Two Second-Site Mutations Alleviate Aggregation of The A4V Variant Human Superoxide Dismutase-1 Associated with Familiar Amyotrophic Lateral Sclerosis

---

A Dissertation

Presented to

the faculty of the School of Engineering and Applied Science

University of Virginia

---

in partial fulfillment

of the requirements for the degree

Master of Science

by

Shaojie Zhang

May

2013

APPROVAL SHEET

The dissertation

is submitted in partial fulfillment of the requirements

for the degree of

Master of Science

_____

AUTHOR

The dissertation has been read and approved by the examining committee:

Erik Fernandez

_____

Advisor

Giorgio Carta

_____

Inchan Kwon

_____

_____

_____

_____

Accepted for the School of Engineering and Applied Science:

Dean, School of Engineering and Applied Science

May

2013

Abstract

Protein aggregation is a serious problem related to not only a variety of human diseases such as Alzheimer's, Parkinson's and prion disease, but also to the biopharmaceutical industry. The destabilization and aggregation of one variant of the homodimer human superoxide dismutase-1 (hSOD1), in which the alanine at the 4th residue is substituted with valine (hSOD1^{A4V}), is believed to result in most of the cases of familiar amyotrophic lateral sclerosis (fALS) in North America. In this work, an introdomain steric clash between Val4 and Phe20 was first found in a structure generated with A4V sequence on a wild type hSOD1 backbone by using a computational protein design program RosettaDesign. We hypothesized the introdomain steric clash between Val4 and Phe20 should be responsible for the destabilization and aggregation propensity of the hSOD1^{A4V} variant. We chose to test this hypothesis two steps further in a collaboration with the Kwon lab: 1) by using RosettaDesign and rational design approaches to identify residues which could relieve the clash caused by the inserted valine, 2) by testing variants *in vitro* and in cultured cells. Consequently, two second-site mutations at the 20th residue of hSOD1^{A4V} molecule were shown to alleviate the adverse effects caused by the A4V mutation both *in vitro* and *in vivo*. were performed in this work, the results strongly supported our hypothesis that the introdomain clash between Val4 and Phe is the actual cause of the destabilization of hSOD1^{A4V} variant. The strategy performed in this work can be used as a potential generalized method to help reveal the cause of other mutant protein misfolding/aggregation related diseases and to help inhibit the aggregation-prone behavior of biopharmaceutical proteins.

Acknowledgements

I would like to thank the following people:

My research advisor, Professor Erik Fernandez and collaborator Professor Inchan Kwon for their continuing support and guidance.

Professor Giorgio Carta for serving as the chairperson on my committee.

My collaborator, Simpson Gregoire from Kwon group for his help.

My labmates from the Fernandez group for their support and help, particularly

Joe Costanzo, Adrian Gospodarek and Jing Guo

The undergraduate research assistants Kelly Wilson and Weitong Sun for their efforts and contributions.

Finally my friends and family, specifically my parents, Quanming Zhang and Linxian Kang, and my best friend, Yun Wei, for their non-stopping love and support.

Table of contents

List of Figures

List of Tables

1. Introduction

Protein unfolding and aggregation is a serious problem in many important applications, including biopharmaceutical manufacturing and several human diseases. Since recombinant insulin was first approved in 1982, the development of recombinant DNA technology has brought more than 120 therapeutic proteins to the market (Dingermann, 2008). However, protein instability and aggregation propensity are common barriers along the road from manufacturing to the end consumer market (W. Wang, 2005). Aggregation of therapeutic proteins not only affects levels of production yield, but also may result in adverse effects, such as immune response (Rosenberg, 2006). Abnormal protein aggregation is also thought to contribute to the development of more than 20 human diseases (Stefani & Dobson, 2003), such as Alzheimer's disease (AD), Parkinson's disease (PD), Prion disease, amyotrophic lateral sclerosis (ALS) and others. The aggregates found in these patients often consist of misfolded protein fibrils, which are usually built from monomers interacting in a β-sheet conformation. Although it is not completely understood, the precursor aggregates are usually thought to cause most of the cytotoxicity (Stefani & Dobson, 2003).

There were two general methods of increasing protein stability and inhibiting protein aggregation over the past several decades. The first is by modifying the protein structure. Proteins altered with site-directed mutagenesis and chemical modification were both shown successfully to have reduced aggregation propensity. Examples are thymidylate synthase (Prasanna et al., 1999), amyloid β peptide (Soto et al., 1995) and insulin conjugated with methoxypoly ethylene glycol (Hinds et al., 2000). An alternative

method is by changing the external environmental properties. Various excipients/additives and other factors such as temperature, pH, ionic strength, type and concentration of the denaturant and protein and other environmental conditions have been shown to reduce protein aggregation successfully (W. Wang, 2005).

ALS is a disease of particular interest because of the critical role played by protein instability and aggregation on disease progression. ALS is a progressive fatal neurological disease characterized by the loss of motor neurons in spinal cord, brainstem and motor cortex (Brown, 1995). ALS affects over 35,000 Americans and over 350,000 people worldwide. Although it is predominantly a sporadic disease, nearly 20% of the cases are inherited (Ray et al., 2005). Around 20% of the familial form of ALS cases (fALS) is caused by point mutations in human superoxide dismutase-1 (hSOD1) (Rosen et al., 1993). Wild-Type (WT) hSOD1 ($hSOD1^{WT}$) is a 32 kDa, β-sheet-rich dimeric metalloenzyme, which contains a copper and a zinc ion in each monomer. hSOD1 catalyzes the dismutation reaction of superoxide to hydrogen peroxide and oxygen and thus protects the body from the harmful effects of superoxide ions (Tainer et. al., 1983; Rosen, 1993).

Over 100 mutations distributed throughout the hSOD1 monomer gene have been identified as fALS-associated (Banci et al., 2008). Although it is clear that mutant hSOD1 is responsible for causing a portion of fALS cases, the mechanism for developing fALS remains unknown. However, there is substantial evidence that shows fALS does not occur from a loss of enzymatic hSOD1 activity. Instead, fALS may result from a ‘‘gain of toxic function’’ due to aggregation of the mutant hSOD1 (Bruijn et al., 1996). *In vitro* experiments (Stathopulos et al., 2003), studies with cultured cells (Koide et al., 1998) and

transgenic mice models (Bruijn et al., 1998; Bruijn et al., 1997) all show that variant hSOD1s aggregate more readily than $hSOD1^{WT}$, and the transgenic mice overexpressing variant hSOD1 also developed the disease phenotypes (Kopito, 2000). Studies with transgenic mice models also suggested a strong connection between the formation of variant hSOD1 and the onset of neurodegeneration (Li et al., 2001; Wang et al., 2002).

In other familial diseases caused by protein misfolding, such as prion diseases and amyloidoses, it is found that single point mutations promote protein aggregation by destabilizing the native state of the protein and populating the unfolded or partially unfolded states (Canet et al., 2002; Dobson, 1999). Similarly, many fALS-associated mutations destabilize the native state of hSOD1 and promote aggregation (Brown, 1995). In particular, the alanine to valine mutation at the 4^{th} residue ($hSOD1^{A4V}$), which occurs near the dimer interface, has been shown to destabilize both apo and holo states of the protein (Hayward et al., 2002; Schmidlin et Al., 2009). The $hSOD1^{A4V}$ causes a rapidly progressing form of fALS, with life expectancy only a year after diagnosis (Ray and Lansbury 2004). Since the A4V mutation occurs near the dimer interface, it is usually believed that the dimer interface destabilization is the cause of the destabilization and aggregation of the $hSOD1^{A4V}$ protein (Cardoso et al., 2002; DiDonato et al., 2003; Hough et al., 2004).

There is strong evidence to support the hypothesis that stabilizing the native dimer of mutant hSOD1 can inhibit aggregate formation. Researchers have observed that an intersubunit disulfide bond between two subunits of $hSOD1^{A4V}$ stabilized the protein dimer and inhibited aggregate formation successfully *in vitro* (Ray et al., 2004). A detailed analysis of hSOD1 aggregation supported the idea that monomerization of

hSOD1 dimer is also needed for *in vivo* aggregate formation (Khare et al., 2004). Another study showed that by cross-linking the mutant hSOD1 dimer using two adjacent cysteines on the 111th residues of each monomer, dramatic melting temperature increases were observed, which suggests that the mutant hSOD1 dimer had been stabilized (Auclair et. al 2010).

Current approaches to reducing aggregation have significant limitations. Modifying protein structure with site-directed mutagenesis involves extensive screening work to identify effective mutations, and chemical modification may reduce protein activity. Altering the environmental properties involves tedious trial and error processes (Clark, 1998). However, computer based protein design is an alternative approach which could facilitate the identification of mutations which reduce or eliminate aggregation. By constraining design to alter structure only outside the active site, protein activities can be retained. In previous studies, it has been shown that introducing new point mutations, predicted by the computational protein design program—RosettaDesign (Kuhlman and Baker 2000), can stabilize a protein (Dantas et al., 2007; Sahin et al., 2010). In another study, Sammond and co-workers used RosettaDesign to design second-site suppressor mutations at protein-protein interfaces, which greatly improved the protein binding specificities caused by the amino acid mutations at the protein-protein interface in signaling pathways (Sammond et al., 2010).

It is apparent that the most affected residues in a mutated protein should be the residues around the single point mutation. By mutating these residues, the adverse effects caused by the mutation might be alleviated. In the previous work, Sammond and co-worker first introduced a point mutation at the protein-protein interface and then did a

second-site design only within a certain distance of this point mutation. They showed that the weaker protein-protein binding affinity caused by the first point mutation was greatly improved by the second point mutation (Sammond et al., 2010).

We were interested to identify which interactions are key to the destabilization of $hSOD1^{A4V}$. Previous investigators had identified several residues as potentially important (Cardoso, 2002). We were interested in determining whether computational design methodology could identify mutations which can suppress the adverse effects caused by A4V mutation in $hSOD1^{A4V}$ protein. This could provide valuable information about the mechanism of $hSOD1^{A4V}$ aggregation as well as a methodology that could be applied to other disease-related and biopharmaceutical proteins. We started with $hSOD1^{A4V}$ crystal structure to identify residues within 5 Å of the 4^{th} residue valine. RosettaDesign was then performed to identify candidate second-site mutations that increase the apparent thermodynamic stability of $hSOD1^{A4V}$. Then both *in vivo* and *in vitro* assays were performed to validate the suppressing mutations.

Likewise, the destabilization of protein folding is believed to be the dominant cause of biopharmaceutical protein aggregation. Therefore, if the native structure can be stabilized, there is a high chance that biopharmaceutical aggregation propensity can be inhibited. Currently, there are several programs available for the determination of aggregation "hot spots" using the primary sequence of protein (Costanzo's unpublished Ph.D. Thesis). Previous work also shows that the Complementarity Determining Regions (CDRs) of antibody might be the "hot spots" for aggregation (X. Wang, Das, Singh, & Kumar, 2009). Although biopharmaceutical antibodies are large compared to hSOD1, the fold of hSOD1 is similar to the individual domains of antibodies. Thus, the strategy

performed in this work could potentially serve as a method to identify key residues responsible for the destabilization and aggregation behavior of certain antibodies. By designing corresponding mutations that could stabilize the antibody molecule without interrupting the activity, there is a high chance that it could inhibit the aggregation propensity of a therapeutic antibody.

2. Background and theoretical development

2.1 RosettaDesign

Computational biomolecular design has been developing rapidly over the past decade. Among all the computational methods available, Rosetta is a molecular modeling software package created at the University of Washington for understanding protein structures, protein design, protein docking, protein-DNA and protein-protein interactions.

RosettaDesign is one module of the Rosetta suite package. RosettaDesign must finish two tasks in each design run. First, RosettaDesign searches through relevant conformational/sequence space mostly with knowledge-guided Metropolis Monte Carlo approaches. Second, Rosetta evaluates the relative favorability of the designed structures mostly with knowledge-based energy functions, which assume that most molecular properties can be derived from available data (Meiler et al., 2010). The core energy function of RosettaDesign is a linear sum of a 6-12 Lennard-Jones potential that favors close packed amino acids (Brian Kuhlman et al., 2003); the Lazaridis-Karplus implicit solvation model which favors hydrophobic residues buried inside of protein and polar residues on the surface (Lazaridis & Karplus, 1999); an empirical based orientation dependent hydrogen bonding potential (Kortemme, Morozov, & Baker, 2003); backbone-dependent rotamer probabilities (Dunbrack & Cohen, 1997); a knowledge-based electrostatics energy potential between charged residues based on the probabilities of seeing two residue types near each other, amino acid probabilities based on particular regions of φ/ψ space and a unique reference energy which approximates the energies of each amino acid in the unfolded state (B Kuhlman & Baker, 2000).

From a given PDB as an initial target structure for design and a "resfile" which specifies which residues will be varied, RosettaDesign performs an iterative process that energetically optimizes both the structure and sequence of the protein. Either "fixed backbone" or "flex backbone" design can be performed. Because of the extensive computer resources and setup required for flexible backbone design (parallel calculations on 100+ node cluster for multiple days), only "fixed backbone" design was used in this work. "Fixed backbone" design has been successfully used to stabilize globular proteins (Dantas et al., 2003) and protein-protein interfaces (Sammond et al., 2010).

