Development and Applications of Computational Protein Design

Thesis by

Eun Jung Choi

In Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

2006

(Defended May 26, 2006)

Acknowledgements

I would like to thank my advisor, Dr. Stephen Mayo, for his support throughout my graduate studies. He is a brilliant scientist, and during my stay in his lab, I learned a lot from him. I am grateful for the opportunity I had in learning about the exciting field of computational protein design.

My committee members, Dr. Pamela Bjorkman, Dr. David Chan and Dr. Bill Goddard, always gave me helpful advice at my annual committee meetings. They cared not only about my research, but also about my long-term goals. I always looked forward to these meetings.

I thank everyone in the Mayo lab, Alex, Barry, Ben, Christina, Dan, Darryl, Deepshika, Eric, Geoff, Heidi, Jennifer, JJ, Julia, Karin, Kyle, Oscar, Mary, Matthew, Peter, Possu, Premal, Sarah, Scott, Shannon, Shira, Tom, and Welison for providing a stimulating but comfortable research environment. Needless to say, Cynthia and Rhonda are the two most indispensable people in the lab. Marie spent endless hours editing my papers and this thesis would not have seen the light of day were it not for her efforts. Jessica M. deserves special mention because it was her fabulous idea to collaborate together. It was a pleasure working with her.

I owe a lot to my valuable friends whom I was lucky to have met during my stay at Caltech. Hyun Jee, In Jung, Jessica E., Ji Young, Keum Joo, Kyung Ah, Mi Sook, Sarina and all of the other people, both past and present, who were always there for me, thank you so much. I would especially like to thank Caglar for his endless capacity to help on anything and everything. Ann Marie, Daphne and Frosso, thanks for being such good housemates.

And last, but not the least, I thank my family, my parents and my sister. It is impossible to express the gratitude I feel for the sacrifice and support my mom and dad have shown me throughout the years. And thank you Eun Kyong for flying to the U.S. multiple times to see me.

Abstract

Success in computational design of proteins requires a good understanding of the physical properties that determine the protein structure. However, computational protein design also provides us the opportunities to test and improve our current knowledge of protein structure. In our lab, the "protein design cycle" is used to improve the energy function and increase our knowledge of protein chemistry. This strategy utilizes a cyclic protocol where first, a modification that is predicted to improve the result is applied to the energy function. Next, proteins are simulated using this modified function. These molecules are then synthesized and analyzed to see if our simulation results correlate with the physical properties of the molecules. The new knowledge from the analysis is incorporated into the next design in the form of a new modification and the whole cycle starts again. Using this technique, an energy function highly successful in designing protein stability has been acquired in our lab.

A similar approach was used to determine whether the introduction of a backbone entropy term allows us to incorporate proline into the pool of amino acids we consider in designs. Using ORBIT, several proline mutants of protein G were simulated and synthesized. The correlation between the experimental results and computational results was analyzed. From this analysis, we learned that the entropic benefit of a proline mutation is usually small compared to the enthalpic disadvantages. However, including a backbone entropy term did increase the correlation between the rank order of the experimental stabilities and computational energies.

The ultimate test of our protein design protocol is its application to biological systems of interest. We have applied ORBIT to three different design problems,

increasing the affinity of peptide-protein interaction of TRAF6, the elimination of disulfide bonds for biosensor design and enhancement of solubility and stability of an anti-htt antibody fragment. The results provided here show that protein design protocol is a fast and efficient way to manipulate and study biological systems.

Table of Contents

Appendix A.

List of Tables and Figures

Chapter 1.

Chapter 2.

Chapter 3.

Chapter 4.

Chapter 5.

Appendix A.

Chapter 1

Introduction

The idea of protein design debuted in the early 1980s (Drexler 1981; Pabo 1983). The protein design problem, which is the search for an amino acid sequence that folds into the desired structure, was thought of as the inverse of the protein folding problem, which aims to find the correct fold for a given amino acid sequence. Unlike the protein folding problem where there is a single solution fold per sequence, multiple solution sequences can be found for each target fold in the protein design problem (Figure 1). Assisted by the increase in computational power and developments in molecular biology, drastic progress has been made in the field of protein design. Today, proteins are engineered for increased stability and solubility, and for new or altered functionalities.

