fluorescence of the SOD1 double mutants is greater than that of $SOD1^{G93A}$ in each cell line, two-sided Students t-tests were applied to the data (* $P<0.05$; ** $P<0.01$) 92

Figure 4.4: The adjusted mean cellular fluorescence of the transfected cells expressing the EGFP fusion of $SOD1^{WT}$, $SOD1^{G93A}$, and $SOD1^{G93A}$ variants containing a second mutation. The values were calculated by dividing the mean cellular fluorescence of each sample by the relative expression level to that of $SOD1^{WT}$ Values = mean ± standard deviation (n = 2). In order to determine whether the mean cellular fluorescence of the SOD1 double mutants is greater than that of $SOD1^{G93A}$, two-sided Student's t-tests were applied to the data (* $p < 0.05$; ** $p < 0.01$)

.....94

Figure 4.5: The fluorescence microscopic images of the transfected HEK293T cells expressing EGFP fusion of SOD1 variants. The images of transfected HEK293T cells expressing EGFP fusion of $SOD1^{WT}$, $SOD1^{G93A}$, and $SOD1^{G93A}$ double mutants were taken at 2 days post-transfection. (white arrows: intracellular SOD1 mutant aggregates; scale bar: 50 µm).....95

Figure 4.6: The fluorescence microscopic images of the transfected HEK293T cells expressing EGFP fusion of SOD1 variants. The images of transfected HEK293T cells expressing EGFP fusion of $SOD1^{G93A}$ double mutants were taken at 2 days post-transfection. Scale Bar: 50 µm.....

.....97

Figure 4.7: The fraction of the transfected HEK293T cells exhibiting SOD1 aggregates. The number of cells exhibiting intracellular SOD1 aggregates was determined by analyzing fluorescence microscopy images of the transfected cells expressing EGFP fusion of four SOD1 variants ($SOD1^{G93A}$, $SOD1^{G93A/K36E}$, $SOD1^{G93A/K91D}$ and $SOD1^{G93A/K91E}$) and $SOD1^{WT}$ at 2 days post-transfection ($p<0.01$ for all samples compared to $SOD1^{G93A}$).....98

Figure 4.8: The relative aggregation propensity of SOD1 proteins at 48 h. In order to determine whether the SOD1 variant aggregation propensity is smaller than that of $SOD1^{G93A}$, two-sided Students t-tests were applied to the data (*, $P<0.05$) (n=3).....99

Figure 4.9: The western-blot bands of the insoluble and soluble fractions of the transfected HEK293T cells expressing the SOD1 variants.....101

Figure 4.10: The fraction of the transfected HEK293T cells exhibiting intracellular SOD1 aggregates determined at 0, 24, 48 and 72 hours post-transfection.....103

Figure A1: Representative overlaid histograms of fluorescence of transfected cells expressing EGFP fusion of four SOD1 variants ($SOD1^{WT}$, $SOD1^{A4V/C111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$) .130

Figure A2: Immuno-reactivity of untransfected HEK293T cell lysate (NC) and transfected HEK293T cell lysate expressing $SOD1^{WT}$-EGFP (WT). Either untransfected or transfected HEK293T cells were centrifuged to obtain a cell pellet. After washing with PBS buffer once, the cell pellet was lysed using RIPA buffer (Thermo Fisher, Pittsburgh, PA). Protein concentration of the cell lysate was determined using BCA kit using bovine serum albumin as a standard. Dot-blot assay of the cell lysate was performed as described earlier (Wong and Kwon 2011; Wong et al. 2011) except anti-SOD1. The blot images were captured using a BioSpectrum imaging system (UVP).....131

Figure A3: Gating strategy for flow cytometric analysis. (A) P3 gate was set in order to analyze only viable cells. (B) A histogram of the negative control cellular fluorescence is shown. Only transfected cells in M3 region were analyzed to determine the mean cellular fluorescence.....132

List of Tables

Table 3.1: Summary of the change in total energy score ($\Delta\Delta G_f$) of the $SOD1^{A4V}$ variants relative to that of $SOD1^{WT}$ using two wild-type starting structures from the PDB (1PU0 and 2C9V) prepared as described in the methods 55

Table 4.1: Scoring summary of $SOD1^{WT}$ and $SOD1^{G93A\_WT}$ variants 87

Table A1: Mean cellular fluorescence intensities 128

Table A2: The percentage of transfected HEK293T cells exhibiting SOD1 aggregates 129

Table B1: Mean cellular fluorescence intensities 134

Table B2: The percentage of transfected HEK293T cells exhibiting SOD1 aggregates 135

Chapter 1: Introduction

Background and Significance

Protein Misfolding/Aggregation

The complex process of protein folding is subject to close examination because of the intertwined networks that participate in the process of folding a protein into its final conformation. The delicate cellular machinery that turns a linear, heteropolymeric chain into a three-dimensional, biologically functional protein must operate such that there is very little error in order to prevent aberrant, abnormal species (Figure 1.1) (Dantas et al. 2003). These off-pathway species are known as protein aggregates. Researchers have determined that various molecular interactions are involved in stabilizing aggregates (Gsponer and Vendruscolo 2006; Jahn and Radford 2008; Roberts 2007). Covalent bonds, electrostatics, Van Der Waals interactions, and other harsh environmental changes (pH, temperature, presence of denaturant) have been shown to initiate and stabilize misfolded species.

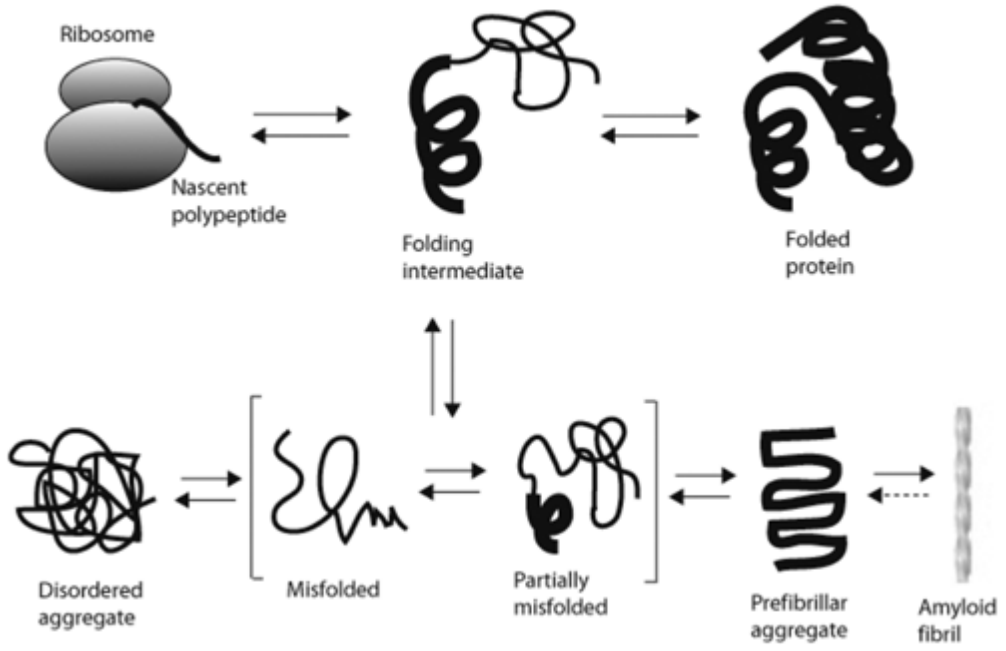


Figure 1.1: Scheme of protein folding and misfolding/aggregation

It is of particular importance to understand protein aggregation and its role in biopharmaceuticals and protein deposition diseases in humans (Chiti et al. 2002; Lee et al. 2007; Tyedmers et al. 2010). As it pertains to biopharmaceuticals, misfolding and subsequent aggregation of protein drugs lead to increase immunogenicity, reduced drug efficacy, and lack of patient trust when administered within the body. Additionally, protein aggregation plays a factor in various purification steps that occur before the protein reaches the customer. Additional purification steps may lead to additional costs and longer production times. Also, intra- and extra-cellular protein aggregation has been associated with at least 20 neurological disorders, also known as conformational disorders. Huntington's, Alzheimer's, Parkinson's, Amyotrophic Lateral Sclerosis (ALS), and other diseases have target proteins that form aggregates that lose activity or gain a toxic function that leads to the development of the disease. Although the proteins associated with various diseases have no sequence homology and varying pathologies, they share similarities in the development of aggregates. Hence, it is worthwhile to understand how these proteins form aggregates because it may lead to insight into aggregation of all proteins associated with conformational diseases.

Protein Stability

Uncovering the mechanism behind the intricate process of protein folding was proven difficult until recently. Substantial progress has been made in understanding the fundamental parameters behind protein folding. As linear chains of amino acids begin to fold, the protein begins to form a dynamic structure that is in its lowest energy state (Dobson 2003). Through molecular evolution, proteins have found more efficient paths to form their lowest energy state (Dobson 2004). During the process of protein folding and reaching particular conformations,

protein stability must be taken into consideration. In particular, protein stability can be observed from a kinetic or thermodynamic standpoint (Pey et al. 2008; Sanchez-Ruiz 2010). Thermodynamic stability is a measure of free energy where native and unfolded structures have reached equilibrium. Thermodynamically stable proteins retain their biological function when placed under harsh environmental conditions and allowed to reach equilibrium. However, kinetic stability is less straightforward. Kinetic stability is the capability of a protein to remain biologically active over a biologically-relevant period of time. Although a protein may not be thermodynamically stable or remain in its native-conformation, it may retain its biological activity because of a high energy barrier that may prevent it from progressing to an irreversible state (aggregated, unfolded, etc.) (Baker and Agard 1994; Rodriguez-Larrea et al. 2006; Sanchez-Ruiz 2010).

Enhancement of protein stability is a desired in many therapeutic or biotechnological applications of proteins. For many biotechnological processes, increased protein stability has substantial economic implications, such as reducing costs of raw materials, longer usage in undesirable conditions (Eijsink et al. 2004). Also, for many of the proteins implicated in human protein deposition diseases, cellular environment and introduction of point mutations play a substantial role in decreasing the stability, kinetic, thermodynamic or both, leading to the development of toxic or non-functional proteins. (Kamata et al. 2009; Lynch et al. 2004; Manning and Colon 2004; McCutchen et al. 1995).

Amyotrophic Lateral Sclerosis and Human Copper/Zinc Superoxide Dismutase

Amyotrophic lateral sclerosis (ALS), often referred to as Lou Gehrig's disease, is a progressive neurodegenerative disorder that attacks and inevitably leads to the death of motor

neurons in the brain, brain stem, and spinal cord. ALS is the most common motor neuron disease in humans, affecting 4 to 6 out of 100,000 (Andersen 2006; Cleveland and Rothstein 2001; Reaume et al. 1996; Valentine et al. 2005). Degeneration of these motor neurons leads to loss of control of the voluntary muscles, muscle spasticity, muscle atrophy and eventually, paralysis and death. Unfortunately, the average lifespan of a patient who has been diagnosed with ALS is between 3-5 years, and there is only one FDA-approved drug on the market that marginally extends the patient's lifespan.

There are two forms of ALS, the more prevalent sporadic ALS and the inherited form, familial ALS (fALS). The first implications that there was a genetic link to ALS were found in 1993. Rosen et al. reported that nearly 20% of fALS cases involve mutations in the human zinc/copper superoxide dismutase (SOD1), eventually leading to the onset of ALS (Rosen et al. 1993). Researchers have since found over 100 mutations in SOD1 that are known to cause ALS (Banci et al. 2008). SOD1, a homodimeric antioxidant protein, is prevalent throughout the body (Gurney et al. 1994). By examining transgenic mice expressing fALS-linked mutant SOD1 genes, researchers could determine what caused the onset of fALS. Researchers observed that SOD1-null mice did not exhibit any ALS-linked symptoms, weakening the hypothesis that ALS was linked to altered superoxide scavenging activity (Bruijn and Cleveland 1996; Reaume et al. 1996). However, when transgenic mice expressed mutant SOD1 despite maintaining normal SOD1 activity, the mice exhibited ALS-like symptoms. From these two findings, researchers now believe that mutant SOD1 has gained an unknown toxic function that affects motor neurons, eventually leading to the development of ALS (Borchelt et al. 1994).

There has been much debate as to where this toxic property emerges. It has been proposed that the gain of toxic function associated with mutant SOD1s is linked to mutant SOD1

aggregation (Karch et al. 2009; Ray et al. 2005; Shaw and Valentine 2007; Wang et al. 2008; Zhang and Zhu 2006). In other neurological disorders such as Alzheimer's disease and Huntington's disease, the hallmark feature is the aggregates development in the nervous system. Therefore, considerable work has been done to determine if aggregate development in the spinal cord also leads to the onset of ALS. Researchers have found variant SOD1 deposits in the spinal cords of transgenic mice, ALS patients, and in cell culture models expressing mutant SOD1, implying a strong relationship between the ALS development and SOD1 aggregation (Jonsson et al. 2004; Karch et al. 2009; Nagai et al. 2007; Prudencio et al. 2009; Zhang and Zhu 2006). Although the mechanism that these aggregates impose a toxic function to motor neurons remains unclear, researchers have proposed various reasons for how these aggregates affect cellular function, including disrupting function of essential proteins or cellular compartments upon the aggregate binding (Cluskey and Ramsden 2001; Pasinelli and Brown 2006).

In other familial forms of protein-misfolding diseases, such as Parkinson's disease and amyloidoses, many mutations are known to promote aggregate formation by destabilizing the native state of the protein. Similarly, over 100 mutations of SOD1 increase aggregation propensity of SOD1 by perturbing the native state of the SOD1. In particular, the alanine to valine mutation in the fourth residue ($SOD1^{A4V}$) leads to local unfolding in a substantial portion of the β-barrel structure of apo-form of $SOD1^{A4V}$ (Shaw and Valentine 2007). $SOD1^{A4V}$ has been frequently studied, because it is the most common mutant in North America and leads to a rapidly progressing form of FALS (Ray and Lansbury 2004). $SOD1^{A4V}$ leads to a protrusion in the dimer interface (Stathopulos et al. 2003). As a result, $SOD1^{A4V}$ forms aggregates. Under oxidative stress, histidine residues involved in zinc binding are selectively oxidized such that a zinc ion can be released from mutant SOD1 to form zinc-deficient mutant SOD1, which has a

high potential for substantial aggregation. Further oxidation of FALS-associated $SOD1^{A4V}$ leads to loss of all metal ions (apo-$SOD1^{A4V}$), with the resulting extremely unstable monomeric SOD1 exhibiting rapid aggregation (Rakhit et al. 2004). When the $SOD1^{A4V}$ fused with enhanced GFP (EGFP) was over-expressed in cultured cells, its aggregates were clearly observed and remaining fluorescence in the cytosol was low, whereas wild type SOD1 fused with EGFP were evenly distributed throughout the cytosol. Similarly, $SOD1^{A4V}$ aggregates and evenly distributed wild-type SOD1 in the cytosol were also observed in cultured NSC-34 motor neuron-like cells (Turner et al. 2005).

In addition to $SOD1^{A4V}$, understanding the misfolding and subsequent aggregation process for the fALS SOD1 mutant $SOD1^{G93A}$ is of particular importance. It has been hypothesized that $SOD1^{G94A}$ aggregation is also implicated in motor neuron death in fALS patients with this particular mutation (Johnston et al. 2000; Kilic et al. 2004). There has been significant advance in understanding the pathological hallmarks of mutant SOD1-associated fALS. One of the key results from SOD1-associated fALS research is the development of a transgenic mouse line which expresses the $SOD1^{G93A}$ (Gurney et al. 1994). This transgenic mouse line has been extensively studied and most closely resembles the pathological hallmarks of the disease in humans.

Stabilizing the folded conformation of mutant SOD1 is a promising therapeutic approach to blocking the formation of toxic aggregates. The considerable therapeutic potential of this strategy has inspired the development of various chaperone molecules that promote the correct folding of target proteins during biological processes. For example, over-expression of the hsp70 protein chaperone reduces aggregate formation of a familial form of ALS-associated mutant

SOD1 in a transgenic mouse (Takeuchi et al. 2002). However, the therapeutic application of a protein chaperone is limited because its effect is not specific to the target protein and hence it might result in unexpected side effects (Bukau and Horwich 1998). Alternatively, small molecules stabilizing the folded $SOD1^{A4V}$ dimer have been identified by *in silico* screening of small molecule libraries (Ray et al. 2005). The small molecules identified slowed aggregation of some SOD1 mutants *in vitro*. This pioneering research strengthens our hypothesis that stabilization of dimer of aggregation-prone SOD1 mutants can significantly inhibit formation of toxic aggregates. However, the small molecules selected (Ray et al. 2005) were tested only *in vitro* using purified SOD1 but not tested in more biologically relevant systems, such as cultured cells or animal models.

Monitoring Protein Aggregation/Misfolding

As discussed earlier, to facilitate a better understanding of protein misfolding and aggregation, researchers have created methods to monitor such the aggregation process *in vitro*, cultured cells and transgenic mice models (Gregoire et al. 2012). *In vitro* characterization of protein aggregation provides quick insight on the mechanism behind the formation of aggregates. Researchers have found that particular small molecular probes such as Thioflavin, and Congo Red bind to fibrillar species and that anilinonaphthalene 8-sulfonate (ANS) is known to bind to hydrophobic surfaces of proteins, areas of a protein known to interact with each other to initiate aggregation. Also, conformation-specific antibodies have been developed to recognize particular species of protein that are formed along the aggregation pathway. Secondary structure analysis using circular dichroism (CD) and Fourier transform infrared spectroscopy (FTIR) provides insight on key structures that are associated with aggregate formation. Finally, direct imaging of

the aggregates using electron microscopy and atomic force microscopy allows for direct visualization of the morphology of aggregates. Unfortunately, many of these techniques cannot be used to monitor protein aggregation in more physiologically relevant systems, such as inside cultured cells and transgenic mice models without additional invasive procedures being performed (lysing cells, euthanasia of transgenic animals harvesting tissue/blood from dead transgenic animals). Also, some of these molecules interact with organelles within cells, making it difficult to de-convolute whether these molecules are mainly interacting with the protein itself, or other compartments within the cell. Also, due to the lack of penetration of those molecules into particular cell types, issues may arise with examining aggregation properties of intracellular proteins.

A more preferred method to monitoring protein misfolding/aggregation is expression of the recalcitrant protein in bacterial and mammalian cells. Using fusion proteins containing fluorescent proteins or enzymes that correlate a phenotype (cellular fluorescence or enzymatic activity) to the state of folding/aggregation of the recalcitrant protein, researchers have been able to visualize intracellular aggregates, determine the stability of the target protein and find small molecule aggregation inhibitors or additional mutations that increase the solubility of the protein (Cabantous et al. 2005a; Cabantous et al. 2005b; Cabantous and Waldo 2006; Foit et al. 2009; Guglielmi et al. 2011; Ignatova 2005; Waldo 2003; Waldo et al. 1999; Wiseman and Kelly 2009).

Computational Design Using RosettaDesign

The screening process to find aggregation-resistant proteins in mammalian cell culture takes numerous rounds of mutagenesis and selection. Such a process is labor-intensive and may

take months prior to finding a promising variant with desired properties. As a complementary approach to help facilitate the process of finding mutations that reduce aggregation, computational design can utilized. In particular, RosettaDesign, a computational protein redesign software has shown extreme promise in aiding the process of finding aggregation-suppressor mutations (Das and Baker 2008; Rohl et al. 2004; Sammond et al. 2007). Using Monte Carlo simulations and knowledge-based energy functions, RosettaDesign performs simulations that energetically optimizes the structure of the protein when given a high-resolution X-ray crystal structure file using either a fixed or flexible protein backbone (Dantas et al. 2003). The core energy function of RosettaDesign is a linear sum of a 6-12 Lennard-Jones potential, the Lazaridis-Karplus implicit solvation model, an empirical hydrogen bonding potential, backbone-dependent rotamer probabilities, a knowledge-based electrostatics energy potential, amino acid probabilities based on particular regions of φ/ψ space and a unique reference energy for each amino acid. Rosetta has been successful in predicting locations for charged amino acids on protein surfaces to reduce aggregation, increasing stability of globular protein in comparison to the WT structure as well as others (Der et al. 2013).

Semi-Rational Protein Engineering

Protein engineering is a multi-disciplinary field where the protein structure is altered to create the desired or novel functions. With advancements in molecular biology and recombinant DNA technology, researchers have developed a toolbox of techniques to improve protein stability, enzymatic activity, binding affinities of antibodies and other vital functions of proteins. In the current paradigm of protein engineering, rational protein design using computational methods and site-directed mutagenesis or directed evolution are used to incorporate amino acid

substitutions into the gene of choice (Romero and Arnold 2009; Tracewell and Arnold 2009). Using a host organism to express the gene, these proteins are then screened or selected to determine which variants contained the desired function.

Although directed evolution is the most efficient manner to locate variants with a desired phenotype, semi-rational protein design can be used to reduce the number of possible mutants that must be screened, reducing the complexity and difficulty of screening millions of mutant variants. An increase in the number of high resolution crystal structures and amino acid conservation of homologous protein sequences have been beneficial for reducing the library size and encourages more hypothesis-driven protein engineering (Lutz 2010).

Goals and Objectives

Non-native protein aggregation within mammalian cells presents unique challenges that have many unmet needs. In an era where many of the best drugs on the market for diseases are biopharmaceuticals, it is important to maintain the native structure of the protein, wherein much of the activity of the protein lies. In addition to biopharmaceutical production, protein misfolding and aggregation are often causative to the onset of numerous neurodegenerative diseases such as Alzheimer's disease, Parkinson’s disease, Huntington’s disease, and Amyotrophic lateral sclerosis (ALS). For some of these diseases, point mutations in the primary amino acid sequence of a target protein causes a substantial destabilization of the native structure of the protein, leading to the development of toxic, off-pathway species. Therefore, it is of great importance to understand how these non-native species are developed as well as designing aggregation-resistant form of protein variants.

The primary goal of this dissertation is to develop several novel semi-rational protein engineering strategies to suppress or inhibit intracellular aggregation. We hypothesize that semi-rational protein engineering can be used to find protein variants that have increased stability (thermodynamically or kinetically), which subsequently leads to a reduction in aggregation propensity in mammalian cells. Since there is no non-invasive and quantitative assay to monitor protein misfolding/aggregation in mammalian cells, we must develop such an assay. Next, we aim to perform the computational stability analysis and crystal structural analysis to find key residues that are important for protein aggregation and to perform mammalian cell-based screening of protein variants to identify mutations that can reduce or inhibit aggregation of the protein in mammalian cell culture. Specifically, two types of destabilizing interactions, side chain/side chain clashing and side chain/protein backbone repulsion will be investigated. The work of this dissertation will concentrate on reducing the aggregation propensity of mutant human copper-zinc superoxide dismutase (SOD1), implicated in the protein-aggregation disease, familial amyotrophic lateral sclerosis (fALS). The following specific tasks are proposed:

1. Develop and validate a mammalian cell-based screening method to determine the relative aggregation/misfolding propensity of proteins (Chapter 2)
2. Develop a rational approach utilizing computational stability analysis and crystal structure examination to determine a key interaction(s) responsible for the aggregation of $SOD1^{A4V}$ and $SOD1^{G93A}$ variants (Chapters 3 and 4)
3. Utilize mammalian cell-based screening to find a mutation(s) that reduces the aggregation propensity of $SOD1^{A4V}$ and $SOD1^{G93A}$ by increasing conformational stability and kinetic stability, respectively (Chapters 3 and 4)

Chapter 2: A Revisited Folding Reporter for Quantitative Assay of Protein Misfolding and Aggregation in Mammalian Cells

Abstract

Protein misfolding and aggregation play important roles in many physiological processes. These include pathological protein aggregation in neurodegenerative diseases and biopharmaceutical protein aggregation during production in mammalian cells. In order to develop a simple non-invasive assay for protein misfolding and aggregation in mammalian cells, the folding reporter green fluorescent protein (GFP) system, originally developed for bacterial cells, was evaluated. As a folding reporter, GFP was fused to the C-terminus of a panel of human copper/zinc superoxide dismutase (SOD1) mutants with varying misfolding/aggregation propensities. Flow cytometric analysis of transfected HEK293T and NSC-34 cells revealed that the mean fluorescence intensities of the cells expressing GFP fusion of SOD1 variants exhibit an inverse correlation with the misfolding/aggregation propensities of the four SOD1 variants. Our results support the hypothesis that the extent of misfolding/aggregation of a target protein in mammalian cells can be quantitatively estimated by measuring the mean fluorescence intensity of the cells expressing GFP fusion. The assay method developed here will facilitate understanding of aggregation process of SOD1 variants and identifying aggregation inhibitors. The method also has great promise for misfolding/ aggregation study of other proteins in mammalian cells.

Introduction

Protein misfolding and aggregation are often causative to the onset of numerous neurodegenerative diseases such as Alzheimer's disease, Parkinson's disease, Huntington's disease, and Amyotrophic lateral sclerosis (ALS) (Bartolini and Andrisano 2010; Colby et al. 2006; Keshet et al. 2010; Lee et al. 2007; Murphy 2002; Pike et al. 1991; Valentine and Hart 2003). Therefore, modulating protein misfolding and aggregation are generally considered a promising strategy to treat neurodegenerative diseases (Colby et al. 2004; Ghanta et al. 1996; Lowe et al. 2001; Masuda et al. 2006; Ray et al. 2005; Wong and Kwon 2011; Wong et al. 2011). Furthermore, intracellular aggregate formation and misfolding of biopharmaceuticals, such as α-galactosidase, α-glucosidase, antithrombine III, and angiopoietin-1 in mammalian cells are often obstacles in achieving high yield production of biopharmaceuticals (Hwang et al. 2011; Ioannou et al. 1992; Ishii et al. 1996; Okumiya et al. 2007; Schröder et al. 2002). Therefore, quantitative analysis of protein misfolding and aggregation in mammalian cells will facilitate enhanced understanding of the molecular mechanisms for various human diseases and improve recombinant protein quality produced in mammalian cells. In his pioneering work, Johnson et al. reported a novel split GFP complementation system to monitor aggregation of mutant tau proteins, which is associated with Alzheimer's disease, in mammalian cells (Chun et al. 2007). The split GFP complementation is based on the assumption that a small GFP fragment fused to a target protein in aggregates has very limited accessibility to a large GFP fragment. Therefore, this method may not be effective when a target protein forms loosely packed aggregates enabling the small GFP fragment access to the large fragment. However, efficient complementation of the split GFP fragments requires a delicate control of relative (and absolute) expression levels of two split GFP fragments (Chun et al. 2007). Therefore, the need remains to

develop a simple but effective quantitative assay for protein misfolding and aggregation in mammalian cells. In order to address this need, intact GFP fusion to a target protein, originally developed to monitor protein aggregation/misfolding in bacterial cells, was re-evaluated in mammalian cells.

For bacterial and yeast cells, this simple assay method for protein misfolding and aggregation has proven effective (Caine et al. 2007; Kim and Hecht 2005; Kim and Hecht 2008; Kim et al. 2006; Waldo et al. 1999). GFP folding for fluorescence is directly affected by folding of a target protein fused to the N-terminus of GFP. Any deviations from a correctly folded target protein structure such as misfolding and aggregation lead to a loss in the fluorescence of cells expressing the GFP fusion of the target protein in bacteria. With the aid of analytic tools that measure cellular fluorescence, including fluorescence microplate reader and flow cytometer, the extent of the deviation from correctly folded target protein structure can be quantitatively determined. However, GFP fusion for quantitative analysis of protein misfolding and aggregation has thus far been restricted to bacterial systems. Considering the general utility of GFP in both bacteria and mammalian cells as a reporter of cellular processes including protein expression, degradation, localization, and proteolysis (Li et al. 1998; Nicholls et al. 2011; Tsien 1998), we hypothesize that GFP fusion to the C-terminus of a target protein is also effective in quantitatively monitoring protein misfolding and aggregation in mammalian cells.

In order to validate the GFP fusion method's capacity to monitor protein misfolding/aggregation in mammalian cells, human copper/zinc superoxide dismutase (SOD1) mutants were chosen due to several features of SOD1. First, SOD1 mutants form intracellular aggregates in mammalian cells in a relatively short period time, usually within two to three days post-transfection (Niwa et al. 2007). Second, SOD1 mutants are known to form loosely packed

aggregates inside mammalian cells (Matsumoto et al. 2006) and so the split GFP complementation may not be effective in monitoring protein aggregation. Third, SOD1 mutants are associated with neurodegenerative disorder familial form of amyotrophic lateral sclerosis (fALS) (Ray et al. 2005; Rosen 1993; Zhang and Zhu 2006). Amyotrophic lateral sclerosis (ALS), often referred to as Lou Gehrig's disease, is a progressive neurodegenerative disorder that attacks and inevitably leads to the death of motor neurons in the brain, brain stem and spinal cord (Rosen 1993). There are over one hundred natural SOD1 mutants with varying misfolding/aggregation propensities, though wild-type SOD1 forms a stable dimer (Banci et al. 2008; Galaleldeen et al. 2009; Rakhit and Chakrabartty 2006; Valentine et al. 2005; Zhang and Zhu 2006). It has been proposed that aggregation of the mutant SOD1 is linked to the gain of toxic function, eventually leading to the development of ALS. The most common mutation among mutant SOD1-associated fALS patients in the United States is alanine to valine at 4th residue (A4V). Researchers have biophysically characterized this mutation and have found that this mutation causes substantial destabilization at the dimer interface and the rapid development of aggregates in vitro and in vivo (Auclair et al. 2010; Banci et al. 2008; Schmidlin et al. 2009). The progression of fALS has been linked to the level of aggregate formation in the spinal cord, implying SOD1 aggregation a potential drug target for preventing or curing ALS (Prudencio et al. 2009). In order to screen molecules that inhibit mutant SOD1 aggregation, a simple, but effective way to monitor mutant SOD1 aggregation needs to be developed. Stevens et al. demonstrated that GFP tagging does not substantially alter properties of SOD1 variants inside cells using fluorescence microscopy (Stevens et al. 2010). Although this result supports the idea that GFP is safe to fuse to SOD1 variants, the fluorescence microscopy technique is not suitable

to quantitatively monitor aggregation of mutant SOD1. Therefore, we investigated whether flow cytometric analysis can be employed to quantitatively monitor aggregation of mutant SOD1s.

We chose wild-type SOD1 ($SOD1^{WT}$) and three aggregation-prone SOD1 mutants ($SOD1^{A4V}$, $SOD1^{A4V/C57S,}$ and $SOD1^{A4V/C111S}$) as model proteins. As described earlier, $SOD1^{A4V}$ is known to form aggregates in vitro and in vivo (Auclair et al. 2010; Banci et al. 2008; Schmidlin et al. 2009). Among four cysteines (Cys6, Cys57, Cys111, and Cys146), two free cysteines Cys6 and Cys111 are known to form inter-molecular disulfide bond network with other SOD1 monomers containing free C6 or C111 during the aggregation process (Cozzolino et al. 2008; de Beus et al. 2004; Fujiwara et al. 2007; Niwa et al. 2007). Therefore, Cys to Ser mutation at residue 111 of SOD1 reduces SOD1 aggregation (de Beus et al. 2004; Niwa et al. 2007). $SOD1^{A4V/C111S}$ mutant is more stable and soluble than $SOD1^{A4V}$ mutant in cultured cells and in vitro (Cozzolino et al. 2008; Fujiwara et al. 2007; Watanabe et al. 2007), indicating that $SOD1^{A4V/C111S}$ has a lower misfolding/aggregation propensity than $SOD1^{A4V}$. The other two cysteines, C57 and C147, are involved in an intra-subunit disulfide bond formation; an important post-translational modification of SOD1. It was reported that removing a Cys residue involved in the intra-subunit disulfide bond formation greatly destabilizes the protein and so enhances misfolding of the protein in cultured cells and in vitro (Cozzolino et al. 2008; Furukawa et al. 2006; Niwa et al. 2007; Watanabe et al. 2007), indicating that the misfolding propensity of $SOD1^{A4V/C57S}$ is higher than that of $SOD1^{A4V}$. In summary, for the four SOD1 mutants used in this study, the order of extent of deviations from correctly folded structure (misfolding/aggregation) is $SOD1^{A4V/C57S} > SOD1^{A4V} > SOD1^{A4V/C111S} >$ and $SOD1^{WT}$.

Herein we investigate whether there is a direct correlation between misfolding/aggregation propensity of the four SOD1 variants and mean cellular fluorescence of

transfected HEK293T cells expressing the corresponding SOD1 variant fused to GFP. Then, in order to generalize our hypothesis that GFP fusion is effective in quantitatively monitoring mammalian cell protein misfolding/aggregation, we also extend our studies to another cell line (neuroblastoma NSC-34 cells) and promoter (ubiquitin-C).