During each run, to simplify and accelerate the optimization procedure, each amino acid side chain is only allowed to adopt a discrete set of conformations, referred to as rotamers, which are constrained to the Dunbrack Library of most probable rotamer values (Kaufmann et al., 2010). By allowing the rotamers to vary chi1 and chi2 one standard deviation (~10 °) away from most probable values, the amino acid side chain can explore more rotamers to accommodate more design changes. A Monte Carlo optimization with simulated annealing is used to find the low energy sequences and side-chain rotamers. The simulation starts at a very high temperature where almost all substitutions are allowed and finishes at 0 °. For per 100 residues being varied, approximately 1 million rotamer substitutions are attempted (Liu & Kuhlman, 2006). The output is a mutated protein structure with the lowest energy score of an assigned number of runs.

2.2 Two-state model for protein denaturation

2.2.1 Two-state model protein unfolding fitting with ΔG_{H_2O} and *m*

For a chemical denaturation assay, the equilibrium constant (K_U) for denaturation at each denaturant concentration is calculated by

$$K_U = \frac{y_F - y}{y - y_U} \tag{2-1}$$

where *y* is the spectroscopic value (e.g. intensity for fluorescence, or molar ellipticity for circular dichroism) at a certain concentration of denaturant, y_F and y_U are the spectroscopic value for the folded and unfolded form, respectively. The apparent free energy of unfolding (ΔG_U) of proteins is:

$$\Delta G_U = -RTlnK_U \tag{2-2}$$

It has been shown experimentally that ΔG_U in the presence of denaturant is linearly related to the concentration of denaturant (Pace, 1986):

$$\Delta G_U = \Delta G_{H_2O} - m[D] \tag{2-3}$$

where ΔG_{H_2O} is the apparent free energy of unfolding in the absence of denaturant, *[D]* is the concentration of the denaturant and *m* is the slope of the transition which reflects the cooperativity of the transition.

With the assumption that both y_F and y_U are linearly dependent on the concentration of denaturant (Horovitz et al., 1992), the spectroscopic data can be fitted to the following equation

$$y = \frac{(y_F + s_F[D]) + (y_U + s_U[D])e^{\frac{m[D] - \Delta G_{H_2O}}{RT}}}{1 + e^{\frac{m[D] - \Delta G_{H_2O}}{RT}}} \quad (2\text{-}4)$$

where y_F and y_U are the intercepts and s_F and s_U are the slopes of the baselines of folded and unfolded states.

2.2.2 Two-state model protein unfolding fitting with *MP* and *m*

The spectroscopic data can be fitted to Equation (2-4) to obtain the values of ΔG_{H_2O} and m. However, when denaturation is not reversible, *MP* values, which are the mid-point values of transition, are often used to compare the apparent stability of wild type and mutant proteins. At *MP*, it is apparent from Equation (2-3) that $\Delta G_{H_2O} = m \times MP$ and thus

$$\Delta G_U = m(MP - [D]) \quad (2\text{-}5)$$

and Equation (2-4) becomes

$$y = \frac{(y_F + s_F[D]) + (y_U + s_U[D])e^{\frac{m([D] - MP)}{RT}}}{1 + e^{\frac{m([D] - MP)}{RT}}} \quad (2\text{-}6)$$

When the spectroscopic value of folded and unfolded states is observed not to be dependent on the denaturant concentrations, the data could be fitted to the following simpler equation (Clarke & Fersht, 1993):

$$y = y_F - \frac{(y_F - y_U)e^{\frac{m([D] - MP)}{RT}}}{1 + e^{\frac{m([D] - MP)}{RT}}} \quad (2\text{-}7)$$

Of note, as seen in Equation (2-7), since y, y_F and y_U are directly proportional to the protein concentration, slight concentration difference between two proteins would not affect the fitting results of either MP or m.

3. Research Objectives

The purpose of this study is to identify second-site mutations which can suppress the destabilization and aggregation effects caused by A4V mutation in $hSOD1^{A4V}$ molecule. The following objectives were completed along with this work:

1. Set up a purification protocol of hSOD1 protein from the *Saccharomyces cerevisiae* EG118 strain of yeast cells containing plasmids with the hSOD1 variants.
    a. Determine the protein yield from one batch of yeast cell growth.
    b. Evaluate the purification protocol by examining the purity and secondary structure maintaining of purified protein.
2. Set up the *in vitro* assays to determine the stability and aggregation behavior of $hSOD1^{WT}$ and $hSOD1^{A4V}$ variants.
    a. Qualitatively compare the stability of WT and A4V hSOD1s by the guanidine hydrochloride-induced denaturation assay with circular dichroism detection.
    b. Evaluate the aggregation propensity of WT and A4V hSOD1s by the EDTA-induced aggregation assay with size exclusion chromatography and *UV* detection at 280nm wavelength.
3. Initial potential second-site mutations were obtained by Joe Costanzo with RosettaDesign version 3.0 (Costanzo's Ph.D. thesis). In this work, we focused to improve the computational design strategy with a higher version of RosettaDesign (version 3.4).

4. Purify select double-point hSOD1 mutants and experimentally evaluate the RosettaDesign recommendations of second-site mutations.
    a. Compare the stability of the double-point hSOD1 mutants with that of $hSOD1^{WT}$ and $hSOD1^{A4V}$ variants.
    b. Compare the aggregation resistance of the double-point hSOD1 mutants with that of $hSOD1^{WT}$ and $hSOD1^{A4V}$ variants.
5. Structurally justify the reason how the suggested second-site mutations work to stabilize the $hSOD1^{A4V}$ molecules.
6. Obtaining a better understanding of the actual cause of the destabilization of $hSOD1^{A4V}$ protein.

4. Materials and Methods

4.1 RosettaDesign

The computational design was performed with the molecular modeling program RosettaDesign, version 3.4 (Rohl, Strauss, Misura, & Baker, 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. 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.

In this work, for all the design runs, only "fixed backbone" design was performed, in which the chi_1 and chi_2 of side chain rotamers, dependent on Dunbrack's backbone-dependent library (Kaufmann et al., 2010), were allowed to vary one standard deviation away from their most probable values. Each RosettaDesign run requires a backbone to be specified and which remains fixed in the "fixed backbone" calculations performed here. The backbone structures were obtained from high resolution X-ray crystal structure files, determined from wild type or variant hSOD1. Two crystal structures of $hSOD1^{WT}$, PDB ID: 1PU0 ($hSOD1^{WT}_{1PU0-WT}$) and 2C9V ($hSOD1^{WT}_{2C9V-WT}$), were used as RosettaDesign starting structures.

RosettaDesign was then used to carry out "scoring runs" or "design runs". In scoring runs, the program was used to examine scores of specific mutations (e.g. A4V) for comparison to a reference (e.g. WT). In these runs, the backbone and sequence were fixed, and the side chain positions were optimized. Alternatively in "design runs", RosettaDesign was set free to simultaneously optimize side chain positions and amino acid sequence to lower the overall energy score.

4.2 Computational strategy for identifying second-site mutations which alleviate the destabilization imposed by A4V mutation

To maintain the enzymatic activity of hSOD1 proteins, all residues chelated to metal ions (His46, His48, His63, His, His80, Asp83, His120) in both chains were fixed to the natural identity and rotamer with "NATRO" command in resfile (see Appendix I, 2a). Protein side chains were first "relaxed" by maintaining the identity of all amino acids but optimizing side chain rotamers, except for the metal ion chelated residues listed above using "NATAA" resfile (see Appendix I, 2a), and the energy scores were recorded. The A4V mutation was then generated and the energy score were also recorded. All residues with any atom(s) within 5 Å of Val4 in $hSOD1^{A4V}$ were allowed to mutate to any amino acid except cysteine with "fixed backbone" design using "ALLAAwc" resfile (see Appendix I, 2c). During this design phase, RosettaDesign was allowed to mutate as many residues as appear to stabilize the structure. The mutations selected by RosettaDesign in every one of 10 generated design structures were recorded. Double-point mutants (A4V and the other potential candidate second-site mutation X) were then generated and the

energy scores were recorded. The promising double-point mutants were selected for the experimental examinations.

4.3 Construction of expression vectors (done by S. Gregoire in collaborative Kwon lab)

The YEP351-SOD1^{WT} and YEP351-SOD1^{A4V} plasmids were kindly provided by Dr. Joan Valentine (University of California Los Angeles). The mutations were introduced into the YEP351-SOD1^{A4V} plasmid using the QuikChange II site-directed mutagenesis kit (Stratagene, Santa Clara, USA). 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. The sequences of the key mutations were confirmed using DNA sequencing.

4.4 Protein expression and purification

All hSOD1 variants were expressed and purified based on previously described (Hough et al., 2004). The plasmids containing the hSOD1 variants were expressed in *Saccharomyces cerevisiae* EG118 strain lacking the endogenous yeast SOD1 gene. Yeast cells were first grown on a plate containing Yeast Synthetic Drop-Out Media Supplement without Leucine (Sigma-Aldrich, Inc., Saint Louis, USA) for two to three days at 30 ℃, followed by inoculation into 50 mL fresh liquid media containing Yeast Synthetic Drop-out Media Supplement without Leucine and grown for two to three days at 30 ℃ 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) batches and cells were grown for another four to five days under the same conditions until the OD_{600} read about 10.

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 ℃ and 11,000 rpm for 30 min. Ammonium sulfate was then slowly added to 60% saturation to the supernatant while stirring in ice bath and centrifugation was performed again at 4 ℃ 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 hSOD1 were pooled and dialyzed against double distilled H_2O (ddH_2O) at 4 ℃. 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. hSOD1 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 ℃ freezer.

4.5 Far UV Circular Dichroism (CD)

The secondary structure of the protein was detected on a JASCO J-710 spectropolarimeter (Jasco Products Company, Oklahoma City, USA). Proteins were in 20 mM potassium phosphate buffer, pH 7.0. A 300 µL cuvette (Hellma, Plainview, USA) with 0.1 cm path length was used for the measurement, and the parameters were set as follows: data pitch, 0.1 nm; scanning speed, 50 nm/min; response, 2 s; and bandwidth, 2 nm. Each spectrum was an average of four measurements. A blank spectrum containing only buffer was deducted.

4.6 Guanidine Hydrochloride-induced denaturation assay

The stability of hSOD1 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 30 s measurements with response time set at 2 s were averaged and buffer adsorption was deducted to obtain the final value. Since no dependence of the ellipticity of the folded and unfolded states with the GdnHCl concentration was seen, a simpler two-state unfolding Equation (4-1) was used to fit the data (Serrano et al., 1994). The midpoint (*MP*) of transition, which can be used to

compare the apparent stability of wild type and mutant hSOD1s, and unfolding *m* value (slope of transition) can be obtained:

$$\theta = \theta_n - \frac{(\theta_n - \theta_d)e^{m([GdnHCl]-MP)/RT}}{1+e^{m([GdnHCl]-MP)/RT}} \quad (4\text{-}1)$$

where θ is the ellipticity at 218nm and certain GdnHCl concentration, θ_n and θ_d are the fitted ellipticity for the native and denatured states at 218 nm, respectively.

For easier comparison of the denaturation curves among hSOD1 variants, the *CD* ellipticities θ for all the samples were converted to the fraction unfolded (f_u) with the following equation:

$$f_u = \frac{(\overline{\theta_n} - \theta)}{(\overline{\theta_n} - \overline{\theta_d})} \quad (4\text{-}2)$$

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

4.7 EDTA-induced aggregation assay

Aggregation behavior of hSOD1 variant was monitored by Size Exclusion Chromatography (SEC) as a function of incubation time at 37 ℃ with a Tosoh SEC-2000 column (Tosoh Bioscience, Grove City, USA) for $hSOD1^{WT}$ or a Superdex 75 column (GE healthcare) for mutant hSOD1s connected to a Waters e2695 HPLC system (Waters Corporation, Milford, USA). Proteins were buffer exchanged to tris-buffered saline (TBS, 20 mM Tris, 150 mM NaCl, pH 7.4) and the concentration (measured as dimer) was 120 μM. Samples were incubated at 37 ℃ in water bath. 3 mM EDTA was added 10 min after putting the samples in water bath to induce aggregate formation. 25 μL aliquots

were periodically taken from the samples and run through the SEC. All chromatography was performed in TBS with a flow rate of 1 mL/min at room temperature and *UV* adsorption at 280 nm was monitored. To account for small differences in the concentrations of protein stock solutions for the different variants, the chromatograms were normalized with respect to the summit reading of the 0 hr sample for each particular hSOD1 variant.