Computational protein design methods use force fields to select the optimal sequence of amino acids for a target fold. They employ an energy potential, statistically significant conformations of amino acids called rotamers and a selection algorithm, which screens through the combinatorial complexity of the sequence space to obtain the lowest energy sequences. In ORBIT (Optimization of Rotamers By Iterative Techniques), the protein design program developed in our lab, the energy function is comprised of 5 terms: van der Waals, hydrogen bonding, electrostatic, solvation and entropy. The energy function is empirical, but based on fundamental physical principles. A simple explanation of each of the terms will be given below.

Description of the energy function in ORBIT

The van der Waals interaction is calculated using a Lennard-Jones 12-6 potential (Figure 2A). This term is critical because it ensures a well-packed protein structure, essential for the stability of folded proteins. Design methods that optimize only this term

are still able to produce stable and folded mutants, which points to the importance of the van der Waals term (Desjarlais and Handel 1995; Dahiyat and Mayo 1996; 1997b; Lazar et al. 1997).

Hydrogen bonds are part electrostatic and part covalent interactions that have been shown to be important in the design of enzymatic activities, binding sites and full sequence designs (Dahiyat and Mayo 1997a; Looger et al. 2003). An angle-dependent, 12-10 hydrogen-bond potential is used in ORBIT (Figure 2B).

Electrostatics, like hydrogen bonds, is thought to be more important in the specificity and functionality of proteins rather than their stability (Tanford et al. 1962). Coulomb's law is used to describe the electrostatic interactions in ORBIT (Figure 2C). Our selection algorithm requires interactions to be described in a pairwise decomposable manner, and Coulomb's law satisfies this requirement. The Coulombic term is scaled down through a distance dependent dielectric constant. This reflects the small contribution of electrostatic interactions to protein stability.

When designing for binding and catalysis, electrostatic interactions in the active site become important. For this purpose, a more accurate electrostatic function based on physical properties is needed. Several different methods for continuum electrostatics were recently modified to fit the need of protein design protocols. These include the Poisson-Boltzman equation (Marshall et al. 2005), the Tanford-Kirkwood model (Havranek and Harbury 1999), and the Generalized Born model (Pokala and Handel 2004).

The solvation term represents the interaction of the protein with water. The protein's aqueous environment influences its stability in two ways: (1) it causes the

hydrophobic effect that provides the major driving force for protein folding (Kauzmann 1959); and (2) the solvent screens the electrostatic interactions. ORBIT utilizes a surface area based solvation method (Figure 2D) (Street and Mayo 1998). This method penalizes polar surface area burial and nonpolar surface area exposure, and benefits nonpolar surface area burial. Recently, the Lazaridis and Karplus excluded volume model (LK model) (Lazaridis and Karplus 1999) was successfully incorporated into ORBIT. Unlike the surface area based solvation method, which is computationally expensive, the LK model assumes that the solvation free energy can be decomposed into a sum of group contributions that makes it computationally efficient. Ideally, an electrostatic term should be able to describe all electrostatic interactions including hydrogen bonds and solvation. Unfortunately, the implementation of such a description in protein design is intractable at the moment, due to limitations in the selection algorithms and the available computational power.