Materials and Methods

Materials. Anti-SOD1 polyclonal antibody was purchased from Santa Cruz Biotechnology (Santa Cruz, CA). Anti-α tubulin antibody was obtained from Sigma-Aldrich (St. Louis, MO). Anti-rabbit antibody was obtained from Invitrogen (Carlsbad, CA). HEK293T and NSC-34 cells were obtained from Invitrogen and CELLutions Biosystems (Burlington, Ontario, Canada), respectively. Restriction enzymes and DNA ligases were obtained from New England Biolabs (Ipswich, MA). All other chemicals were obtained from Sigma-Aldrich, unless otherwise noted.

Construction of Expression Vectors. The EGFP gene is located to 3' of either $SOD1^{WT}$ or $SOD1^{A4V}$ gene in pEGFP-N3-$SOD1^{WT}$ or pEGFP-N3-$SOD1^{A4V}$ plasmid, respectively (Zhang and Zhu 2006). In both plasmids, SOD1-EGFP fusion protein expression is under the control of Cytomegalovirus (CMV) constitutive promoter. Other detailed procedures to construct expression vectors are described in Appendix A.

Transfection of HEK293T and NSC-34 cells. HEK293T cells were cultured on 6-well plates at 37° C in Dulbecco's modified Eagle's medium/High Glucose (DMEM/High Glucose, Thermo

Scientific, Pittsburg, PA) supplemented with 10% fetal bovine serum (FBS), 100 µg/mL of streptomycin sulfate and 100 units/ml of penicillin. NSC-34 cells were maintained in DMEM/12:1:1 modified containing 10% FBS. When the cells grew to 80-90% confluency, they were transfected with appropriate plasmids via calcium phosphate precipitation (Graham and van der Eb 1973; Jang et al. 2011). All samples were tested in triplicate, unless otherwise described.

Fluorescence Microscopy. At two days after transfection, the transfected HEK293T cells cultured on the 6-well plates were visualized by fluorescence microscopy using a VistaVision Inverted Fluorescence Microscope (VWR, Radnor, PA). Fluorescence microscopic images of the cells were taken with a DV-2C digital camera equipped on the microscope at the same magnification and camera settings. The fluorescence excitation wavelength is between 420 and 485 and the emission wavelength is 515 nm. The number of cells in the fluorescence microscopy images was manually counted. The percentage of transfected cells exhibiting SOD1-EGFP aggregates was determined by dividing the number of aggregates-containing cells by the number of total cells analyzed.

Flow Cytometric Analysis of Cellular Fluorescence. After incubating the desired amount of time, the fluorescence intensities of the HEK293T and NSC-34 cells expressing SOD1 variant-EGFP fusion protein were measured using the C6 flow cytometer (Accuri, Ann Arbor, MI). To prepare the cells for flow cytometric analysis, the HEK 293T cells and NSC-34 cells were trypsinized, washed twice with phosphate buffered saline buffer (PBS; pH 7.4) and diluted ten-fold in PBS to prevent cell aggregation. The excitation wavelength was 488 nm and the fluorescence emission was detected at 585 nm. Only GFP positive cells were used to calculate

the mean cellular fluorescence. Transfection efficiency was determined by dividing the number of fluorescence positive cells by the number of total cells analyzed. The mean cellular fluorescence intensities were obtained by analyzing 20,000 to 50,000 cells. Each sample was prepared in triplicate and cellular fluorescence indicates mean cellular fluorescence, unless otherwise noted. Student's paired t-test was used for statistical analysis of fluorescence data. Gating strategy is described in Appendix A (Figure A6).

Dot-blotting. $5x10^5$ transfected HEK293T cells were trypsinized and centrifuged to obtain a cell pellet. The pellet was then washed with 1X PBS once and centrifuged to remove the buffer. The cell pellet was lysed in 100 μL of lysis buffer containing 100mM NaH2PO4, 10mM Tris-Cl, 1x protease inhibitor cocktail (Sigma), and 8 M Urea at room temperature for 1 hour. The cell lysates were centrifuged at 14,000g for 30 minutes to remove the cell debris and the supernatant was collected. Protein concentration of the cell lysate was determined using BCA kit (Pierce, Rockford, IL) using bovine serum albumin as a standard. Dot-blot assay of the cell lysate was performed as described earlier (Wong and Kwon 2011; Wong et al. 2011) except antibodies used (anti-α tubulin and HRP-conjugated anti-GFP antibodies in this study). The blot images were captured using a BioSpectrum imaging system (UVP).

Results and Discussion

Fluorescence Microscopic Observation of Transfected HEK293T Cells Expressing SOD1 Variants.

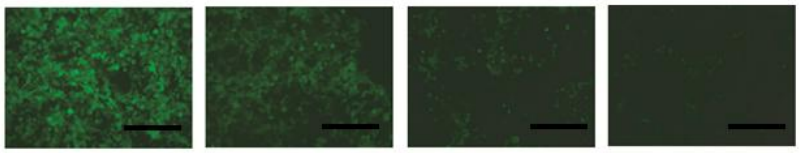
HEK293T cells were transfected with plasmids encoding each of the four SOD1-EGFP variants genes ($SOD1^{WT}$-EGFP, $SOD1^{A4V/C111S}$-EGFP, $SOD1^{A4V}$-EGFP, and $SOD1^{A4V/C57S}$-

EGFP). At two days post-transfection, fluorescence images of the transfected cells were taken (Figure 2.1A). Transfected HEK293T cells expressing $SOD1^{WT}$-EGFP were brightest among the four different transfected cells. The order of brightness of the transfected cells expressing four SOD1 variants fused to EGFP is $SOD1^{WT}$ > $SOD1^{A4V/C111S}$ > $SOD1^{A4V}$ > and $SOD1^{A4V/C57S}$, which is consistent with the inverse order of misfolding/aggregation propensity of the four SOD1 variants in vitro and in cultured cells previously reported (Cozzolino et al. 2008; Fujiwara et al. 2007; Watanabe et al. 2007).

Flow Cytometric Analysis of Transfected HEK293T Cells Expressing SOD1 Variants.

Fluorescence microscopic observation of cells expressing GFP fusion protein is a non-invasive method to qualitatively compare the fluorescence intensities of different cell samples but is not suitable for quantitative analysis. Therefore, we employed flow cytometric analysis of the transfected HEK293T cells to evaluate cellular fluorescence. At one and two days post-transfection of HEK293T cells with plasmids encoding four SOD1-EGFP fusion genes respectively, the mean fluorescence intensities of the four transfected cells were measured by flow cytometry (Figure 2.2A). The representative overlaid histograms of fluorescence of cells are shown in Figure A1 in Appendix A. In this article, the fluorescence intensities measured by flow cytometry indicate the mean cellular fluorescence intensities. As expected, the order of fluorescence intensity of the four transfected cells is $SOD1^{WT}$ > $SOD1^{A4V/C111S}$ > $SOD1^{A4V}$ > $SOD1^{A4V/C57S}$, matching the inverse order of the misfolding/aggregation propensities of the four SOD1 variants. The cellular fluorescence intensity of $SOD1^{A4V/C111S}$ was significantly higher than that of $SOD1^{A4V}$ ($P < 0.05$), whereas the cellular fluorescence intensity of $SOD1^{A4V/C57S}$ was significantly lower than that of $SOD1^{A4V}$ ($P < 0.001$) (Figure 2.2A).

A



SOD1 variant	$SOD1^{WT}$	$SOD1^{A4V/C111S}$	$SOD1^{A4V}$	$SOD1^{A4V/C57S}$
Transfection efficiency (%)	55 ± 5	55 ± 4	50 ± 3	42 ± 2

B

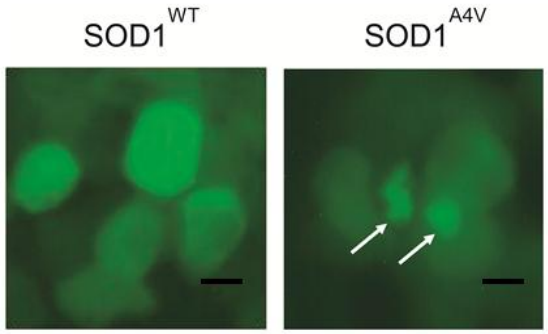


Figure 2.1: Fluorescence microscopy images of the transfected HEK293T cells expressing SOD1 variants fused to EGFP. (A) Images of transfected HEK293T cells expressing EGFP fusion of $SOD1^{WT}$, $SOD1^{A4VC111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$ taken at two days post-transfection. Transfection efficiency for each SOD1 variant was determined by calculating the percentage of GFP positive cells out of total cells analyzed using flow cytometry. Values represent means ± standard deviation (n = 3). Scale bars are 250 μm. (B) Magnified fluorescence microscopic images of the transfected HEK293T cells expressing EGFP fusion of $SOD1^{WT}$ and $SOD1^{A4V}$. Arrows indicate the aggregates within the cell. Scale bars are 50 μm.

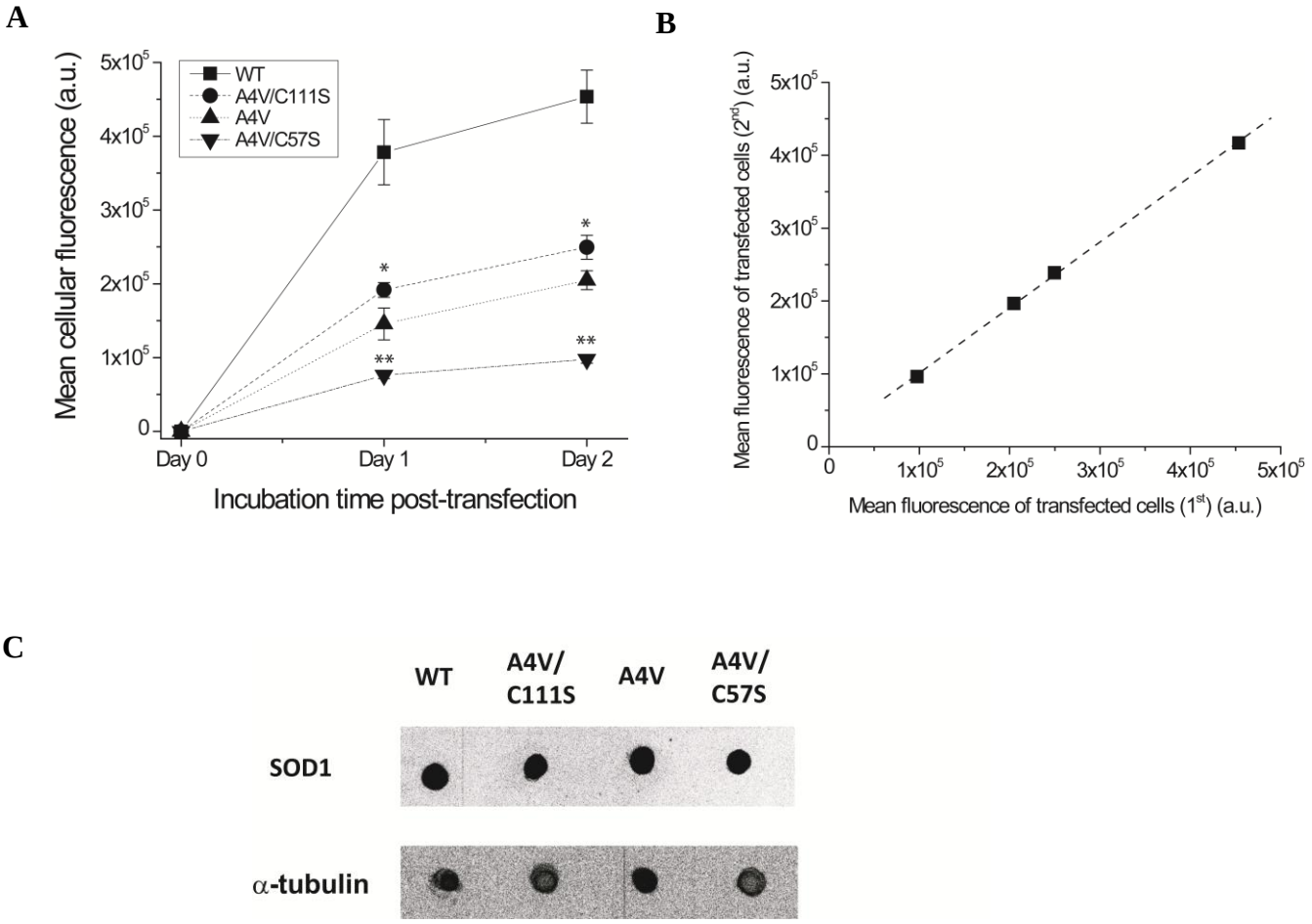


Figure 2.2: Comparison of the cellular fluorescence intensities and expression levels of EGFP fusion of SOD1 variants. (A) Time course of fluorescence intensities of the transfected HEK293T cells expressing EGFP fusion of SOD1 variants. Fluorescence intensities of untransfected cells (day 0) and transfected cells expressing EGFP fusion of four SOD1 variants ($SOD1^{WT}$, $SOD1^{A4VC111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$) at day 1 and 2 were determined by flow cytometry. Values and error bars represent mean cellular fluorescence and standard deviations, respectively (n = 3). Two-sided Student's t-tests were applied to the data (*: $P < 0.05$; **: $P < 0.001$). (B) Comparison of the mean cellular fluorescence of HEK293T cells expressing SOD1

variants that were transfected at two separate days. At two days post-transfection, cellular fluorescence of transfected HEK293T cells expressing EGFP fusion of SOD1 variants ($SOD1^{WT}$, $SOD1^{A4VC111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$) was determined by flow cytometry. The data were fitted to a straight line ($R^2 = 0.99$). (a.u. = arbitrary unit) Numerical values of the mean cellular fluorescence intensities are shown in Table A1 in Appendix A. (C) Dot-blot images of the transited cell lysates of EGFP fusion of $SOD1^{WT}$ (WT), $SOD1^{A4V/C111S}$ (A4V/C111S), $SOD1^{A4V}$. (A4V), and $SOD1^{A4V/C57S}$ (A4V/C57S) using anti-SOD1 polyclonal antibody. (α-tubulin; loading control).

Although flow cytometric analysis of cellular fluorescence is a very convenient way to obtain information on cellular processes, caution should be taken to directly correlate the cellular fluorescence intensities to protein misfolding/aggregation. Other factors such as transfection efficiency and expression level of a target protein can also affect the fluorescence intensity of transfected cells. First, effects of transfection efficiencies on cellular fluorescence were investigated. The transfection efficiencies of the four SOD1 variants were determined by flow cytometry. The transfection efficiency was calculated by dividing the number of GFP positive cells by the number of total cells analyzed by flow cytometry. Although the transfection efficiencies vary from 42% to 55%, the differences are not large enough to explain the differences in cellular fluorescence of the four different transfected cells, supporting the idea that intensity is related to SOD1 properties. The transfection efficiencies of $SOD1^{A4V}$ and $SOD1^{A4V/C57S}$ were 5 to 13% lower than those of $SOD1^{WT}$ and $SOD1^{A4V/C111S}$, which is most likely due to the exclusion of some weak fluorescent cells in the samples. If the fluorescence intensity of the transfected cells are overlapped with that of untransfected cells (negative control), such weak fluorescent cells are not counted in determining the mean cellular fluorescence intensity after the background subtraction in flow cytometric analysis. According to the transfection efficiencies measured by flow cytometry (Figure 2.1), the maximum difference among the transfection efficiencies was 13%. However, this difference is too small to explain the more than 4-fold difference in the fluorescence intensities of cells expressing $SOD1^{WT}$ and $SOD1^{A4V/C57S}$ (Figure 2.2A). Furthermore, the lower transfection efficiency of $SOD1^{A4V/C57S}$ cell sample is most likely made by the exclusion of weakly fluorescent cells from the transfection efficiency calculation due to the background subtraction in data analysis. If the weak fluorescent cells were included to correct the transfection efficiency, the cellular

fluorescence intensities would decrease, making the difference in the cellular fluorescence intensities between $SOD1^{WT}$ and $SOD1^{A4V/C57S}$ larger. Therefore, our results suggest that the low transfection efficiencies of $SOD1^{A4V}$ and $SOD1^{A4V/C57S}$ are not the major cause of their low cellular fluorescence.

Second, we investigated whether the differences in the cellular fluorescence intensities resulted from differences in the expression levels of SOD1 variants; because the mean fluorescence intensities of cells can change according to SOD1-GFP fusion protein expression level. Therefore, expression levels of four different SOD1-EGFP fusion proteins were compared using dot-blot assay using anti-SOD1 polyclonal antibody. Since HEK293T cells also express endogenous $SOD1^{WT}$, we first compared the endogenous $SOD1^{WT}$ level in untransfected HEK293T cells and the total $SOD1^{WT}$ level in transfected HEK293T cells over-expressing $SOD1^{WT}$-EGFP. As expected, anti-SOD1 immuno-reactivity of the untransfected cells was negligible compared to that of the transfected cells over-expressing $SOD1^{WT}$-EGFP (Figure A2). Therefore, we ignored endogenous $SOD1^{WT}$ expression level when over-expression levels of the four SOD1-EGFP fusion proteins are compared. Next, transfected cells expressing each of four SOD1-EGFP fusion proteins ($SOD1^{WT}$-EGFP, $SOD1^{A4V/C111S}$-EGFP, $SOD1^{A4V}$-EGFP, and $SOD1^{A4V/C57S}$-EGFP) were lysed and the total cell lysates were used for dot-blot assay. Anti-α tubulin antibody was used to ensure loading a comparable amount of the protein extracts to a nitrocellulose membrane (Figure 2.2C bottom). The immuno-reactivity of all four SOD1-EGFP fusion proteins was comparable (Figure 2.2C top), strongly indicating that the substantial differences in the cellular fluorescence do not result from differences in their expression levels.

Third, besides transfection efficiency and expression level, there might be other factors that affect cellular fluorescence intensities. If there are other critical factors affecting the cellular

fluorescence intensities, the results may not be reproducible. In order to confirm the reproducibility of the results, transfection and cellular fluorescence measurement were performed on two separate days with a month gap. Then, the mean fluorescence intensities of the transfected cells obtained from two independent experiments were compared (Figure 2.2B). For the four SOD1-EGFP variants ($SOD1^{WT}$-EGFP, $SOD1^{A4V/C111S}$-EGFP, $SOD1^{A4V}$-EGFP, and $SOD1^{A4V/C57S}$-EGFP), there is a linear correlation between the cellular fluorescence intensities of the cells transfected on two separate days ($R^2 = 0.99$), though there was a slight difference in the absolute fluorescence intensities (around 5%). Our results strongly indicate that the correlation between the cellular fluorescence intensities and the misfolding/aggregation propensities of the SOD1 variants is quite reproducible. However, since absolute fluorescence intensities of the transfected cells expressing EGFP fusion protein may vary each time, an accurate estimation of misfolding/aggregation status of a target protein requires a calibration using both highly stable and misfolding/aggregation-prone protein controls.

In order to compare the efficacy of EGFP fusion to that of split GFP complementation, split GFP complementation was performed on $SOD1^{WT}$ and $SOD1^{A4V}$ (experimental details are described in the Appendix A). The mean cellular fluorescence intensities of $SOD1^{WT}$ and $SOD1^{A4V}$ were not significantly different (Figure 2.3), which is likely due to loosely packed structure of $SOD1^{A4V}$ aggregates (Matsumoto et al. 2006) allowing access of a small fragment of GFP fused to $SOD1^{A4V}$ to a large GFP fragment. This finding suggests the idea that EGFP fusion is more effective in monitoring protein aggregation, which is loosely packed, than split GFP complementation.

Aggregate Formation of SOD1^{A4V} Variants in Transfected HEK293T Cells.

Our results indicate that transfected cells expressing EGFP fusion of SOD1 variants exhibited significant disparity in the mean cellular fluorescence intensity; but the disparity was not caused by differences in the transfection efficiency or expression level of the SOD1 variants. Therefore, it is reasonable to assume that there are substantial differences in the structures of the SOD1 variants. We hypothesized that the disparity in fluorescence intensities of cells expressing the SOD1 variants result from differences in the extent of misfolding, aggregation, or both for the SOD1 variants. Transfected cells expressing four SOD1-EGFP variants were re-examined using fluorescence microscopy at a higher magnification (400X). Transfected cells expressing SOD1^{A4V} exhibited visible aggregates inside cells whereas cellular fluorescence of the transfected cells expressing SOD1^{WT} is evenly distributed in the cytosol (Figure 2.1B), consistent with results previously reported (Witan et al. 2008). Similar intracellular aggregates were found in both transfected cells expressing SOD1$^{A4V/C111S}$ and SOD1$^{A4V/C57S}$, respectively (Figure 2.4). The percentage of cells containing intracellular aggregates was determined by examining the fluorescence microscopic images of the transfected cells. Over five hundred cells for each SOD1 variant were analyzed and the results were plotted (Figure 2.5). Around 12% of the transfected cells expressing SOD1^{A4V} exhibited intracellular aggregates, whereas none of the transfected cells expressing SOD1^{WT} showed any intracellular aggregates. However, approximately 5% of the cells expressing SOD1$^{A4V/C111S}$ showed aggregates, consistent with the reduced aggregation of SOD1$^{A4V/C111S}$ compared to SOD1^{A4V} in mammalian cells (Cozzolino et al. 2008). Therefore, the significant increase ($P < 0.05$) in the cellular fluorescence intensity of cells expressing SOD1$^{A4V/C111S}$ over that of SOD1^{A4V} (Figure 2.2A) is attributed to the reduced aggregation of SOD1$^{A4V/C111S}$ compared to SOD1^{A4V} (Figure 2.1B).

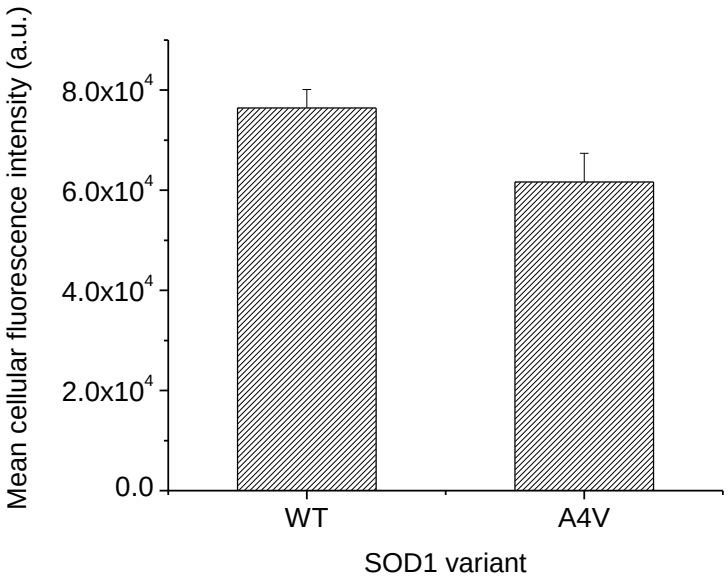


Figure 2.3: Mean cellular fluorescence intensities of transfected HEK293T cells co-expressing split GFP fragments fused to $SOD1^{WT}$ or $SOD1^{A4V}$. In order to achieve split GFP complementation, both pCMV-mGFP_1-10 and pCMV-mGFP_Cterm_S11 plasmids encoding large and small GFP fragment genes, respectively, which were originally developed by Waldo group, and obtained from Theranostech, Inc. (Albuquerque, NM) (Cabantous et al. 2005b; Cabantous and Waldo 2006). PCR-amplified $SOD1^{WT}$ and $SOD1^{A4V}$ genes were inserted between *Nhe*I and *BamH*I restriction sites of pCMV-mGFP_Cterm_S11 to generate pCMV_$SOD1^{WT}$-mGFP_S11 and pCMV-$SOD1^{A4V}$-mGFP_S11, respectively. HEK293T cells were co-transfected with pCMV-mGFP_1-10 and either pCMV_$SOD1^{WT}$-mGFP_S11 or pCMV-$SOD1^{A4V}$-mGFP_S11 plasmid. Mean cellular fluorescence intensities of the transfected HEK293T cells were determined by flow cytometry at two days post-transfection. Values indicate the mean cellular fluorescence intensities and standard deviations (n = 3).

In the case of $SOD1^{A4V/C57S}$, 10% of the transfected cells exhibited intracellular aggregates but the difference from $SOD1^{A4V}$ is not statistically significant ($P > 0.05$). Therefore, the percentage of cells exhibiting SOD1-EGFP aggregates determined by fluorescence microcopy does not correlate well to the cellular fluorescence intensity. However, the enhanced misfolding propensity of $SOD1^{A4V/C57S}$ over $SOD1^{A4V}$ (Cozzolino et al. 2008), which severely inhibits correct folding of GFP, is better attributed to a loss of fluorescence. In Figure 2.2A, the change in the cellular fluorescence caused by C57S mutation was greater than the change caused by C111S, which can be explained by the greater extent of structural perturbations made by aggregation and misfolding. Considering that the loss of an intra-subunit disulfide bond caused by C57S mutation generates very unstable monomeric SOD1s, structural perturbation of SOD1 caused by C57S mutation is likely greater than that of SOD1 caused by intermolecular disulfide network-mediated aggregation. In summary, mean cellular fluorescence intensity of cells expressing EGFP fusion of SOD1 variant is a good indicator of deviations (combined effects of misfolding and aggregation) from correctly folded structure. However, caution should be taken when correlating the cellular fluorescence intensity to only either aggregation or misfolding of SOD1 variants, because both aggregation and misfolding of SOD1 variant may not occur simultaneously.

Comparison of the Cellular Fluorescence of Transfected NSC-34 and HEK293T Cells Expressing SOD1 Variants.

In the transfected HEK293T cells expressing four SOD1-EGFP variants, there was good quantitative correlation between the cellular fluorescence intensity and the combined misfolding and aggregation propensities.

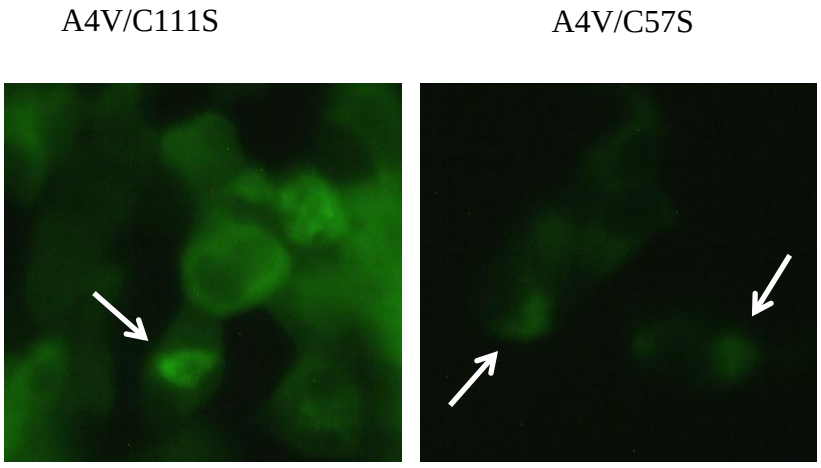


Figure 2.4: Aggregate formation of SOD1$^{A4V/C111S}$ and SOD1$^{A4V/C57S}$ in transfected HEK293T cells. Fluorescence microscopy images of the transfected HEK293T cells expressing EGFP fusion of SOD1^{A4VC111S} and SOD1$^{A4V/C57S}$ taken at two days post-transfection. White arrows indicate intracellular aggregates.

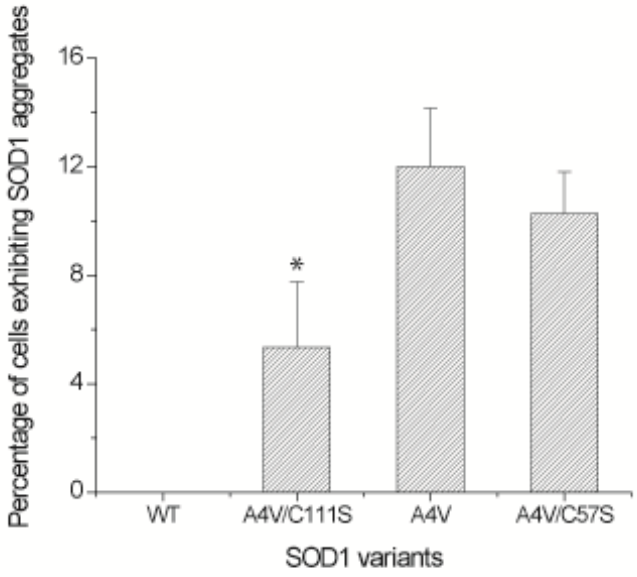


Figure 2.5: The percentage of transfected HEK293T cells exhibiting SOD1 aggregates. The number of cells exhibiting intracellular SOD1 aggregates was determined by analyzing fluorescence microscopy images of the transfected cells expressing EGFP fusion of four SOD1 variants ($SOD1^{WT}$, $SOD1^{A4VC111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$) at day two post-transfection. Values and error bars represent mean and standard deviations, respectively (n = 3). Two-sided Student's t-tests were applied to the data (*: $P < 0.05$). Numerical values of the mean cellular fluorescence intensities are shown in Table A2 in Appendix A.

In order to confirm that the correlation is valid in cells other than HEK293T cells, mouse neuroblastoma NSC-34 cells (Cashman et al. 1992) were also examined at two days post-transfection using flow cytometric analysis similar to the transfected HEK293T cells. NSC-34 cell is a widely used cell line to study SOD1 expression in cultured cells (Babetto et al. 2005; Kupershmidt et al. 2009; Raimondi et al. 2006; Rizzardini et al. 2006; Tartari et al. 2009). Even in the transfected NSC-34 cells, the order of cellular fluorescence intensities matched well with the inverse order of misfolding/aggregation propensities of the SOD1 variants (Figure 2.6A). In order to determine whether there is a direct correlation between the mean fluorescence intensities of HEK293T and NSC-34 cells expressing the four SOD1-EGFP variants, the mean cellular fluorescence intensities of both cells were plotted (Figure 2.6B). The fluorescence ratios of SOD1 variant over $SOD1^{WT}$ in NSC-34 cells were similar to those in HEK293T cells. Furthermore, a linear correlation between the fluorescence intensities of both transfected cells was observed (R^2 = 0.98), indicating that an inverse relationship between the cellular fluorescence intensity and the misfolding/aggregation propensity of SOD1 variants is valid in both cell lines. However, the fluorescence intensities of the transfected NSC-34 cells are about one half of that of the corresponding transfected HEK293T cells, which can be explained by weaker biosynthesis activities in slow growing NSC-34 cells. Therefore, in order to estimate extent of the deviation of a target protein from its correctly folded structure in a specific cell line, cellular fluorescence intensities should be calibrated using appropriate positive and negative control samples.

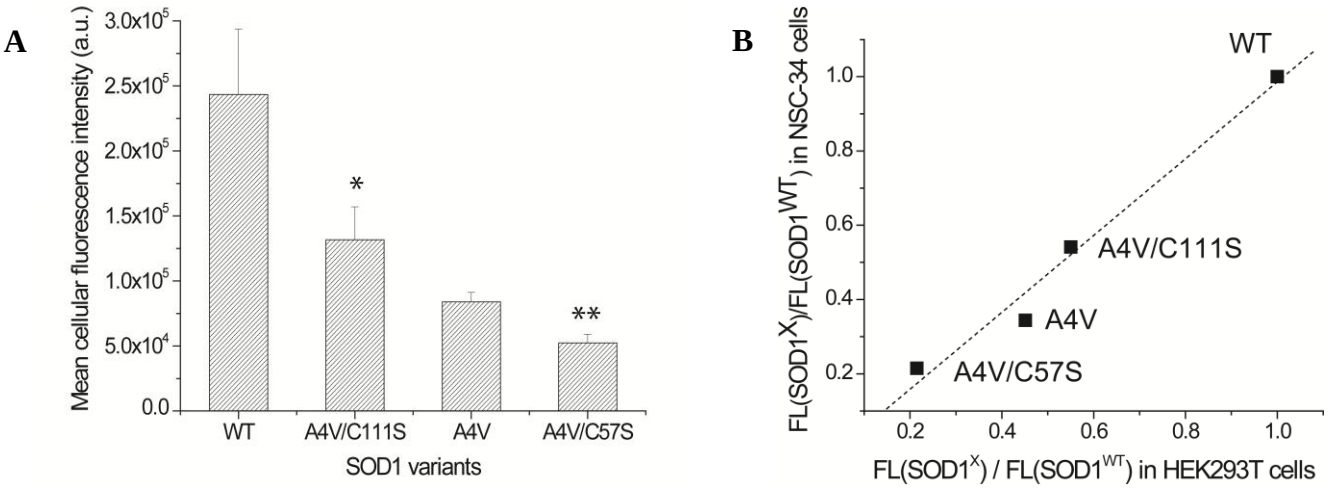


Figure 2.6. Comparison of the orders of fluorescence intensities of transfected HEK293T and NSC-34 cells expressing EGFP fusion of SOD1 variants. (A) Fluorescence intensities of the transfected NSC-34 cells expressing EGFP fusion of SOD1 variants ($SOD1^{WT}$, $SOD1^{A4VC111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$) at two days post-transfection were determined by flow cytometry. Values and error bars represent mean cellular fluorescence and standard deviations, respectively (n = 3). (a.u. = arbitrary unit) Two-sided Student's t-tests were applied to the data (*: $P < 0.05$; **: $P < 0.01$). Numerical values of the mean cellular fluorescence intensities are shown in Table A.1 in Appendix A. (B) Comparison of the cellular fluorescence of the transfected HEK293T at day two and NSC-34 cells at day three post-transfection. The cellular fluorescence of transfected HEK293T and NSC-34 cells expressing EGFP fusion of SOD1 variants ($SOD1^{WT}$, $SOD1^{A4VC111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$) was determined by flow cytometry. The data were fitted to a straight line ($R^2 = 0.98$). $FL(SOD1^X)/FL(SOD1^{WT})$ indicates the ratio of fluorescence intensity of cells expressing a SOD1 variant ($SOD1^X$) over the fluorescence intensity of cells expressing $SOD1^{WT}$.