5. Results and Discussion

5.1 Identifying second-site mutations that computationally stabilize hSOD1^{A4V}

Due to the advantages of time efficiency and budget saving over traditional experimental screening, computational protein design has recently been used more extensively. RosettaDesign is a program with many documented design successes (Dantas, Kuhlman, Callender, Wong, & Baker, 2003; Dantas et al., 2007) and an easy to use interface for non-experts. We used RosettaDesign in an effort to computationally identify second-site mutations that can thermodynamically stabilize hSOD1^{A4V} molecule. In this work, initial design protocol was set up by Joe Costanzo with RosettaDesign version 3.0 (Costanzo's Ph.D. thesis).

There are three important issues to address before employing RosettaDesign. The first is which design method to use. The second is what starting structure ("design target") to use. The third is what criteria will be used to screen the results.

In terms of methodology, previous work has shown that "fixed backbone" design was able to identify second-site mutations that can greatly increase the protein-protein binding affinity at the interface (Sammond et al., 2010). In addition, multiple globular proteins have been successfully designed (Dantas et al., 2003) and stabilized (Dantas et al., 2007). Since "flexible backbone" design requires much greater computational resources, "fixed backbone" design was performed in this work.

In choosing a starting structure for design, it is important to avoid using crystal structures with low resolution, or alternate conformations, etc., as large experimental

errors (e.g. that are common with PDB files with resolution of 2 Å or more) can result in steric clashes. Such artifacts can lead to RosettaDesign making corresponding artifactual design changes to relieve such clashes. Three structures available in the PDB were considered for use as targets for hSOD1 "fixed backbone" design: two wild type entries (1PU0 and 2C9V), and one structure of the A4V variant (1UXM).

$\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{1PU0/2C9V-WT}}$ (WT sequence with backbones determined from crystallography with 1PU0 (chain A and B) and 2C9V (chain A and F)) and $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1UXM-A4V}}$ (A4V sequence with backbones determined from crystallography with 1UXM (chain A and B)) were scored. The top section of Table 1 shows that the Rosetta scores for the three files were substantially different. The much higher scores for $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{2C9V-WT}}$ and $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1UXM-A4V}}$ were due to the presence of steric clashes present in the structures, revealed by high overall Lennard-Jones repulsion (LJ-rep) contributions to the energy score (200+ units). The location of such clashes can be identified by elevated residue LJ-rep scores. For example, in $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{2C9V-WT}}$, an alternate conformation at residue 1 in chain F resulted in a near 200 units contribution to the overall LJ-rep score (see Table 1). Such alternate conformations are not uncommon in surface exposed loops or N- or C-termini, which may be very flexible in a crystal. Such flexibility can result in low electron density and high local uncertainty in atomic positions, which the crystallographer can acknowledge by including alternate positions consistent with the uncertain data. Eliminating one of two alternate conformations reduced the overall LJ-rep score of to 103.5, bringing the Rosetta energy score for $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{2C9V-WT}}$ to -553.6, very comparable to that of $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{1PU0-WT}}$. 1UXM, however, was a lower resolution (1.9 Å) structure than 2C9V (1.07 Å) or 1PU0 (1.7Å), with clashes in multiple

hSOD1 variant	ΔG kcal/mol	LJ-rep Kcal/mol	LJ-rep contribution of Ala1 in Chain F/kcal/mol
$hSOD1^{WT}_{1PU0-WT}$	-538.7	118.0	-
$hSOD1^{WT}_{2C9V-WT}$	-141.5	500.5	198.1
[1]$hSOD1^{WT}_{2C9V-WT}$	-553.6	103.5	0.04
$hSOD1^{A4V}_{1UXM-A4V}$	-463.4	204.2	-
$hSOD1^{A4V}_{1PU0-WT}$	-321.6	348.6	-
[1]$hSOD1^{A4V}_{2C9V-WT}$	-303.9	362.2	-

[1]clash caused by the uncertain conformation of Ala1 in chain F was eliminated

Table 1. Overall Rosetta score, scoring of LJ-rep term and residue contribution to LJ-rep term of potential starting structures and A4V variant hSOD1. The potential starting structures were scored with "NATAA" resfile (see Appendix I, 2a), the results are shown in the top section of the table. The bottom section lists the scores of A4V variant generated with "PIKAA" resfile (see Appendix I, 2b) and wily type backbones.

locations that could not so easily be ascribed to specific artifacts. Not surprisingly, $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1UXM-A4V}}$ yielded much higher overall energy scores. Consequently, only the two wild type structures were used for design runs—$\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{1PU0-WT}}$ and the "modified" $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{2C9V-WT}}$ in which the clash mentioned above was eliminated.

With $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{1PU0/2C9V-WT}}$ as the starting structures, "fixed backbone" design was performed to generate the A4V sequence with wild type backbones ($\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1PU0/2C9V-WT}}$) and the score was listed in the bottom section of Table 1. Since Rosetta scoring units are usually taken to be approximately equal to 1 kcal/mol of protein unfolding free energy (ΔG), the dramatic score increase of $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1PU0/2C9V-WT}}$ compared to $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{1PU0/2C9V-WT}}$ can be thought of as a $\Delta\Delta G_{unf}$ for the mutation A4V. Since a 200+ scoring unit $\Delta\Delta G$ was seen for both wild type backbones, this consistent, very large score increase reflects a large destabilization imposed by A4V mutation—when the backbone is constrained to be fixed. Physically, 200+ scoring units $\Delta\Delta G$ indicates that $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1PU0/2C9V-WT}}$ cannot maintain the wild type backbone under physiological conditions.

The steric clash caused by the substitution of the alanine for a valine in $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1PU0/2C9V-WT}}$ molecule should be responsible for the dramatic ΔG increase. In reality, however, $\mathrm{hSOD1}^{\mathrm{A4V}}$ may assume a slight different backbone from $\mathrm{hSOD1}^{\mathrm{WT}}$ to accommodate the steric clash, which is consistent with previous observation that though considerable similarity is observed between $\mathrm{hSOD1}^{\mathrm{A4V}}$ and $\mathrm{hSOD1}^{\mathrm{WT}}$, there is shift in the monomer-monomer orientation and opening of the dimer interface of $\mathrm{hSOD1}^{\mathrm{A4V}}$ (Cardoso, 2002; Hough et al., 2004)

It is seen in Table 1 that despite the clash present due to low resolution in $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1UXM-A4V}}$ structure, the LJ-rep term of $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1PU0/2C9V-WT}}$ is still 100+ scoring units larger than that of $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1UXM-A4V}}$, while all other terms are relatively the same (not shown). For well packed native proteins, the total LJ-rep term works out to be 0.5-1 scoring units per amino acid. However, when the A4V mutation is made and scored on a wild type backbone, the total LJ-rep term for the entire protein increases by 200+ units. A careful inspection of the individual residue contribution to the LJ-rep term of $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1PU0/2C9V-WT}}$ was performed trying to identify if there were specific residues responsible for the increase of the LJ-rep term. Not surprisingly, as shown in Table 2, the Val4 had a LJ-rep contribution of 50+ units in both chains. Except for Val4, Phe20 was found to be the only other residue that had over 50 units contribution to the LJ-rep term in both chains. The LJ-rep contributions from all other residues were less than 3 units with most of them in between of 0.5 to 1. Notably, despite the clash present in the structure elsewhere, $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1UXM-A4V}}$ still shows an elevated LJ-rep term in Val4 and Phe20 (3+ units).

PyMOL-visualized structures around Ala4/Val4 and Phe20 in $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{1PU0-WT}}$, $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1PU0-WT}}$ and $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1UXM-A4V}}$ are shown in Fig. 1A-C. Both the score and structure comparisons indicate that the steric clash between Val4 and Phe20 should be the main cause of the increase of LJ-rep term of $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1PU0/2C9V-WT}}$. Therefore, it is reasonable to believe that by mutating Phe20 to other residues, there is a high chance that the steric clash could be eliminated and thus the $\mathrm{hSOD1}^{\mathrm{WT}}$ backbone (or very close to) can be maintained by the introduction of the second-site mutation. We chose to test this

hSOD1 variant	Total LJ-rep	Rosetta score of LJ-rep term/kcal/mol				
		Ala4	Val4	Phe20	Gly20	Ala20
$hSOD1^{WT}_{1PU0-WT}$	118.0	0.29	-	0.55	-	-
$hSOD1^{WT}_{2C9V-WT}$	103.5	0.33	-	0.58	-	-
$hSOD1^{A4V}_{1UXM-A4V}$	204.2	-	3.6	3.1	-	-
$hSOD1^{A4V}_{1PU0-WT}$	348.6	-	56.6	57.5	-	-
$hSOD1^{A4V}_{2C9V-WT}$	362.2	-	65.9	65.8	-	-
$hSOD1^{A4V/F20G}_{1PU0-WT}$	120.8	-	1.0	-	0.28	-
$hSOD1^{A4V/F20G}_{2C9V-WT}$	104.5	-	1.3	-	0.35	-
$hSOD1^{A4V/F20A}_{1PU0-WT}$	352.0	-	58.7	-	-	58.0
$hSOD1^{A4V/F20A}_{2C9V-WT}$	352.6	-	64.5	-	-	63.5

Table 2. Scoring of LJ-rep term of starting structures and RosettaDesigned hSOD1s and individual residue contribution to LJ-rep term. The top section lists the scoring of starting structures. The scoring of promising double-point variant hSOD1 is listed in the bottom section.

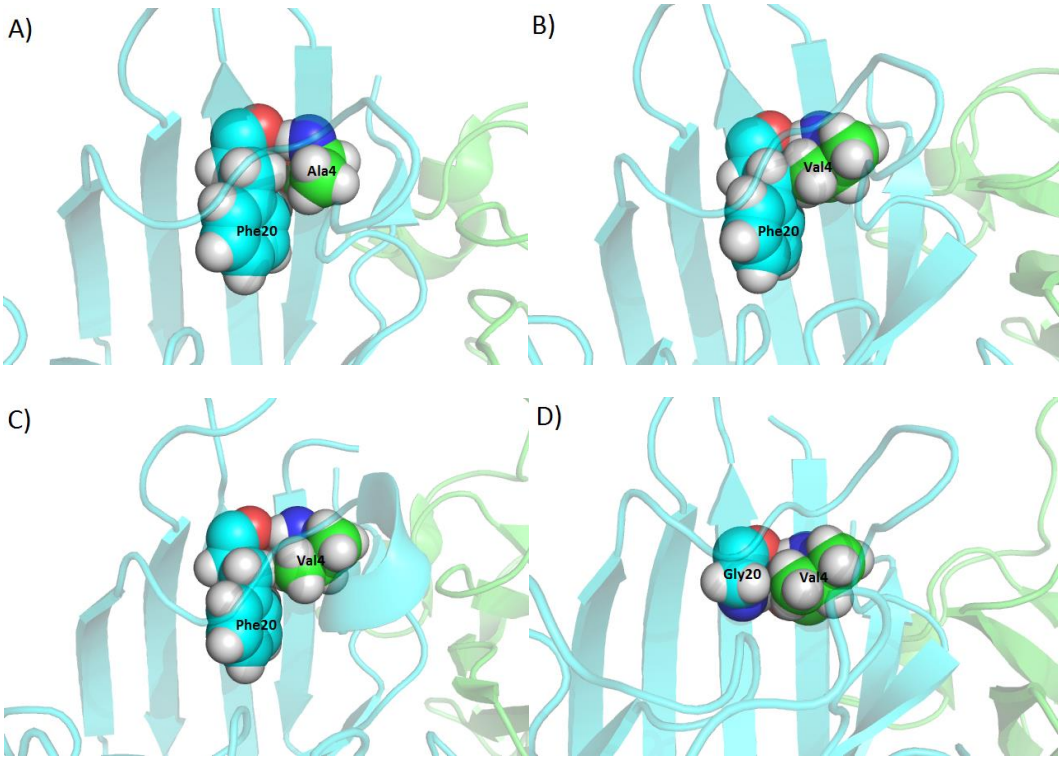

Fig. 1. PyMOL-visualized structures around interested residues in hSOD1 molecules. A) Ala4 and Phe20 in $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{1PU0-WT}}$, B) Val4 and Phe20 in $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1PU0-WT}}$, C) Val4 and Phe in $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1UXM-A4V}}$, D) Val4 and Gly20 in $\mathrm{hSOD1}^{\mathrm{A4V/F20G}}_{\mathrm{1PU0-WT}}$.

hypothesis two steps further in a collaboration with the Kwon lab: 1) by using RosettaDesign and rational design approaches to identify residues which could relieve the clash caused by the inserted valine, 2) by testing variants *in vitro* and in cultured cells.

Since the most affected residues by the A4V mutation should be those around Val4 in hSOD1^{A4V} molecule, any residues containing atom(s) within 5 Å of Val4 would be identified and allowed to mutate during the computational design. A good second-site mutation should be able to "fix" the steric clash by decreasing the LJ-rep term in $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1PU0/2C9V-WT}}$ molecule to a comparable value of that of $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{1PU0/2C9V-WT}}$. Since this improvement would be dramatic, a good second-site mutation should always be favored by RosettaDesign, regardless how many design runs would be carried out. Therefore, to rule out mutations designed to help better "fit" the good second-site mutations, 10 designed structures, each obtained by keeping the lowest score of 10 design runs, would be generated and only the mutations occurring in all the 10 structures would be considered as candidates and later introduced to $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1PU0/2C9V-WT}}$ to generate $\mathrm{hSOD1}^{\mathrm{A4V/X}}_{\mathrm{1PU0/2C9V-WT}}$.