When a protein folds into its native structure from its unfolded state, it loses conformational entropy. The native structure of the protein is the structure where the enthalpic contributions from the native interactions outweigh the entropic loss of folding to the maximum extent. Several different methods have been used in protein design to describe this effect. Side chain entropy, calculated by considering all movable side chain torsional angles at each residue, has been incorporated successfully (Hellinga and Richards 1994). Backbone and side chain entropy values derived from the distribution of rotamers in crystal structures have also been used (Filikov et al. 2002). The mean field selection algorithm in which multiple rotamers are considered at each position simultaneously provides an indirect way of incorporating side chain entropy into the

design protocol (Koehl and Delarue 1994). Although these results show an improvement in accuracy through the inclusion of a side chain entropy term, other studies have found no improvement (Dahiyat and Mayo 1996; Kortemme and Baker 2002; Hu and Kuhlman 2006). Certain assumptions made in the entropy calculations might be responsible for these contradictory results. In some methods, the conformational freedom is assumed to be completely restricted in the folded state, while in other methods the unfolded state is assumed to be completely unfolded and has no structure.

Even though an explicit entropy term is not included in ORBIT, there are negative design terms that reflect some portions of the entropy term. For instance, the nonpolar exposure penalty represents the hydrophobic effect that results from changes in the entropy of the protein and water molecules during folding. An effect commonly seen in proteins designed in the absence of an entropy term is the increase in the number of methionines introduced into the core and the number of long positively-charged residues on the surface. Our lab has made successful attempts to overcome these unnatural effects. For the methionines, a penalty of 8kcal/mol was applied for any methionines that were considered in the design. To reduce the amount of large positively charged residues on the surface, we employed a baseline strategy in which the distribution of residues placed on the protein surface was biased to be similar to the distribution obtained from a statistical survey of the PDB database.

Proline is another residue that has been omitted in our designs due to the lack of an entropic term. Proline residue is the only amino acid where the side chain is covalently linked to the backbone, restricting its mobility. Thus, the backbone conformational entropy loss upon folding is much less for prolines than all the other amino acids, thereby

providing a stabilizing effect. In chapter 2, the effect of proline on protein stability is analyzed by mutational studies on protein G. It is shown that prolines can both stabilize and destabilize the protein depending on their local environment. This reconfirms the general understanding that mutations to proline are stabilizing because of entropic factors, although this behavior can be masked and even reversed by destabilizing enthalpic changes. In the study, ORBIT was shown to predict the stability of the proline mutants considerably well, likely due to the dominance of enthalpic contributions over entropic contributions. Including a backbone entropy term with a proper weighing factor, however, did increase the correlation between the experimental and the calculated energies.

Applications of computational protein design

Computational protein design has two major applications: it is a convenient test for our understanding of the physical properties of proteins; another is the more obvious application of designing novel proteins for various purposes. Since it would be impossible to list all of the applications of computational protein design here, I will discuss three applications related to the projects in this thesis.

One application is the use of computational protein design to engineer peptides that bind to proteins. Peptides have a high potential to be used as therapeutics because of their small size, and a high affinity peptide for any receptor can be used as a competitor in small molecule drug screening. Protein design algorithms have been utilized to design high affinity peptides, and although their increase in affinity was modest, it was shown that backbone flexibility gave better results than fixed backbone (Wollacott and

Desjarlais 2001; Sood and Baker 2006). In chapter 3, we used ORBIT, which utilizes a fixed backbone, to design a high affinity peptide for TRAF6 protein. All of our designed peptides had affinities similar to the CD40 peptide. Surprisingly, the best peptide designed had about a two-fold increase in its affinity, similar to the increase in affinity of the peptides designed with flexible backbones. Thus, from our results it is not clear whether the backbone flexibility is necessary for high affinity peptide design, although in protein-peptide docking studies, the flexibility of both peptide and protein was necessary to achieve good correlation between calculated binding energies and experimental affinities (Liu et al. 2004).