Comparison of the CMV and Ubiquitin-C Promoters.

Thus far SOD1-EGFP fusion proteins were expressed under control of CMV constitutive promoter. We chose CMV promoter because it is widely used to express recombinant proteins in mammalian cells due to its relatively strong promoter activity in diverse cell lines (Qin et al. 2010). However, we can imagine situations where a different promoter should be used. Therefore, we examined whether the choice of promoter affects the correlation between the cellular fluorescence intensity and SOD1 variant misfolding/aggregation. In order to compare promoters, we chose Ubc promoter which is also widely used in mammalian cells (Qin et al. 2010). $SOD1^{WT}$-EGFP and $SOD1^{A4V}$-EGFP genes were sub-cloned into pFUG-IP lentiviral vector plasmid (Jang et al. 2011) where a target gene is under the control of ubiquitin-C (Ubc) promoter and an internal ribosome entry site/puromycin resistance gene cassette is located at 3' of a target gene. HEK293T cells were transfected with the plasmids encoding either $SOD1^{WT}$-EGFP or $SOD1^{A4V}$-EGFP gene under the control of Ubc promoter. At two days post-transfection, fluorescence intensities of the transfected cells were determined by flow cytometry. The cellular fluorescence intensity of cells expressing $SOD1^{A4V}$-EGFP was about six-fold lower than that of cells expressing $SOD1^{WT}$-EGFP (Figure 2.7), indicating that Ubc promoter also allows distinction of cellular fluorescence intensities depending on their misfolding/aggregation propensity of SOD1 variants. The six-fold difference between the cellular fluorescence intensities of $SOD1^{WT}$-EGFP and $SOD1^{A4V}$-EGFP with Ubc promoter is greater than the two-fold difference with CMV promoter. Such a greater difference might be caused by the change of promoter from CMV to Ubc and/or the change of plasmid backbone from pEGFP-N3 to pFUG-IP. In order to investigate this, we replaced the Ubc promoter in the pFUG-IP plasmid backbone with CMV promoter. The cellular fluorescence intensity of cells transfected with the modified

pFUG-IP plasmid carrying CMV-SOD1^{WT}-EGFP cassette was around four-fold greater than that of CMV-SOD1^{A4V}-EGFP cassette in the same plasmid backbone, which suggests that the pFUG-IP plasmid backbone plays an important role in increasing the difference between the cellular fluorescence intensities of SOD1^{WT}-EGFP and SOD1^{A4V}-EGFP. It is noteworthy that the pFUG-IP plasmid encodes a puromycin resistant gene as well as a target protein gene. Therefore, both puromycin resistant gene and SOD1-EGFP genes are over-expressed when pFUG-IP plasmid is used, whereas only SOD-EGFP is over-expressed when pEGFP-N3-SOD1 plasmid is used. Therefore, it is possible that chaperone proteins facilitating folding of unstable SOD1 are less available to SOD1 when pFUG-UP plasmid is used, because the chaperone proteins are also involved in the folding of co-expressed puromycin resistant protein. Furthermore, it was reported that several SOD1 mutants inhibit chaperone activities (Tummala et al. 2005).

Therefore, we hypothesize that over-expression of another gene, as well as SOD1-EGFP, limits chaperone activity required for SOD1-EGFP folding. Such a shortage of chaperone activity is more detrimental to mutant SOD1 folding and leads to the greater reduction in the cellular fluorescence intensity. In order to investigate this hypothesis, we co-transfected HEK293T cells with both pEGFP-N3-SOD1^{A4V} and pAAV-Luc (a plasmid encoding luciferase gene under the control of CMV promoter) plasmids and also observed six-fold difference between the fluorescence intensities of SOD1^{WT}-EGFP and SOD1^{A4V}-EGFP with co-expression of luciferase protein (Figure 2.8). Therefore, when another protein is over-expressed, as well as the EGFP fusion of a target protein, caution should be taken to use the EGFP fusion to monitor aggregation/misfolding.

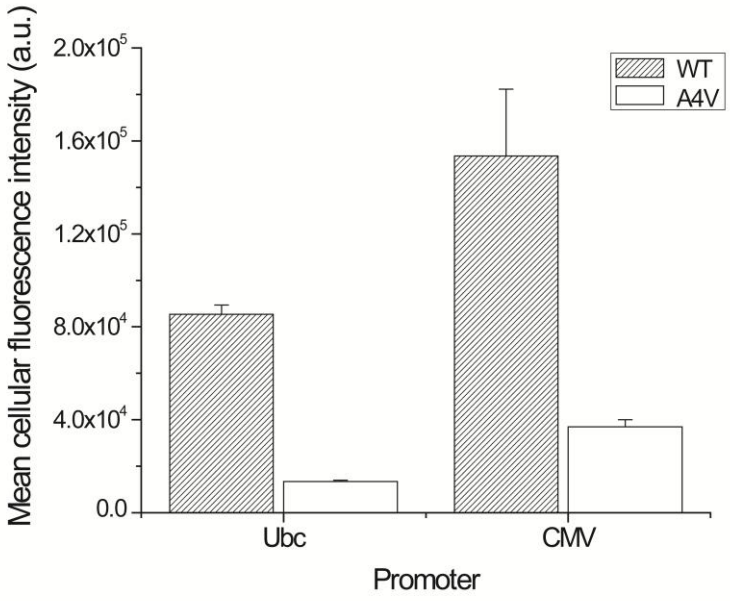


Figure 2.7. Comparison of CMV and Ubc promoter activities for the fluorescence intensities of transfected HEK293T cells expressing EGFP fusion of SOD1 variants. Cellular fluorescence of HEK293Tcells expressing EGFP fusion of $SOD1^{WT}$ and $SOD1^{A4V}$ under the control of either CMV or Ubc promoter was determined by flow cytometry. Values and error bars represent mean cellular fluorescence and standard deviations, respectively (n = 3). (a.u. = arbitrary unit) Numerical values of the mean cellular fluorescence intensities are shown in Table A1 in Appendix A.

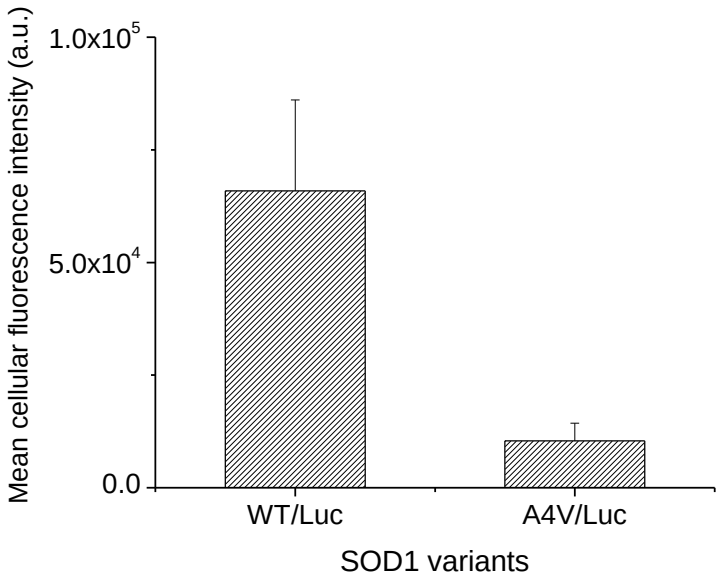


Figure 2.8: Mean cellular fluorescence intensities of transfected HEK293T cells co-expressing firefly luciferase (Luc) and either $SOD1^{WT}$-EGFP or $SOD1^{A4V}$-EGFP. pAAV-Luc plasmid encodes Luc gene under the control of CMV promoter (Jang et al. 2010). HEK293T cells were co-transfected with pAAV-Luc and either pEGFP-N3-$SOD1^{WT}$ or pEGFP-N3-$SOD1^{A4V}$ plasmid. Mean cellular fluorescence intensities of the transfected HEK293T cells were determined by flow cytometry at two days post-transfection. Values indicate the mean cellular fluorescence intensities and standard deviations (n = 3).

The fluorescence intensity of cells expressing $SOD1^{WT}$-EGFP under the control of Ubc promoter was about 5-fold lower than that of CMV promoter (Figure 2.7), likely due to weaker activity of Ubc promoter than CMV promoter in HEK293T cells (Qin et al. 2010). As expected, when the Ubc promoter in the pFUG-IP backbone was replaced with CMV promoter, the cellular fluorescence intensities of cells expressing both $SOD1^{WT}$-EGFP and $SOD1^{A4V}$-EGFP increased two fold (Figure 2.7). In summary, both CMV and Ubc promoters enable distinction of cellular fluorescence depending on misfolding/aggregation propensity of SOD1 variants.

Conclusion

Our investigation has conclusively established that EGFP fused to SOD1 variant C-terminus allows the extent of deviations to be determined from the correctly folded structure of SOD1 via cellular fluorescence measurement. The order of fluorescence intensities of transfected cells expressing four SOD1 variants correlates well to the inverse order of misfolding/aggregation propensities of the four SOD1 variants in both HEK293 and NSC-34 cells. Therefore, mean cellular fluorescence of cells expressing EGFP fusion of a target protein is a good indicator of misfolding/aggregation of the target protein. Flow cytometric monitoring of mutant SOD1 aggregation in cultured cells is a simple, but effective method that provides an array of structural information. This method could be used to screen for small molecule or protein-based drugs that inhibit mutant SOD1 aggregation. Screening of small molecule inhibitors in cultured cells is desirable to identify small molecule inhibitors of mutant SOD1 aggregation. If the small molecules identified were to inhibit mutant SOD1 aggregation, the fluorescence intensity of the cells will be near that of $SOD1^{WT}$. Combined with an automatic sampler in a 96-well plate format, large small molecule inhibitor libraries can be screened with

moderate to high throughput. Protein-based therapeutics that inhibit mutant SOD1 aggregation can be also identified by screening protein libraries in a high-throughput manner using fluorescence activated cell sorter. With the rapid advancements in ALS research, this assay would serve as a powerful tool in understanding the pathological mechanism and developing treatments for fALS.

Generality of GFP tagging method will easily allow extension of this method to monitor other cytosolic proteins that readily form aggregates. Aggregation of recombinant protein expressed in mammalian cells can reduce recombinant protein production yield. Therefore, the folding reporter can be used to determine residues causing protein aggregation and to identify less aggregation-prone protein variants by screening protein libraries. Furthermore, the folding reporter can be used to develop a fusion tag/protein to enhance solubility of target proteins in mammalian cells.

Chapter 3: Cis-Suppression to Arrest Protein Aggregation in Mammalian Cells

ABSTRACT

Protein misfolding and aggregation are implicated in numerous human diseases and significantly lower production yield of proteins expressed in mammalian cells. Despite the importance of understanding and suppressing protein aggregation in mammalian cells, a protein design and selection strategy to modulate protein misfolding/aggregation in mammalian cells has not yet been reported. In this work, we address the particular challenge presented by mutation-induced protein aggregation in mammalian cells. We hypothesize that an additional mutation(s) can be introduced in an aggregation-prone protein variant, spatially near the original mutation, to suppress misfolding and aggregation (cis-suppression). As a model protein, we chose human copper, zinc superoxide dismutase mutant ($SOD1^{A4V}$) containing an alanine to valine mutation at residue 4, associated with the familial form of amyotrophic lateral sclerosis. We used the program RosettaDesign to identify Phe20 in $SOD1^{A4V}$ as a key residue responsible for $SOD1^{A4V}$ conformational destabilization. This information was used to rationally develop a pool of candidate mutations at the Phe20 site. After two rounds of mammalian-cell based screening of the variants, three novel $SOD1^{A4V}$ variants with a significantly reduced aggregation propensity inside cells were selected. The enhanced stability and reduced aggregation propensity of the three novel $SOD1^{A4V}$ variants were verified using cell fractionation and *in vitro* stability assays.

INTRODUCTION

Mutations in cellular proteins spontaneously arise as a genetic variation (Arnheim and Calabrese 2009; Lynch 2010). Certain mutations are intentionally introduced to enzymes or therapeutic proteins in order to improve their catalytic or pharmaceutical activity, respectively (Lazar et al. 2003; Tracewell and Arnold 2009). In nature, about 70% of the spontaneous amino acid mutations are deleterious (Sawyer et al. 2007). Further, even those mutations that benefit one protein property usually lead to a loss of other properties. In particular, the mutation(s) often destabilizes proteins resulting in protein misfolding and aggregation.

Protein misfolding and aggregation are implicated in numerous human diseases, such as Alzheimer's disease, Parkinson's disease, and amyotrophic lateral sclerosis (ALS)(Bartolini and Andrisano 2010; Colby et al. 2006; Keshet et al. 2010; Lee et al. 2007; Murphy 2002; Pike et al. 1991; Valentine and Hart 2003). Furthermore, misfolding and intracellular aggregation of biopharmaceuticals, such as α-galactosidase, α-glucosidase, antithrombine III, and angiopoietin-1, in mammalian cells significantly lower protein production yields when expressed in mammalian cells (Hwang et al. 2011; Ioannou et al. 1992; Ishii et al. 1996; Okumiya et al. 2007; Schröder et al. 2002). Therefore, modulating protein misfolding and aggregation is a very important biotechnology topic.

Design strategies for increasing folding stability and reducing the aggregation propensity of proteins and antibodies have been extensively explored for *in vitro* and bacterial systems. Sidhu and co-workers developed a sound design strategy for developing soluble and stable V_H domains in antibodies (Barthelemy et al. 2008). Christ and co-workers have developed aggregation-resistant CDR regions using combinatorial assembly of pre-selected CDR regions (Dudgeon et al. 2009). However, such a strategy has not been actively explored for

understanding and engineering aggregation-resistant proteins when expressed within mammalian cells. In order to investigate the aggregation of disease-associated proteins or human-origin therapeutic proteins, mammalian cells are a physiologically more relevant environment than *in vitro* conditions. However, a protein design strategy to modulate protein misfolding/aggregation in mammalian cells has not yet been exploited.

Here we report a novel, systematic approach to arrest the mutation-induced aggregation of a protein expressed in mammalian cells. We hypothesize that the introduction of an additional mutation(s) to an aggregation-prone protein variant can suppress the destabilizing effects of the original mutation and therefore suppress the protein aggregation within mammalian cells (cis-suppression). As a model protein, we chose human copper, zinc superoxide dismutase (SOD1) mutants. Wild-type SOD1 ($SOD1^{WT}$) forms a very stable soluble dimer inside cells and catalyzes the conversion of superoxide anion to oxygen and hydrogen peroxide. It has been reported that over 100 mutations in SOD1 are associated with familial form of ALS, the most common motor neuron disease in humans (Reaume et al. 1996; Valentine et al. 2005).

The mechanism of toxicity of the SOD1 mutants to motor neurons is not fully understood. However, there is some evidence supporting the idea that motor neuron death is not caused by loss of SOD1 activity (Reaume et al. 1996), but rather by neurotoxic SOD1 mutant aggregates (Rakhit et al. 2002; Watanabe et al. 2001; Wong et al. 1995). In particular, $SOD1^{A4V}$ mutant, the most common mutant in North America, contains the alanine to valine mutation in the fourth residue (A4V) and leads to intracellular aggregate formation and eventually a rapidly progressing form of fALS (Ray and Lansbury 2004). Therefore, significant efforts have been made to modulate aggregation of $SOD1^{A4V}$. Since intermolecular disulfide bond formation is thought to be involved in $SOD1^{A4V}$ aggregation, cysteine to serine mutations were introduced to

SOD^{A4V} resulting in reduced aggregation in mammalian cells (Cozzolino et al. 2008; Deng et al. 2006). Alternatively, small molecules stabilizing the $SOD1^{A4V}$ dimer have been identified by *in silico* screening of small molecule libraries (Ray et al. 2005). Although the small molecules identified slowed aggregation of some SOD1 mutants *in vitro*, their aggregation modulatory capacities were not verified in mammalian cells.

In this study, the Phe20 was identified as a key residue responsible for the $SOD1^{A4V}$ destabilization first by computational analysis of $SOD1^{A4V}$ using RosettaDesign. Mutations that mitigated the structural perturbation caused by the A4V mutation were selected using the mammalian cell-based protein aggregation assays (Gregoire and Kwon 2012b). Further, *in vitro* characterization of the SOD1 mutants were performed to investigate the link between SOD1 mutant stability and aggregation in cells.

MATERIAL AND METHODS

Materials. The HRP-conjugated anti-rabbit antibody was obtained from Invitrogen (Carlsbad, CA). The anti-SOD1antibody was obtained from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA). HEK293T and NSC-34 cells were obtained from Invitrogen and CELLutionsBiosystems (Burlington, Ontario), respectively. Primers used for construction of expression vectors were obtained from Invitrogen. Gibco Certified Fetal Bovine Serum (FBS) was obtained from Invitrogen.

Construction of Expression Vectors. The pEGFP-N3-$SOD1^{WT}$ and pEGFP-N3-$SOD1^{A4V}$ plasmid were kindly provided by Dr. Haining Zhu (University of Kentucky). The YEP351-$SOD1^{WT}$ and YEP351-$SOD1^{A4V}$ plasmids were kindly provided by Dr. Joan Valentine

(University of California at Los Angeles). The mutations were introduced into the pEGFP-N3-SOD1^{A4V} and YEP351-SOD1^{A4V} plasmids using the QuikChange II site-directed mutagenesis kit (Stratagene). The construction of the pEGFP-N3-SOD1$^{A4V/C111S}$ plasmid was described previously (Gregoire and Kwon 2012a). The primer pairs for each mutant are as follows: 5'-TCCTTCTGCTCGGCATTGATGATGCCCTGCACTGGGC-3' and 5' GCCCAGTGCAGGGCATCATCAATGCCGAGCAGAAGGA-3' for F20A; 5'-GCCCAGTGCAGGGCATCATCAATGGCGAGCAGAAGGA-3 and 5'-TCCTTCTGCTCGCCATTGATGATGCCCTGCACTGGGC-3' for F20G; 5'-CAGTGCAGGGCATCATCAATTTAGAGCAGAAGGAA-3' and 5'-TTCCTTCTGCTCTAAATTGATGATGCCCTGCACTG-3' for F20L; 5'-GCCCAGTGCAGGGCATCATCAATATGGAGCAGAAGGAAA-3' and 5'-TTTCCTTCTGCTCCATATTGATGATGCCCTGCACTGGGC-3' for F20M; 5'-CCAGTGCAGGGCATCATCAATGTCGAGCAGAAG-3' and 5'-CTTCTGCTCGACATTGATGATGCCCTGCACTGG-3' for F20V. The sequences of the mutations were confirmed using DNA sequencing.

Computational Analysis of Protein Structure. The computational analysis of SOD1 stability was performed with the molecular modeling program RosettaDesign, version 3.0 or 3.4 (Rohl et al. 2004). The core energy function of RosettaDesign is a linear sum of a 6-12 Lennard-Jones potential, the Lazaridis-Karplus implicit solvation model, an empirical hydrogen bonding potential, backbone-dependent rotamer probabilities, a knowledge-based electrostatics energy potential, amino acid probabilities based on particular regions of φ/ψ space and a unique reference energy for each amino acid (Rohl et al. 2004). For a given crystal structure,

RosettaDesign uses simulated annealing to scan through a large number of rotamers to minimize the energy score with Monte Carlo optimization (Kuhlman and Baker 2000).

A fixed backbone computation protein design protocol was used in this study to identify the key residue(s) responsible for $SOD1^{A4V}$ destabilization. Two crystal structures of $SOD1^{WT}$ (PDB IDs: 1PU0 and 2C9V) were used (DiDonato et al. 2003; Strange et al. 2006). In order to generate a homodimeric $SOD1^{A4V}$ structure from these $SOD1^{WT}$ structures, all protein side chains were first repacked by allowing all side chains, except those chelated to metal ions, to adopt favorable conformations. The A4V mutation was then inserted into the protein sequence, and the side chains were repacked again. The resulting structure was used as a starting point for computational screening of all possible second-site mutations to identify a key residue(s) responsible for $SOD1^{A4V}$ destabilization.

Transfection of HEK293T and NSC-34 Cells. HEK293T cells were maintained at 37°C and 5% CO_2 in Dulbecco's modified Eagle's medium/High Glucose (DMEM/High Glucose, Thermo Scientific, Pittsburgh, PA) supplemented with 10% FBS, 100 µg/mL of streptomycin sulfate and 100 units/mL of penicillin. NSC-34 cells were maintained in DMEM supplemented with 10% FBS, 4500 mg/L glucose, 0.584 g/L L-glutamine, 3.7 g/L sodium bicarbonate, 100 µg/mL of streptomycin sulfate and 100 units/ml of penicillin. Cells were seeded on 6-well plates one day prior to transfection. Once the cells grew to 80-90% confluency, they were transfected with 3.5 µg of the appropriate plasmid via the calcium phosphate precipitation method. All samples were done in triplicate, unless explicitly stated.

Fluorescence Microscopy. Two days post-transfection, HEK293T or NSC-34 cells were visualized via fluorescence microscopy using a VistaVision Inverted Fluorescence Microscope (VWR, Radnor, PA). Images were captured using a DC-2C digital camera. The fluorescence excitation wavelength range was between 420 and 485 nm and the emission wavelength was 515 nm. The number of cells in the fluorescence microscopy images was manually counted. The percentage of transfected cells exhibiting SOD1-EGFP aggregates was determined by dividing the number of aggregate-containing cells by the number of total cells analyzed. Only images containing greater than 50 cells were used. Three images per well of transfected cells were used.

Flow Cytometric Analysis of Cellular Fluorescence. Two days post-transfection, the transfected HEK293T and NSC-34 cells were prepared for flow cytometry. To prepare the cells for flow cytometric analysis, the HEK 293T cells and NSC-34 cells were trypsinized, washed twice with phosphate buffered saline buffer (PBS; pH 7.4) and resuspended in 500µL of 1X PBS. The fluorescence intensities of the HEK293T and NSC-34 cells expressing SOD1 variant-EGFP fusion protein were measured using the C6 flow cytometer (Accuri, Ann Arbor, MI). The excitation wavelength was 488 nm and the fluorescence emission was detected at 585 nm. Only GFP positive cells were used to calculate the mean cellular fluorescence. Transfection efficiency was determined by dividing the number of fluorescence positive cells by the number of total cells analyzed. Each sample was prepared in triplicate and cellular fluorescence indicates mean cellular fluorescence, unless otherwise noted. Two-sided, Student's paired t-test was used for statistical analysis of fluorescence data.

Total Protein Extraction and Protein Fractionation of HEK293T cells. $2x10^6$ transfected HEK293T cells were centrifuged to obtain a cell pellet. The cell pellet was washed with PBS buffer once and was lysed using 100 µL of RIPA buffer (Thermo Fisher, Pittsburgh, PA). To remove the detergent soluble fraction of proteins, the cells were incubated at 4 °C for 30 minutes. The samples were centrifuged at 4 °C and 15,500 g for 15 minutes. The supernatant was collected as a soluble fraction and kept on ice. The pellet was collected as an insoluble fraction and was then resuspended in 100 µL of RIPA buffer at 4 °C for 30 minutes. The samples were centrifuged at 4 °C and 15,500 g. The supernatant was discarded and the procedure was repeated once more. To extract the detergent-insoluble fraction of proteins, the pellet was incubated at room temperature in 100 µL of RIPA Buffer/8M Urea for 1 hour. The sample was centrifuged at 15,500 g for 15 minutes and the supernatant was collected. For total protein lysate extraction, cell pellets were resuspended in 100 µL of RIPA Buffer/8M Urea and incubated at room temperature overnight. Protein concentrations of the soluble fraction and total protein lysate were determined using BCA kit (Pierce, Rockford, IL) using bovine serum albumin as a standard.

Western Blotting. Soluble fractions were diluted using RIPA and 8M Urea to 1 µg/µL after the protein concentration was determined. The final sample buffer composition of the soluble fraction was 1:3 vol/vol (RIPA:8M Urea). The insoluble fractions were diluted using the same amount of volume of buffer used to dilute the soluble fraction for that particular mutant. 3µL of soluble and insoluble protein lysate were boiled in sample loading buffer containing DTT and electrophoresed on 12% SDS-PAGE Gel. Samples were transferred onto a nitrocellulose membrane for one hour and washed with 1X TBS-T three 3 times for 5 minutes. The membrane

probed with the primary anti-SOD1 antibody diluted in TBS-T for 1 hour at RT (1:5000). Samples were washed three times for 5 minutes in TBS-T and probed with the secondary antibody conjugated with HRP for 1 hour (1:10000). Bands were detected using ECL Prime (GE Healthcare, Sweden).

Expression and Purification of SOD1 Mutants. For *in vitro* studies, all SOD1 variants were expressed and purified based as previously described(Hough et al., 2004). *Saccharomyces cerevisiae* EG118 strain lacking the endogenous yeast SOD1 gene was kindly provided by Dr. Joan Valentine (University of California at Los Angeles). The plasmids containing the SOD1 variants were expressed in yeast cells. These cells were first grown on a plate containing yeast synthetic drop-out media without leucine (Sigma-Aldrich, Inc., Saint Louis, US) for two to three days at 30 °C, followed by inoculation into 50 mL fresh liquid yeast synthetic drop-out media without leucine and grown for two to three days at 30 °C and 220 rpm until the optical density at 600 nm wavelength (OD_{600}) read about 5. The resulting culture was used to inoculate 2 L Yeast Extract Peptone Dextrose (YPD) media, and the cells were grown for another four to five days under the same conditions until the OD_{600} reached 10.

The yeast cells were then pelleted and resuspended in two pellet volumes of lysis buffer (250 mM Tris, 150 mM NaCl, 0.1 mM EDTA, pH 8.0). Agitation with 0.5 mm diameter glass beads was performed in ice bath, followed by centrifugation at 4 °C and 11,000 rpm for 30 min. Ammonium sulfate was then slowly added to 60% saturation to the supernatant while stirring in ice bath. Centrifugation was performed again at 4 °C and 8,500 rpm for 45 min. The supernatant was then applied to a HiPrep Phenyl FF (high sub) 16/10 column (GE Healthcare Biosciences, Pittsburgh, USA). Proteins were eluted in 5 column volumes with a linear gradient of $(NH_4)_2SO_4$

(2.0 down to 0 M) in the buffer of 50 mM NaH_2PO_4, 150 mM NaCl, 0.3 mM EDTA, pH 7.0. Fractions containing mainly SOD1 were pooled and dialyzed against double distilled H_2O (ddH_2O) at 4 °C. Samples were then loaded to a HiTrap DEAE FF column (GE Healthcare Biosciences) and proteins were eluted in 3 column volumes with a linear gradient of potassium phosphate (0.0025-0.25 M), pH 7.0. The SOD1 fractions were pooled and concentrated to a small volume (less than 1 mL) and applied to a Superdex 75 10/30 GL column (GE Healthcare Biosciences) in the buffer of 20 mM potassium phosphate, 80 mM NaCl, pH 7.4. Fractions were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Finally, aliquots of 10 mg/mL protein in 10 mM potassium phosphate, pH 7.0 were frozen in -80 °C freezer.

Guanidine Hydrochloride-induced denaturation assay. The stability of SOD1 variants was measured by guanidine hydrochloride (GdnHCl)-induced denaturation assay. Proteins were buffer exchanged to phosphate buffered saline (PBS, 137 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 2 mM KH_2PO_4, pH 7.4) followed by mixing with increasing concentrations of GdnHCl. The final dimer concentration of each protein sample was between 10 to 15 μM. Samples were gently vortexed, followed by incubation for 16 hr at room temperature and shaking at 100 rpm. CD ellipticities at 218 nm were recorded on a JASCO J-710 spectropolarimeter with a 0.1 cm path length cuvette at room temperature. Three 30s measurements with response time set at 2 s were averaged and buffer adsorption was deducted to obtain the final value.

For easier comparison of the denaturation curves among SOD1 variants, the CD ellipticities (θ) for all the samples were converted to the fraction unfolded (f_u) with the following equation:

$$f_u = \frac{(\overline{\theta_n} - \theta)}{(\overline{\theta_n} - \overline{\theta_d})}$$

Where $\overline{\theta_n}$ and $\overline{\theta_d}$ are the ellipticity averaged from the elliplicities in the native and denatured states, respectively.

RESULTS AND DISCUSSION

Identification of key residues responsible for SOD1^{A4V} conformational destabilization using computational protein design

So far, SOD1^{A4V} misfolding and aggregation has been thought to be mediated by a perturbation in dimer interface. However, our computational protein design results strongly suggest that F20 may also be an important residue for SOD1^{A4V} conformational destabilization, though it is not in the dimer interface.

As a first step to understand and prevent intracellular aggregation of SOD1^{A4V}, computational protein design using RosettaDesign was performed to identify key residues responsible for SOD1^{A4V} conformational destabilization. In order to investigate the structural changes caused by the A4V mutation more clearly, the A4V mutation was introduced into the SOD1^{WT} structures (1PU0 and 2C9V) to generate two pseudo-SOD1^{A4V} structures as described in the Materials and Methods section. Using RosettaDesign, each of the 12 amino acids within a 5-Å distance from the Val4 was substituted with another amino acid (Figure 3.1), resulting in 342 variants. By excluding the variants with a disrupted conformation or with problematic interactions with V4, 33 variants were obtained, and their energy scores (the change in the energy score of each variant relative to that of the SOD1^{WT}; $\Delta\Delta G_f$) are shown in Table 3.1. The mutation of Phe at residue 20 (F20) into amino acids with a smaller side chain leads to a

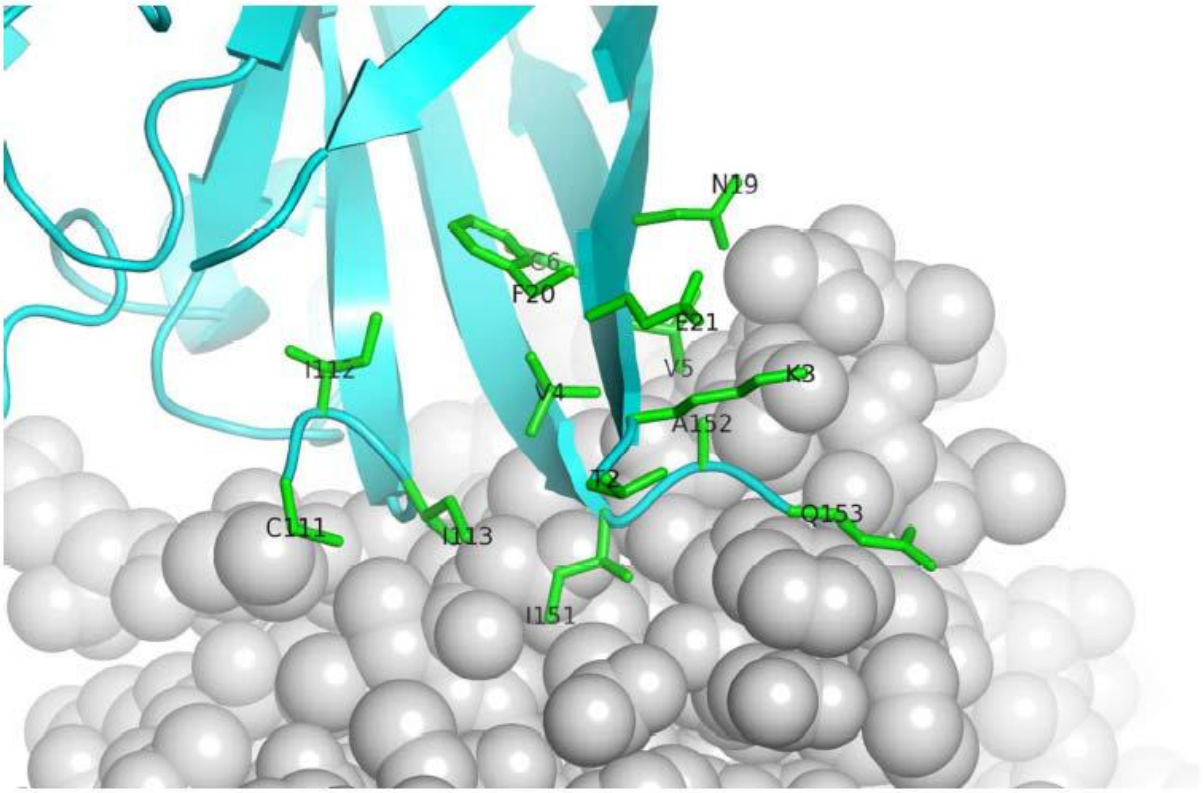


Figure 3.1: The crystal structure of $SOD1^{A4V}$ (PDB ID: 1U70). The residues near Val4 in one subunit (cyan color) are shown in green color. The other subunit is shown in gray spheres. The structure is generated using PyMol graphic software (Schrodinger 2010).