$\Delta\Delta G_{\mathrm{A4V/X\text{-}WT}}$=($\Delta G$($\mathrm{hSOD1}^{\mathrm{A4V/X}}_{\mathrm{1PU0/2C9V-WT}}$)-$\Delta G$($\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{1PU0/2C9V-WT}}$)) is thought to be a good computational screening criterion because it directly indicates how thermodynamically close a variant sequence is to the wild type sequence when scored on the same backbone. Since the design purpose is to recover the wild type-like stability by the introduction of second-site mutation to hSOD1^{A4V} molecule, if the value of $\Delta\Delta G_{\mathrm{A4V/X\text{-}WT}}$ is small, it is reasonable to believe that experimentally the hSOD1$^{A4V/X}$ would assume a wild type-like backbone and have a close stability to hSOD1^{WT}.

With all the three issues resolved, computational design was carried out on the 17 residues (Thr2, Lys3, Val5, Ala6, Ile18, Asn19, Phy20, Glu21, Val29, Leu106, Ile112, Ile113, Ile149, Gly150, Ile151, Ala152, Gln153) which were identified to contain atom(s) within 5Å of the Val4 in chain A of hSOD1^{A4V} crystal structure. The 17 residues were allowed to mutate to any amino acid except cysteine. The same residues in chain B/F were also allowed to mutate in the design runs. 10 redesigned structures were generated. The amino acid sequences of these 10 multi-point mutant hSOD1s were compared to that of hSOD1^{A4V} and the mutations occurring in all 10 structures were shown in the middle section of Table 3. When generating $\mathrm{hSOD1}^{\mathrm{A4V/X}}_{\mathrm{1PU0/2C9V-WT}}$, for the mutations that occurred only in one of the two chains, the same mutation was made for the other chain. After comparing the scores of the 10 generated structures of each $\mathrm{hSOD1}^{\mathrm{A4V/X}}_{\mathrm{1PU0/2C9V-WT}}$, only the lowest score was listed in Table 3, together with $\Delta\Delta G_{\mathrm{A4V/X\text{-}WT}}$ values.

Among all $\mathrm{hSOD1s}^{\mathrm{A4V/X}}_{\mathrm{1PU0/2C9V-WT}}$ listed in Table 3, $\mathrm{hSOD1}^{\mathrm{A4V/F20G}}_{\mathrm{1PU0/2C9V-WT}}$ stands out immediately because its $\Delta\Delta G_{\mathrm{A4V/X\text{-}WT}}$ values are 100+ scoring units smaller compared with other variants. The F20G mutation was identified as eliminating most of the $\Delta\Delta G_{\mathrm{A4V/X\text{-}WT}}$ penalty imposed by A4V. Table 3 shows the $\Delta\Delta G_{\mathrm{A4V/F20G\text{-}WT}}$ is very small with both starting structures—2.5 and 8.0 kcal/mol for $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{1PU0-WT}}$ and $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{2C9V-WT}}$, respectively, which indicates a close thermodynamic stability of $\mathrm{hSOD1}^{\mathrm{A4V/F20G}}_{\mathrm{1PU0/2C9V-WT}}$ to $\mathrm{hSOD1}^{\mathrm{WT}}_{\mathrm{1PU0/2C9V-WT}}$. The inspection of the individual residue contribution to the LJ-rep term showed that the dramatically increased LJ-rep terms of Val4 and Phe20 in $\mathrm{hSOD1}^{\mathrm{A4V}}_{\mathrm{1PU0/2C9V-WT}}$ were dropped back to less than a few scoring unites in $\mathrm{hSOD1}^{\mathrm{A4V/F20G}}_{\mathrm{1PU0/2C9V-WT}}$ as shown in the bottom section of Table 2. Fig. 1D is a

hSOD1 variant	[1]1PU0		[1]2C9V	
	ΔG kcal/mol	[2]$\Delta\Delta G$ kcal/mol	ΔG kcal/mol	$\Delta\Delta G$ kcal/mol
$hSOD1^{WT}$	-538.7	0.0	-553.6	0.0
$hSOD1^{A4V}$	-321.6	217.0	-303.9	249.7
$hSOD1^{A4V/T2L}$	-326.3	212.4	-	-
$hSOD1^{A4V/K3R}$	-	-	-307.0	246.6
$hSOD1^{A4V/C6A}$	-326.2	212.5	-307.6	246.0
$hSOD1^{A4V/N19W}$	-329.9	208.8	-	-
$hSOD1^{A4V/N19F}$	-	-	-310.4	243.2
$hSOD1^{A4V/F20G}$	-536.1	2.5	-545.6	8.0
$hSOD1^{A4V/I151Y}$	-325.0	213.7	-306.0	247.6
$hSOD1^{A4V/A152H}$	-	-	-305.5	248.1
$hSOD1^{A4V/Q153G}$	-323.5	215.1	-	-
[3]$hSOD1^{A4V/F20A}$	-310.6	228.1	-303.4	250.2
[3]$hSOD1^{A4V/F20V}$	-293.3	245.4	-287.5	266.1
[3]$hSOD1^{A4V/F20I}$	-292.8	245.8	-291.7	261.9
[3]$hSOD1^{A4V/F20L}$	-419.1	119.6	-412.4	141.2
[3]$hSOD1^{A4V/F20M}$	-416.3	122.3	-402.6	151.0

[1]PDB IDs of Wild-Type starting structures used for RosettaDesign

[2]calculated by $\Delta\Delta G_{A4V/X\text{-}WT}=(\Delta G(hSOD1^{A4V/X}_{1PU0/2C9V-WT})-\Delta G(hSOD1^{WT}_{1PU0/2C9V-WT}))$

[3]Double-point mutations obtained from rational design

Table 3. Overall RosettaDesign energy scores of potential starting structures and potential second-site variant hSOD1s. All mutants were generated by the "PIKAA" command (see Appendix I, 2b). Prior to design, the alternate conformation for residue 1 (labeled AALA/F/1) in the PDB file of 2C9V was removed to relieve a clash with residue 2 as described in the text. The potential starting structures were scored as shown in the top section of the table. The middle section of the table lists all the residues that were identified by RosettaDesign in every one of 10 generated design structures. Each one of the designed residues listed was separately scored as a double mutation with A4V, and the resulting energy score is shown. In the bottom section of the table, potential second-site site mutations suggested by rational design are listed and scored. All scores were produced with the same protocol detailed in Appendix I.

PyMOL-visualized structure around Val4 and Gly20 in $\mathrm{hSOD1}^{\mathrm{A4V/F20G}}_{\mathrm{1PU0-WT}}$. When comparing it with Fig. 1C, it clearly shows that the steric clash between Val4 and Phe20 is removed. The computational result confirmed the initial guess that the steric clash introduced by A4V mutation in $\mathrm{hSOD1}^{\mathrm{A4V}}$ molecule could be eliminated by mutating Phe20 to a smaller residue.

Based on the computational result, a rational design was performed at the key residue 20. As seen in Fig. 1, since the side chain of residue 20 of hSOD1 is buried in the hydrophobic core, only amino acids with hydrophobic side chains were considered. The side chain should also be smaller than a phenylalanine to avoid the steric clash with Val4. As a result, alanine, valine, isoleucine, leucine and methionine were mutated into residue 20 and scored with RosettaDesign. The scores of these $\mathrm{hSOD1s}^{\mathrm{A4V/X}}_{\mathrm{1PU0/2C9V-WT}}$ were listed in the bottom section of Table 3. Although none of the $\mathrm{hSOD1s}^{\mathrm{A4V/X}}_{\mathrm{1PU0/2C9V-WT}}$ showed a $\Delta\Delta G_{\mathrm{A4V/X\text{-}WT}}$ as promising as $\mathrm{hSOD1s}^{\mathrm{A4V/F20G}}_{\mathrm{1PU0/2C9V-WT}}$ did, we wanted to test the predictions of more fully, since it is known that RosettaDesign is not always right at predicting mutations with lower scoring (Costanzo's unpublished Ph.D. thesis). Thus, based simply on the fact that residue 20 was identified by RosettaDesign as the key residue interacting with Val4, the broader set of smaller hydrophobic amino acid variants (Ala, Val, Ile, Met, Leu,) were still sent to the collaborative lab for the *in vivo* aggregation screening.

The *in vivo* aggregation assay showed that in both HEK293T kidney cells and NSC-34 spinal cells expressing $\mathrm{hSOD1}^{\mathrm{A4V/F20G}}$ and $\mathrm{hSOD1}^{\mathrm{A4V/F20A}}$, aggregate formation was reduced compared to both types of cells expressing $\mathrm{hSOD1}^{\mathrm{A4V}}$ (S. Gregoire's unpublished results) . Consequently, the genes containing the above two double-point

mutations were introduced into the yeast cell plasmids for protein purification and further *in vitro* assays.

5.2 Protein Expression and Purification

The protein yield was about 10 mg from 50 grams of wet yeast cells for $hSOD1^{WT}$ and $hSOD1^{A4V}$ and about 7 mg from 50 grams for $hSOD1^{A4V/F20G}$ and $hSOD1^{A4V/F20A}$. The purity was first estimated from the SEC chromatogram (not shown) of the purification procedure and then evaluated by the SDS-PAGE after the purification, which is shown in Fig. 2A. The single band located above 20 kDa is indicative of hSOD1 and shows that there is high purity for all purified hSOD1 variants. Mass spectrometry (*MS*) was used to verify the molecular mass of all the hSOD1 variants. The deconvolution of *MS* chromatograms gave molecular masses of 15847 Da, 15875 Da, 15782 Da and 15796 Da for $hSOD1^{WT}$, $hSOD1^{A4V}$, $hSOD1^{A4V/F20G}$ and $hSOD1^{A4V/F20A}$, respectively (data not shown). These masses correspond to the expected masses of wild type and mutant hSOD1 monomers devoid of metal ions within the error range of a few Daltons (not shown).

Far *UV CD* (195-240 nm) was used to analyze the secondary structure of the proteins under physiological conditions. Fig. 2B shows the *CD* spectra for wild type and mutant hSOD1s in 20 mM potassium phosphate, pH 7.0. For easier comparison, all the data points were normalized to the lowest *CD* reading of that particular protein. All hSOD1 proteins show a peak with a minimum at 208 nm. The only differences between the spectra were at low wavelengths below 200 nm where cuvette effects increase the uncertainty of the results. Previous work shows the same *CD* spectra for both $hSOD1^{WT}$

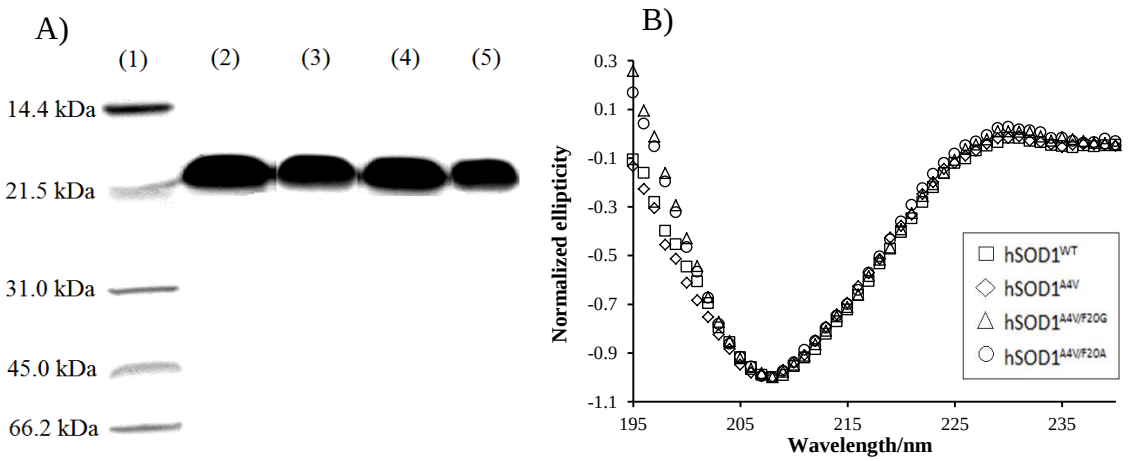

Fig. 2. Characterization of wild type and mutant hSOD1s. A) SDS-PAGE of purified wild type and mutant hSOD1s. B) *CD* spectra of wild type and mutant hSOD1s. In panel A, lane contains standard proteins with the molecular masses marked; lane 2 to lane 4 contains purified hSOD1^{WT}, hSOD1^{A4V}, hSOD1$^{A4V/F20G}$, hSOD1$^{A4V/F20A}$ respectively.

and hSOD1^{A4V} under physiological conditions (Arnesano et al., 2004). Thus, the CD spectra indicate that all hSOD1 variants maintain the native secondary structure after the purification procedures.