Another protein design application that has recently gained much attention is the design of protein biosensors. Proteins, because of their specificity in molecular recognition of ligands, have a high potential to be developed into reagentless detectors for a variety of molecules. The only deterrence is detecting the binding event, which can be resolved by attaching fluorescent probes onto the protein at positions where conformational change occurs with ligand binding. Cass and co-workers engineered the maltose binding protein (MBP) for reagentless optical sensing of maltose by attaching different thio-reactive fluorophores to it. They suggested that other periplasmic binding proteins (PBPs) could be engineered as biosensors using a similar approach (Gilardi et al. 1994). Hellinga and coworkers took this suggestion and, using protein design methods, designed a variety of PBPs into biosensors with novel specificity such as to Zn(II), TNT, L-lactate, serotonin, PMPA and dihydroxyacetone phosphate (Dwyer and Hellinga 2004). In chapter 4, the lipid transfer protein (LTP) from maize is developed into a scaffold for the design of biosensors for hydrophobic ligands. The binding pocket of PBPs are small

and hydrophilic, to accommodate their small and hydrophilic natural ligands. Although PBPs were designed to recognize small non-native hydrophilic ligands, it would be an extremely difficult task to design the PBP binding site to accept large hydrophobic ligands and still retain their native fold. To expand the chemical scope of target ligands, a protein that naturally binds to large hydrophobic molecules, the maize LTP, was selected as a scaffold for biosensor design. To achieve this goal, the four disulfide bonds which exist in all LTPs were designed out individually and the fluorophore attached LTP variant, which shows the largest fluorescence intensity change after ligand binding, was selected as a putative biosensor scaffold. With this scaffold, it will be possible to design biosensors for hydrophobic ligands of interest.

Engineered proteins and peptides have high potential for use as therapeutics for a variety of disease. This is because the design of protein therapeutics integrates the traditional goals of protein design that include stability, solubility and binding specificity, with biological and clinical parameters that include half-life, immunogenicity, toxicity and degradation susceptibility (Lazar et al. 2003). Antibodies are epitope-specific, which enable them to easily be used as therapeutics. By binding to specific antigens, antibodies can block interactions; tag target molecules; lock proteins into specific conformation, and even act as catalysts (Stocks 2004). Recently, the Fc region of an anti-tumor antibody was designed using ORBIT for affinity and specificity for Fcγ receptors, increasing cytotoxicity (Lazar et al. 2006). In chapter 5, a project is outlined where an intrabody that binds to huntingtin protein is solubilized and stabilized by using ORBIT. This antibody variant has the potential to be used as a therapeutic for the treatment of Huntington's disease.

In Appendix A, a project involving the structure prediction of G protein coupled receptors (GPCRs) is described. As explained above, the protein folding problem is much more complex than the protein design problem (Figure 1). Predicting the structure of a protein from its sequence has fascinated scientists for many years. Currently, with CASP7 coming up, there is more interest in this problem than ever before. Structure prediction methods can be divided into two broad categories. One category uses first principles to predict the fold. The other category uses methods based on the identification and utilization of related sequences and structures, such as in homology modeling and fold recognition. As can be seen in the analysis of CASP6, the former gives more accurate results and seems to be improving at a faster rate than the latter (Kryshtafovych et al. 2005). Nevertheless, novel folds cannot be predicted by comparative modeling, thus predictions based on the physical and chemical principles are important. The structure of endothelial differentiation gene receptor, a GPCR that binds to sphingolipids, was predicted using an *ab initio* computational structure prediction method. This method was previously validated by predicting the three-dimensional structures of bacteriorhodopsin and bovine rhodopsin for which crystal structural data are available (Vaidehi et al. 2002). Docking procedures were used to obtain relative affinities for positive and negative ligands and these values were compared to experimental results to validate our structure. Although the model is a crude representation of the actual structure, it can be used to predict positions for mutation to improve ligand binding.

References

Dahiyat, B.I., and Mayo, S.L. 1996. Protein design automation. *Protein Sci* **5:** 895-903.

Dahiyat, B.I., and Mayo, S.L. 1997a. *De novo* protein design: fully automated sequence selection. *Science* **278:** 82-87.

Dahiyat, B.I., and Mayo, S.L. 1997b. Probing the role of packing specificity in protein design. *Proc Natl Acad Sci USA* **94:** 10172-10177.

Desjarlais, J.R., and Handel, T.M. 1995. *De novo* design of the hydrophobic cores of proteins. *Protein Sci* **4:** 2006-2018.