Table 3.1. Summary of the change in total energy score ($\Delta\Delta G_f$) of the $SOD1^{A4V}$ variants relative to that of $SOD1^{WT}$ using two wild-type starting structures from the PDB (1PU0 and 2C9V) prepared as described in the methods.

Mutation	1PU0	2C9V
	$\Delta\Delta G_f$ kcal/mol	$\Delta\Delta G_f$ kcal/mol
A1G	0.3	-7.4
K3E	0.7	0.9
K3R	-1.0	-1.6
K3T	-0.5	-0.6
K3V	-0.8	0.3
V5S	-0.5	-0.2
C6A	0.5	0.6
C6V	0.6	0.6
N19F	-3.5	-2.9
N19W	-3.5	-2.6
N19Y	-3.5	-2.9
F20G	-37.9	-42.8
F20L	-14.9	-13.9
E21K	-0.2	0.3
E21M	0.1	0.1
E21R	0.0	0.4
E21V	-0.2	2.0
E21Y	1.7	11.7
I112V	1.3	1.5
I113Q	1.2	1.0
I113R	-0.6	-0.1
I113V	0.2	0.2
I151D	0.9	0.8
I151V	0.2	0.2
I151Y	-0.2	-0.1
A152C	-1.0	-0.8
A152H	75.6	1.6
A152S	-1.0	-0.7
Q153A	0.1	0.2
Q153C	0.1	-0.1
Q153D	-0.3	0.1
Q153S	-0.1	0.1

substantial increase in the stability in both crystal structures. Consistent with this observation, the repulsive Lennard-Jones contribution within the energy function was adversely affected in A4V and the enhanced stability increase made by the mutations at F20 site reduced this steric effect (data not shown). It was also reported A4V mutation reduces the contact time of F20 with the various residues near the dimer interface (Cardoso et al. 2002; Schmidlin et al. 2009). Therefore, we hypothesized that F20 is a key residue responsible for $SOD1^{A4V}$ conformational destabilization.

Screening of $SOD1^{A4V}$ variants containing a mutation at the F20 site to identify mutations eliminating the $SOD1^{A4V}$destabilization

We hypothesized that introducing an additional mutation(s) can mitigate the conformationally destabilizing effects of the A4V mutation generating more stable $SOD1^{A4V}$ variants (cis-suppression). Therefore, we generated $SOD1^{A4V}$ variants containing a mutation at the F20 site followed by the cell-based screening to identify mutations eliminating the destabilizing effects of the A4V in $SOD1^{A4V}$ structure. Hypothesizing that intracellular $SOD1^{A4V}$ aggregation is attributed to the $SOD1^{A4V}$ conformational destabilization, such stabilizing mutations were expected to arrest the $SOD1^{A4V}$ aggregation. To alleviate the steric clash between V4 and F20, we mutated F20 into five amino acids with a smaller, hydrophobic side chain (Gly, Ala, Val, Leu, and Met) generating five $SOD1^{A4V}$ variants containing a mutation at the F20 site ($SOD1^{A4V/F20G}$,$SOD1^{A4V/F20A}$, $SOD1^{A4V/F20V}$, $SOD1^{A4V/F20L}$,and $SOD1^{A4V/F20M}$). Among the five variants, the two variants ($SOD1^{A4V/F20G}$ and $SOD1^{A4V/F20L}$) were also predicted by RosettaDesign to be conformationally stabilizing (Table 3.1.).

To compare the effect of these mutations (F20X) on the $SOD1^{A4V}$ aggregation, we utilized the cell-based misfolding/aggregation assay that we developed (Gregoire and Kwon 2012a). We previously demonstrated that the cellular fluorescence intensities of transfected mammalian cells (HEK293T and NSC-34 cells) expressing the enhanced green fluorescent protein (EGFP) fusion of SOD1 variants are inversely correlated to the misfolding/aggregation propensities of the SOD1 variants (Gregoire and Kwon 2012). The two cell lines (HEK293T and NSC-34 cells) were also used for the studies described here. HEK293T cell is one of the most common human-origin cell lines to study gene expression. NSC-34 cell is a hybrid cell line made by a fusion of mouse spinal cord neurons with mouse neuroblastoma and is widely used to study SOD1 aggregation associated with fALS (Babetto et al. 2005; Cashman et al. 1992; Kupershmidt et al. 2009; Raimondi et al. 2006; Rizzardini et al. 2006; Tartari et al. 2009). Both HEK293T and NSC-34 cells were transfected with the plasmid encoding each of the five SOD1 double mutants ($SOD1^{A4V/F20G}$, $SOD1^{A4V/F20A}$, $SOD1^{A4V/F20V}$, $SOD1^{A4V/F20L}$, and $SOD1^{A4V/F20M}$). As a positive control with a reduced aggregation-propensity, another double mutant containing C111S mutation as well as A4V ($SOD^{1A4V/C111S}$) was also tested. The C111S mutation is known to reduce $SOD1^{A4V}$ aggregation inside mammalian cells by inhibiting intramolecular disulfide-bond formation (Cozzolino et al. 2008; Deng et al. 2006). At two days post-transfection, the transfected cells were harvested, and the mean cellular fluorescence of the cells was determined using a flow cytometer. As reported previously (Gregoire and Kwon 2012), the transfected HEK293T cells expressing EGFP fusion of the $SOD1^{WT}$ exhibited the mean cellular fluorescence higher than that of the $SOD1^{A4V}$ variant by about 2.5-fold (Figure 3.2A). Numerical values of the cellular fluorescence are shown in Table B1 in Appendix B. The HEK293T cells expressing the $SOD1^{A4V/C111S}$ variant also exhibited an enhanced mean cellular fluorescence compared to that of

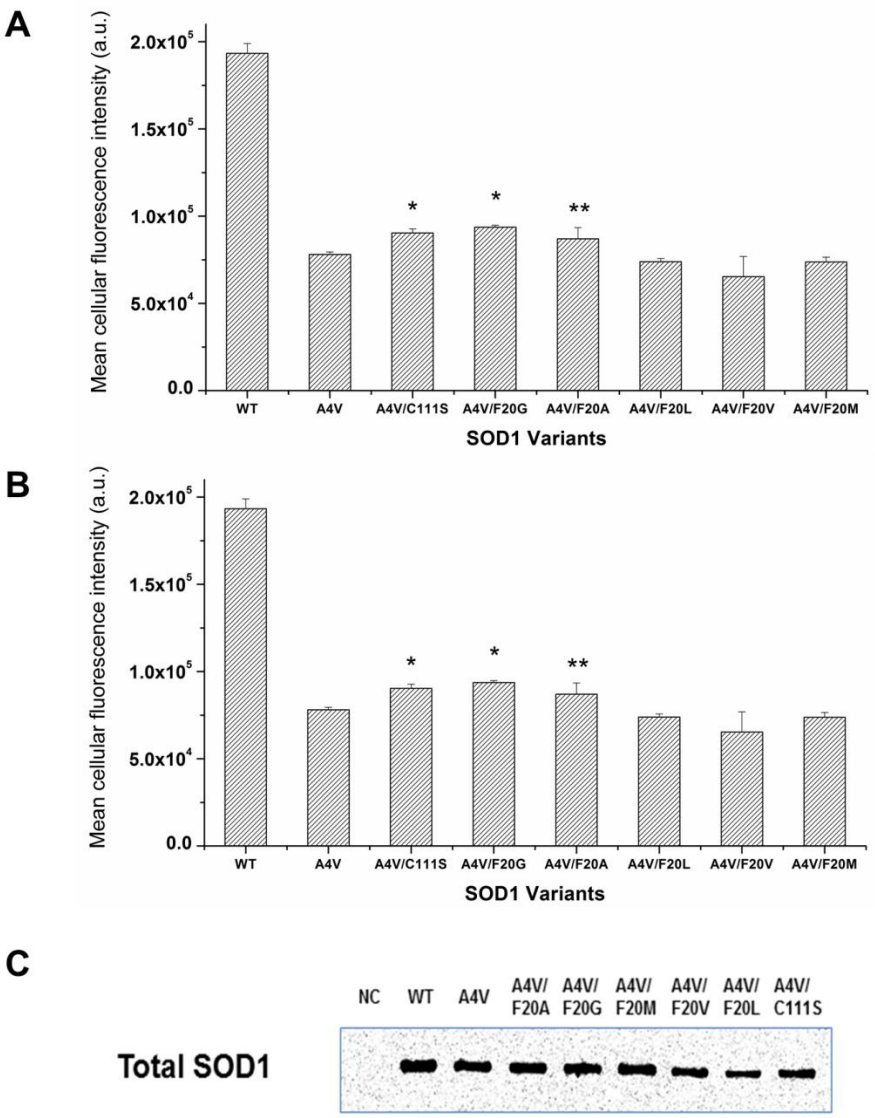


Figure 3.2: The mean cellular fluorescence of the transfected HEK293T (A) and NSC-34 (B) cells expressing EGFP fusion of SOD1 variants. (C) The SOD1 bands of the transfected HEK293T cells expressing the SOD1 variants. Values and error bars represent mean cellular fluorescence and standard deviations, respectively (n = 3). In order to determine whether the mean cellular fluorescence of the SOD1 double mutants is greater than that of $SOD1^{A4V}$, two-sided Student's t-tests were applied to the data (* $p < 0.05$; ** $p < 0.01$).

$SOD1^{WT}$ ($p < 0.05$) suggesting a reduced aggregation propensity. Among the five $SOD1^{A4V}$ double mutants containing a F20X mutation ($SOD1^{A4V/F20X}$), the HEK293T cells expressing the EGFP fusion of the $SOD1^{A4V/F20A}$ and $SOD1^{A4V/F20G}$ showed a significant increase in cellular fluorescence compared to that of $SOD1^{A4V}$ variant. However, the HEK293T cells expressing the other three $SOD1^{A4V/F20X}$ double mutants exhibited a reduced or similar mean cellular fluorescence compared to that of $SOD1^{A4V}$. The relative mean cellular fluorescence of the transfected neuroblastoma NSC-34 cells expressing SOD1 variants showed a similar trend with that of the transfected HEK293T cells (Figure 3.2B), supporting the idea that the cellular fluorescence trend found in the SOD1 variants is valid. Since mean cellular fluorescence can be affected by the expression level of a fusion protein, the relative expression levels of the EGFP fusion of SOD1 variants were determined by comparing their western-blot band intensities (Figure 3.2C). The expression levels of the $SOD1^{A4V/F20V}$ and $SOD1^{A4V/F20L}$ variants were substantially lower than those of the $SOD1^{WT}$ and $SOD1^{A4V}$, whereas the expression levels of the $SOD1^{A4V/F20A}$, $SOD1^{A4V/F20G}$, and $SOD1^{A4V/F20M}$ variants were comparable to those of the $SOD1^{WT}$ and $SOD1^{A4V}$. When the mean cellular fluorescence was adjusted using the relative expression level of each SOD1 variant, the adjusted mean cellular fluorescence of the $SOD1^{A4V/F20V}$ and $SOD1^{A4V/F20L}$ variants as well as the $SOD1^{A4V/F20A}$ were significantly higher than that of the $SOD1^{A4V}$ (Figure 3.3). The relatively low expression level of the he $SOD1^{A4V/F20V}$ and $SOD1^{A4V/F20L}$ variants is likely an indication of a significant structural deviation from the $SOD1^{WT}$ resulting in an enhanced degradation rate inside cells (Kabuta et al. 2006). Therefore, the $SOD1^{A4V/F20A}$ and $SOD1^{A4V/F20G}$ variants were considered promising candidates with a reduced aggregation propensity.

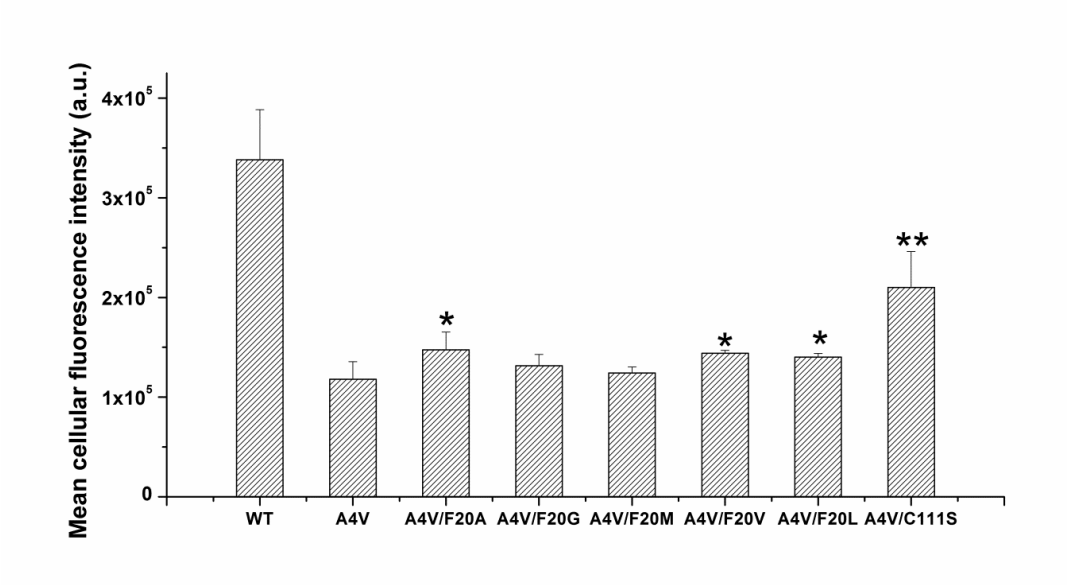


Figure 3.3: The adjusted mean cellular fluorescence of the transfected cells expressing the EGFP fusion of $SOD1^{WT}$, $SOD1^{A4V}$, and $SOD1^{A4V}$ variants containing a second mutation. The values were calculated by dividing the mean cellular fluorescence of each sample by the relative expression level to that of $SOD1^{WT}$. Values = mean ± standard deviation (n = 3). In order to determine whether the mean cellular fluorescence of the SOD1 double mutants is greater than that of $SOD1^{A4V}$, two-sided Student's t-tests were applied to the data (* $p < 0.05$; ** $p < 0.01$).

Fluorescence microscopic analysis of the transfected HEK293T cells expressing the SOD1^{A4V} variants.

To confirm whether the increased mean cellular fluorescence of the transfected HEK293T cells expressing the SOD1$^{A4V/F20A}$ and SOD1$^{A4V/F20G}$ double mutants is attributed to the reduced aggregation inside cells, the fluorescence images of the transfected cells were obtained using a fluorescence microscope (Figure 3.4). The transfected HEK293T cells expressing the EGFP fusion of the SOD1^{WT} did not show any intracellular aggregates (*WT Panel* in Figure 3.4) consistent with the findings in the literature (Gregoire and Kwon 2012; Witan et al. 2008; Zhang and Zhu 2006). However, a substantial portion of the transfected cells expressing the SOD1^{A4V} exhibited aggregate formation within the cytosol (*A4V Panel* in Figure 3.4). Similar to the SOD1^{A4V}, the cells expressing the SOD1$^{A4V/F20V}$, SOD1$^{A4V/F20M}$, and SOD1$^{A4V/F20L}$ exhibited intracellular aggregates *(A4V/F20A, A4V/F20L and A4V/F20M Panels* in Figure 3.4). Therefore, the combination of the cellular fluorescence measurement and the fluorescence microscopic observation indicate that the SOD1$^{A4V/F20V}$, SOD1$^{A4V/F20M}$, and SOD1$^{A4V/F20L}$ variants are misfolded or aggregated as much as the SOD1^{A4V}.

However, intracellular aggregates are rarely found in the cells expressing the SOD1$^{A4V/F20A}$ and SOD1$^{A4V/F20G}$ (*A4V/F20A and A4V/F20G* Panels in Figure 3.4), strongly indicating reduced aggregation propensity of the double mutants SOD1$^{A4V/F20A}$ and SOD1$^{A4V/F20G}$ compared to that of SOD1^{A4V}. For a more quantitative analysis, we determined the percentage of cells that exhibit intracellular aggregates using image analysis described in the Material and Methods section. In the nine images, at least 50 cells were analyzed for each of the SOD1 variants (SOD1^{WT}, SOD1^{A4V}, SOD1$^{A4V/C111S}$, SOD1$^{A4V/F20A}$, and SOD1$^{A4V/F20G}$) (Figure 3.5). Around 14% of the transfected HEK293T cells expressing the SOD1^{A4V} exhibited intracellular

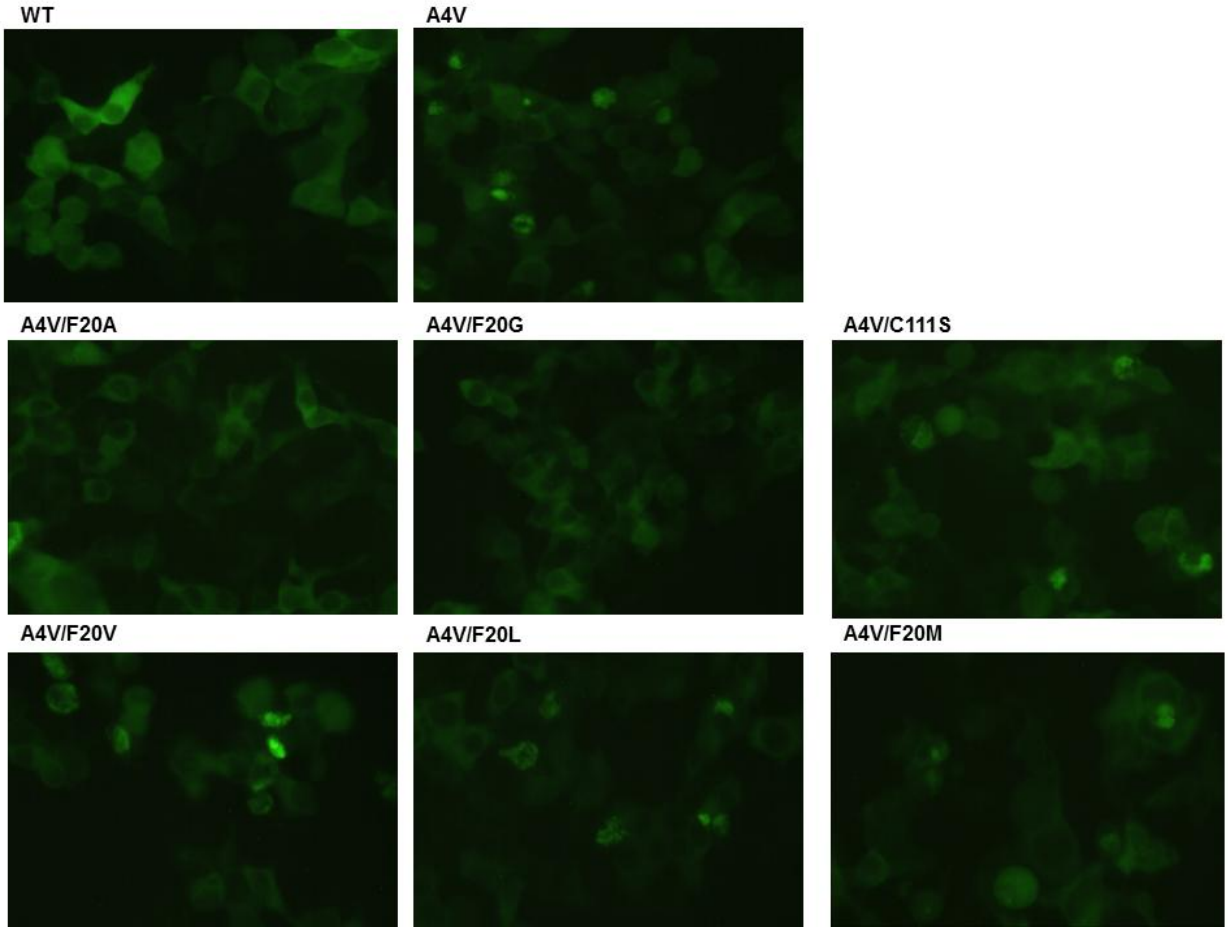


Figure 3.4: The fluorescence microscopic images of the transfected HEK293T cells expressing EGFP fusion of SOD1 variants. The images of transfected HEK293T cells expressing EGFP fusion of $SOD1^{WT}$, $SOD1^{A4V}$, and $SOD1^{A4V}$ double mutants were taken at two days post-transfection.

aggregates, consistent with our previous results(Gregoire and Kwon 2012). In case of the less aggregation-prone SOD1$^{A4V/C111S}$ versus the SOD1^{A4V}, the percentage of the cells exhibiting intracellular aggregates decreased to 5%. More importantly, only 3% and 0.1% of the cells expressing SOD1$^{A4V/F20A}$ and SOD1$^{A4V/F20G}$ respectively exhibited intracellular aggregates (Figure 3.5). Numerical values of the percentage of the transfected cells exhibiting aggregates are shown in Table B2 in Appendix B. To our knowledge, the SOD1$^{A4V/F20A}$ and SOD1$^{A4V/F20G}$ are the first SOD1^{A4V} variants with a substantially reduced aggregation propensity without any change in the cysteine residues or metal-binding regions. The SOD1$^{A4V/F20A}$ and SOD1$^{A4V/F20G}$ double mutants were used for further mutagenesis in order to reduce the aggregation propensity.

Identification of SOD1 triple mutants with a significantly reduced misfolding/aggregation propensity

The addition of F20A or F20G mutation to SOD1^{A4V} led to a substantial reduction in intracellular aggregate formation in mammalian cells. However, our initial screen using flow cytometric analysis showed only a slight increase in the mean cellular fluorescence, implying that the misfolding/aggregation was not completely avoided. Therefore, there is still a need to further reduce the misfolding/aggregation propensity of the F20 double mutants. We hypothesized that the introduction of the C111S mutation, proven to reduce the SOD1^{A4V}, into the SOD1$^{A4V/F20A}$and SOD1$^{A4V/F20G}$ will have a synergic effect on the aggregation reduction. Therefore, two triple mutants, SOD1$^{A4V/F20A/C111S}$ and SOD1$^{A4V/F20G/C111S}$ were generated. When the EGFP fusion of the SOD1$^{A4V/F20A/C111S}$ and SOD1$^{A4V/F20G/C111S}$ were expressed in HEK293T and NSC-34 cells, the SOD1$^{A4V/F20A/C111S}$ showed a significant increase in cellular fluorescence from that of the SOD1^{A4V} (Figure 3.6A and B). Numerical values of the cellular fluorescence are

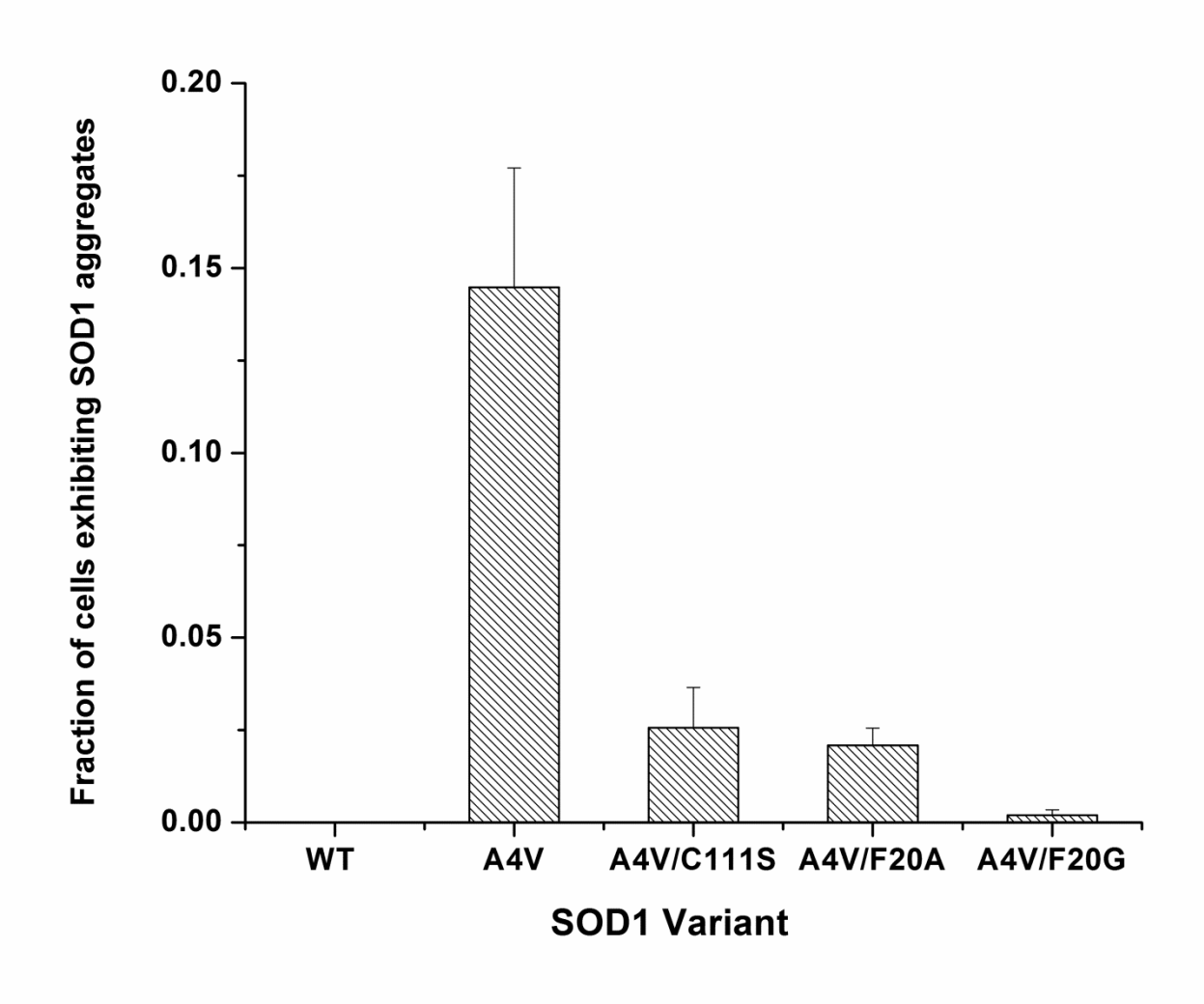


Figure 3.5: The fraction of the transfected HEK293T cells exhibiting SOD1 aggregates. The number of cells exhibiting intracellular SOD1 aggregates was determined by analyzing fluorescence microscopy images of the transfected cells expressing EGFP fusion of five SOD1 variants ($SOD1^{WT}$, $SOD1^{A4V}$, $SOD1^{A4V/C111S}$, $SOD1^{A4V/F20A}$, and $SOD1^{A4V/F20G}$) at two days post-transfection.

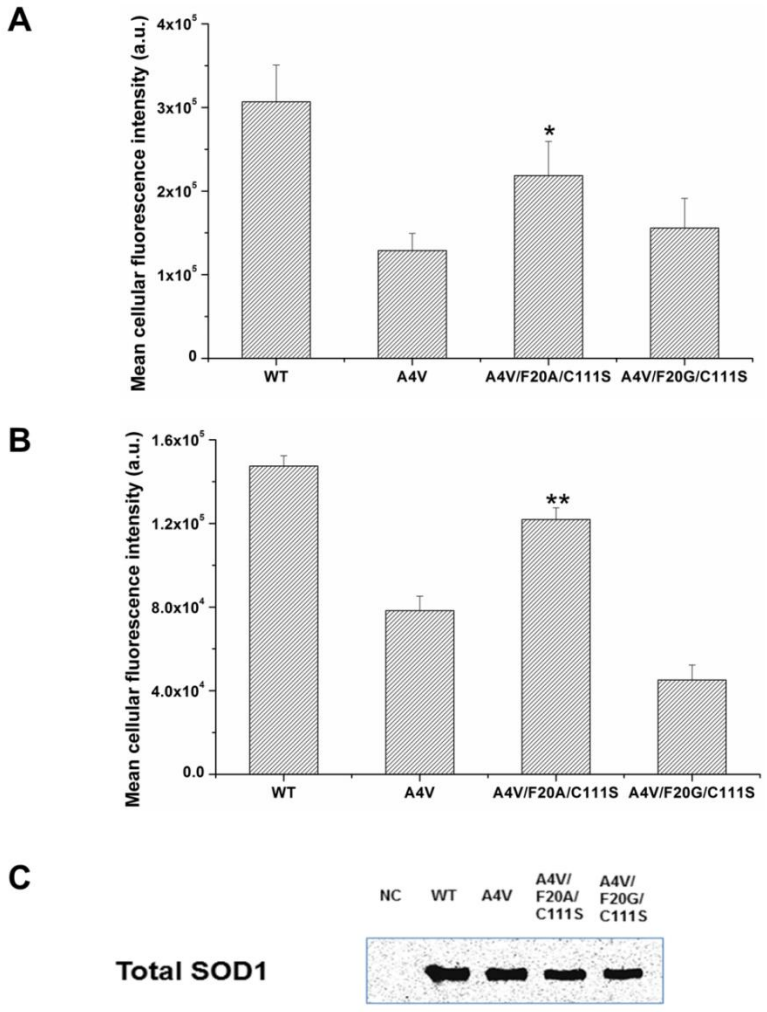


Figure 3.6: The mean cellular fluorescence of the transfected (A) HEK293T cells and (B) NSC-34 expressing EGFP fusion of the SOD1 variants ($SOD1^{WT}$, $SOD1^{A4V}$, $SOD1^{A4V/F20A/C111S}$, and $SOD1^{A4V/F20G/C111S}$). (C) The SOD1 bands of the transfected HEK293T cells expressing the SOD1 variants. Values and error bars represent mean cellular fluorescence and standard deviations, respectively (n = 3). In order to determine whether the mean cellular fluorescence of the SOD1 triple mutants is greater than that of $SOD1^{A4V}$, two-sided Student's t-tests were applied to the data (* $p < 0.05$; ** $p < 0.01$).

shown in Table B1 in Appendix B. When the mean cellular fluorescence of the SOD1 variants were adjusted using the relative expression levels (Figure 3.6C and Figure 3.7), the adjusted mean cellular fluorescence of the $SOD1^{A4V/F20A/C111S}$ in the transfected HEK293T cells was comparable to that of the $SOD1^{WT}$ ($p = 0.79$ in two-sided the Student's t-test) suggesting that the misfolding/aggregation of the $SOD1^{A4V/F20A/C111S}$ was substantially eliminated. However, the transfected HEK293T or NSC-34 cells expressing the $SOD1^{A4V/F20G/C111S}$ showed no substantial increase or a reduction in mean cellular fluorescence, respectively (Figure 3.6 A and B). When compared to that of the $SOD1^{A4V}$, the introduction of C111S to $SOD1^{A4V/F20G}$ is potentially destabilizing. Even the adjusted mean fluorescence of the $SOD1^{A4V/F20G/C111S}$ was not significantly higher than that of the $SOD1^{A4V}$ (Figure 3.7) ($p > 0.05$).

To confirm whether the increased mean cellular fluorescence of the $SOD1^{A4V/F20A/C111S}$ is attributed to a reduction in aggregate formation, fluorescence microscopy images of the HEK293T cells expressing the SOD1 variants were taken at two days post-transfection (Figure 3.8). The number of cells exhibiting intracellular aggregates was manually counted (Figure 3.9). The cells expressing the $SOD1^{A4V/F20A/C111S}$ rarely exhibited intracellular aggregates (*A4V/F20A/C111S* Panel in Figure 3.8). In case of the cells expressing the $SOD1^{A4V/F20A/C111S}$, the percentage of cells exhibiting intracellular aggregates was 0.1%, which is a remarkable decrease when compared to the cells expressing the $SOD1^{A4V}$ (13.5 %). Numerical values of the percentage of the transfected cells exhibiting aggregates are shown in Table B2 in Appendix B. The percentage of cells exhibiting intracellular aggregates which expressed $SOD1^{A4V/F20A/C111S}$ was even lower than that of the $SOD1^{A4V/F20G}$ variant (0.4%) (Figure 3.9).