5.3 GdnHCl-induced denaturation assay

GdnHCl-induced *in vitro* denaturation assay was carried out to examine the apparent thermodynamic stability of hSOD1$^{A4V/F20G}$ and hSOD1$^{A4V/F20A}$ relative to hSOD1^{WT} and hSOD1^{A4V}. 10-15 μM (dimer) wild type or variant hSOD1 in PBS was mixed with different GdnHCl concentrations. *CD* readings at 218 nm were recorded to measure the percentage of protein become unfolded after 16 hr incubation at room temperature and 100 rpm gentle rotation. For easier comparison among hSOD1 variants with slightly different concentrations used for the denaturation assay, raw *CD* data were converted to fraction unfold (f_u), and plotted against GdnHCl concentrations as shown in Fig. 3A.

To test the reproducibility of the assays, the experiment with hSOD1$^{A4V/F20A}$ was carried out twice. For both hSOD1^{WT} and hSOD1^{A4V}, the second set of data was obtained by S. Gregorie using the same conditions as above. However, there was only one set of data available for hSOD1$^{A4V/F20G}$. All the data points, converted to f_u according to Equation (4-2), were shown in Fig. 3A.

Both hSOD1^{WT} and hSOD1^{A4V} show a single transition consistent with two-state transition behavior, but the transition range shifted to lower GdnHCl concentrations for hSOD1^{A4V}. hSOD1$^{A4V/F20G}$ and hSOD1$^{A4V/F20A}$ also show a typical two-state transition

behavior and the transition ranges of both fall in between that of $hSOD1^{WT}$ and $hSOD1^{A4V}$.

A renaturation assay was done simultaneously with a denaturation assay with $hSOD1^{WT}$ to assess the reversibility of the denaturation process. The result is shown in Fig. 3B. For the renaturation assay, the $hSOD1^{WT}$ sample was incubated in 6 M GdnHCl for 12 hr and then diluted to desired GdnHCl concentrations. Another $hSOD1^{WT}$ sample, of the same concentration, was incubated in PBS buffer during this 12 hr period for the denaturation assay. The denaturation assay was then prepared as before, with the addition of varying concentrations of GdnHCl. Both samples were then incubated for another 16 hr. The result shows that under the same GdnHCl concentrations between 1.6 M and 5.6 M, there was a greater fraction of $hSOD1^{WT}$ unfolded in the renaturation assay than in the denaturation assay. For example, only 3% of $hSOD1^{WT}$ was unfolded for the denaturation assay compared to 75% for the renaturation assay at GdnHCl concentration of 2.4 M; also, at 3.6 M, the percentages were 35% and 87% for denaturation and renaturation assay, respectively.

The different values obtained for the same final denaturant concentration via renaturation and denaturation assay indicates that the denaturation process of $hSOD1^{WT}$ is not fully reversible under the conditions used. Therefore, we cannot treat the chemical denaturation process as one at equilibrium, and consequently, estimating the free energy change was avoided in this work. Instead, we fitted transition middle point (*MP*) value to compare the "apparent stability" among different hSOD1 variants (Clarke & Fersht, 1993). f_u were used to fit for the *MP* values for wild type and mutant hSOD1s and the

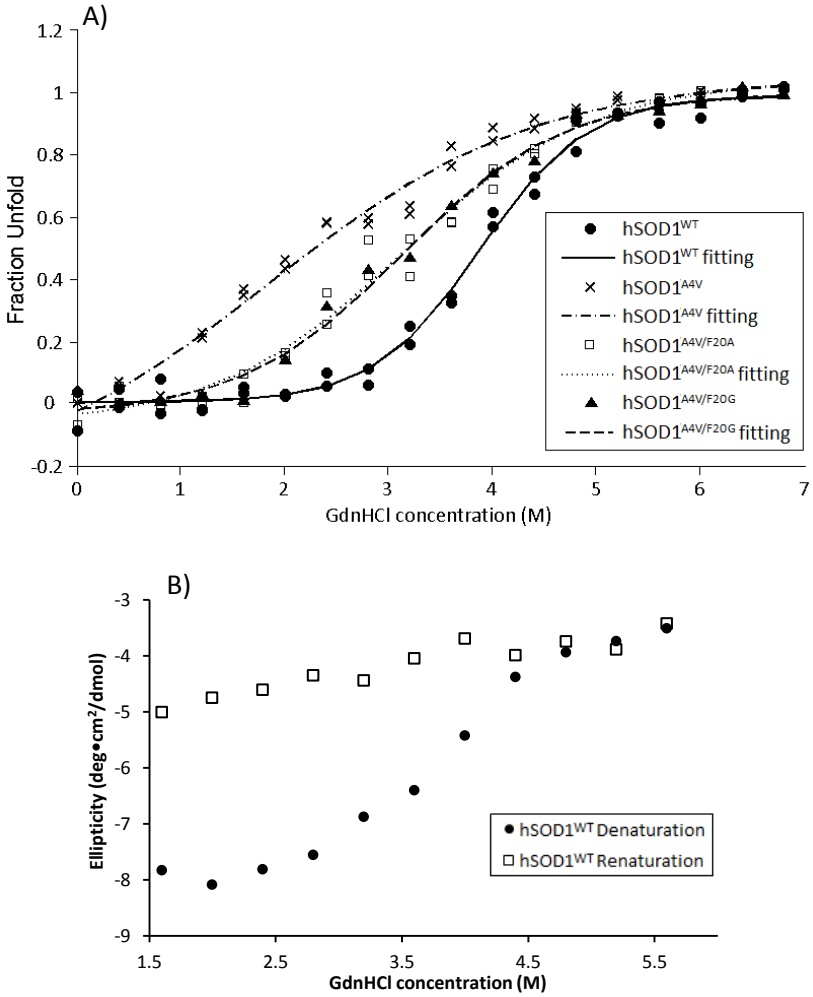

Fig. 3. A) Denaturation curves of wild type and mutant hSOD1s. B) Raw CD data with buffer adsorption subtracted of hSOD1^{WT} denaturation and renaturation assay. In panel A, fraction unfolded (f_u) was calculated according to Equation (4-2) and was plotted against GdnHCl concentrations. f_u were fitted to Equation (4-1) for different hSOD1 variants. The curves were obtained from best fit parameters shown in Table 4.

fitting results were listed in Table 4. The curves in Fig. 3A were obtained from fitting the f_u to Equation (4-1).

The confidence interval for all the fitting results in Table 4 was obtained with a 95% confidence. For hSOD1^{WT}, hSOD1^{A4V} and hSOD1$^{A4V/F20A}$, two sets of data as stated above were used for fitting. For hSOD1$^{A4V/F20G}$, only one set of data were used. Though no concrete conclusion can be drawn because of limited replicates, the preliminary results show that there is no significant difference between hSOD1$^{A4V/F20G}$ and hSOD1$^{A4V/F20A}$. However, both *MP* values are much larger than that of hSOD1^{A4V} and closer to hSOD1^{WT}. This indicates that both hSOD1$^{A4V/F20G}$ and hSOD1$^{A4V/F20A}$ molecules are experimentally more stable than hSOD1^{A4V}. Therefore, the result of the denaturation assay shows that both of the introduced second-site mutations, F20G and F20A, successfully increased the stability of the hSOD1^{A4V} variant.

5.4 EDTA-induced Aggregation Behavior

In vivo aggregation assay done by S. Gregoire in Kwon lab showed reduced aggregate formation in both HEK293T kidney cell lines and NSC-34 spinal cells expressing either hSOD1$^{A4V/F20G}$ or hSOD1$^{A4V/F20A}$ compared to hSOD1^{A4V} (see Appendix II). Although *in vitro* assays cannot replicate all the complicated environmental factors in human cells, at the same time, they provide useful complementary information under more defined conditions. For example, chemical denaturation experiments can conceivably provide information about the effects of mutation on free energies of unfolding, and thereby tell us if our use of RosettaDesign positively affected the thermodynamic stability of the protein. Thus, the *in vitro* aggregation assay with purified

hSOD1 variants	*MP* M	*m* kcal/(mol•M)
$hSOD1^{WT}$	3.88 ±0.09[1]	1.15 ±0.17
$hSOD1^{A4V}$	1.84 ±0.47	0.46 ±0.13
$hSOD1^{A4V/F20G}$	3.13 ±0.20	0.75 ±0.19
$hSOD1^{A4V/F20A}$	3.14 ±0.18	0.65 ±0.14

[1]with 95% confidence intervals determined from the Matlab (version R2010b) used for fitting

Table 4. The two-state unfolding fitting results of *MP* and *m* values for wild type and mutant hSOD1s. Equation (4-1) was used to fit the f_u data. The errors for both *MP* and *m* are regression errors.

wild type and mutant hSOD1s were performed to support the *in vivo* aggregation assay results.

Previous reports have shown that hSOD1^{WT} and hSOD1^{A4V} are extremely stable and resistant to both trifluoroethanol and thermal induced aggregation *in vitro* (Stathopulos et al., 2003). We have also observed that the native dimer peak of SEC of hSOD1^{WT} and hSOD1^{A4V} remain unchanged even after incubation at 37 ℃ for 140 hr (data not shown). It is believed that the loss of metal ions plays an important role in hSOD1 aggregation (Rakhit et al., 2004; Banci et al., 2010). Thus, in order to accelerate the *in vitro* aggregation process of hSOD1 to more practical time scales, 3 mM EDTA was added to induce the aggregate formation.

Fig. 4 shows the SEC chromatograms that indicate effects of mutation on *in vitro* aggregation behavior. For each panel, all data points were normalized relative to the summit reading of the 0 hr incubation sample of that particular protein. The native dimer form of hSOD1 molecules elutes as the peak at 8.7 mL for hSOD1^{WT} with a Tosoh SEC2000 column and 11.4 mL for the other three mutants for which a Superdex 75 column was used. The dimer peak of hSOD1^{WT} remains stable even after 96 hr incubation, while the dimer peak of hSOD1^{A4V} decays along with the increase of incubation time. For hSOD1^{A4V}, some high molecular weight species (HMWS) eluted before the dimer peak is noticeable after 6 hr incubation. After 48 hr incubation, even higher molecular weight species are observed, and the signal intensity for the HMWS increases with the incubation time. For hSOD1$^{A4V/F20G}$ and hSOD1$^{A4V/F20A}$, though dimer peak decay is observed, no HMWS are seen as the situation with hSOD1^{A4V}, even after 96 hr incubation.

Broad and possibly multiple peaks were observed eluting after the dimer peak in the SEC chromatograms of hSOD1$^{A4V/F20G}$ and hSOD1$^{A4V/F20A}$, and other techniques, like light scattering, need to be used to identify these peaks. However, since the peaks absorb at 280 nm, they appear to be protein. Further, total protein mass calculated as sum of the all the integrated peak areas was well conserved for different time points for each variant, with the variation of less than 10% (data not shown). Thus, we hypothesize that, the peaks eluting after the native dimer peak for hSOD1$^{A4V/F20G}$ and hSOD1$^{A4V/F20A}$ may be partially folded or unfolded dimer and/or monomer species that may be exhibiting some affinity for the SEC resin.

The coordination with metal ions of hSOD1 molecule is believed important for not only acquiring hSOD1 activity but also maintaining the native structure (Strange et al., 2006; Khare, Caplow, & Dokholyan, 2004; Doucette et al., 2004). Especially, previous work has shown that zinc ion plays an important role in the maintaining of hSOD1 native structure (Roberts et al., 2007). Therefore, since EDTA was added in the aggregation assay, the metal ions may be extracted from native dimer, which may cause the protein molecule to unfold. Protein unfolding would expose the hydrophobic patches buried inside of the folded native dimer. These hydrophobic patches may have interactions with the resin in the SEC column and hence cause the protein to elute later than the native dimer. Therefore, the peaks eluting later than the native dimer may be the indications of the spread of different unfolded stages of hSOD1$^{A4V/F20G}$ and hSOD1$^{A4V/F20A}$ molecules along the unfolding pathway from loss of metal ions to unfolded monomers.