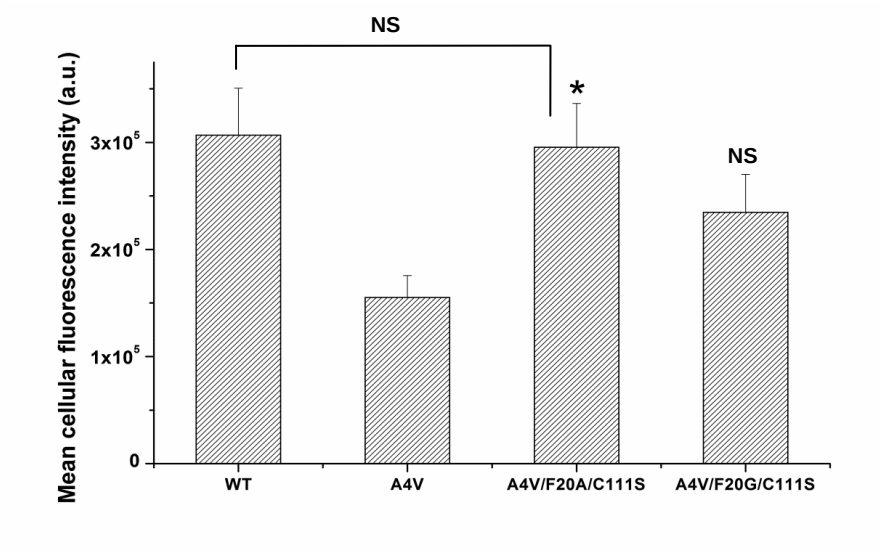


Figure 3.7: The adjusted mean cellular fluorescence of the transfected cells expressing the EGFP fusion of $SOD1^{WT}$, $SOD1^{A4V}$, $SOD1^{A4V/F20A/C111S}$, and $SOD1^{A4V/F20G/C111S}$. The values were calculated by dividing the mean cellular fluorescence of each sample by the relative expression level to that of $SOD1^{WT}$. Values = mean ± standard deviation (n = 3). In order to determine whether the mean cellular fluorescence of the SOD1 triple mutants is greater than that of $SOD1^{A4V}$, two-sided Student's t-tests were applied to the data (* $p < 0.05$; NS: not significant). The mean cellular fluorescence of the $SOD1^{A4V/F20A/C111S}$ is not significantly different from that of $SOD1^{WT}$.

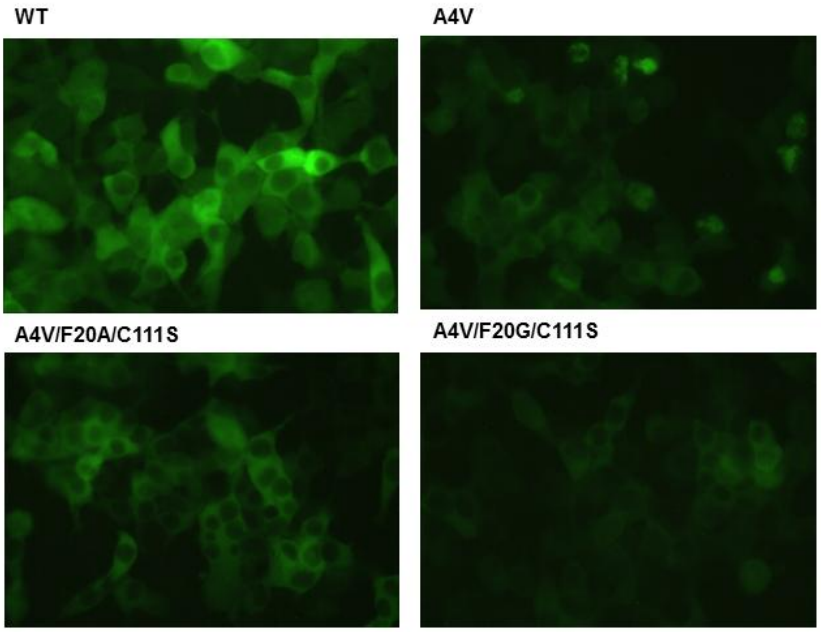


Figure 3.8: The fluorescence microscopic images of the transfected HEK293T cells expressing EGFP fusion of SOD1 variants. The images of the transfected HEK293T cells expressing EGFP fusion of the $SOD1^{WT}$, $SOD1^{A4V}$ and $SOD1^{A4V}$ triple mutants were taken at two days post-transfection.

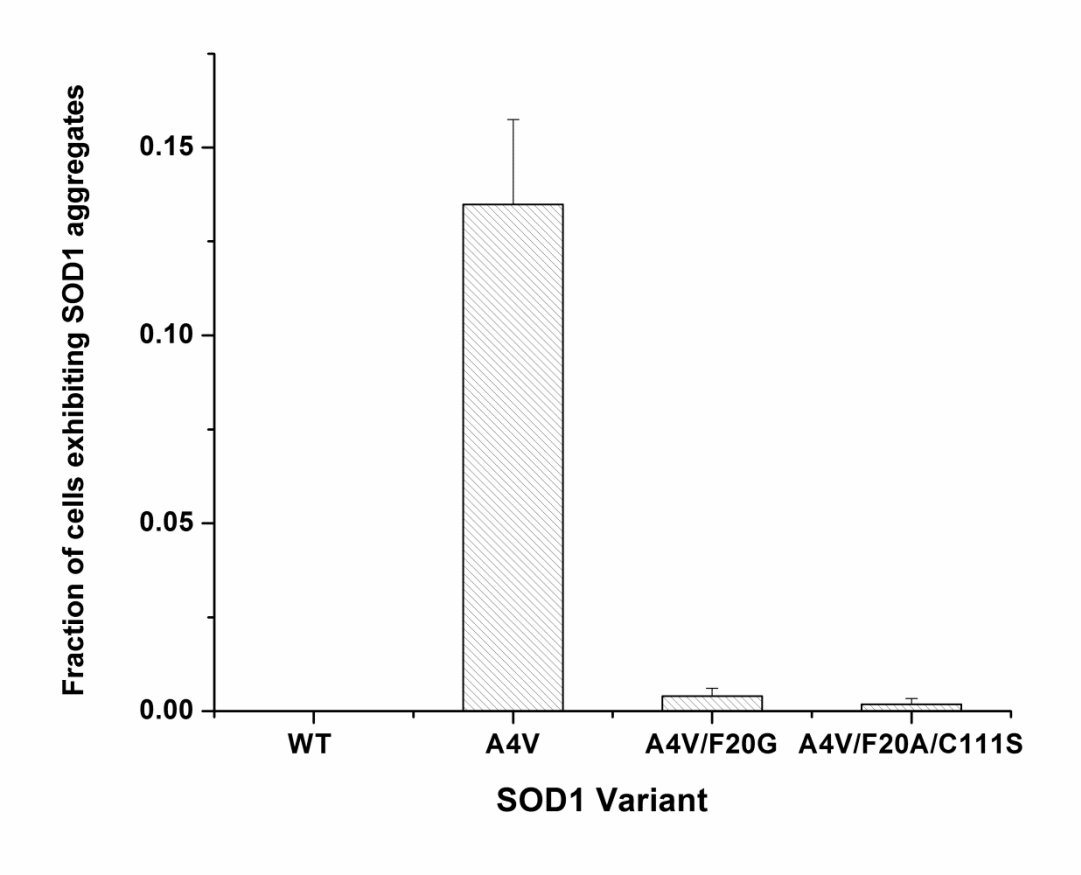


Figure 3.9: The fraction of the transfected HEK293T cells exhibiting SOD1 aggregates for the SOD1 variants ($SOD1^{WT}$, $SOD1^{A4V}$, $SOD1^{A4V/F20G}$, and $SOD1^{A4V/F20A/C111S}$).

Although the cells expressing the $SOD1^{A4V/F20G/C111S}$ also exhibited a decrease in intracellular aggregate formation (*A4V/F20G/C111S* Panel in Figure 3.6), no significant increase in mean cellular fluorescence (Figure 3.6 A and B) suggests that the $SOD1^{A4V/F20A/C111S}$ destabilization was not significantly eliminated.

Solubility assay of the novel $SOD1^{A4V}$variants to estimate their relative aggregation propensities

The evaluation of the SOD1 double or triple mutants led to the discovery of three SOD1 variants ($SOD1^{A4V/F20G}$, $SOD1^{A4V/F20A}$, and $SOD1^{A4V/F20A/C111S}$) with reduced aggregation inside the cells. It was previously reported that the aggregation propensity of the SOD1 variants correlates to their solubility inside mammalian cells (Brotherton et al. 2012; Karch et al. 2009; Prudencio and Borchelt 2011; Prudencio et al. 2010; Prudencio et al. 2009; Prudencio et al. 2012). Therefore, complementary to cellular fluorescence measurement and fluorescence microscopic analysis, we also performed cell fractionation analysis in order to estimate the relative solubility and eventually the relative aggregation propensity of the SOD1 variants. Five SOD1 variants ($SOD1^{WT}$, $SOD1^{A4V}$, $SOD1^{A4V/C111S}$, $SOD1^{A4V/F20G}$, $SOD1^{A4V/F20A}$, and $SOD1^{A4V/F20A/C111S}$) were used for the solubility assay. From the transfected HEK293T cells expressing the SOD1 variants, the soluble and insoluble fractions of proteins were extracted and quantified by comparing their western-blot band intensities in triplicate (Figure 3.10A).The aggregation propensity of each SOD1 variant is represented as the ratio of the insoluble amount and the soluble amount of the SOD1 variant in the cell lysate (Figure 3.10B). The band intensity of the $SOD1^{A4V}$ was used as an internal standard in each western-blot image to adjust the band intensities of the other SOD1 variants obtained from three independent western-blot images.

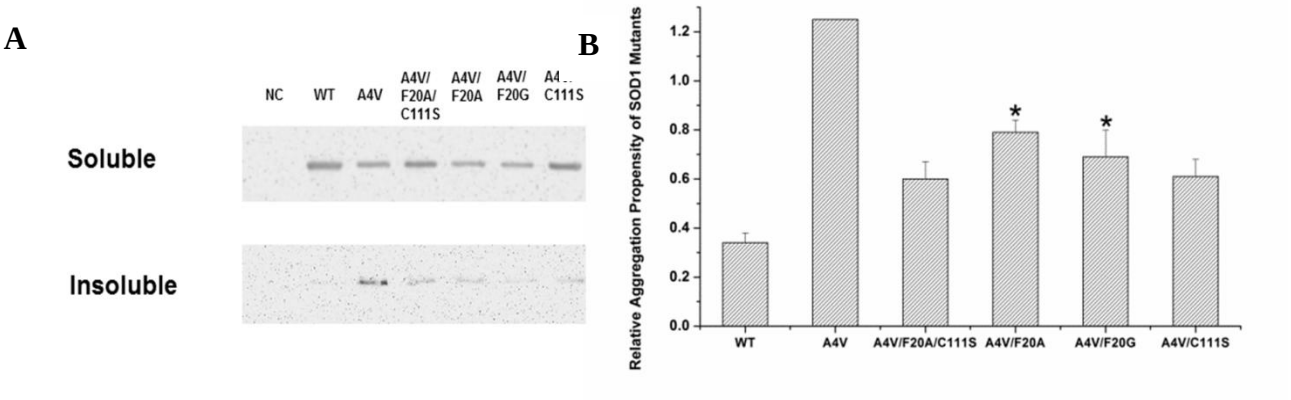


Figure 3.10: (A) The western-blot bands of the insoluble and soluble fractions of the transfected HEK293T cells expressing the SOD1 variants. (B) The relative aggregation propensity of SOD1 proteins at 48 hrs. NC: untransfected cells. Values = mean ± standard deviation (n = 3). In order to determine whether the SOD1 variant aggregation propensity is smaller than that of $SOD1^{A4V}$, two-sided Student's t-tests were applied to the data (*; $p < 0.05$).

The three novel SOD1 variants as well as the $SOD1^{A4V/C111S}$ variant showed a significantly reduced aggregation propensity compared to that of the $SOD1^{A4V}$ (Figure 3.10B). Considering that the aggregation propensities of the $SOD1^{A4V/F20A/C111S}$ and $SOD1^{A4V/C111S}$ are lowest among the $SOD1^{A4V}$ variants, the C111S mutation is thought to be a major contributor to the reduced aggregation propensity.

Denaturation assay of the novel SOD1 variants to estimate their relative stabilities

Although the aforementioned solubility assays provide substantial insight on the aggregation properties of the novel SOD1 variants, it is very difficult to directly measure the conformational stability of the protein in cultured cells. Therefore, *in vitro* denaturation assay using the purified SOD1 variants was performed to compare stabilities of the SOD1 variants. The three novel SOD1 variants ($SOD1^{A4V/F20A}$, $SOD1^{A4V/F20G}$ and $SOD1^{A4V/F20A/C111S}$) as well as the $SOD1^{WT}$and $SOD1^{A4V}$ were expressed in yeast and purified as described in the Materials and Methods section. The fraction unfolded of the five SOD1 variants at varying guanidine hydrochloride (GdnHCl) concentrations was plotted (Figure 3.11). Samples were done in duplicates, except for $SOD1^{A4V/F20G}$ and $SOD1^{A4V/F20A/C111S}$. Due to the length of time it takes to make starting material, we could not perform a duplicate experiment for these two samples. The transition range of the $SOD1^{A4V}$ shifted to lower GdnHCl concentration compared to that of $SOD1^{WT}$ indicating that the $SOD1^{A4V}$ is less stable than the $SOD1^{WT}$. The transition range of the novel three SOD1 variants ($SOD1^{A4V/F20A}$, $SOD1^{A4V/F20G}$ and $SOD1^{A4V/F20A/C111S}$) were very similar and fall in between that of the $SOD1^{WT}$ and $SOD1^{A4V}$ supporting the idea that the three novel SOD1 variants are more stable than the $SOD1^{A4V}$ but less than $SOD1^{WT}$.

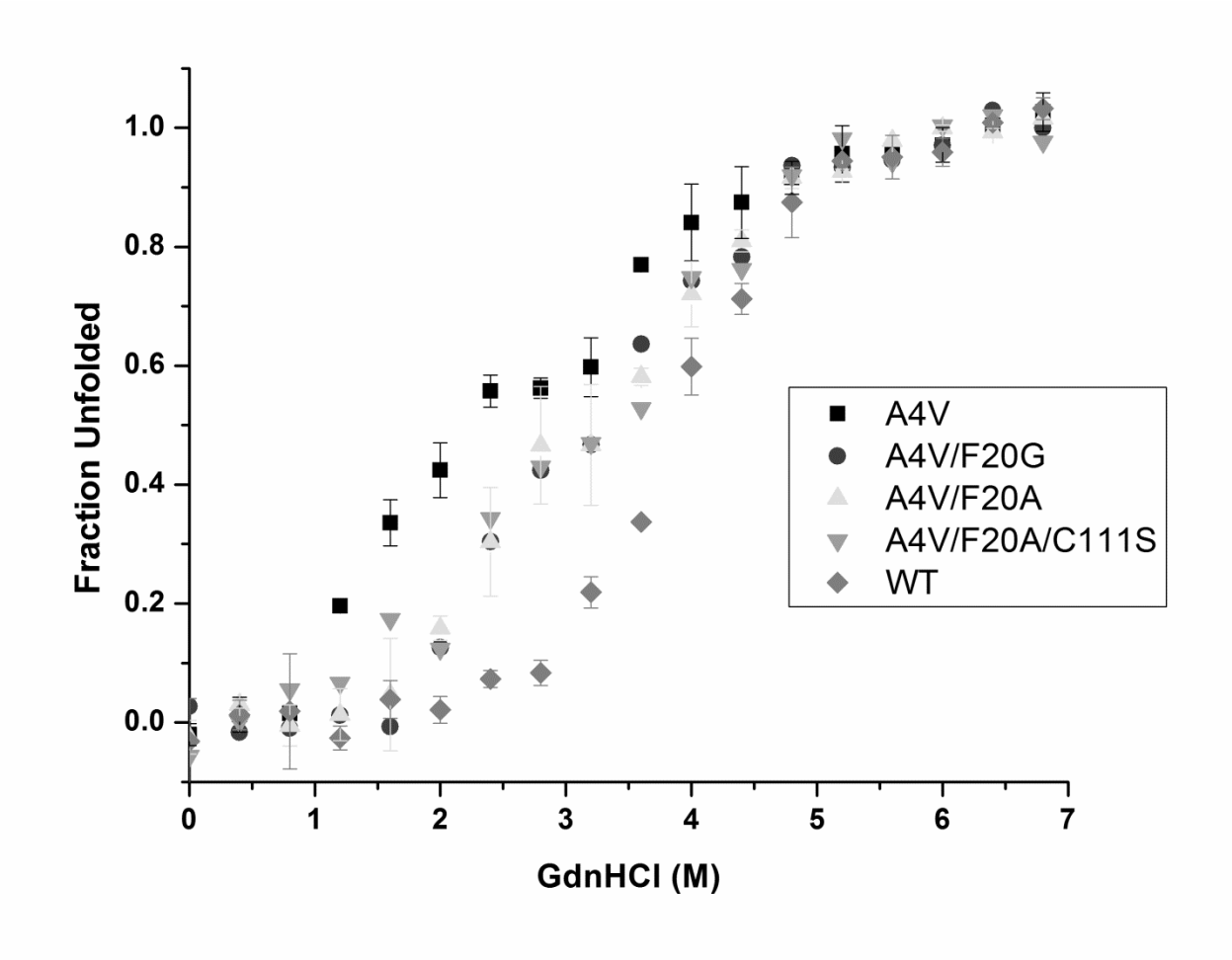


Figure 3.11: The denaturation curves of the $SOD1^{WT}$, $SOD1^{A4V}$, $SOD1^{A4V/F20G}$, $SOD1^{A4V/F20A}$, and $SOD1^{A4V/F20A/C111S}$.

According to the denaturation assay results, the C111S mutation in the $SOD1^{A4V/F20A/C111S}$ has little effect on the stability.

Combined with the solubility assay results, the denaturation assay results show that the novel three SOD1 variants ($SOD1^{A4V/F20A}$, $SOD1^{A4V/F20G}$ and $SOD1^{A4V/F20A/C111S}$) have lower aggregation propensity and are more stable compared to the $SOD1^{A4V}$.

CONCLUSIONS

In this paper, we have discovered that F20 is a key residue responsible for the $SOD1^{A4V}$ conformational destabilization/aggregation using computational protein design via RosettaDesign, which was supported by characterization results of $SOD1^{A4V}$ double and triple mutants *in vitro* and in cultured cells. In order to suppress the conformationally destabilizing effects of A4V mutation, we utilized a "cis-suppression" strategy and generated three novel $SOD1^{A4V}$ variants ($SOD1^{A4V/F20G}$, $SOD1^{A4V/F20A}$, and $SOD1^{A4V/F20A/C111S}$) with a significantly reduced aggregation propensity in mammalian cells and an increased conformational stability *in vitro*. To our knowledge, this is the first time that the aggregation of $SOD1^{A4V}$ has been arrested by using an additional mutation(s) to eliminate the destabilizing effects of $SOD1^{A4V}$. Therefore, we successfully demonstrated that protein aggregation in mammalian cells caused by a destabilizing mutation can be suppressed using a streamlined approach consisting of four steps: 1) identifying a key residue responsible for the protein destabilization using computational protein design, 2) using that information to prepare protein variants containing candidate mutations that improve the conformational stability, 3) selecting protein variants with reduced aggregation using the cell-based assays, and 4) verifying aggregation reduction and stability enhancement of the protein variants selected using cell-based and *in vitro* assays. Although

mutations introduced suppress the aggregation of a target protein, the mutations may affect protein activity (Rothlisberger et al. 2008). As far as we know, it is still very difficult to predict catalytic property changes once a mutation is introduced into enzyme. In the future, when more advanced computational tools are developed, it is desirable to consider catalytic property changes as well as conformational stability change in protein design. Furthermore, in order to avoid catalytic property loss, another activity assay needs to be combined with the cell-based aggregation assay described here. These are important topics that require more extensive studies. Considering the generality of the approach and techniques that were employed in this paper, we believe they can be applied to suppress the aggregation of other proteins containing a conformationally destabilizing mutation. Furthermore, the discovery of more stable variants than the $SOD1^{A4V}$ provides promise in understanding the molecular mechanism of SOD1-mediated fALS.

Chapter 4: Suppressing the protein aggregation expressed in mammalian cells by increasing kinetic stability

ABSTRACT

Protein aggregation within mammalian cells has significant implications in the biologics production and disease pathology making the modulation of protein aggregation within mammalian cells a very important engineering topic. Previously, we showed that the semi-rational design approach can be used to reduce the intracellular aggregation of a protein by enhancing the conformational stability. However, this approach has the limited utility when no rational design approach to enhance conformational stability is readily available. In order to overcome this limitation, we investigated modulating protein aggregation by enhancing kinetic stability. As a model system, human copper, zinc superoxide dismutase mutant containing glycine to alanine mutation at position 93 ($SOD1^{G93A}$) was used. The computational stability analysis and crystal structure examination revealed that the $SOD1^{G93A}$ is destabilized by the repulsive interaction between a side chain and protein backbone. A panel of conservative mutations was introduced into residues substantially displaced upon the G93A mutation. By using cell-based aggregation assays, we identified several novel variants of $SOD1^{G93A}$ with an increased kinetic stability and reduced aggregation propensity within mammalian cells. Our findings successfully demonstrate that the protein aggregation within mammalian cells can be suppressed using the combination of rational design and screening strategies to increase the protein kinetic stability.

INTRODUCTION

Protein misfolding and aggregation have been implicated in many human diseases (Kaytor and Warren 1999). Therefore, suppressing protein aggregation is considered as an effective strategy to treat numerous protein aggregation-associated diseases (Bartolini and Andrisano 2010; Colby et al. 2006; Keshet et al. 2010; Lee et al. 2007; Murphy 2002; Valentine and Hart 2003). Furthermore, many biologics approved for human use have been produced using mammalian cells, such as Chinese hamster ovary cells. From 2006 to 2010, around 58 biologics gained FDA approval, and 32 of those were produced solely in mammalian cells, intimating the increased importance of understanding protein production in mammalian cells (Walsh 2010). In particular, non-native aggregation of proteins presents a significant challenge at all stages of biologics production (Bondos and Bicknell 2003; Chi et al. 2003; Gregoire et al. 2013; Rodriguez et al. 2005; Wang et al. 2006). Besides monoclonal antibody aggregation in the formulation step, the aggregation issue should be resolved in the early stages of new antibody development. Wild-type antibodies are subjected to extensive engineering in order to improve their binding affinity and/or reduce their side effects. In the course of such antibody engineering, it is possible to identify antibody mutants with improved therapeutic properties, but reduced stability due to the mutation(s) introduced. Although strategies were developed to modulate protein aggregation *in vitro* or in bacterial cells (Barthelemy et al. 2008; Beerten et al. 2012; Dudgeon et al. 2009), it is not straightforward to apply them to mammalian cells due to differences in protein synthesis and folding machinery (Hartl and Hayer-Hartl 2009). Therefore, the screening of protein variants in mammalian cells is beneficial to identify protein variants with a reduced aggregation propensity in mammalian cells.

Recently it was demonstrated that soluble intrabodies can be obtained by screening the intrabody libraries in mammalian cells (Guglielmi et al. 2011). Complementary to the high-throughput screening, we utilized the semi-rational design approach to arrest the aggregation of a human copper, zinc superoxide dismutase mutant containing alanine to valine mutation at position 4 ($SOD1^{A4V}$) inside mammalian cells by enhancing the thermodynamic conformational stability (Gregoire et al. 2013). Although wild-type SOD1 ($SOD1^{WT}$) is very stable and does not form intracellular aggregates, the A4V mutation greatly increases the aggregation-propensity. In our previous studies, the mutation of a bulky side chain of phenylalanine at position 20 (F20), in direct contact with aside chain of valine at position 4 (V4), into smaller ones successfully eliminated the steric hindrance between the side chains of F20 and V4 resulting in the enhanced thermodynamic conformational stability and reduced protein aggregation (Gregoire *et al.*, 2013). This approach is applicable in cases where protein destabilization is caused by steric hindrance between side chains, but not to cases where the steric hindrance between a side chain of mutated residue and backbone protein induces protein destabilization. To our knowledge, no rational design approach to redesign protein backbones to eliminate such destabilizing effects has been reported. Therefore, alternative strategies to reduce protein aggregation are required.

Even though the conformational stability of a folded protein is lower than that of its aggregates, it may not form aggregates over a certain period of time due to a high energy barrier for aggregate formation (kinetic stability) (Baker and Agard 1994; Manning and Colon 2004; Sanchez-Ruiz 2010). As evidenced with protein deposition diseases, the process of aggregate formation is a higher-order kinetic process and can be controlled by kinetic stability (Cohen and Kelly 2003). For example, a mutant transthyretin (TTR) found in people affected by familial amyloid polyneuropathy was kinetically stabilized by another TTR mutant found in nature by

forming extracellular heterodimers with a reduced aggregation rate (Cohen and Kelly 2003; Hammarström et al. 2001). Taking advantage of the concept of kinetic stabilization to suppress protein aggregation found in nature, we hypothesize that the aggregation of a target protein within mammalian cells can be reduced using a protein engineering strategy increasing kinetic stability. The combination of rational design and cell-based aggregation assay can be used to identify such a mutation(s) to increase kinetic stability. As a model protein, we chose a SOD1 variant containing glycine to alanine mutation at position 93 ($SOD1^{G93A}$) associated with the familial form of amyotrophic lateral sclerosis (fALS), a fatal muscular neurodegenerative disease. In the crystal structure of $SOD1^{G93A}$ (PDB ID:2KWO), a side chain of the mutated residue (Ala93) does not contact with any side chain of neighboring residues but contacts with neighboring backbones (Galaleldeen et al. 2009). Furthermore, the high aggregation propensity of $SOD1^{G93A}$ inside mammalian cells makes it an attractive candidate for our studies on suppressing intracellular aggregation (Galaleldeen et al. 2009; Karch et al. 2009; Prudencio et al. 2009).

MATERIAL AND METHODS

Materials. The HRP-conjugated anti-rabbit antibody was obtained from Invitrogen (Carlsbad, CA). The anti-SOD1antibody was obtained from Santa Cruz Biotechnology, Inc. (Santa Cruz, CA). HEK293T and NSC-34 cells were obtained from Invitrogen and CELLutions Biosystems (Burlington, Ontario, CA), respectively. Primers used for construction of expression vectors were obtained from Invitrogen. Gibco Certified Fetal Bovine Serum (FBS) was obtained from Invitrogen. All other chemicals were purchased from Sigma-Aldrich Corporation (St. Louis, MO).

Construction of Expression Vectors. The pEGFP-N3-SOD1^{WT} plasmid was kindly provided by Dr. Haining Zhu (University of Kentucky). The mutations were introduced into the pEGFP-N3-SOD1^{WT} using the QuikChange II site-directed mutagenesis kit (Stratagene). The primer pairs for each mutant are as follows: 5'-GACTGCTGACAAAGATGCTGTGGCCGATGTGTC-3' and 5' GACACATCGGCCACAGCATCTTTGTCAGCAGTC-3' for G93A; 5'-GCCTTCAGTCAGTCCCTCAATGCTTCCCCACACCTTCAC-3 and 5'-GTGAAGGTGTGGGGAAGCATTGAGGGACTGACTGAAGGC-3' for K36E; 5'-AGTGAAGGTGTGGGGAAGCATTTTAGGACTGACTGAAGG-3' and 5'-CCTTCAGTCAGTCCTAAAATGCTTCCCCACACCTTCACT-3' for K36L; 5'-GGCAATGTGACTGCTGACGCAGATGCTGTGGCCGATGTGTCTAT-3' and 5'-ATAGACACATCGGCCACAGCATCTGCGTCAGCAGTCACATTGCC-3' for K91A/G93A; 5'- GGCAATGTGACTGCTGACGATGATGCTGTGGCCGATGTGTCTAT-3' and 5'-ATAGACACATCGGCCACAGCATCATCGTCAGCAGTCACATTGCC-3' for K91D/G93A;5'-ATAGACACATCGGCCACAGCATCCTCGTCAGCAGTCACATTGCC -3' and 5' GGCAATGTGACTGCTGACGAGGATGCTGTGGCCGATGTGTCTAT -3' for K91E/G93A; 5'-GGCAATGTGACTGCTGACAGAGATGCTGTGGCCGAT -3' and 5' ATCGGCCACAGCATCTCTGTCAGCAGTCACATTGCC -3' for K91R/G93A; 5'-CATTAAAGGACTGACTGCAGGCCTGCATGGATTCC-3 and 5'-GGAATCCATGCAGGCCTGCAGTCAGTCCTTTAATG-3' for E40A; 5'-GAATCCATGCAGGCCTTTAGTCAGTCCTTTAATGC-3' and 5'-GCATTAAAGGACTGACTAAAGGCCTGCATGGATTC-3' for E40K; 5'-GGAAGCATTAAAGGACTGACTGATGGCCTGCATGGA -3' and 5'-TCCATGCAGGCCATCAGTCAGTCCTTTAATGCTTCC -3' for E40D;5'-

CCATGCAGGCCTCTAGTCAGTCCTTTAATGCTTCCCC -3' and 5'-GGGGAAGCATTAAAGGACTGACTAGAGGCCTGCATGG-3' for E40R. The sequences of the mutations were confirmed using DNA sequencing.

Computational Analysis of Protein Structure. The computational analysis of SOD1 conformational stability was performed using the molecular modeling program RosettaDesign version 3.4 (Rohl et al. 2004). The core energy function of RosettaDesign is a linear sum of a 6-12 Lennard-Jones potential, the Lazaridis-Karplus implicit solvation model, an empirical hydrogen bonding potential, backbone-dependent rotamer probabilities, a knowledge-based electrostatics energy potential, amino acid probabilities based on particular regions of φ/ψ space and a unique reference energy for each amino acid (Rohl et al. 2004). For a given crystal structure, RosettaDesign uses simulated annealing to scan through a large number of rotamers to minimize the energy score with Monte Carlo optimization (Kuhlman and Baker 2000).

A fixed backbone protein design protocol was used in this study to identify the key interactions responsible for $SOD1^{G93A}$ destabilization. The holo-form of $SOD1^{WT}$crystal structure (PDB ID: 2C9V) was used (DiDonato et al. 2003; Strange et al. 2006). A homodimeric holo-form of $SOD1^{G93A\_WT}$ structure was constructed from the $SOD1^{WT}$ structure by repacking all protein side chains except those chelated to metal ions (H46, H63, H71, H80, D83 and H120) to adopt favorable conformations followed by the insertion of the G93A mutation and repeat of side chain repacking. Using the $SOD1^{G93A\_WT}$ structure generated from the $SOD1^{WT}$ structure, folding energy scores of $SOD1^{G93A\_WT}$ and its variants were determined using RosettaDesign.

Homology Modeling and Crystal Structure Analysis The crystal structures of holo-form $SOD1^{WT}$ and $SOD1^{G93A}$ (PDB ID: 2C9V and 2WKO, respectively) (Galaleldeen et al. 2009; Strange et al. 2006) were visualized using PyMol (Schrodinger 2010).

Transfection of HEK293T and NSC-34 Cells. Transfection of HEK293T and NSC-34 cells was performed as previously reported (Gregoire et al. 2013). HEK293T cells were maintained at 37°C and 5% CO_2 in Dulbecco's modified Eagle's medium/High Glucose (DMEM/High Glucose, Thermo Scientific, Pittsburgh, PA) supplemented with 10% FBS, 100 µg/mL of streptomycin sulfate and 100 units/mL of penicillin. NSC-34 cells were maintained in DMEM supplemented with 10% FBS, 4500 mg/L glucose, 0.584 g/L L-glutamine, 3.7 g/L sodium bicarbonate, 100 µg/mL of streptomycin sulfate and 100 units/ml of penicillin. Cells were seeded on 6-well plates one day prior to transfection. Once the cells grew to 80-90% confluency, they were transfected with 3.5 µg of the appropriate plasmid via the calcium phosphate precipitation method. All samples were done in triplicate, unless explicitly stated.

Fluorescence Microscopy. Fluorescence microscopy was performed as previously reported (Gregoire et al. 2013). Two days post-transfection, HEK293T or NSC-34 cells were visualized via fluorescence microscopy using a VistaVision Inverted Fluorescence Microscope (VWR, Radnor, PA). Images were captured using a DC-2C digital camera. The fluorescence excitation wavelength range was between 420 and 485 nm and the emission wavelength was 515 nm. The number of cells in the fluorescence microscopy images was manually counted. The percentage of transfected cells exhibiting SOD1-EGFP aggregates was determined by dividing the number

of aggregate-containing cells by the number of total cells analyzed. Only images containing greater than 50 cells were used. Three images per well of transfected cells were used.