Since no HMWS are observed in the SEC chromatograms of hSOD1$^{A4V/F20G}$ and

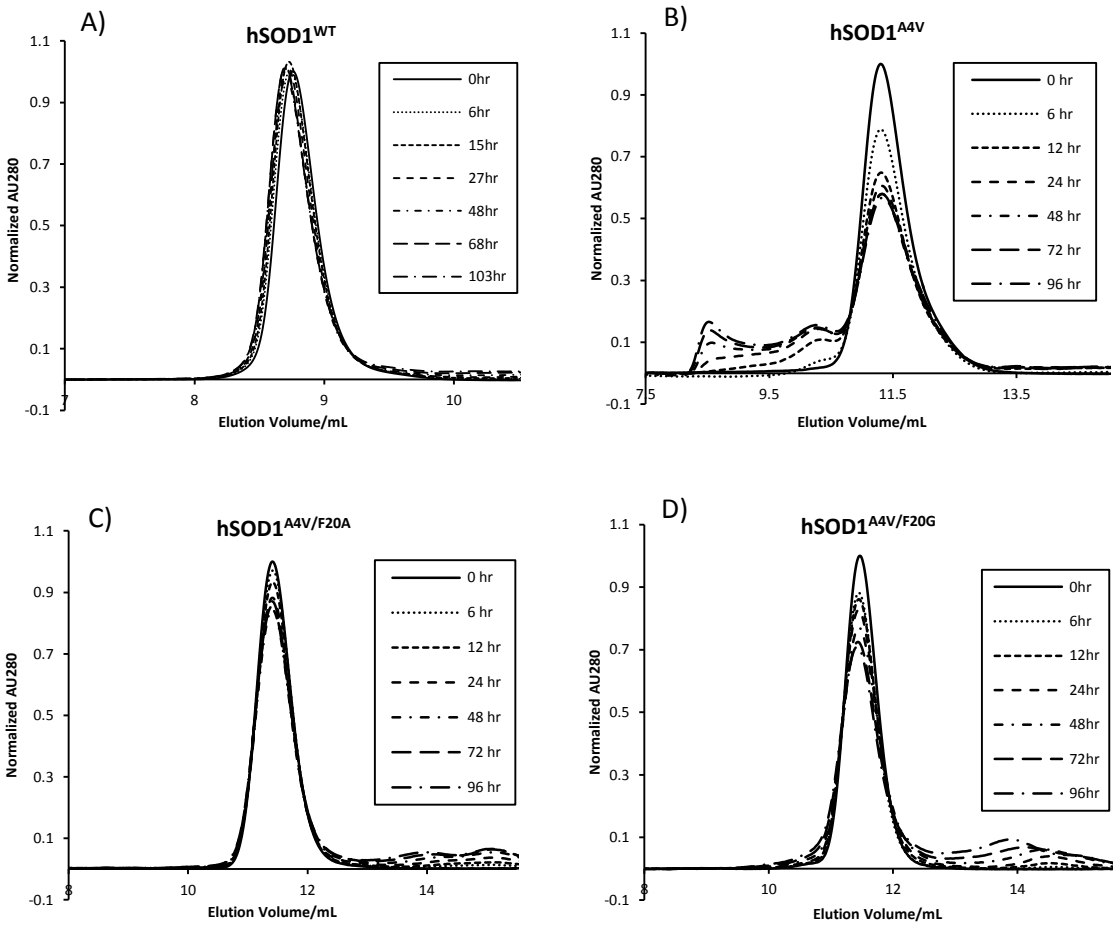

Fig. 4. SEC chromatograms of EDTA-induced aggregation assay. A) SEC chromatograms of hSOD1^{WT} with a Tosoh SEC2000 column, B-D) SEC chromatograms of hSOD1^{A4V}, hSOD1$^{A4V/F20A}$ and hSOD1$^{A4V/F20G}$ with a Superdex G75 column, respectively. UV adsorption at 280 nm was used to detect protein elution. TBS was used as mobile phase and the flow rate was 1mL/min. For each panel, all data points were normalized relative to the summit reading of 0 hr incubation sample of that particular protein.

hSOD1$^{A4V/F20A}$, the dimer fraction decay curve, as shown in Fig. 5, was plotted to show the aggregation resistance of wild type and mutant hSOD1s. Fig. 5 shows that the hSOD1^{WT} dimer fraction remains unchanged along the incubation time course, while the hSOD1^{A4V} dimer decays fast for the first 12 hr incubation and only 64% of dimer fraction remains after 12 hr incubation. However, the dimer decay slows down after 24 hr incubation and after 96 hr incubation, 57% dimer fraction still remains. hSOD1$^{A4V/F20G}$ and hSOD1$^{A4V/F20A}$ fall in between of hSOD1^{WT} and hSOD1^{A4V}. A fast phase of dimer decay is also observed for the first 12 hr and 24 hr for hSOD1$^{A4V/F20G}$ and hSOD1$^{A4V/F20A}$, respectively. After 96 hr incubation, 71% and 85% dimer fraction remains for hSOD1$^{A4V/F20G}$ and hSOD1$^{A4V/F20A}$, respectively. Previous results (Banci et al., 2008; Furukawa, Kaneko, Yamanaka, O'Halloran, & Nukina, 2008) have shown that the aggregate formation of hSOD1 molecules is a very complicated process and a detailed analysis of the kinetic mechanism of aggregate formation was not the focus of this work. Nonetheless, though no replicates of the assay were performed, the result of EDTA-induced aggregation assay strongly indicates that although hSOD1$^{A4V/F20G}$ and hSOD1$^{A4V/F20A}$ behave not as aggregation resistant as hSOD1^{WT}, they are more aggregation resistant than hSOD1^{A4V}. This result correlates well with the result of *in vivo* aggregation assay (Gregoire's unpublished results, see Appendix II), which is a strong indication that the aggregation propensity introduced by the A4V mutation in hSOD1^{A4V} is successfully alleviated by the second-site mutation of F20G and F20A.

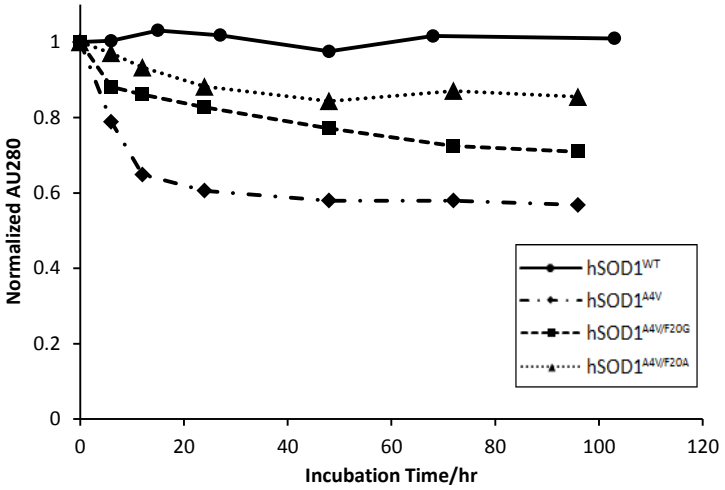

Fig. 5. Decay of native dimer of wild type and mutant hSOD1s. hSOD1 samples were incubated in TBS at 37 ℃ and 25 μL samples were periodically taken for SEC analysis. The native dimer peak summit reading of each SEC chromatogram corresponding to each incubation time in Fig. 4 was used as the dimer fraction. Data points have been connected by lines for clarity.

5.5 In vitro experiments show apparent "stability" and aggregation behavior which agrees with cellular assays

Both *in vitro* denaturation and aggregation assays showed that the added second-site mutation F20G or F20A successfully alleviated the destabilization and aggregation propensity of $hSOD1^{A4V}$. The results of *in vitro* assays were in good agreement with the results of *in vivo* aggregation assays done in both HEK293T kidney cell lines and NSC-34 spinal cells, where both the introduced second-site mutations were shown to help reduce the aggregate formation of $hSOD1^{A4V}$ (Gregoire's unpublished results, see Appendix II). This indicates that while other elements in the cellular assay could play a role (variable or different redox state, involvement of foldases or chaperone protein, et al.), just the elements presented in the *in vitro* assay are enough to qualitatively produce the same irreversible behavior observed in the *in vivo* assay.

5.6 Why did RosettaDesign not adequately flag F20A?

Although experimentally both the second-site mutation F20G and F20A were shown to suppress the instability and aggregation prone behavior of $hSOD1^{A4V}$, RosettaDesign only succeeded in identifying the F20G mutation. The F20A mutation did not occur in any of the 10 structures generated with the 5 Å design around Val4 (data not shown).

A possible explanation for RosettaDesign overlooking the F20A mutant could be the "fixed backbone" design method used here. With the assumption that the backbone was unchanged from $hSOD1^{WT}$, RosettaDesign could find side chain rotamers that could accommodate valine inserted in residue 4 when glycine replaced Phe20. However, the

combined constraints of a fixed backbone unchanged from hSOD1^{WT} and making single amino acid changes appears to have prevented RosettaDesign from identifying side chain rotamers that could accommodate valine inserted at position 4 when alanine rather than glycine replaced Phe20. It may be that allowing for further mutations (Dantas et al., 2003) or the use of "flexible backbone" design would enable RosettaDesign to productively identify improvements like hSOD1$^{A4V/F20A}$.

We hypothesize that the backbone of hSOD1$^{A4V/F20G}$ is very close to hSOD1^{WT}, while hSOD1$^{A4V/F20A}$ may assume a backbone in between of hSOD1^{WT} and hSOD1^{A4V}. In this hypothesized structure, in order to fit the Ala20, Val4 slightly "pushes" the backbone a little away from hSOD1^{WT} backbone. The slight difference in backbone structure between hSOD1$^{A4V/F20A}$ and hSOD1^{WT} is enough to fit the alanine but not big enough to cause as large a decrease in stability of the molecule as hSOD1^{A4V} does. In order to determine backbone adopted by hSOD1$^{A4V/F20A}$, a much more CPU-intensive "flexible backbone" design may be performed with RosettaDesign, but was not pursued here because of the extensive computational resources required.

5.7 The steric clash between Val4 and Phe20 might be the actual cause of the adverse effects of hSOD1^{A4V}.

Although previous structure analysis has narrowed the dominant amino acids which are responsible for the destabilization of hSOD1^{A4V} down to several amino acids (Cardoso, 2002), by mutating Phe20 to a smaller and less hydrophobic glycine or alanine and hence alleviated the adverse effects caused by A4V mutation, our results suggest that the steric clash between Val4 and Phe20 is the actual cause of the destabilization of

hSOD1^{A4V} molecule. In hSOD1^{A4V} molecule, the local area around Val4 and Phe20 is "expanded" to accommodate the added two methyl groups and therefore causes the local backbone alteration from hSOD1^{WT}. This local conformation change in turn causes the destabilization of the dimer interface, as reported by several groups (Cardoso et al., 2002; DiDonato et al., 2003; Hough et al., 2004).

5.8 Why did hSOD1$^{A4V/F20A}$ perform a little better than hSOD1$^{A4V/F20G}$?

In Fig. 5, the hSOD1$^{A4V/F20A}$ variant shows a somewhat slower initial rate of aggregation and higher fraction of native dimer at long times. This would suggest that kinetically, at least, the hSOD1$^{A4V/F20A}$ variant is somewhat less aggregation prone than the hSOD1$^{A4V/F20G}$ variant.

A possible explanation for the observation that hSOD1$^{A4V/F20A}$ was slightly more resistant against both irreversible denaturation and aggregation than hSOD1$^{A4V/F20G}$ may be because compared to hSOD1$^{A4V/F20G}$, there are more hydrophobic interactions in the hydrophobic core around Val4 in hSOD1$^{A4V/F20A}$ variant. When an alanine is mutated to a valine, two methyl groups are introduced, which would add buried hydrophobicity to the hSOD1^{A4V} variant. However, mutating a highly hydrophobic phenylalanine to a non-hydrophobic glycine would lose even more buried hydrophobicity. Therefore, compared to hSOD1^{WT}, the net effect of A4V and F20G mutations may cause hSOD1$^{A4V/F20G}$ to lose some buried hydrophobicity. The inspection of the Rosetta score terms shows that the Lennard-Jones attraction (LJ-attr) term of hSOD1$^{A4V/F20G}$ is about 10 kcal/mol more positive than that of hSOD1^{WT} and hence supported the loss of buried hydrophobicity in the hSOD1$^{A4V/F20G}$ variant. However, by adding one methyl group to Gly20, more

hydrophobicity would be expected in the hSOD1$^{A4V/F20A}$ variant. Other more hydrophobic second-site mutation F20V, F20I, F20L and F20M would contribute more to the LJ-attr term. As listed in Table 3, though RosettaDesgin does not strongly favor any other mutations other than F20G, F20L and F20M are the closest in energy to F20G. The reason that except F20A, other mutations did not pass the *in vivo* aggregation assay screening may be because the side chain of these residues is too big to fit in the hydrophobic core. Therefore, the hypothesis is that compared with F20G and other hydrophobic mutations, F20A may be a better balance in which the side chain is small enough to fit in the hydrophobic core and more buried hydrophobicity is added to further stabilize the molecule.

5.9 Can multi-point mutant hSOD1 further improve the score over hSOD1$^{A4V/F20G}$?

To test if RosettaDesign could identify an A4V-containing variant if more mutations were allowed, the full cluster of mutations found to be beneficial in Table 3, were scored as multi-point hSOD1 variant. In addition, all triple mutants involving the proposed mutations shown in Table 3 were also constructed and scored. The results were shown in Table 5.

The cluster variant does even better than hSOD1^{WT} with over 10 kcal/mol improvement! Such an achievement is more relevant to aggregation of protein pharmaceuticals, where we could consider engineering the protein than for human disease, where it is hard to “re-engineer” proteins in the body. For this reason alone, it would be of interest in future work to express the multi-point hSOD1 mutant and test the *in vitro* stability and aggregation-resistant behavior. However, while the mutations listed in Table

5 predicted by RosettaDesign might indeed improve the thermodynamic stability, it may turn out that their hydrophobic, surface exposed nature might actually exacerbate aggregation through kinetic traps that slow folding kinetics, and/or stabilize associated states.

hSOD1 variant[1]	*ΔG*/kcal/mol
$hSOD1^{WT}$	-538.7
$hSOD1^{A4V/F20G/T2L}$	-538.9
$hSOD1^{A4V/F20G/C6A}$	-538.7
$hSOD1^{A4V/F20G/N19W}$	-542.5
$hSOD1^{A4V/F20G/I151Y}$	-537.7
$hSOD1^{A4V/F20G/Q153G}$	-536.1
hSOD1(cluster)[2]	-550.4

[1]all the variants were generated with PDB ID 1PU0 as starting structure
[2]this variant includes all the mutations listed above it

Table 5. Scoring of RosettaDesigned multi-point hSOD1 variants.

6. Conclusions

In this work, though limited replicate assays were performed, the results were very strong indication that certain mutants can "suppress" the effects of the destabilizing effects of the deadly A4V mutation in hSOD1 associated with fALS. In particular, F20G and F20A were shown to suppress much of the apparent reduction in stability as well as reducing the rate of aggregation. These two mutants were identified by a cellular assay, which subjected a number of residues neighboring residue 20 to testing. As an alternative, we showed that a computational protein design program, RosettaDesign, identified F20G as the most effective second-site mutation for restoring the thermodynamic stability of the protein. *In vitro* measurements showed that the added F20G and F20A both improved apparent hSOD1^{A4V} stability to chemical denaturation key to the destabilization. Though less elegant and definitive than the cellular assay, RosettaDesign calculations could be completed in far less time, and while they did not score specific mutations quantitatively, it did successfully identify residue 20 as key to the destabilization, suggesting that intradomain destabilization may be the dominant interaction that drives the destabilization of hSOD1^{A4V} variant. The ability of the approach to identify mutations that can increase apparent stability and reduce aggregation also suggests this computational design approach could be useful for biopharmaceutical development.

7. Recommendations

The success of the computational design/simulation tool at revealing potential molecular mechanisms for the action of mutations important in human disease, suggests that this approach should be explored as a general strategy to identify the cause of other mutations in hSOD1 molecule and even other proteins responsible for other protein misfolding/aggregation diseases, such as Alzheimer's, Parkinson's and the prion diseases.

Though the introduction of the second-site mutation F20G and F20A showed promising results, both the *in vitro* and *in vivo* assays suggested that $hSOD1^{A4V/F20G}$ and $hSOD1^{A4V/F20A}$ did not behave as well as $hSOD1^{WT}$. More mutations, likely in close proximity to the Val4, must be introduced to further improve the performance of either $hSOD1^{A4V/F20G}$ or $hSOD1^{A4V/F20A}$. As seen in Table 5, several multi-point mutants already have better scores than $hSOD1^{WT}$. Besides, since Table 3 only listed the mutations that occurred in all 10 generated structures during the design runs around 5 Å of Val4 in $hSOD1^{A4V}$, it would be helpful to include all the mutations suggested by RosettaDesign to the candidate mutation list, even for the mutations that occurred only once. The theory behind this is that more mutations in the vicinity of Val4 can help add more interactions in the local structure and therefore further improve the stability of the local conformation.

However, discretion should be taken about what the mutations are and where they occur. For example, as shown in Table 5, N19W/F involves mutating a polar side chain to a highly hydrophobic side chain on the surface of the molecule. It may still be worth a try, however, as the mutation clearly adds to beneficial hydrophobic contacts with other

neighboring residues (LJ-attr effects and PyMOL visualization data not shown), and sits in something of a cleft protecting it from the solvent to some degree.

Although the simulation/design part of this work has revealed some of the details of how the A4V mutation may act to destabilize hSOD1^{A4V}, there is no direct therapeutic application of this work. Both the second-site mutation F20G and F20A occur inside of each subunit and currently there is no available way to directly make the second-site mutation to the existing hSOD1^{A4V} molecules in patients.

However, this work suggests promise for biopharmaceuticals, where proteins initially engineered for biological activity, might undergo a round of computational design to improve thermodynamic stability and reduce aggregation. Specially, the aggregation "hot spots" of the aggregation prone antibody molecules have been identified by previous works (Wang et al., 2009). Although antibodies are big molecules, the aggregation "hot spot" may often consist of only several small segments of polypeptides. Therefore, we anticipate that RosettaDesign work performed around the "hot spot" can identify the key residues responsible for the instability and/or aggregation propensity of the antibody. Combining computational predictions and rational design could increase the probability that redesigned proteins are both more thermodynamically stable and aggregation-resistant. This could be a promising strategy to solve this critical problem in the biopharmaceutical industry.

8. Reference

Arnesano, F., Banci, L., Bertini, I., Martinelli, M., Furukawa, Y., & O'Halloran, T. V. (2004). The unusually stable quaternary structure of human Cu,Zn-superoxide dismutase 1 is controlled by both metal occupancy and disulfide status. *The Journal of biological chemistry*, *279*(46), 47998–8003. doi:10.1074/jbc.M406021200

Auclair, J. R., Boggio, K. J., Petsko, G. a, Ringe, D., & Agar, J. N. (2010). Strategies for stabilizing superoxide dismutase (SOD1), the protein destabilized in the most common form of familial amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(50), 21394–9. doi:10.1073/pnas.1015463107

Banci, L., Bertini, I., Boca, M., Girotto, S., Martinelli, M., Valentine, J. S., & Vieru, M. (2008). SOD1 and amyotrophic lateral sclerosis: mutations and oligomerization. *PloS one*, *3*(2), e1677. doi:10.1371/journal.pone.0001677

Banci, L., Bertini, I., Ciofi-Baffoni, S., Kozyreva, T., Zovo, K., & Palumaa, P. (2010). Affinity gradients drive copper to cellular destinations. *Nature*, *465*(7298), 645–8. doi:10.1038/nature09018

Bruijn, L. I. (1998). Aggregation and Motor Neuron Toxicity of an ALS-Linked SOD1 Mutant Independent from Wild-Type SOD1. *Science*, *281*(5384), 1851–1854. doi:10.1126/science.281.5384.1851

Brown, R. H. (1995). Amyotrophic lateral sclerosis: recent insights from genetics and transgenic mice. *Cell*, *80*(5), 687–92. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/7889564

Bruijn, L. I., Becher, M. W., Lee, M. K., Anderson, K. L., Jenkins, N. a, Copeland, N. G., Sisodia, S. S., et al. (1997). ALS-linked SOD1 mutant G85R mediates damage to astrocytes and promotes rapidly progressive disease with SOD1-containing inclusions. *Neuron*, *18*(2), 327–38. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9052802

Bruijn, L. I., & Cleveland, D. W. (1996). Mechanisms of selective motor neuron death in ALS: insights from transgenic mouse models of motor neuron disease. *Neuropathology and Applied Neurobiology*, *22*(5), 373–387. doi:10.1111/j.1365-2990.1996.tb00907.x

Canet, D., Last, A. M., Tito, P., Sunde, M., Spencer, A., Archer, D. B., Redfield, C., et al. (2002). Local cooperativity in the unfolding of an amyloidogenic variant of human lysozyme. *Nature structural biology*, *9*(4), 308–15. doi:10.1038/nsb768

Cardoso, R. (2002). Insights into Lou Gehrig's Disease from the Structure and Instability of the A4V Mutant of Human Cu,Zn Superoxide Dismutase. *Journal of Molecular Biology*, *324*(2), 247–256. doi:10.1016/S0022-2836(02)01090-2

Clark, E. (1998). Refolding of recombinant proteins. *Current opinion in biotechnology*, *9*(2), 157–63. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9664053

Clarke, J., & Fersht, a R. (1993). Engineered disulfide bonds as probes of the folding pathway of barnase: increasing the stability of proteins against the rate of denaturation. *Biochemistry*, *32*(16), 4322–9. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/8476861

Dantas, G., Corrent, C., Reichow, S. L., Havranek, J. J., Eletr, Z. M., Isern, N. G., Kuhlman, B., et al. (2007). High-resolution structural and thermodynamic analysis of extreme stabilization of human procarboxypeptidase by computational protein design. *Journal of molecular biology*, *366*(4), 1209–21. doi:10.1016/j.jmb.2006.11.080

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. *Journal of Molecular Biology*, *332*(2), 449–460. doi:10.1016/S0022-2836(03)00888-X

Dingermann, T. (2008). Does Size Matter? Facts and Thoughts on Recombinant Drugs. *Towards Drugs of the Future*, doi:10.3233/978-1-58603-949-3-81

Doucette, P. a, Whitson, L. J., Cao, X., Schirf, V., Demeler, B., Valentine, J. S., Hansen, J. C., et al. (2004). Dissociation of human copper-zinc superoxide dismutase dimers using chaotrope and reductant. Insights into the molecular basis for dimer stability. *The Journal of biological chemistry*, *279*(52), 54558–66. doi:10.1074/jbc.M409744200

Dunbrack, R. L., & Cohen, F. E. (1997). Bayesian statistical analysis of protein side-chain rotamer preferences. *Protein science : a publication of the Protein Society*, *6*(8), 1661–81. doi:10.1002/pro.5560060807

Furukawa, Y., Kaneko, K., Yamanaka, K., O'Halloran, T. V, & Nukina, N. (2008). Complete loss of post-translational modifications triggers fibrillar aggregation of SOD1 in the familial form of amyotrophic lateral sclerosis. *The Journal of biological chemistry*, *283*(35), 24167–76. doi:10.1074/jbc.M802083200

Hayward, L. J., Rodriguez, J. a, Kim, J. W., Tiwari, A., Goto, J. J., Cabelli, D. E., Valentine, J. S., et al. (2002). Decreased metallation and activity in subsets of mutant superoxide dismutases associated with familial amyotrophic lateral sclerosis. *The Journal of biological chemistry*, *277*(18), 15923–31. doi:10.1074/jbc.M112087200

Hinds, K., Koh, J. J., Joss, L., Liu, F., Baudys, M., & Kim, S. W. (n.d.). Synthesis and characterization of poly(ethylene glycol)-insulin conjugates. *Bioconjugate chemistry*. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10725096

Horovitz, A., Matthews, J. M., & Fersht, A. R., (1992) J. *Mol. Biol.* 227, 560-568.

Hough, M. a, Grossmann, J. G., Antonyuk, S. V, Strange, R. W., Doucette, P. a, Rodriguez, J. a, Whitson, L. J., et al. (2004). Dimer destabilization in superoxide dismutase may result in disease-causing properties: structures of motor neuron disease mutants. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(16), 5976–81. doi:10.1073/pnas.0305143101

Kaufmann, K. W., Lemmon, G. H., Deluca, S. L., Sheehan, J. H., & Meiler, J. (2010). Practically useful: what the Rosetta protein modeling suite can do for you. *Biochemistry*, *49*(14), 2987–98. doi:10.1021/bi902153g

Khare, S. D., Caplow, M., & Dokholyan, N. V. (2004). The rate and equilibrium constants for a multistep reaction sequence for the aggregation of superoxide dismutase in amyotrophic lateral sclerosis. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(42), 15094–9. doi:10.1073/pnas.0406650101

Koide, T., Igarashi, S., Kikugawa, K., Nakano, R., Inuzuka, T., Yamada, M., Takahashi, H., et al. (1998). Formation of granular cytoplasmic aggregates in COS7 cells expressing mutant Cu/Zn superoxide dismutase associated with familial amyotrophic lateral sclerosis. *Neuroscience letters*, *257*(1), 29–32. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/9857958

Kopito, R. R. (2000). Inclusion bodies and protein aggregation. *Cell*, *68*(1999), 524–530.

Kortemme, T., Morozov, A. V., & Baker, D. (2003). An Orientation-dependent Hydrogen Bonding Potential Improves Prediction of Specificity and Structure for Proteins and Protein–Protein Complexes. *Journal of Molecular Biology*, *326*(4), 1239–1259. doi:10.1016/S0022-2836(03)00021-4

Kuhlman, B., & Baker, D. (2000). Native protein sequences are close to optimal for their structures. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(19), 10383–8. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=27033&tool=pmcentrez&rendertype=abstract

Kuhlman, B., Dantas, G., Ireton, G. C., Varani, G., Stoddard, B. L., & Baker, D. (2003). Design of a novel globular protein fold with atomic-level accuracy. *Science (New York, N.Y.)*, *302*(5649), 1364–8. doi:10.1126/science.1089427

Lazaridis, T., & Karplus, M. (1999). Effective energy function for proteins in solution. *Proteins*, *35*(2), 133–52. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10223287

Li, J. D., W. G. Tang, et al. (2001). "Phages Displaying Peptides with beta Turn Conformation." *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)* 33(3): 335-339.