Flow Cytometric Analysis of Cellular Fluorescence. Flow cytometric analysis of cellular fluorescence was performed as previously reported (Gregoire and Kwon 2012; Gregoire et al. 2013). Two days post-transfection, the transfected HEK293T and NSC-34 cells were trypsinized, washed twice with 1X PBS and resuspended in 500µL of 1X PBS. The fluorescence intensities of the HEK293T and NSC-34 cells expressing SOD1 variant-EGFP fusion protein were measured using the C6 flow cytometer (BD Biosciences, San Jose, California). The excitation wavelength was 488 nm and the fluorescence emission was detected at 585 nm. Only GFP positive cells were used to calculate the mean cellular fluorescence. Transfection efficiency was determined by dividing the number of fluorescence positive cells by the number of total cells analyzed. Each sample was prepared in triplicate and cellular fluorescence indicates mean cellular fluorescence, unless otherwise noted. Two-sided, Student's paired t-test was used for statistical analysis of fluorescence data.

Total Protein Extraction and Protein Fractionation of HEK293T cells. Total protein extraction and protein fractionation of HEK293T cells were performed as previously reported (Gregoire et al. 2013). $2x10^6$ transfected HEK293T cells were centrifuged to obtain a cell pellet. The cell pellet was washed once with PBS buffer and lysed using 100 µL of RIPA buffer (Thermo Fisher, Pittsburgh, PA). To remove the detergent soluble fraction of proteins, the cells were incubated at 4 °C for 10 minutes. The samples were centrifuged at 4 °C and 15,500 g for 15 minutes. The supernatant was collected as a soluble fraction and kept on ice. The pellet was

collected as an insoluble fraction and was then resuspended in 100 µL of RIPA buffer at 4 °C for 10 minutes. The samples were centrifuged at 4 °C and 15,500 g. The supernatant was discarded and the procedure was repeated once more. To extract the detergent-insoluble fraction of proteins, the pellet was incubated at room temperature in 100 µL of RIPA Buffer/8M Urea overnight. The sample was centrifuged at 15,500 g for 15 minutes and the supernatant was collected. For total protein lysate extraction, cell pellets were resuspended in 100 µL of RIPA Buffer/8M Urea and incubated at room temperature overnight. Protein concentrations of the soluble fraction and total protein lysate were determined using BCA kit (Pierce, Rockford, IL) using bovine serum albumin as a standard.

Western Blotting. Western Blotting was performed as previously reported with minor adjustments (Gregoire and Kwon 2012; Gregoire et al. 2013). Soluble fractions were diluted using RIPA and 8M Urea to 1 µg/µL after the protein concentration was determined. The final sample buffer composition of the soluble fraction was 1:3 vol/vol (RIPA:8M Urea). The insoluble fractions were diluted using the same amount of volume of buffer used to dilute the soluble fraction for that particular mutant. For total SOD1 determination, the final concentration for the lysate was adjusted to account for varying transfection efficiencies among the variants. 3 µL of total protein, soluble or insoluble protein lysate were boiled in sample loading buffer containing DTT and electrophoresed on 12% SDS-PAGE Gel. Samples were transferred onto a nitrocellulose membrane for one hour and washed with 1X TBS-T three times for 5 minutes. The membrane probed with the primary anti-SOD1 antibody diluted in TBS-T for 1 hour at RT (1:5000). Samples were washed three times for 5 minutes in TBS-T and probed with the

secondary antibody conjugated with HRP for 1 hour (1:10000). Bands were detected using ECL Prime (GE Healthcare, Sweden).

RESULTS AND DISCUSSIONS

Determination of key factors responsible for the destabilization of $SOD1^{G93A}$ using *in silico* techniques

It is well known that $SOD1^{G93A}$ forms intracellular aggregates in mammalian cells likely due to the reduced conformational stability caused by glycine to alanine mutation at position 93 (G93A) (Galaleldeen et al. 2009; Prudencio et al. 2009). As the first step to reduce $SOD1^{G93A}$ aggregation, we investigated key factors responsible for $SOD1^{G93A}$ destabilization. In order to evaluate effects of G93A mutation on the conformational stability, we constructed a structural model of $SOD1^{G93A}$ ($SOD1^{G93A\_WT}$) by introducing G93A mutation into $SOD1^{WT}$ structure. Using RosettaDesign, we calculated the conformational stability (folding energy score; $\Delta\Delta G_f$) of $SOD1^{WT}$and $SOD1^{G93A\_WT}$ (Table 4.1). The folding energy score of $SOD1^{G93A\_WT}$ relative to that of $SOD1^{WT}$ substantially increased, indicating that $SOD1^{G93A\_WT}$ is conformationally less stable than $SOD1^{WT}$. Then, we investigated which score plays a key role to increase the folding energy score among twenty individual scores, summed for the final folding energy score. It is noteworthy that the G93A mutation substantially increases fa_rep score (about 440 kcal/mol) comparable to the increase in the final folding energy score (Table 4.1). Since fa_rep score is a measure of repulsive portion of the Lennard-Jones Potential, such an increase in fa_rep score indicates certain atoms are closer than their optimal distance leading to the $SOD1^{G93A}$ destabilization. Complementary to the computational stability analysis, we also performed visual inspection of the crystal structures of $SOD1^{WT}$ (PDB ID:2C9V) and $SOD1^{G93A}$ (PDB ID:2KWO).

Table 4.1: Scoring summary of $SOD1^{WT}$ and $SOD1^{G93A\_WT}$ variants

SOD1 variant	**$\Delta\Delta G_f$ (kcal/mol)**	**fa_rep (kcal/mol)**
$SOD1^{WT}$	-541	114
$SOD1^{G93A\_WT}$	-122	511
$SOD1^{G93A/K36E\_WT}$	-132	511
$SOD1^{G93A/K91D\_WT}$	-127	510
$SOD1^{G93A/K91E\_WT}$	-128	511

Although G93 in $SOD1^{WT}$ does not have any close contact with a neighboring protein backbone, a methyl group in the side chain of A93 in $SOD1^{G93A}$ comes in close contact with the neighboring protein backbone (Figure 4.1 A and B). Therefore, the combination of computational stability analysis and visual inspection of protein crystal structure strongly support the fact that the repulsive interaction between the side chain of A93 and the protein backbone is responsible for the $SOD1^{G93A}$ destabilization.

Combination of rational design and cell-based screening in order to identify to identify $SOD1^{G93A}$ variants with a reduced aggregation propensity

To our knowledge, no rational design strategy to redesign a protein backbone in order to eliminate repulsive interaction between a side chain and a protein backbone has been reported. Therefore, instead of enhancing protein conformational stability, we sought an alternative way to reduce $SOD1^{G93A}$ aggregation by increasing kinetic stability. When the protein backbones of $SOD1^{WT}$ and $SOD1^{G93A}$ are overlaid, there are displacements at multiple sites, including the $SOD1^{G93A}$ backbone shift from the $SOD1^{WT}$ backbone (Figure 4.1A) and side chain displacements (Figure 4.2A to F) likely due to the repulsive interaction between the side chain of A93 and a protein backbone (Figure 4.1B). Therefore, we hypothesize that these side chain displacements facilitate $SOD1^{G93A}$ aggregation, and so mutation of the amino acids substantially displaced upon the G93A mutation will reduce protein aggregation. We focused on amino acid side chains substantially displaced upon G93A within a 5Å distance from the mutated A93. By examining the crystal structures of both $SOD1^{WT}$ and $SOD1^{G93A}$ (Figure 4.1C and D), there are three residues that fall under these specific criteria, K36, E40 and K91 (Figure 4.2A to F). It is noteworthy that no side chain of these residues directly contacts the side chain of A93. In order

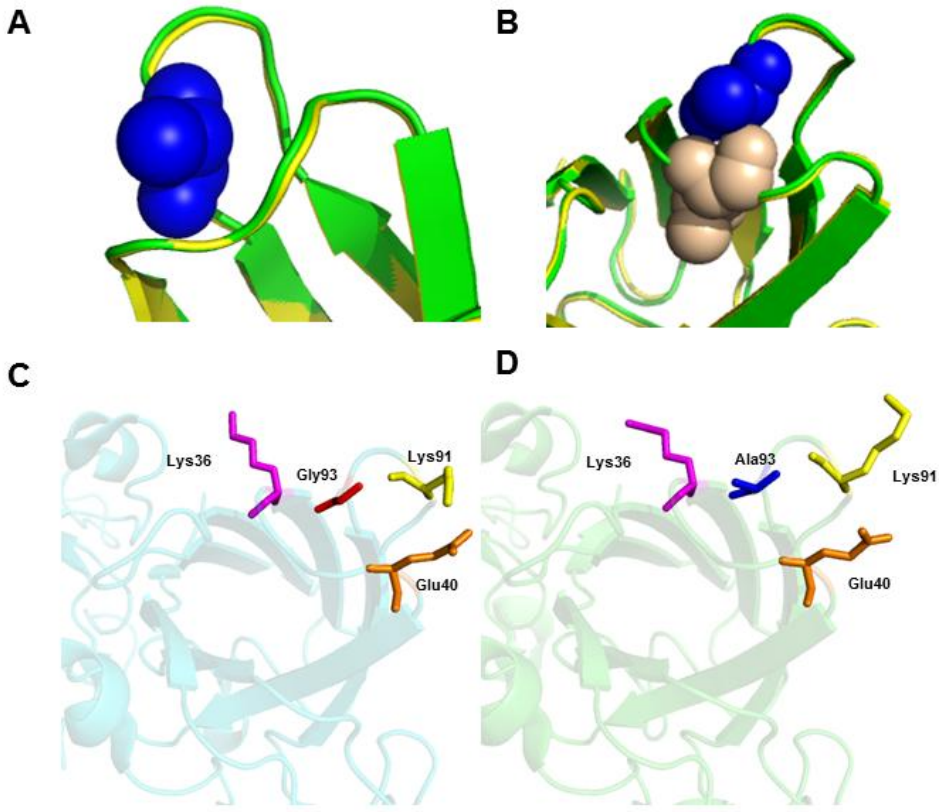


Figure 4.1: Structural analysis of SOD1^{WT} (PDB ID: 2C9V) and SOD1^{G93A} (PDB ID:2WKO). (**A**) Ala93 (Blue) mutation causes shift in backbone of SOD1^{G93A} in comparison to SOD1^{WT} (SOD1^{WT} backbone in yellow; SOD1^{G93A} backbone in green) (**B**) Ala93 (blue) shifts the amino acid back bone of Leu38 (gray) (**C**) The crystal structure (PDB ID: 2C9V) of the holo-form of SOD1^{WT}. Residues of interest within 5 angstroms were highlighted as specified. (**D**) The crystal structure (PDB ID: 2WKO) of the holo-form of SOD1^{G93A}. Residues of interest within 5 angstroms were highlighted as specified.

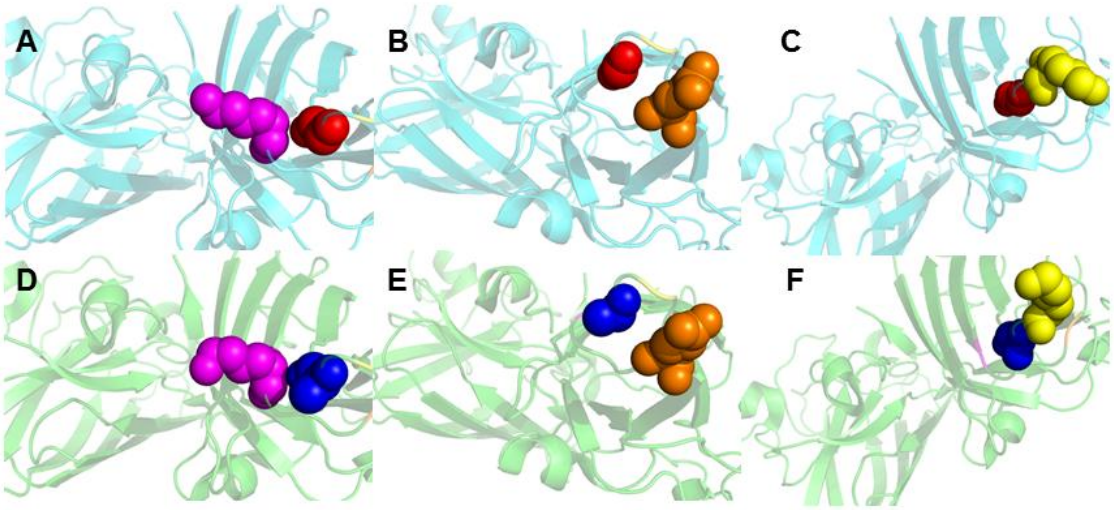


Figure 4.2: Residues of SOD1^{G93A} substantially displaced compared to those of SOD1^{WT}. Residues K36 (*purple*) and G93 (*red*) (**A**); E40 (*orange*) and G93 (*red*) (**B**); and K91 (*yellow*) and G93 (*red*) (**C**) within crystal structure of the holo-form of SOD1^{WT} (PDB ID: 2C9V). Residues K36 (*purple*) and A93 (*blue*) (**D**); E40 (*orange*) and A93 (*blue*) (**E**); and K91 (*yellow*) and A93 (*blue*) (**F**) within crystal structure of the holo-form of SOD1^{G93A} (PDB ID: 2WKO).

to avoid the perturbation of local polarity, we generated seven $SOD1^{G93A}$ variants containing a mutation of each of three amino acids (K36, E40, and K91) into different charged amino acids (K36E; E40R, E40K, and E40D; K91E, K91D, and K91R). One additional mutation from a charged amino acids into a bulky, hydrophobic one (K36L) was also as a non-conservative mutation.

In order to identify $SOD1^{G93A}$ variants with a reduced aggregation propensity, we utilized the mammalian cell-based aggregation assay correlating misfolding/aggregation propensity to the extent of reduction in cellular fluorescence (Gregoire and Kwon 2012; Gregoire et al. 2013). Briefly, an EGFP was fused to C-terminus of $SOD1^{WT}$, $SOD1^{G93A}$, or each of $SOD1^{G93A}$ variants containing the aforementioned additional mutation. Both cell lines (HEK293T and NSC-34) were used to express an EGFP fusion of $SOD1^{variant}$. HEK293T is a cell line used extensively in cell biology experiments and biologics production, and NSC-34 is a mouse neuroblastoma commonly used for fALS research (Babetto et al. 2005; Cashman et al. 1992; Kupershmidt et al. 2009; Raimondi et al. 2006; Rizzardini et al. 2003; Tartari et al. 2009). Both HEK293T and NSC-34 cells were transfected and harvested two days post-transfection for flow cytometric analysis. Only three $SOD1^{G93A}$ variants ($SOD1^{G93A/K36E}$, $SOD1^{G93A/K91E}$ and $SOD1^{G93A/K91D}$) exhibited a significant increase in the cellular fluorescence in both cell lines ($P<0.01$ or $P < 0.05$) (Figure 4.3 A and B). As demonstrated previously, such an increase in the cellular fluorescence is very likely attributed to a significant reduction in the aggregation propensity (Gregoire and Kwon 2012; Gregoire et al. 2013; Guglielmi et al. 2011). Although absolute cellular fluorescence is a key indicator of whether the target protein has exhibited a significant increase in aggregation propensity, total protein expression must also be taken into account to dismiss any notion that the cellular fluorescence increase resulted from an increase in protein expression. To determine

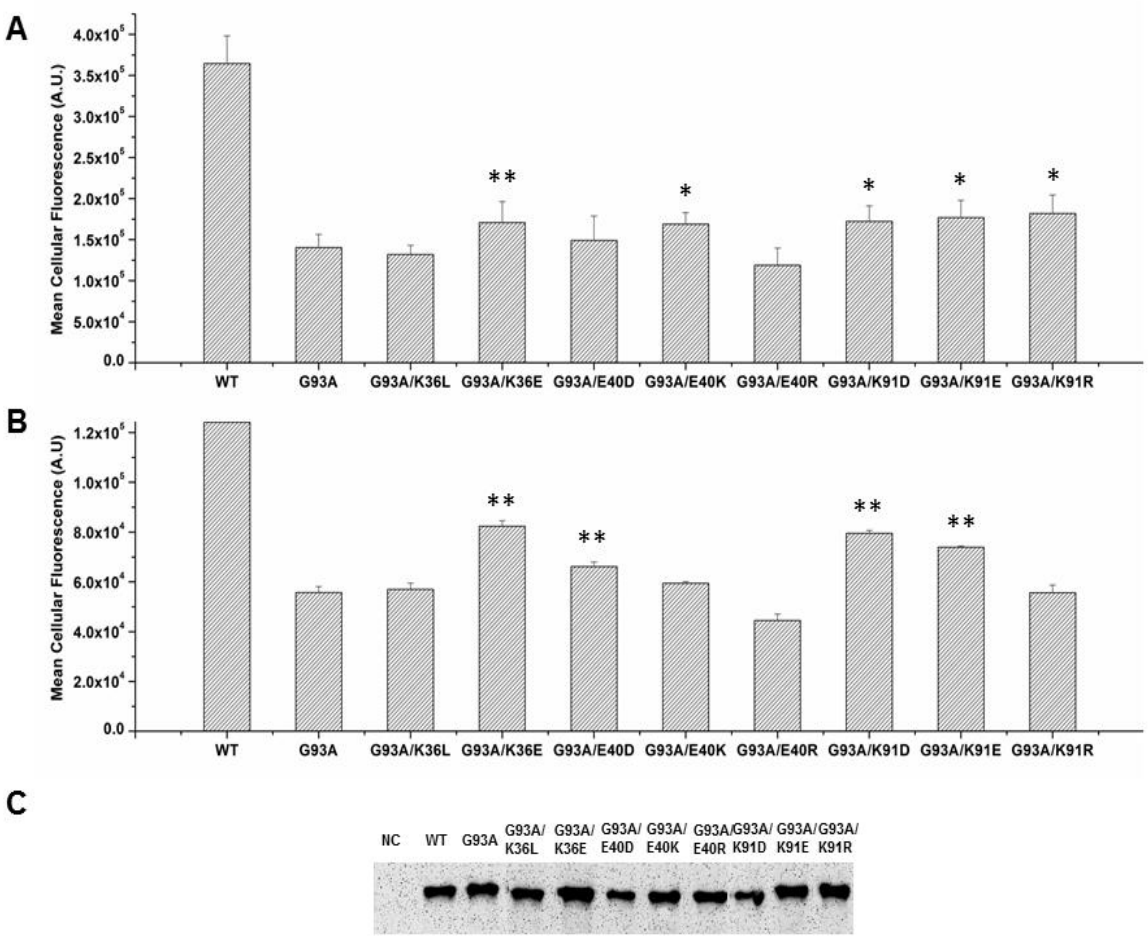


Figure 4.3: The mean cellular fluorescence of the transfected HEK293T (**A**) and NSC-34 (**B**) cells expressing EGFP fusion of SOD1 variants. (**C**) The SOD1 bands of the transfected HEK293T cells expressing the SOD1 variants. Values and error bars represent mean cellular fluorescence and standard deviations, respectively (n=3). In order to determine whether the mean cellular fluorescence of the SOD1 double mutants is greater than that of $SOD1^{G93A}$ in each cell line, two-sided Students t-tests were applied to the data (* $P<0.05$; ** $P<0.01$).

relative total SOD1 protein expression for each $SOD1^{variant}$, the total protein lysate was extracted from HEK293T cells and analyzed using western blotting. The mean cellular fluorescence values for HEK293T were adjusted by determining the ratio of the band intensities of each variant to $SOD1^{WT}$to determine the effect of total SOD1 expression (Figure 4.4). Even when accounting for variation in total protein expression, all three variants ($SOD1^{G93A/K36E}$, $SOD1^{G93A/K91E}$ and $SOD1^{G93A/K91D}$) exhibited a substantial increase in mean fluorescence. In contrast, $SOD1^{G93A/K36L}$ did not exhibit any increase in the cellular fluorescence in both cell lines (Figure 4.3A and B) suggesting that the mutation of a charged amino acid (K36) into a bulky, hydrophobic amino acid is not effective in reducing the aggregation propensity of $SOD1^{G93A}$. The three promising $SOD1^{G93A}$ variants ($SOD1^{G93A/K36E}$, $SOD1^{G93A/K91E}$ and $SOD1^{G93A/K91D}$) and several other variants were further subjected to fluorescence microscopic analysis.

Fluorescence microscopic analysis of the promising $SOD1^{G93A}$variants expressed in HEK293T cells

Fluorescence microscopy of HEK293T cells expressing an EGFP fusion of $SOD1^{variant}$at two days post-transfection was performed to confirm whether the trends observed from the fluorescence intensity measurements were due to changes in intracellular aggregation of the SOD1 variants. Similar approaches have been performed by other laboratories in order to visualize intracellular aggregation of mutant SOD1 (Gregoire et al. 2012; Gregoire et al. 2013; Prudencio and Borchelt 2011; Stevens et al. 2010; Witan et al. 2008). From visualization of the EGFP fusion of $SOD1^{variant}$ within HEK293T cells, it is apparent that $SOD1^{G93A}$ exhibits significant intracellular aggregation $SOD1^{G93A}$ (Figure 4.5 *G93A Panel*), while SOD1WT does not exhibit any intracellular aggregation (Figure 4.5 *WT Panel*). As expected, the HEK293T cells

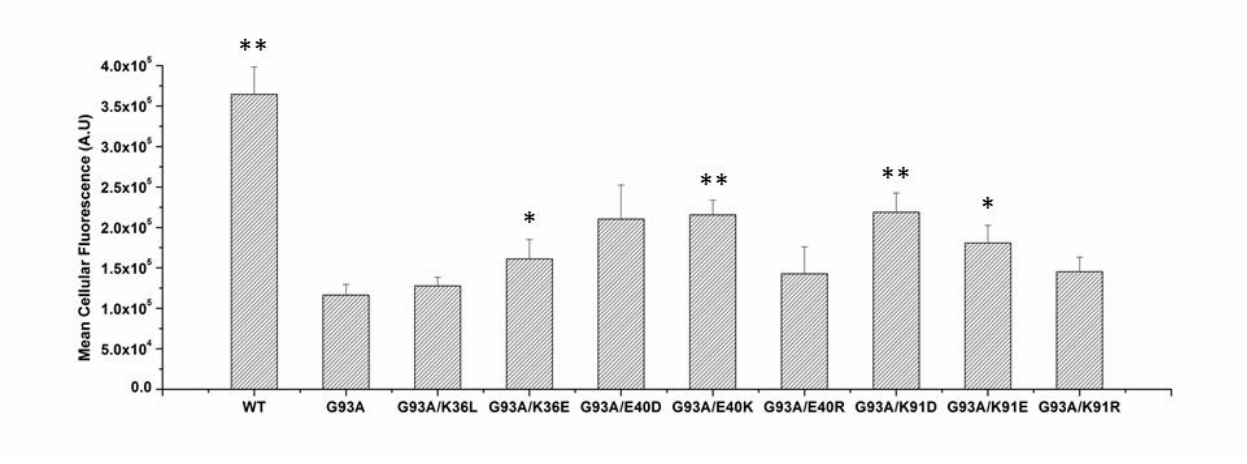


Figure 4.4: The adjusted mean cellular fluorescence of the transfected cells expressing the EGFP fusion of $SOD1^{WT}$, $SOD1^{G93A}$, and $SOD1^{G93A}$ variants containing a second mutation. The values were calculated by dividing the mean cellular fluorescence of each sample by the relative expression level to that of $SOD1^{WT}$ Values = mean ± standard deviation (n = 2). In order to determine whether the mean cellular fluorescence of the SOD1 double mutants is greater than that of $SOD1^{G93A}$, two-sided Student's t-tests were applied to the data (* $p < 0.05$; ** $p < 0.01$).

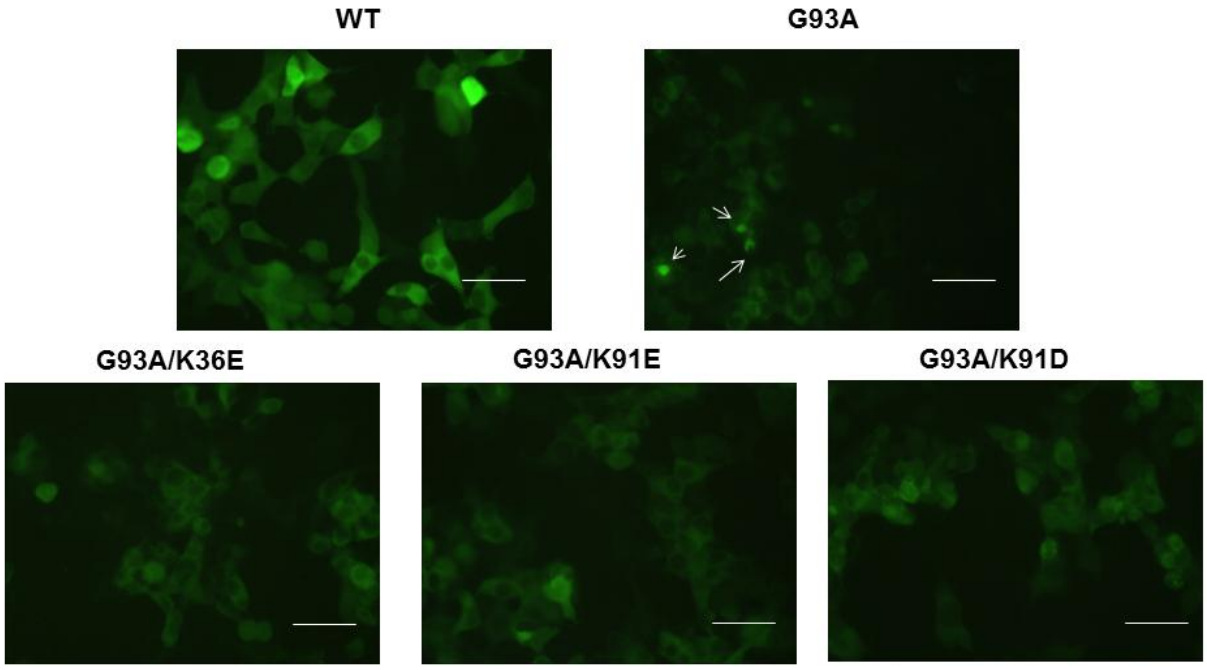


Figure 4.5: The fluorescence microscopic images of the transfected HEK293T cells expressing EGFP fusion of SOD1 variants. The images of transfected HEK293T cells expressing EGFP fusion of $SOD1^{WT}$, $SOD1^{G93A}$, and $SOD1^{G93A}$ double mutants were taken at 2 days post-transfection. (white arrows: intracellular SOD1 mutant aggregates; scale bar: 50 µm)

expressing each of three promising $SOD1^{G93A}$ variants ($SOD1^{G93A/K36E}$, $SOD1^{G93A/K91D}$, and $SOD1^{G93A/K91E}$) exhibited substantially reduced levels of intracellular aggregation in comparison to $SOD1^{G93A}$ (Fig. 4.5 *G93A/K36E, G93A/K91E, and G93A/K91D Panels*), which is consistent with the results found in flow cytometric analysis (Figure 4.3 A and B). The other $SOD1^{G93A}$ variants that do not exhibit a significant increase in the cellular fluorescence in both cell lines, including $SOD1^{G93A/K36L}$, show significant levels of intracellular aggregation (Figure 4.6). For $SOD1^{G93A}$ and the three promising $SOD1^{G93A}$variants ($SOD1^{G93A/K36E}$, $SOD1^{G93A/K91D}$, and $SOD1^{G93A/K91E}$), the fraction of cells exhibiting intracellular aggregation was determined (Figure 4.7). For the three promising $SOD1^{G93A}$variants, there was at least a 2.5 fold decrease in the fraction of cells exhibiting intracellular SOD1 aggregation ($p<0.01$ for all samples compared to $SOD1^{G93A}$) (Figure 4.7). In particular, the introduction of K36E into $SOD1^{G93A}$ almost completely eliminates intracellular aggregation in HEK293T cells.

Determination of the aggregation propensity of three $SOD1^{G93A}$ variants by using cell lysate fractionation

Using rational protein design, mammalian-cell based screening, and fluorescence microscopic analysis, three $SOD1^{G93A}$ variants ($SOD1^{G93A/K36E}$, $SOD1^{G93A/K91D}$, and $SOD1^{G93A/K91E}$) with a reduced aggregation inside mammalian cells were identified. In order to more quantitatively evaluate relative aggregation propensity of SOD1 variants, we performed cell fractionation analysis of HEK293T cells expressing the EGFP fusion of $SOD1^{WT}$, $SOD1^{G93A}$, $SOD1^{G93A/K36E}$, $SOD1^{G93A/K91D}$, and $SOD1^{G93A/K91E}$ (Figure 4.8). As described previously (Gregoire et al. 2013), detergent soluble and detergent-insoluble proteins for the aforementioned SOD1 constructs were extracted from HEK293T cells two days post transfection

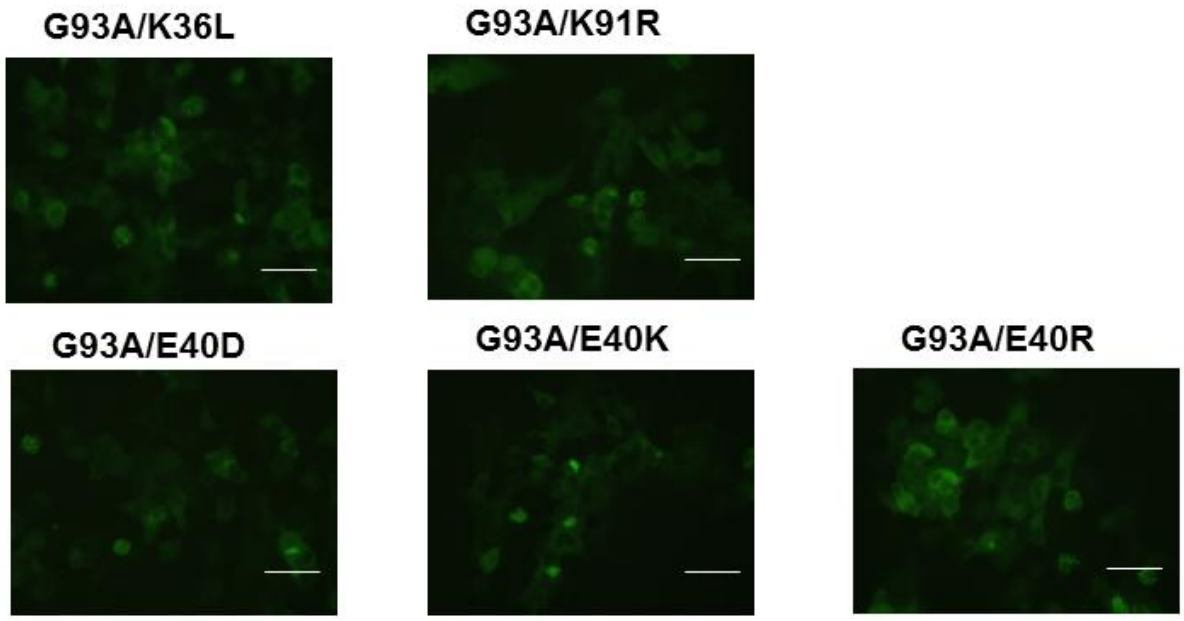


Figure 4.6: The fluorescence microscopic images of the transfected HEK293T cells expressing EGFP fusion of SOD1 variants. The images of transfected HEK293T cells expressing EGFP fusion of $SOD1^{G93A}$ double mutants were taken at 2 days post-transfection. Scale Bar: 50 µm

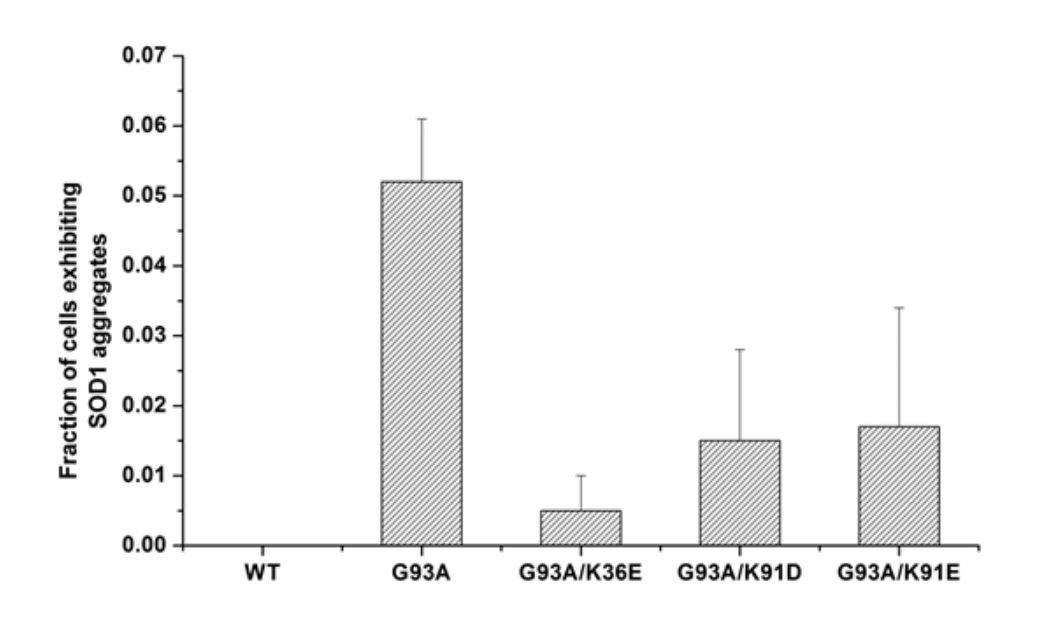


Figure 4.7: The fraction of the transfected HEK293T cells exhibiting SOD1 aggregates. The number of cells exhibiting intracellular SOD1 aggregates was determined by analyzing fluorescence microscopy images of the transfected cells expressing EGFP fusion of four SOD1 variants ($SOD1^{G93A}$, $SOD1^{G93A/K36E}$, $SOD1^{G93A/K91D}$ and $SOD1^{G93A/K91E}$) and $SOD1^{WT}$ at 2 days post-transfection ($p<0.01$ for all samples compared to $SOD1^{G93A}$).