Liu, Y., & Kuhlman, B. (2006). RosettaDesign server for protein design. *Nucleic acids research*, *34*(Web Server issue), W235–8. doi:10.1093/nar/gkl163

Munoz V., Lopez E. M., Jager M., Serrano L. (1994). Kinetic Characterization of the Chemotactic Protein from Escherichia coli, CheY. Kinetic Analysis of the Inverse Hydrophobic Effect. *Biochemistry* 33, 5858-5866. doi: 10.1021/bi00185a025

Pace, C. N. (1986) *Methods Enzymol.* 131, 266-279

Prasanna, V., Gopal, B., Murthy, M. R., Santi, D. V, & Balaram, P. (1999). Effect of amino acid substitutions at the subunit interface on the stability and aggregation properties of a dimeric

protein: role of Arg 178 and Arg 218 at the Dimer interface of thymidylate synthase. *Proteins*, *34*(3), 356–68. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10024022

Rakhit, R., Crow, J. P., Lepock, J. R., Kondejewski, L. H., Cashman, N. R., & Chakrabartty, A. (2004). Monomeric Cu,Zn-superoxide dismutase is a common misfolding intermediate in the oxidation models of sporadic and familial amyotrophic lateral sclerosis. *The Journal of biological chemistry*, *279*(15), 15499–504. doi:10.1074/jbc.M313295200

Ray, S. S., Nowak, R. J., Brown, R. H., & Lansbury, P. T. (2005). Small-molecule-mediated stabilization of familial amyotrophic lateral sclerosis-linked superoxide dismutase mutants against unfolding and aggregation. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(10), 3639–44. doi:10.1073/pnas.0408277102

Ray, S. S., Nowak, R. J., Strokovich, K., Brown, R. H., Walz, T., & Lansbury, P. T. (2004). An intersubunit disulfide bond prevents in vitro aggregation of a superoxide dismutase-1 mutant linked to familial amytrophic lateral sclerosis. *Biochemistry*, *43*(17), 4899–905. doi:10.1021/bi030246r

Ray, S. S., & Lansbury, P. T. (2004). A possible therapeutic target for Lou Gehrig's disease. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(16), 5701–2. doi:10.1073/pnas.0401934101

Roberts, B. R., Tainer, J. a, Getzoff, E. D., Malencik, D. a, Anderson, S. R., Bomben, V. C., Meyers, K. R., et al. (2007). Structural characterization of zinc-deficient human superoxide dismutase and implications for ALS. *Journal of molecular biology*, *373*(4), 877–90. doi:10.1016/j.jmb.2007.07.043

Rohl, C. a, Strauss, C. E. M., Misura, K. M. S., & Baker, D. (2004). Protein structure prediction using Rosetta. *Methods in enzymology*, *383*(2003), 66–93. doi:10.1016/S0076-6879(04)83004-0

Rosen D. R., Siddique T., Patterson D., Figlewicz D. A., Sapp P., Hentati A., Donaldson D., Goto J., O'Regan J. P. and Deng H. X. (1993). Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis. *Nature* 362, 59–62.

Rosenberg, A. S. (2006). Effects of protein aggregates: an immunologic perspective. *The AAPS journal*, *8*(3), E501–7. doi:10.1208/aapsj080359

Sahin, E., Jordan, J. L., Spatara, M. L., Naranjo, A., Costanzo, J. a, Weiss, W. F., Robinson, A. S., et al. (2011). Computational design and biophysical characterization of aggregation-resistant point mutations for γD crystallin illustrate a balance of conformational stability and intrinsic aggregation propensity. *Biochemistry*, *50*(5), 628–39. doi:10.1021/bi100978r

Sammond, D. W., Eletr, Z. M., Purbeck, C., & Kuhlman, B. (2010). Computational design of second-site suppressor mutations at protein-protein interfaces. *Proteins*, *78*(4), 1055–65. doi:10.1002/prot.22631

Schmidlin, T., Kennedy, B. K., & Daggett, V. (2009). Structural Changes to Monomeric CuZn Superoxide Dismutase Caused by the Familial Amyotrophic Lateral Sclerosis-Associated Mutation A4V. *Biophysj*, *97*(6), 1709–1718. doi:10.1016/j.bpj.2009.06.043

Soto, C., Castano, E.M., et al., 1995. The alpha-helical to betastrand transition in the amino-terminal fragment of the amyloid beta-peptide modulates amyloid formation. *J. Biol. Chem.* 270, 3063–3067.

Stathopulos, P. B., Rumfeldt, J. a O., Scholz, G. a, Irani, R. a, Frey, H. E., Hallewell, R. a, Lepock, J. R., et al. (2003). Cu/Zn superoxide dismutase mutants associated with amyotrophic lateral sclerosis show enhanced formation of aggregates in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(12), 7021–6. doi:10.1073/pnas.1237797100

Stefani, M., & Dobson, C. M. (2003). Protein aggregation and aggregate toxicity: new insights into protein folding, misfolding diseases and biological evolution. *Journal of molecular medicine (Berlin, Germany)*, *81*(11), 678–99. doi:10.1007/s00109-003-0464-5

Tainer, J. A., Getzoff, E. D., Richardson, J. S. & Richardson, D. C. (1983). Structure and mechanism of copper, zinc superoxide dismutase. *Nature* 306, 284 - 287 (17 November 1983); doi:10.1038/306284a0

Wang, J., Xu, G., & Borchelt, D. R. (2002). High molecular weight complexes of mutant superoxide dismutase 1: age-dependent and tissue-specific accumulation. *Neurobiology of disease*, *9*(2), 139–48. doi:10.1006/nbdi.2001.0471

Wang, W. (2005). Protein aggregation and its inhibition in biopharmaceutics. *International journal of pharmaceutics*, *289*(1-2), 1–30. doi:10.1016/j.ijpharm.2004.11.014

Wang, X., Das, T. K., Singh, S. K., & Kumar, S. (2009). Potential aggregation prone regions in biotherapeutics: A survey of commercial monoclonal antibodies. *mAbs*, *1*(3), 254–67. Retrieved from http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2726584&tool=pmcentrez&rendertype=abstract

Appendix I

1. Command line for RosettaDesgin runs (resfile and PDB input need to be assigned for a particular run):

   ./fixbb.macosgccrelease -in:ignore_unrecognized_res -database ~/rosetta3.4-build/rosetta_database/ -use_input_sc -resfile resfile_hSOD.txt -s 1PU0.pdb -ndruns 10 -nstruct 10 -minimize_sidechains -out:file:scorefile hSOD_score

2. Resfile for RosettaDesign runs in this work

   a. "NATAA" resfile for relaxing molecule

```
NATAA
EX 1 EX 2
USE_INPUT_SC

start
46 A NATRO
48 A NATRO
120 A NATRO
63 A NATRO
71 A NATRO
80 A NATRO
83 A NATRO
46 B NATRO
48 B NATRO
120 B NATRO
63 B NATRO
71 B NATRO
80 B NATRO
83 B NATRO
```

   b. "ALLAAwc" resfile for the 5 Å around Val4 design runs

```
NATAA
EX 1 EX 2
USE_INPUT_SC

start
46 A NATRO
48 A NATRO
120 A NATRO
63 A NATRO
71 A NATRO
80 A NATRO
83 A NATRO
```

```
46 B NATRO
48 B NATRO
120 B NATRO
63 B NATRO
71 B NATRO
80 B NATRO
83 B NATRO
2 A ALLAAwc EX 1 EX 2
3 A ALLAAwc EX 1 EX 2
5 A ALLAAwc EX 1 EX 2
6 A ALLAAwc EX 1 EX 2
18 A ALLAAwc EX 1 EX 2
19 A ALLAAwc EX 1 EX 2
20 A ALLAAwc EX 1 EX 2
21 A ALLAAwc EX 1 EX 2
29 A ALLAAwc EX 1 EX 2
106 A ALLAAwc EX 1 EX 2
112 A ALLAAwc EX 1 EX 2
113 A ALLAAwc EX 1 EX 2
149 A ALLAAwc EX 1 EX 2
150 A ALLAAwc EX 1 EX 2
151 A ALLAAwc EX 1 EX 2
152 A ALLAAwc EX 1 EX 2
153 A ALLAAwc EX 1 EX 2
2 B ALLAAwc EX 1 EX 2
3 B ALLAAwc EX 1 EX 2
5 B ALLAAwc EX 1 EX 2
6 B ALLAAwc EX 1 EX 2
18 B ALLAAwc EX 1 EX 2
19 B ALLAAwc EX 1 EX 2
20 B ALLAAwc EX 1 EX 2
21 B ALLAAwc EX 1 EX 2
29 B ALLAAwc EX 1 EX 2
106 B ALLAAwc EX 1 EX 2
112 B ALLAAwc EX 1 EX 2
113 B ALLAAwc EX 1 EX 2
149 B ALLAAwc EX 1 EX 2
150 B ALLAAwc EX 1 EX 2
151 B ALLAAwc EX 1 EX 2
152 B ALLAAwc EX 1 EX 2
153 B ALLAAwc EX 1 EX 2
```

c. “PIKAA” resfile for generating mutants (for example, generating hSOD1^{A4V})

```
NATAA
USE_INPUT_SC
EX 1 EX 2
```

start
4 A PIKAA V EX 1 EX 2
4 B PIKAA V EX 1 EX 2
20 A PIKAA G EX 1 EX 2
20 B PIKAA G EX 1 EX 2
46 A NATRO
48 A NATRO
120 A NATRO
63 A NATRO
71 A NATRO
80 A NATRO
83 A NATRO
46 B NATRO
48 B NATRO
120 B NATRO
63 B NATRO
71 B NATRO
80 B NATRO
83 B NATRO

Appendix II

Cellular assays with wild type and mutant hSOD1s were done by Simpson Gregoire in our collaborative Kwon lab in both HEK293T kidney and NSC-34 spinal cell lines. Mean cellular fluorescence intensity was used to indicate aggregate formation. The more fluorescence intensity the cell shows, the less aggregates the cell contains.

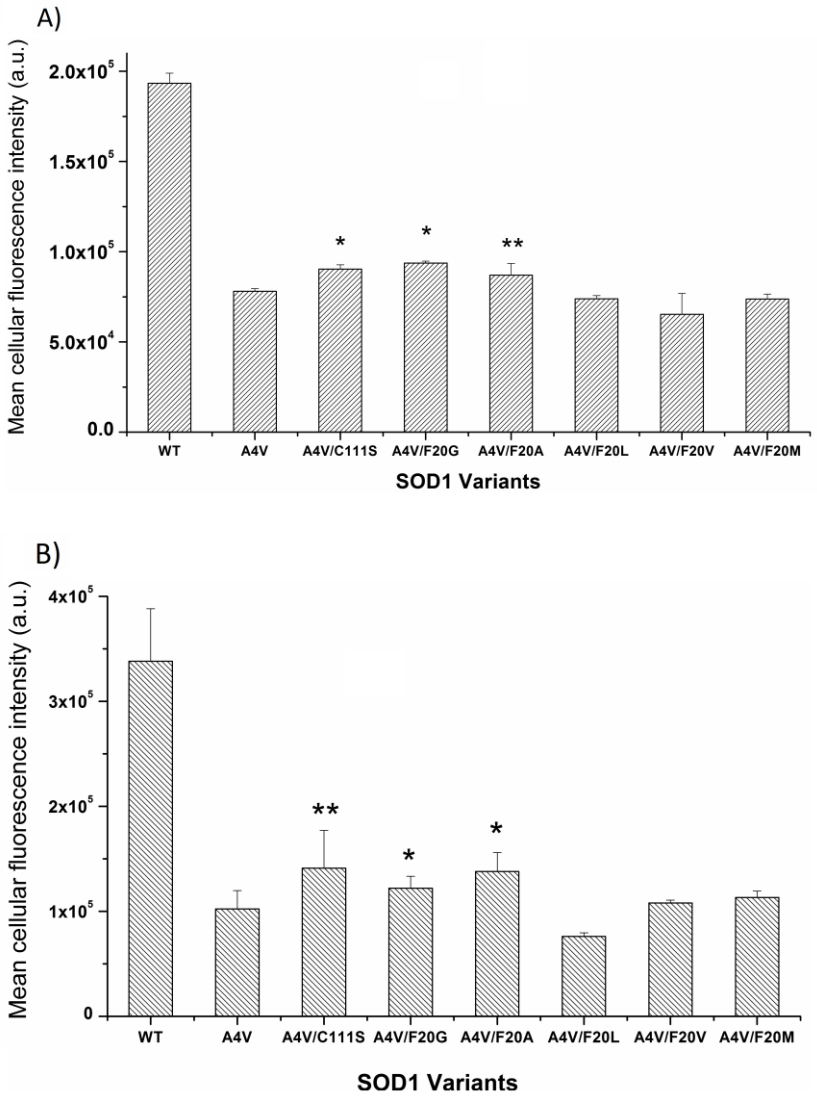

Fig. A) Aggregation assay in HEK293T kidney cells expressing wild type and mutant hSOD1s. B) Aggregation assay in NSC-34 spinal cells expressing wild type and mutant hSOD1s. In both assays, mean cellular fluorescence intensity was used to indicate aggregate formation is cells expressing each hSOD1 variant.