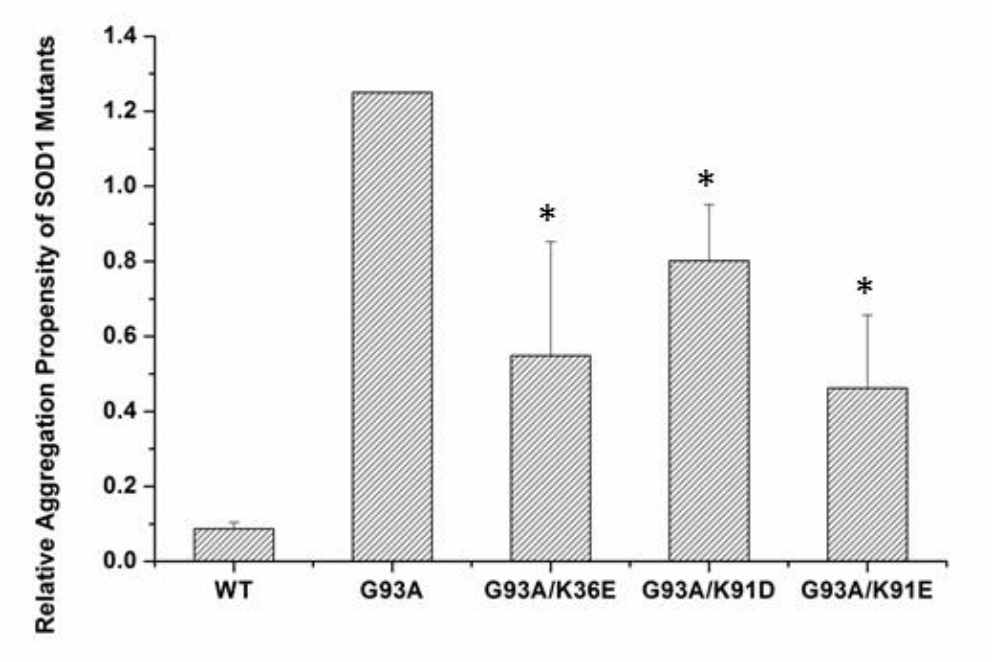


Figure 4.8: The relative aggregation propensity of SOD1 proteins at 48 h. In order to determine whether the SOD1 variant aggregation propensity is smaller than that of $SOD1^{G93A}$, two-sided Students t-tests were applied to the data (*, $P<0.05$) (n=3).

and their western-blot band intensities were compared (Figure 4.9). Relative aggregation propensity is the ratio of insoluble fraction band intensity and the soluble fraction band intensity, as described previously (Gregoire et al. 2013; Prudencio and Borchelt 2011; Prudencio et al. 2009; Prudencio et al. 2012). To normalize bands from different membranes, the $SOD1^{G93A}$ aggregate propensity was set to 1.25 and the insoluble band intensities for all other samples were adjusted. All three novel $SOD1^{G93A}$ variants showed a significant reduction ($p<0.05$) in the relative aggregation propensity compared to that of $SOD1^{G93A}$, consistent with the trend observed in the fluorescence microscopic and flow cytometric analyses described earlier.

Enhanced kinetic stability of the novel $SOD1^{G93A}$ variants with a reduced aggregation-propensity

So far, all three variants ($SOD1^{G93A/K36E}$, $SOD1^{G93A/K91E}$ and $SOD1^{G93A/K91D}$) showed a reduced propensity to aggregate within mammalian cells compared to $SOD1^{G93A}$. As stated in the Introduction, we expect to increase kinetic stability of $SOD1^{G93A}$ by introducing an additional mutation. In order to confirm this, we first investigated whether the addition of each of those mutations enhances the conformational stability of $SOD1^{G93A}$. Therefore, utilizing the RosettaDesign, we determined whether those mutations (K36E, K91E, and K91D) affect the folding energy score ($\Delta\Delta G_f$) and repulsive portion of Lennard-Jones Potential (fa_rep). Compared to a significant change in both scores upon G93A mutation in $SOD1^{WT}$ (Table 4.1), the change in both scores made by each of the mutations (K36E, K91E, and K91D) is negligible (Table 4.1). Considering that neither the side chain of K36 nor K91 directly interact with the side chain of A93 causing the repulsive interaction with the neighboring protein backbone, it is not

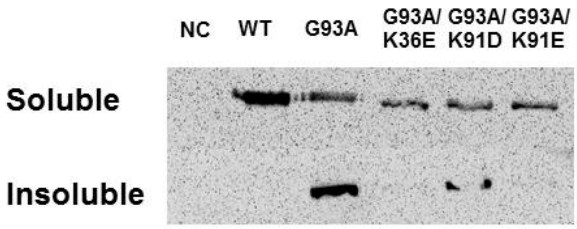


Figure 4.9: The western-blot bands of the insoluble and soluble fractions of the transfected HEK293T cells expressing the SOD1 variants

surprising to find out that none of the mutations (K36E, K91E, and K91D) eliminates such repulsive interaction and improve the conformational stability.

Next, we investigated whether the three variants ($SOD1^{G93A/K36E}$, $SOD1^{G93A/K91E}$ and $SOD1^{G93A/K91D}$) are kinetically more stable than $SOD1^{G93A}$. In order to determine the rate of aggregation formation within HEK293T cells, the fraction of cells exhibiting intracellular aggregates of SOD1 variant was determined every 24 hours for three days using fluorescence microscopy (Figure 4.10). As expected, the three variants ($SOD1^{G93A/K36E}$, $SOD1^{G93A/K91E}$ and $SOD1^{G93A/K91D}$) form intracellular aggregates significantly slower than $SOD1^{G93A}$, suggesting that there is less inclination for the three variants to aggregate over the given period of time.

CONCLUSION

Understanding how to systematically reduce intracellular protein aggregation has substantial implications for biopharmaceutical production and disease pathology. To mitigate complications that arise from protein aggregate development within mammalian cells, we employed rational protein design and cell-based aggregation assay to identify mutations that enhance protein kinetic stability, resulting in reduction of the aggregation propensity of the protein within cells. The aggregation-prone mutant $SOD1^{G93A}$ was used as a model system. The G93A mutation causes the repulsive interaction with the neighboring protein backbone leading to backbone shift and multiple side chain displacements. A panel of $SOD1^{G93A}$ variants containing a mutation at three residues (K36, E40, and K91) of which side chain were significantly displaced upon the G93A mutation were generated and subjected to the cell-based screening. Three novel $SOD1^{G93A}$ variants ($SOD1^{G93A/K36E}$, $SOD1^{G93A/K91D}$ and $SOD1^{G93A/K91E}$) with a reduced aggregation propensity inside mammalian cells were identified.

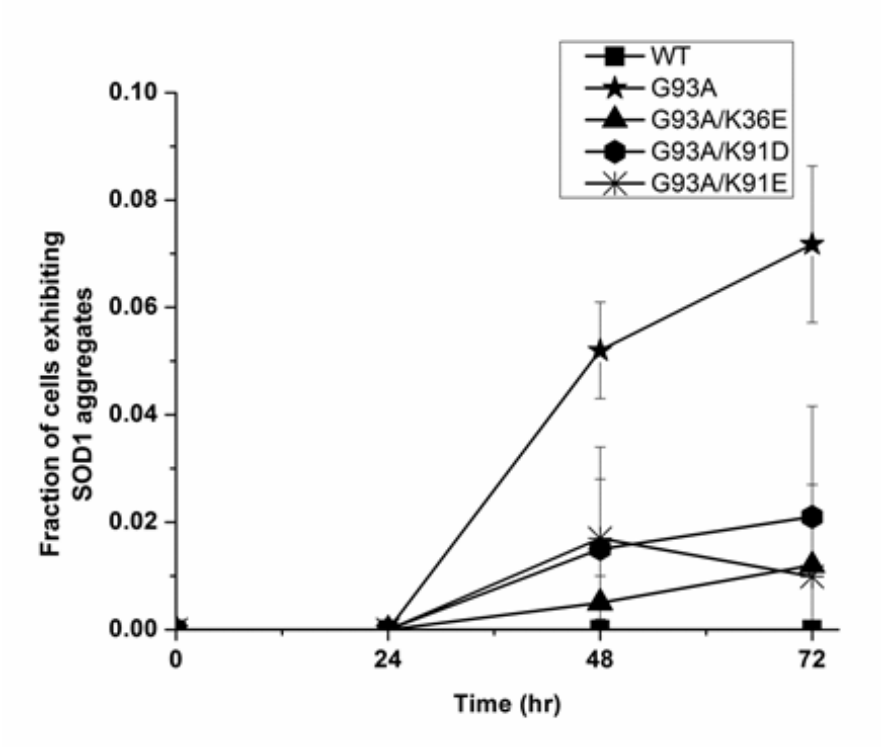


Figure 4.10 The fraction of the transfected HEK293T cells exhibiting intracellular SOD1 aggregates determined at 0, 24, 48 and 72 hours post-transfection

Although the conformational stability of these variants is just comparable to that of $SOD1^{G93A}$, the aggregation rate of these variants is substantially lower than that of $SOD1^{G93A}$ supporting that the three variants are kinetically more stable than $SOD1^{G93A}$. To our knowledge, this is the first report of a semi-rational protein design approach being used to reduce the aggregation propensity of a target protein by increasing the kinetic stability. This approach provided substantial information on the mechanism behind the destabilization of the protein and areas of the structure that are substantially perturbed by the addition of the destabilizing mutation. Not only has this technique provided information about the mechanism for $SOD1^{G93A}$ aggregation, but it has also opened a new door for engineering other proteins to reduce their aggregation.

Chapter 5: Project Summary and Avenues for Future Work

Project Summary

In this dissertation, two protein engineering strategies (increasing the conformational stability or kinetic stability) to reduce the aggregation propensity of destabilized and aggregation-prone proteins within mammalian cells were investigated. Two SOD1 mutants, $SOD1^{A4V}$ and $SOD1^{G93A}$ were used as model proteins. Of particular importance, misfolding and aggregation mutant SOD1 is heavily implicated in the familial form of ALS. Although there have been substantial efforts to reduce SOD1 aggregation to understand the mechanism behind mutant SOD1 aggregation and as a potential therapeutic strategy, there have been minimal success in developing a reproducible and effective strategy to reduce aggregation/misfolding of mutant SOD1. The semi-rational protein design strategy utilizing the computational stability analysis and mammalian-cell based aggregation assay was developed to reduce the aggregation of the two SOD1 variants.

In order to identify the second or third site suppressor mutations to suppressing the SOD1 variants aggregation, there was an unmet need to develop a non-invasive, cell-based screening assay for monitoring protein aggregation and misfolding in mammalian cells. Utilizing a fusion protein that linked EGFP to variants of $SOD1^{WT}$, the relative level protein misfolding/aggregation was determined to be inversely correlated to mean cellular fluorescence when expressed in mammalian cells. This simple, but effective monitoring system of aggregation and misfolding of SOD1 provides an array of structural information on mutant SOD1. More broadly, generality of GFP tagging method will easily allow extension of this method to monitoring other cytosolic proteins that readily form aggregates.

Next, RosettaDesign (a computational design program) along with visual inspection of the crystal structure was used to locate the origin of the protein destabilization induced by the mutation in $SOD1^{A4V}$ and $SOD1^{G93A}$. For $SOD1^{A4V}$, the bulky, hydrophobic residue, F20, interacts with the large, hydrophobic, V4, which was identified to destabilize and play a substantial factor in aggregation of the protein. Therefore, it was hypothesized that mutations at F20 would be beneficial for increasing the conformational stability of the protein and in return, reduces the aggregation propensity of the protein in mammalian cell culture. Only small, hydrophobic residues were considered for mutational analysis. However, it was found out that $SOD1^{G93A}$ has a vastly different mechanism for protein destabilization from that of $SOD1^{A4V}$. The larger side chain A93 residue directly contacts the backbone of the protein, causing a shift in the protein backbone and several residues adjacent to (but not in contact with) A93. Since there are no residue side chains in direct contact with the A93 side chain, it is a big challenge to select residues increasing the conformational stability. Instead, residues within a 5 angstrom distance from the G93A mutation that showed a substantial distortion upon introduction of the mutation were selected for mutational analysis to increase the kinetic stability of the protein. Since these residues were charged (K36, K91 and E40), only other charged residues were using for mutational analysis.

For both $SOD1^{A4V}$ and $SOD1^{G93A}$, only small pools of SOD1 variants (less than 15 variants) were generated. Using the protein misfolding/aggregation monitoring assays described previously, the SOD1 variants were screened to identify ones with a reduced aggregation-propensity. The SOD1 variants that showed a significant increase in mean cellular fluorescence in both cell lines were further characterized using various *in vitro* and cellular assays. It was

found out that the novel SOD1 variants with a reduced aggregation-propensity has an increased thermodynamic or kinetic stability.

This work demonstrates the synergistic effect of *in silico* analyses and mammalian cell-based screening on the identification of mutations to reduce the aggregation of proteins expressed in mammalian cells. The steps are as followed: 1) identifying a key residue(s) responsible for the protein destabilization using computational stability analysis and protein structure examination, 2) using that information to prepare protein variants containing candidate mutations that improve the stability, 3) selecting protein variants with a reduced aggregation-propensity using the mammalian cell-based assays, and 4) verifying the reduction in the aggregation propensity and enhancement in the stability of the protein variants selected using cell-based and *in vitro* assays. To date, there has been no such strategy dedicated to reducing the aggregation propensity and increasing the stability of recalcitrant proteins in mammalian cells. Such a systematic approach can be utilized for understanding mechanisms for $SOD1^{mutant}$ aggregation or any other intracellular protein that aggregates when destabilized by a mutation.

Future Work and Suggestions

Understanding the effect of reducing mutant SOD1 aggregation on cellular toxicity

So far, researchers have not been able to pinpoint specific events that lead to the onset of ALS. However, researchers suggest that the formation of SOD1 aggregation is implicated in the development and progression of SOD1-associated fALS (Karch et al. 2009; Prudencio et al. 2009). In transgenic ALS mice models, levels of insoluble SOD1 increase drastically for the duration of the disease. It is apparent that the development of insoluble SOD1 aggregates may

play a role in the development of the "toxic function" SOD1 gains during the development of fALS.

With our capabilities to reduce aggregation of SOD1 in mammalian cell culture, it is worthwhile to understand how the reduction in SOD1 aggregation impacts the inherent toxicity of the mutant SOD1. The Kwon lab has extensive experience with the MTT assay, a cytotoxicity assay that relates cell healthiness to a reduction in color of a particular dye (Wong and Kwon 2011). To reduce the effect of transgene expression on toxicity, lentiviral vectors will be utilized to introduce the mutant SOD1 transgene into NSC-34, a neuroblastoma cell line primarily used to mimic motor neurons. Utilizing viral gene delivery vehicles have been utilized before to understand the toxicity of the SOD1 transgene expressed (Zimmerman et al. 2007). We predict that in cells expressing mutant SOD1 containing other suppressor mutants will exhibit less cytotoxicity than cells that express solely mutant SOD1.

Directed evolution of fALS-related SOD1 mutants to increase stability

The research performed in this dissertation was substantially aided by having a crystal structure for each SOD1 mutant under observation. However, one may not have the benefit of having a crystal structure to analyze how mutations may alter the native structure. As an alternative strategy for incorporating mutations into the protein without prior knowledge of the protein crystal structure, directed evolution of the target protein could be utilized. Mimicking and accelerating nature's capability to adapt to certain environmental pressures, directed evolution is a tool protein engineers exploit to create a vast library of proteins with an assortment of properties by performing several iterative rounds of random mutagenesis using error-prone polymerase chain reaction (Romero and Arnold 2009; Tracewell and Arnold 2009). Using a

robust screen to analyze the properties of these proteins and select the proteins with optimized of novel properties, researchers can circumvent the issue of having little to no prior knowledge of the protein structure.

If directed evolution were applied instead of using a semi-rational protein design method, the transition would be seamless. The most apparent change in the design process is the removal of the RosettaDesign and crystal structure analysis. These steps would be replaced with a lentiviral library construction using error-prone PCR. Lentiviral infection of cells allows for long-term expression of the SOD1-EGFP protein because the gene is integrated into the host cell's chromosomal DNA (Azzouz 2006). To select the best variants for further analysis, fluorescence assisted cell sorting (FACS) would be used to select and collect the highly fluorescence mammalian cells and will continue to culture them after selection. It would be interesting to see whether using directed evolution would select similar variants of $SOD1^{A4V}$ or $SOD1^{G93A}$ that were selected using the semi-rational protein design strategy in this dissertation.

Trans-suppression of SOD1

Although the identification and analysis of novel cis-acting mutations for reducing aggregation and increasing stability of $SOD1^{A4V}$ and $SOD1^{G93A}$ was extremely useful for understanding the mechanism behind aggregation of specific intracellular proteins, it will be difficult to use such an approach for treatment of the disease. Although there are molecular biology techniques that researchers can use to target and edit specific sequences in the human genome (TALENs, zinc finger nucleases, etc.), their efficacy for long-term expression and gene editing have yet to be established (Joung and Sander 2013). Currently, there are no effective treatments for preventing or curing ALS. The only FDA approved drug for ALS, riluzole,

marginally increases the lifespan of an ALS patient. It is also important to mention that SOD1 has no natural ligand that would normally stabilize it in vivo, making library screening of compounds that could potentially inhibit SOD1 aggregation very difficult.

As a safer therapeutic strategy for fALS, trans-suppression of fALS mutants like $SOD1^{A4V}$ or $SOD1^{G93A}$ is proposed. The concept of "trans-suppression" was proposed to explain the stabilization of aggregation-prone protein by a complementary mutant of the same protein. Previously observed in familial amyloidotic polyneuropathy (FAP), we will provide another mutant subunit protein containing a complementary mutation. The mutated protein in FAP is transthyretin (TTR), a 55 kDa protein that exists as a tetramer with identical subunits (Hammarström et al. 2001). Typically, patients with the most aggressive case of FAP have a valine to methionine mutation at the 30th residue (V30M). This point mutation on the subunits destabilizes the entire tetramer, causing it to misfold and dissociate. These V30M-TTR monomers become amyloidogenic and form toxic aggregates that deposit throughout the body. However, researchers found that Portuguese patients who had the V30M mutation on one allele as well as the threonine to methionine mutation at the 119th residue (T119M) on another allele did not develop the disease. In vitro experiments have also supported this finding. It is believed that T119M mutation on at least one of the subunits kinetically stabilizes the other mutated subunit, allowing stable heterotetramers formation and suppressing toxic amyloid fibrils formation (Hammarström et al. 2001; Kamata et al. 2009).

There are striking similarities between TTR and SOD1 proteins and we believe that this method could be utilized to stabilize the heterodimer of mutant SOD1. First, both proteins are composed of subunits that are stabilized by a hydrophobic surface. Also, mutant forms of each protein are unstable in the multi-unit form, making the monomer the most readily available form.

Each protein will exist as a multiunit protein before quickly monomerizing. Then, these monomers rapidly form toxic aggregates that deposit themselves throughout the body. It is hypothesized that the trans-suppression strategy can be applied to stabilize the SOD1 heterodimer, control SOD1 aggregation as well as reduce cytotoxicity of the aggregates on motor neurons.

REFERENCES

Andersen PM. 2006. Amyotrophic lateral sclerosis associated with mutations in the CuZn superoxide dismutase gene. *Curr Neurol Neurosci Rep* 6(1):37-46.

Arnheim N, Calabrese P. 2009. Understanding what determines the frequency and pattern of human germline mutations. *Nat Rev Genet* 10(7):478-488.

Auclair JR, Boggio KJ, Petsko GA, Ringe D, Agar JN. 2010. Strategies for stabilizing superoxide dismutase (SOD1), the protein destabilized in the most common form of familial amyotrophic lateral sclerosis. *Proc Natl Acad Sci U S A* 107(50):21394-21399.

Azzouz M. 2006. Gene Therapy for ALS: progress and prospects. *Biochim Biophys Acta* 1762(11-12):1122-7.

Babetto E, Mangolini A, Rizzardini M, Lupi M, Conforti L, Rusmini P, Poletti A, Cantoni L. 2005. Tetracycline-regulated gene expression in the NSC-34-tTA cell line for investigation of motor neuron diseases. *Mol Brain Res* 140(1-2):63-72.

Baker D, Agard DA. 1994. Kinetics versus thermodynamics in protein folding. *Biochemistry* 33(24):7505-9.

Banci L, Bertini I, Boca M, Girotto S, Martinelli M, Valentine JS, Vieru M. 2008. SOD1 and amyotrophic lateral sclerosis: mutations and oligomerization. *PLoS ONE* 3(2):e1677.

Barthelemy PA, Raab H, Appleton BA, Bond CJ, Wu P, Wiesmann C, Sidhu SS. 2008. Comprehensive analysis of the factors contributing to the stability and solubility of autonomous human VH domains. *J Biol Chem* 283(6):3639-54.

.

Bartolini M, Andrisano V. 2010. Strategies for the Inhibition of Protein Aggregation in Human Diseases. *ChemBioChem* 11(8):1018-1035.

Beerten J, Jonckheere W, Rudyak S, Xu J, Wilkinson H, De Smet F, Schymkowitz J, Rousseau F. 2012. Aggregation gatekeepers modulate protein homeostasis of aggregating sequences and affect bacterial fitness. *Protein Eng Des Sel* 25(7):357-66.

Bondos SE, Bicknell A. 2003. Detection and prevention of protein aggregation before, during, and after purification. *Anal Biochem* 316(2):223-31.

Borchelt DR, Lee MK, Slunt HS, Guarnieri M, Xu ZS, Wong PC, Brown RH, Price DL, Sisodia SS, Cleveland DW. 1994. Superoxide dismutase 1 with mutations linked to familial amyotrophic lateral sclerosis possesses significant activity. *Proc Natl Acad Sci U S A* 91(17):8292-6.

Brotherton TE, Li Y, Glass JD. 2012. Cellular toxicity of mutant SOD1 protein is linked to an easily soluble, non-aggregated form in vitro. *Neurobiol Dis* 49C:49-56.

Bruijn LI, Cleveland DW. 1996. Mechanisms of selective motor neuron death in ALS: insights

from transgenic mouse models of motor neuron disease. *Neuropathol Appl Neurobio*l 22(5):373 87.

Bukau B, Horwich AL. 1998. The Hsp70 and Hsp60 chaperone machines. *Cell* 92(3):351-366.

Cabantous S, Pédelacq JD, Mark BL, Naranjo C, Terwilliger TC, Waldo GS. 2005a. Recent advances in GFP folding reporter and split-GFP solubility reporter technologies. Application to improving the folding and solubility of recalcitrant proteins from Mycobacterium tuberculosis. *J Struct Funct Genomics* 6(2-3):113-9.

Cabantous S, Terwilliger TC, Waldo GS. 2005b. Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol* 23(1):102-107.

Cabantous S, Waldo GS. 2006. In vivo and in vitro protein solubility assays using split GFP. *Nat Meth* 3(10):845-854.

Caine J, Sankovich S, Antony H, Waddington L, Macreadie P, Varghese J, Macreadie I. 2007. Alzheimer's Aβ fused to green fluorescent protein induces growth stress and a heat shock response. *FEMS Yeast Res* 7(8):1230-1236.

Cardoso RMF, Thayer MM, DiDonato M, Lo TP, Bruns CK, Getzoff ED, Tainer JA. 2002. Insights into Lou Gehrig's Disease from the Structure and Instability of the A4V Mutant of Human Cu,Zn Superoxide Dismutase. *J Mol Biol* 324(2):247-256.

Cashman NR, Durham HD, Blusztajan JK, Oda K, Tabira T, Shaw IT, Dahrouge S, Antel JP. 1992. Neuroblastoma x spinal cord (NSC) hybrid cell lines resemble developing motor neurons. *Dev Dynam* 194(3):209-221.

Chi EY, Krishnan S, Randolph TW, Carpenter JF. 2003. Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation. *Pharm Res* 20(9):1325-36.

Chiti F, Calamai M, Taddei N, Stefani M, Ramponi G, Dobson CM. 2002. Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Proc Natl Acad Sci U S A* 99 Suppl 4:16419-26.

Chun WJ, Waldo GS, Johnson GVW. 2007. Split GFP complementation assay: a novel approach to quantitatively measure aggregation of tau in situ: effects of GSK3 beta activation and caspase 3 cleavage. *J Neurochem* 103(6):2529-2539.

Cleveland DW, Rothstein JD. 2001. From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS. *Nat Rev Neurosci* 2(11):806-19.

Cluskey S, Ramsden DB. 2001. Mechanisms of neurodegeneration in amyotrophic lateral sclerosis. *Mol Pathol* 54(6):386-92.

Cohen FE, Kelly JW. 2003. Therapeutic approaches to protein-misfolding diseases. *Nature* 426(6968):905-9.

Colby DW, Cassady JP, Lin GC, Ingram VM, Wittrup KD. 2006. Stochastic kinetics of intracellular huntingtin aggregate formation. *Nat Chem Bio* 2(6):319-323.

Colby DW, Chu YJ, Cassady JP, Duennwald M, Zazulak H, Webster JM, Messer A, Lindquist S, Ingram VM, Wittrup KD. 2004. Potent inhibition of huntingtin and cytotoxicity by a disulfide bond-free single-domain intracellular antibody. *Proc Natl Acad Sci U S A* 101(51):17616-17621.

Cozzolino M, Amori I, Pesaresi MG, Ferri A, Nencini M, Carrì MT. 2008. Cysteine 111 affects aggregation and cytotoxicity of mutant Cu,Zn-superoxide dismutase associated with familial amyotrophic lateral sclerosis. *J Biol Chem* 283(2):866-74.

Dantas G, Kuhlman B, Callender D, Wong M, Baker D. 2003. A large scale test of computational protein design: folding and stability of nine completely redesigned globular proteins. *J Mol Biol* 332(2):449-60.

Das R, Baker D. 2008. Macromolecular modeling with rosetta. *Annu Rev Biochem* 77:363-82.

de Beus MD, Chung J, Colón W. 2004. Modification of cysteine 111 in Cu/Zn superoxide dismutase results in altered spectroscopic and biophysical properties. *Protein Sci* 13(5):1347 1355.

Deng HX, Shi Y, Furukawa Y, Zhai H, Fu R, Liu E, Gorrie GH, Khan MS, Hung WY, Bigio EH and others. 2006. Conversion to the amyotrophic lateral sclerosis phenotype is associated with intermolecular linked insoluble aggregates of SOD1 in mitochondria. *Proc Natl Acad Sci U S A* 103(18):7142-7.

Der BS, Kluwe C, Miklos AE, Jacak R, Lyskov S, Gray JJ, Georgiou G, Ellington AD, Kuhlman B. 2013. Alternative computational protocols for supercharging protein surfaces for reversible unfolding and retention of stability. *PLoS One* 8(5):e64363.

DiDonato M, Craig L, Huff ME, Thayer MM, Cardoso RMF, Kassmann CJ, Lo TP, Bruns CK, Powers ET, Kelly JW and others. 2003. ALS Mutants of Human Superoxide Dismutase Form Fibrous Aggregates Via Framework Destabilization. *J Mol Biol* 332(3):601-615.

Dobson CM. 2003. Protein folding and misfolding. *Nature* 426(6968):884-90.

Dobson CM. 2004. Experimental investigation of protein folding and misfolding. *Methods* 34(1):4-14.

Dudgeon K, Famm K, Christ D. 2009. Sequence determinants of protein aggregation in human VH domains. *Protein Eng Des Sel* 22(3):217-20.

Eijsink VG, Bjork A, Gaseidnes S, Sirevag R, Synstad B, van den Burg B, Vriend G. 2004. Rational engineering of enzyme stability. *J Biotechnol* 113(1-3):105-20.

Foit L, Morgan GJ, Kern MJ, Steimer LR, von Hacht AA, Titchmarsh J, Warriner SL, Radford SE, Bardwell JC. 2009. Optimizing protein stability in vivo. *Mol Cell*. 36(5):861-71.

Fujiwara N, Nakano M, Kato S, Yoshihara D, Ookawara T, Eguchi H, Taniguchi N, Suzuki K. 2007. Oxidative Modification to Cysteine Sulfonic Acid of Cys111 in Human Copper-Zinc Superoxide Dismutase. *J Bio Chem* 282(49):35933-35944.

Furukawa Y, Fu R, Deng HX, Siddique T, O'Halloran TV. 2006. Disulfide cross-linked protein represents a significant fraction of ALS-associated Cu, Zn-superoxide dismutase aggregates in spinal cords of model mice. *Proc Natl Acad Sci U S A* 103(18):7148-53.

Galaleldeen A, Strange RW, Whitson LJ, Antonyuk SV, Narayana N, Taylor AB, Schuermann JP, Holloway SP, Hasnain SS, Hart PJ. 2009. Structural and biophysical properties of metal-free pathogenic SOD1 mutants A4V and G93A. *Arch Biochem and Biophys*. 492(1-2):40-7.

Ghanta J, Shen CL, Kiessling LL, Murphy RM. 1996. A strategy for designing inhibitors of beta amyloid toxicity. *J Biol Chem* 271(47):29525-29528.

Graham FL, van der Eb AJ. 1973. A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology* 52(2):456-67.

Gregoire S, Irwin J, Kwon I. 2012. Techniques for monitoring protein misfolding and aggregation in vitro and in living cells. *Korean J. Chem. Eng.* 29(6):693-702.

Gregoire S, Kwon I. 2012. A revisited folding reporter for quantitative assay of protein misfolding and aggregation in mammalian cells. *Biotechnol J* 7(10):1297-1307.

Gregoire S, Zhang S, Costanzo J, Wilson K, Fernandez EJ, Kwon I. 2013. Cis-suppression to arrest protein aggregation in mammalian cells. *Biotechnol and Bioeng* doi: 10.1002/bit.25119

Gsponer J, Vendruscolo M. 2006. Theoretical approaches to protein aggregation. *Protein Pept Lett* 13(3):287-93.

Guglielmi L, Denis V, Vezzio-Vie N, Bec N, Dariavach P, Larroque C, Martineau P. 2011. Selection for intrabody solubility in mammalian cells using GFP fusions. *Protein Eng Des Sel* 24(12):873-881.

Gurney ME, Pu H, Chiu AY, Dal Canto MC, Polchow CY, Alexander DD, Caliendo J, Hentati A, Kwon YW, Deng HX. 1994. Motor neuron degeneration in mice that express a human Cu,Zn superoxide dismutase mutation. *Science* 264(5166):1772-5.

Hammarström P, Schneider F, Kelly JW. 2001. Trans-suppression of misfolding in an amyloid disease. *Science* 293(5539):2459-62.

Hartl FU, Hayer-Hartl M. 2009. Converging concepts of protein folding in vitro and in vivo. *Nat Struct Mol Biol* 16(6):574-81.

Hwang S-J, Yoon S, Koh G, Lee G. 2011. Effects of culture temperature and pH on flag-tagged COMP angiopoietin-1 (FCA1) production from recombinant CHO cells: FCA1 aggregation. *Appl Microbiol and Biotechnol* 91(2):305-315.

Ignatova Z. 2005. Monitoring protein stability in vivo. *Microb Cell Fac* 4:23.

Ioannou YA, Bishop DF, Desnick RJ. 1992. Overexpression of human alpha-galactosidase-A results in its intracellular aggregation, crystallization in lysosomes, and selective secretion. *J Cell* 119(5):1137-1150.

Ishii S, Kase R, Okumiya T, Sakuraba H, Suzuki Y. 1996. Aggregation of the Inactive Form of Human α-Galactosidase in the Endoplasmic Reticulum. *Biochem and Biophys Res Com* 220(3):812-815.

Jahn TR, Radford SE. 2008. Folding versus aggregation: Polypeptide conformations on competing pathways. *Arch Biochem and Biophys* 469(1):100-117.

Jang J-H, Koerber JT, Kim J-S, Asuri P, Vazin T, Bartel M, Keung A, Kwon I, Park KI, Schaffer DV. 2011. An Evolved Adeno-associated Viral Variant Enhances Gene Delivery and Gene Targeting in Neural Stem Cells. *Mol Ther* 19(4):667-75

Jang JH, Koerber JT, Gujraty K, Bethi SR, Kane RS, Schaffer DV. 2010. Surface immobilization of hexa-histidine-tagged adeno-associated viral vectors for localized gene delivery. *Gene Ther* 17(11):1384-1389.

Johnston JA, Dalton MJ, Gurney ME, Kopito RR. 2000. Formation of high molecular weight complexes of mutant Cu, Zn-superoxide dismutase in a mouse model for familial amyotrophic lateral sclerosis. *Proc Natl Acad Sci U S A* 97(23):12571-6.

Jonsson PA, Ernhill K, Andersen PM, Bergemalm D, Brännström T, Gredal O, Nilsson P, Marklund SL. 2004. Minute quantities of misfolded mutant superoxide dismutase-1 cause amyotrophic lateral sclerosis. *Brain* 127(Pt 1):73-88.

Joung JK, Sander JD. 2013. TALENs: a widely applicable technology for targeted genome editing. *Nat Rev Mol Cell Biol* 14(1):49-55.

Kabuta T, Suzuki Y, Wada K. 2006. Degradation of amyotrophic lateral sclerosis-linked mutant Cu,Zn-superoxide dismutase proteins by macroautophagy and the proteasome. *J Biol Chem* 281(41):30524-30533.

Kamata M, Susanto MT, Chen IS. 2009. Enhanced transthyretin tetramer stability following expression of an amyloid disease transsuppressor variant in mammalian cells. *J Gene Med*

11(2):103-11.

Karch CM, Prudencio M, Winkler DD, Hart PJ, Borchelt DR. 2009. Role of mutant SOD1 disulfide oxidation and aggregation in the pathogenesis of familial ALS. *Proc Natl Acad Sci U S A* 106(19):7774-9.

Kaytor MD, Warren ST. 1999. Aberrant protein deposition and neurological disease. *J Biol Chem* 274(53):37507-10.

Keshet B, Yang IH, Good TA. 2010. Can Size Alone Explain Some of the Differences in Toxicity Between beta-Amyloid Oligomers and Fibrils. *Biotechnol Bioeng* 106(2):333-337.

Kilic E, Weishaupt JH, Kilic U, Rohde G, Yulug B, Peters K, Hermann DM, Bahr M. 2004. The superoxide dismutase1 (SOD1) G93A mutation does not promote neuronal injury after focal brain ischemia and optic nerve transection in mice. *Neuroscience* 128(2):359-64.

Kim W, Hecht MH. 2005. Sequence determinants of enhanced amyloidogenicity of Alzheimer A beta 42 peptide relative to A beta 40. *J Biol Chem* 280(41):35069-35076.

Kim W, Hecht MH. 2008. Mutations enhance the aggregation propensity of the Alzheimer's A beta peptide. *J Mol Biol* 377(2):565-574.

Kim W, Kim Y, Min J, Kim DJ, Chang YT, Hecht MH. 2006. A high-throughput screen for compounds that inhibit aggregation of the Alzheimer's peptide. *Acs Chem Bio*l 1(7):461-469.

Kuhlman B, Baker D. 2000. Native protein sequences are close to optimal for their structures. *Proc Natl Acad Sci U S A* 97(19):10383-8.

Kupershmidt L, Weinreb O, Amit T, Mandel S, Carri MT, Youdim MBH. 2009. Neuroprotective and neuritogenic activities of novel multimodal iron-chelating drugs in motor-neuron-like NSC 34 cells and transgenic mouse model of amyotrophic lateral sclerosis. *FASEB J* 23(11):3766 3779.

Lazar GA, Marshall SA, Plecs JJ, Mayo SL, Desjarlais JR. 2003. Designing proteins for therapeutic applications. *Curr Opin Struct Biol* 13(4):513-518.

Lee S, Fernandez EJ, Good TA. 2007. Role of aggregation conditions in structure, stability, and toxicity of intermediates in the Abeta fibril formation pathway. *Protein Sci* 16(4):723-32.

Li XQ, Zhao XN, Fang Y, Jiang X, Duong T, Fan C, Huang CC, Kain SR. 1998. Generation of destabilized green fluorescent protein transcription reporter. *J Biol Chem* 273(52):34970-34975.

Lowe TL, Strzelec A, Kiessling LL, Murphy RM. 2001. Structure-function relationships for inhibitors of beta-amyloid toxicity containing the recognition sequence KLVFF. *Biochemistry* 40(26):7882-7889.

Lutz S. 2010. Beyond directed evolution--semi-rational protein engineering and design. *Curr Opin Biotechnol* 21(6):734-43.

Lynch M. 2010. Rate, molecular spectrum, and consequences of human mutation. *Proc Natl Acad Sci U S A.* 107(3):961-968.

Lynch SM, Boswell SA, Colon W. 2004. Kinetic stability of Cu/Zn superoxide dismutase is dependent on its metal ligands: implications for ALS. *Biochemistry* 43(51):16525-31.

Manning M, Colon W. 2004. Structural basis of protein kinetic stability: resistance to sodium dodecyl sulfate suggests a central role for rigidity and a bias toward beta-sheet structure. *Biochemistry* 43(35):11248-54.

Masuda M, Suzuki N, Taniguchi S, Oikawa T, Nonaka T, Iwatsubo T, Hisanaga S, Goedert M, Hasegawa M. 2006. Small molecule inhibitors of alpha-synuclein filament assembly. *Biochemistry* 45(19):6085-6094.

Matsumoto G, Kim S, Morimoto RI. 2006. Huntingtin and Mutant SOD1 Form Aggregate Structures with Distinct Molecular Properties in Human Cells. *J Biol Chem* 281(7):4477-4485.

McCutchen SL, Lai Z, Miroy GJ, Kelly JW, Colón W. 1995. Comparison of lethal and nonlethal transthyretin variants and their relationship to amyloid disease. *Biochemistry* 34(41):13527-36.

Murphy RM. 2002. Peptide aggregation in neurodegenerative disease. *Ann Rev Biomed Eng* 4:155-174.

Nagai M, Re DB, Nagata T, Chalazonitis A, Jessell TM, Wichterle H, Przedborski S. 2007. Astrocytes expressing ALS-linked mutated SOD1 release factors selectively toxic to motor neurons. *Nat Neurosci* 10(5):615-22.

Nicholls SB, Chu J, Abbruzzese G, Tremblay KD, Hardy JA. 2011. Mechanism of a Genetically Encoded Dark-to-Bright Reporter for Caspase Activity. *J Biol Chem* 286(28):24977-24986.

Niwa J, Yamada S, Ishigaki S, Sone J, Takahashi M, Katsuno M, Tanaka F, Doyu M, Sobue G. 2007. Disulfide bond mediates aggregation, toxicity, and ubiquitylation of familial amyotrophic lateral sclerosis-linked mutant SOD1. *J Biol Chem* 282(38):28087-28095.

Okumiya T, Kroos MA, Van Wet L, Takeuchi H, Van der Ploeg AT, Reuser AJJ. 2007. Chemical chaperones improve transport and enhance stability of mutant alpha-glucosidases in glycogen storage disease type II. *Mol Genet Metab* 90(1):49-57.

Pasinelli P, Brown RH. 2006. Molecular biology of amyotrophic lateral sclerosis: insights from genetics. *Nat Rev Neurosci* 7(9):710-23.

Pey AL, Rodriguez-Larrea D, Bomke S, Dammers S, Godoy-Ruiz R, Garcia-Mira MM,

Sanchez-Ruiz JM. 2008. Engineering proteins with tunable thermodynamic and kinetic stabilities. *Proteins* 71(1):165-74.

Pike CJ, Walencewicz AJ, Glabe CG, Cotman CW. 1991. In vitro aging of beta-amyloid protein causes peptide aggregation and neurotoxicity. *Brain Res* 563(1-2):311-314.

Prudencio M, Borchelt DR. 2011. Superoxide dismutase 1 encoding mutations linked to ALS adopts a spectrum of misfolded states. *Mol Neurodegener* doi: 10.1186/1750-1326-6-77.

Prudencio M, Durazo A, Whitelegge JP, Borchelt DR. 2010. An examination of wild-type SOD1 in modulating the toxicity and aggregation of ALS-associated mutant SOD1. *Hum Mol Genet* 19(24):4774-89.

Prudencio M, Hart PJ, Borchelt DR, Andersen PM. 2009. Variation in aggregation propensities among ALS-associated variants of SOD1: correlation to human disease. *Hum Mol Genet* 18(17):3217-26.

Prudencio M, Lelie H, Brown HH, Whitelegge JP, Valentine JS, Borchelt DR. 2012. A novel variant of human superoxide dismutase 1 harboring amyotrophic lateral sclerosis-associated and experimental mutations in metal-binding residues and free cysteines lacks toxicity in vivo. *J Neurochem* 121(3):475-85.

Qin JY, Zhang L, Clift KL, Hulur I, Xiang AP, Ren B-Z, Lahn BT. 2010. Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. *PLoS ONE* 5(5):e10611.

Raimondi A, Mangolini A, Rizzardini M, Tartari S, Massari S, Bendotti C, Francolini M, Borgese N, Cantoni L, Pietrini G. 2006. Cell culture models to investigate the selective vulnerability of motoneuronal mitochondria to familial ALS-linked G93ASOD1. *Eur J Neurosc*i 24(2):387-399.

Rakhit R, Chakrabartty A. 2006. Structure, folding, and misfolding of Cu,Zn superoxide dismutase in amyotrophic lateral sclerosis. Biochim Biophys Acta (BBA) - *Mol Basis Dis* 1762(11-12):1025-1037.

Rakhit R, Crow JP, Lepock JR, Kondejewski LH, Cashman NR, Chakrabartty A. 2004. Monomeric Cu,Zn-superoxide dismutase is a common misfolding intermediate in the oxidation models of sporadic and familial amyotrophic lateral sclerosis. *J Biol Chem* 279(15):15499 15504.

Rakhit R, Cunningham P, Furtos-Matei A, Dahan S, Qi XF, Crow JP, Cashman NR, Kondejewski LH, Chakrabartty A. 2002. Oxidation-induced misfolding and agagaregation of superoxide dismutase and its implications for amyotrophic lateral sclerosis. *J Biol Chem* 277(49):47551-47556.

Ray SS, Lansbury PT. 2004. A possible therapeutic target for Lou Gehrig's disease. *Proc Natl*

Acad Sci U S A 101(16):5701-5702.

Ray SS, Nowak RJ, Brown RH, Lansbury PT. 2005. Small-molecule-mediated stabilization of familial amyotrophic lateral sclerosis-linked superoxide dismutase mutants against unfolding and aggregation. *Proc Natl Acad Sci U S A* 102(10):3639-44.

Reaume AG, Elliott JL, Hoffman EK, Kowall NW, Ferrante RJ, Siwek DF, Wilcox HM, Flood DG, Beal MF, Brown RH and others. 1996a. Motor neurons in Cu/Zn superoxide dismutase deficient mice develop normally but exhibit enhanced cell death after axonal injury. *Nat Genet* 13(1):43-47.

Rizzardini M, Lupi M, Bernasconi S, Mangolini A, Cantoni L. 2003. Mitochondrial dysfunction and death in motor neurons exposed to the glutathione-depleting agent ethacrynic acid. *J Neurolo Sci* 207(1-2):51-8.

Rizzardini M, Lupi M, Mangolini A, Babetto E, Paolo U, Cantoni L. 2006. Neurodegeneration induced by complex I inhibition in a cellular model of familial amyotrophic lateral sclerosis. *Brain Res Bull* 69(4):465-474.

Roberts CJ. 2007. Non-native protein aggregation kinetics. *Biotechnol Bioeng* 98(5):927-38.

Rodriguez J, Spearman M, Huzel N, Butler M. 2005. Enhanced production of monomeric interferon-beta by CHO cells through the control of culture conditions. *Biotechnol Prog* 21(1):22-30.

Rodriguez-Larrea D, Minning S, Borchert TV, Sanchez-Ruiz JM. 2006. Role of solvation barriers in protein kinetic stability. *J Mol Biol* 360(3):715-24.

Rohl CA, Strauss CEM, Misura KMS, Baker D. 2004. Protein Structure Prediction Using Rosetta. In: Ludwig B, Michael LJ, editors. *Methods in Enzymol* 383:66-93

Romero PA, Arnold FH. 2009. Exploring protein fitness landscapes by directed evolution. *Nat Rev Mol Cell Biol* 10(12):866-76.

Rosen DR. 1993. Mutations in Cu/Zn superoxide-dismutase gene are associated with familial amyotrophic lateral scrlerosis. *Nature* 364(6435):362-362.

Rosen DR, Siddique T, Patterson D, Figlewicz DA, Sapp P, Hentati A, Donaldson D, Goto J, O'Regan JP, Deng HX. 1993. Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 362(6415):59-62.

Rothlisberger D, Khersonsky O, Wollacott AM, Jiang L, DeChancie J, Betker J, Gallaher JL, Althoff EA, Zanghellini A, Dym O and others. 2008. Kemp elimination catalysts by computational enzyme design. *Nature* 453(7192):190-5.

Sammond DW, Eletr ZM, Purbeck C, Kimple RJ, Siderovski DP, Kuhlman B. 2007. Structure

based protocol for identifying mutations that enhance protein-protein binding affinities. *J Mol Biol* 371(5):1392-404.

Sanchez-Ruiz JM. 2010. Protein kinetic stability. *Biophys Chem*148(1-3):1-15.

Sawyer SA, Parsch J, Zhang Z, Hartl DL. 2007. Prevalence of positive selection among nearly neutral amino acid replacements in Drosophila. *Proc Natl Acad Sci U S A* 104(16):6504-6510.

Schmidlin T, Kennedy BK, Daggett V. 2009. Structural Changes to Monomeric CuZn Superoxide Dismutase Caused by the Familial Amyotrophic Lateral Sclerosis-Associated Mutation A4V. *Biophys J* 97(6):1709-1718.

Schrodinger, LLC. 2010. The PyMOL Molecular Graphics System, Version 1.3r1.

Schröder M, Schäfer R, Friedl P. 2002. Induction of protein aggregation in an early secretory compartment by elevation of expression level. *Biotechnol Bioeng* 78(2):131-140.

Shaw BF, Valentine JS. 2007. How do ALS-associated mutations in superoxide dismutase 1 promote aggregation of the protein? *Trends Biochem Sci* 32(2):78-85.

Stathopulos PB, Rumfeldt JAO, Scholz GA, Irani RA, Frey HE, Hallewell RA, Lepock JR, Meiering EM. 2003. Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis show enhanced formation of aggregates in vitro. *Proc Natl Acad Sci U S A* 100(12):7021-7026.

Stevens JC, Chia R, Hendriks WT, Bros-Facer V, van Minnen J, Martin JE, Jackson GS, Greensmith L, Schiavo G, Fisher EMC. 2010. Modification of Superoxide Dismutase 1 (SOD1) Properties by a GFP Tag – Implications for Research into Amyotrophic Lateral Sclerosis (ALS). *PLoS ONE* 5(3):e9541.

Strange RW, Antonyuk SV, Hough MA, Doucette PA, Valentine JS, Hasnain SS. 2006. Variable Metallation of Human Superoxide Dismutase: Atomic Resolution Crystal Structures of Cu–Zn, Zn–Zn and As-isolated Wild-type Enzymes. *J Mol Biol* 356(5):1152-1162.

Takeuchi H, Kobayashi Y, Yoshihara T, Niwa J, Doyu M, Ohtsuka K, Sobue G. 2002. Hsp70 and Hsp40 improve neurite outgrowth and suppress intracytoplasmic aggregate formation in cultured neuronal cells expressing mutant SOD1. *Brain Res* 949(1-2):11-22.

Tartari S, D'Alessandro G, Babetto E, Rizzardini M, Conforti L, Cantoni L. 2009. Adaptation to G93A superoxide dismutase 1 in a motor neuron cell line model of amyotrophic lateral sclerosis. *FEBS J* 276(10):2861-2874.

Tracewell CA, Arnold FH. 2009. Directed enzyme evolution: climbing fitness peaks one amino acid at a time. *Curr Opin Chem Biol* 13(1):3-9.

Tsien RY. 1998. The green fluorescent protein. *Annu Rev Biochem* 67:509-544.

Tummala H, Jung C, Tiwari A, Higgins CMJ, Hayward LJ, Xu Z. 2005. Inhibition of Chaperone Activity Is a Shared Property of Several Cu,Zn-Superoxide Dismutase Mutants That Cause Amyotrophic Lateral Sclerosis. *J Biol Chem* 280(18):17725-17731.

Turner BJ, Atkin JD, Farg MA, Zang DW, Rembach A, Lopes EC, Patch JD, Hill AF, Cheema SS. 2005. Impaired extracellular secretion of mutant superoxide dismutase 1 associates with neurotoxicity in familial amyotrophic lateral sclerosis. *J Neurosci* 25(1):108-117.

Tyedmers J, Mogk A, Bukau B. 2010. Cellular strategies for controlling protein aggregation. *Nat Rev Mol Cell Biol* 11(11):777-788.

Valentine JS, Doucette PA, Zittin Potter S. 2005. Copper-zinc superoxide dismutase and amyotrophic lateral sclerosis. *Ann Rev Biochem* 74(1):563-593.

Valentine JS, Hart PJ. 2003. Misfolded CuZnSOD and amyotrophic lateral sclerosis. *Proc Natl Acad Sci U S A* 100(7):3617-3622.

Waldo GS. 2003. Improving protein folding efficiency by directed evolution using the GFP folding reporter. *Methods Mol Biol* 230:343-59.

Waldo GS, Standish BM, Berendzen J, Terwilliger TC. 1999. Rapid protein-folding assay using green fluorescent protein. *Nat Biotechnol* 17(7):691-695.

Walsh G. 2010. Biopharmaceutical benchmarks 2010. *Nat Biotechnol* 28(9):917-24.

Wang J, Xu G, Borchelt DR. 2006. Mapping superoxide dismutase 1 domains of non-native interaction: roles of intra- and intermolecular disulfide bonding in aggregation. *J Neurochem* 96(5):1277-88.

Wang Q, Johnson JL, Agar NY, Agar JN. 2008. Protein aggregation and protein instability govern familial amyotrophic lateral sclerosis patient survival. *PLoS Biol* 6(7):e170.

Watanabe M, Dykes-Hoberg M, Culotta VC, Price DL, Wong PC, Rothstein JD. 2001. Histological evidence of protein aggregation in mutant SOD1 transgenic mice and in amyotrophic lateral sclerosis neural tissues. *Neurobiol Dis* 8(6):933-941.

Watanabe S, Nagano S, Duce J, Kiaei M, Li Q-X, Tucker SM, Tiwari A, Brown Jr RH, Beal MF, Hayward LJ and others. 2007. Increased affinity for copper mediated by cysteine 111 in forms of mutant superoxide dismutase 1 linked to amyotrophic lateral sclerosis. *Free Radical Biol Med* 42(10):1534-1542.

Wiseman RL, Kelly JW. 2009. Evolving protein stability through genetic selection. *Mol Cell* 36(5):730-1.

Witan H, Kern A, Koziollek-Drechsler I, Wade R, Behl C, Clement AM. 2008. Heterodimer

formation of wild-type and amyotrophic lateral sclerosis-causing mutant Cu/Zn-superoxide dismutase induces toxicity independent of protein aggregation. *Hum Mol Genet* 17(10):1373-85.
Wong HE, Kwon I. 2011. Xanthene Food Dye, as a Modulator of Alzheimer's Disease Amyloid beta Peptide Aggregation and the Associated Impaired Neuronal Cell Function. *PLoS ONE* 6(10):e25752.

Wong HE, Qi W, Choi H-M, Fernandez EJ, Kwon I. 2011. A Safe, Blood-Brain Barrier Permeable Triphenylmethane Dye Inhibits Amyloid-β Neurotoxicity by Generating Nontoxic Aggregates. *Acs Chem Neurosci* 2(11):645-657.

Wong PC, Pardo CA, Borchelt DR, Lee MK, Copeland NG, Jenkins NA, Sisodia SS, Cleveland DW, Price DL. 1995. An adverse property of a familial ALS-linked SOD1 mutation causes motor-neuron disease characterized by vacuolar degeneration of mitochondria. *Neuron* 14(6):1105-1116.

Zhang F, Zhu H. 2006. Intracellular conformational alterations of mutant SOD1 and the implications for fALS-associated SOD1 mutant induced motor neuron cell death. *Biochim Biophys Acta* 1760(3):404-14.

Zimmerman MC, Oberley LW, Flanagan SW. 2007. Mutant SOD1-induced neuronal toxicity is mediated by increased mitochondrial superoxide levels. *J Neurochem* 102(3):609-18.

Appendix A: Chapter 2 Supplemental Information

Construction of Expression Vectors

Cysteine (Cys) to serine (Ser) mutation was introduced into either Cys57 or Cys111 residue of $SOD1^{A4V}$ by PCR mutagenesis using pEGFP-N3-$SOD1^{A4V}$ as a template in order to generate pEGFP-N3-$SOD1^{A4V/C57S}$ or pEGFP-N3-$SOD1^{A4V/C111S}$ plasmid, respectively. The primer pairs for each mutant are as follows: 5'-GGAGATAATACAGCAGGCAGCACCAGTGCAGGTCCTCAC 3' and 5' GTGAGGACCTGCACTGGTGCTGCCTGCTGTATTATCTCC-3' for the C57S mutation and 5'- CTCACTCTCAGGAGACCATAGCATCATTGGCCGCACACTG-3'; 5'-CAGTGTGCG GCCAATGATGCTATGGTCTCCTGAGAGTGAG for the C111S mutation. In the pFUG-IP plasmid, expression of GFP protein is under the control of ubiquitin-C constitutive promoter (Ubc) (15). In order to express SOD1-EGFP fusion protein using Ubc promoter, the SOD1-EGFP genes were amplified from the pEGFP-N3-SOD1 plasmids using two pairs of primers flanked by *Age*I and *EcoR*I sites at both ends (forward primer for $SOD1^{WT}$: 5'-GATCACACCGGTATGGCGACGAAGGCCGTGTGCGTGCTGA-3'; forward primer for $SOD1^{A4V}$: 5'-GATCACACCGGTATGGCGACGAAGGTCGTGTG-3'; reverse primer for both SOD1s: 5'-GTCATCGAATTCTTACTTGTACAGCTCGTCCA). Then, the SOD1-EGFP genes were sub-cloned into the pFUG-IP plasmid 3' of the Ubc promoter. In order to sub-clone the CMV-SOD1-EGFP cassette into pFUG-IP plasmid backbone, a unique *Pac*I restriction site in the pFUG-IP plasmid was mutated into an *EocR*I restriction site via PCR mutagenesis using the following primers (forward: 5'-CGCATGGACGAGCTGTACAAGTAAGGCGCGCCCCCTCTCCCTCCCCCC-3'; reverse: 5'-GGGGGGAGGGAGAGGGGGCGCGCCTTACTTGTACAGCTCGTCCATGCCG-3'). The CMV-SOD1-EGFP cassette was amplified from the pEGFP-N3-SOD1 plasmid using two

primers flanked with *Asc*I and *Pac*I restriction sites at both ends (forward: 5'-GCATTAATTAACCTGATTCTGTGGATAACCG-3'; reverse: 5'-CAGGGCGCGCCTTTATTTGTAACCATTATAA-3') and were sub-cloned into the pFUG-IP. All DNA fragments were purified using either DNA agarose gel extraction or spin-column desalting.

Table A1: Mean cellular fluorescence intensities

Figure	SOD1	Promoter	Cell Line	Time (day)[a]	(Fluorescence Intensity[b])/10^5
Figure 2.2A	WT	CMV	HEK293T	1	3.78±0.44
	A4V/C111S	CMV	HEK293T	1	1.92±0.10
	A4V	CMV	HEK293T	1	1.46±0.22
	A4V/C57S	CMV	HEK293T	1	0.76±0.05
	WT	CMV	HEK293T	2	3.78±0.44
	A4V/C111S	CMV	HEK293T	2	2.50±0.16
	A4V	CMV	HEK293T	2	2.05±0.13
	A4V/C57S	CMV	HEK293T	2	0.975±0.05
Figure 2.6A	WT	CMV	NSC-34	2	2.43±0.503
	A4V/C111S	CMV	NSC-34	2	1.32±0.253
	A4V	CMV	NSC-34	2	0.83±0.07
	A4V/C57S	CMV	NSC-34	2	0.52±0.06
Figure 2.7	WT	Ubc	HEK293T	2	0.85±0.04
	A4V	Ubc	HEK293T	2	0.13±0.01
	WT	CMV	HEK293T	2	1.54±0.29
	A4V	CMV	HEK293T	2	0.37±0.03

a: incubation time post-transfection
b: mean value ± standard deviation

Table A2: The percentage of transfected HEK293T cells exhibiting SOD1 aggregates

	SOD1	Promoter	Cell Line	Time (day)[a]	Cells exhibiting SOD1 aggregates (%)
Figure 2.5	WT	CMV	HEK293T	2	N.D.[b]
	A4V/C111S	CMV	HEK293T	2	5.35±2.42
	A4V	CMV	HEK293T	2	12.00±2.16
	A4V/C57S	CMV	HEK293T	2	10.26±1.56

a: incubation time post-transfection
b: N.D.= Not Detected

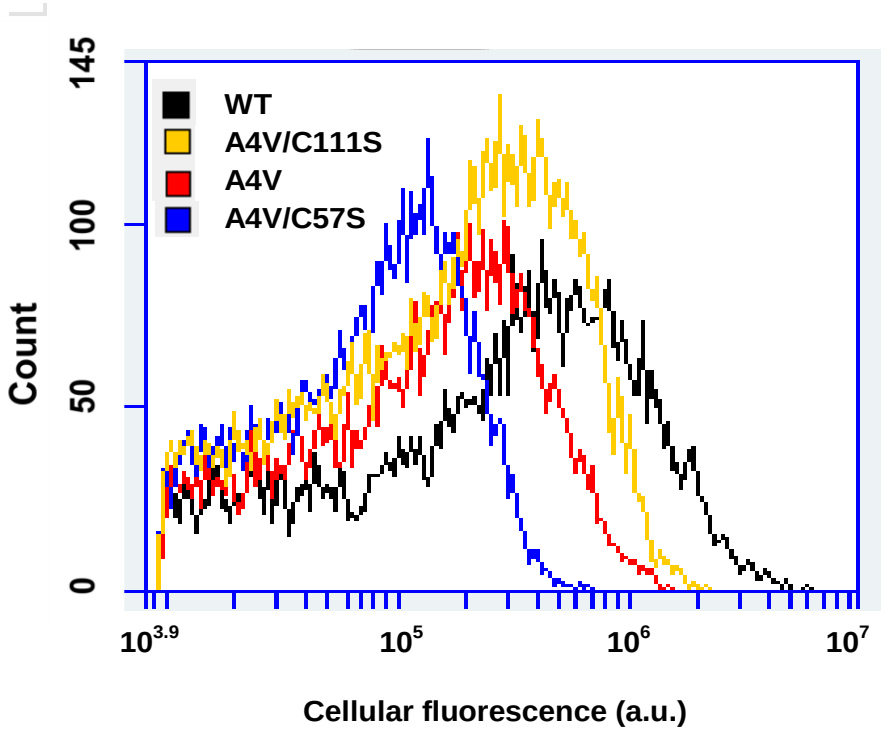


Figure A1: Representative overlaid histograms of fluorescence of transfected cells expressing EGFP fusion of four SOD1 variants ($SOD1^{WT}$, $SOD1^{A4V/C111S}$, $SOD1^{A4V}$, and $SOD1^{A4V/C57S}$).

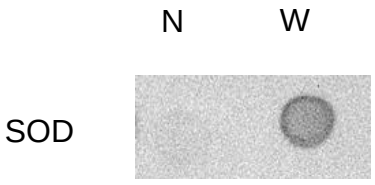


Figure A2: Immuno-reactivity of untransfected HEK293T cell lysate (NC) and transfected HEK293T cell lysate expressing $SOD1^{WT}$-EGFP (WT). Either untransfected or transfected HEK293T cells were centrifuged to obtain a cell pellet. After washing with PBS buffer once, the cell pellet was lysed using RIPA buffer (Thermo Fisher, Pittsburgh, PA). Protein concentration of the cell lysate was determined using BCA kit using bovine serum albumin as a standard. Dot-blot assay of the cell lysate was performed as described earlier (Wong and Kwon 2011; Wong et al. 2011) except anti-SOD1. The blot images were captured using a BioSpectrum imaging system (UVP).

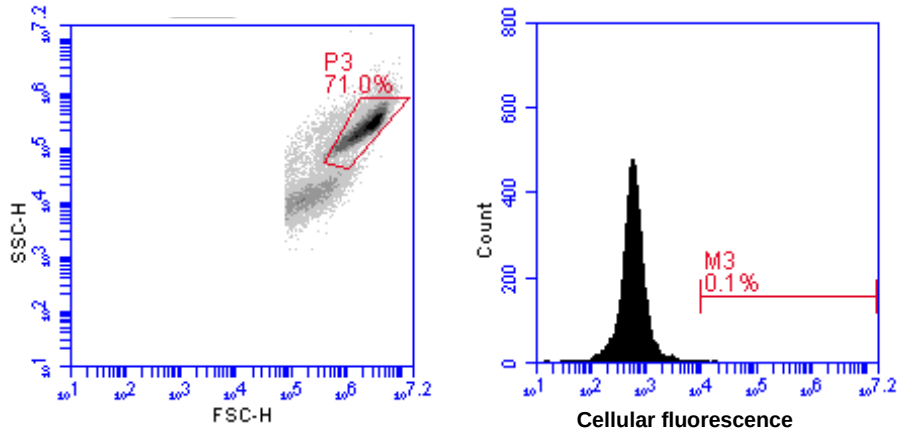


Figure A3: Gating strategy for flow cytometric analysis. (A) P3 gate was set in order to analyze only viable cells. (B) A histogram of the negative control cellular fluorescence is shown. Only transfected cells in M3 region were analyzed to determine the mean cellular fluorescence.

Appendix B: Chapter 3 Supplemental Information

Table B1: Mean cellular fluorescence intensities

Figure	SOD1	Cell Line	(Fluorescence Intensity[b])/10^5
Figure 3.2A	WT	HEK293T	3.38±0.50
	A4V	HEK293T	1.02±0.17
	A4V/C111S	HEK293T	1.41±0.35
	A4V/F20G	HEK293T	1.22±0.11
	A4V/F20A	HEK293T	1.38±0.18
	A4V/F20L	HEK293T	0.78±0.03
	A4V/F20V	HEK293T	1.07±0.03
	A4V/F20M	HEK293T	1.13±0.06
Figure 3.2B	WT	NSC-34	1.93±0.05
	A4V	NSC-34	0.78±0.01
	A4V/C111S	NSC-34	0.90±0.02
	A4V/F20G	NSC-34	0.93±0.01
	A4V/F20A	NSC-34	0.86±0.06
	A4V/F20L	NSC-34	0.74±0.02
	A4V/F20V	NSC-34	0.65±0.11
	A4V/F20M	NSC-34	0.74±0.03
Figure 3.6A	WT	HEK293T	3.06±0.44
	A4V	HEK293T	1.29±0.20
	A4V/F20A/C111S	HEK293T	2.18±0.40
	A4V/F20G/C111S	HEK293T	1.55±0.35
Figure 3.6B	WT	NSC-34	1.47±0.05
	A4V	NSC-34	0.78±0.07
	A4V/F20A/C111S	NSC-34	1.21±0.06
	A4V/F20G/C111S	NSC-34	0.45±0.07

a: incubation time post-transfection
b: mean value ± standard deviation

Table B2: The percentage of transfected HEK293T cells exhibiting SOD1 aggregates

	SOD1	Cell Line	Cells exhibiting SOD1 aggregates (%)
Figure 3.5	WT	HEK293T	N.D.[b]
	A4V	HEK293T	13.3±2.26
	A4V/C111S	HEK293T	2.57±1.08
	A4V/F20A	HEK293T	2.09±0.08
	A4V/F20G	HEK293T	0.40±0.10
Figure 3.9	WT	HEK293T	N.D.[b]
	A4V	HEK293T	13.3±2.26
	A4V/F20G	HEK293T	0.40±0.20
	A4V/F20A/C111S	HEK293T	0.10±0.10

a: incubation time post-transfection
b: N.D.= Not Detected