Drexler, K.E. 1981. Molecular engineering: An approach to the development of general capabilities for molecular manipulation. *Proc Natl Acad Sci U S A* **78:** 5275-5278.

Dwyer, M.A., and Hellinga, H.W. 2004. Periplasmic binding proteins: a versatile superfamily for protein engineering. *Curr Opin Struct Biol* **14:** 495-504.

Filikov, A.V., Hayes, R.J., Luo, P., Stark, D.M., Chan, C., Kundu, A., and Dahiyat, B.I. 2002. Computational stabilization of human growth hormone. *Protein Sci* **11:** 1452-1461.

Gilardi, G., Zhou, L.Q., Hibbert, L., and Cass, A.E. 1994. Engineering the maltose binding protein for reagentless fluorescence sensing. *Anal Chem* **66:** 3840-3847.

Havranek, J.J., and Harbury, P.B. 1999. Tanford-Kirkwood electrostatics for protein modeling. *Proc Natl Acad Sci U S A* **96:** 11145-11150.

Hellinga, H.W., and Richards, F.M. 1994. Optimal sequence selection in proteins of known structure by simulated evolution. *Proc Natl Acad Sci U S A* **91:** 5803-5807.

Hu, X., and Kuhlman, B. 2006. Protein design simulations suggest that side-chain conformational entropy is not a strong determinant of amino acid environmental preferences. *Proteins* **62:** 739-748.

Kauzmann, W. 1959. Some factors in the interpretation of protein denaturation. *Adv Protein Chem* **14:** 1-63.

Koehl, P., and Delarue, M. 1994. Application of a self-consistent mean field theory to predict protein side-chains conformation and estimate their conformational entropy. *J Mol Biol* **239:** 249-275.

Kortemme, T., and Baker, D. 2002. A simple physical model for binding energy hot spots in protein-protein complexes. *Proc Natl Acad Sci U S A* **99:** 14116-14121.

Kryshtafovych, A., Venclovas, C., Fidelis, K., and Moult, J. 2005. Progress over the first decade of CASP experiments. *Proteins* **61 Suppl 7:** 225-236.

Lazar, G.A., Dang, W., Karki, S., Vafa, O., Peng, J.S., Hyun, L., Chan, C., Chung, H.S., Eivazi, A., Yoder, S.C., et al. 2006. Engineered antibody Fc variants with enhanced effector function. *Proc Natl Acad Sci U S A* **103:** 4005-4010.

Lazar, G.A., Desjarlais, J.R., and Handel, T.M. 1997. *De novo* design of the hydrophobic core of ubiquitin. *Protein Sci* **6:** 1167-1178.

Lazar, G.A., Marshall, S.A., Plecs, J.J., Mayo, S.L., and Desjarlais, J.R. 2003. Designing proteins for therapeutic applications. *Curr Opin Struct Biol* **13:** 513-518.

Lazaridis, T., and Karplus, M. 1999. Effective energy function for proteins in solution. *Proteins* **35:** 133-152.

Liu, Z., Dominy, B.N., and Shakhnovich, E.I. 2004. Structural mining: self-consistent design on flexible protein-peptide docking and transferable binding affinity potential. *J Am Chem Soc* **126:** 8515-8528.

Looger, L.L., Dwyer, M.A., Smith, J.J., and Hellinga, H.W. 2003. Computational design of receptor and sensor proteins with novel functions. *Nature* **423:** 185-190.

Marshall, S.A., Vizcarra, C.L., and Mayo, S.L. 2005. One- and two-body decomposable Poisson-Boltzmann methods for protein design calculations. *Protein Sci* **14:** 1293.

Pabo, C. 1983. Molecular technology. Designing proteins and peptides. *Nature* **301:** 200.

Pokala, N., and Handel, T.M. 2004. Energy functions for protein design I: efficient and accurate continuum electrostatics and solvation. *Protein Sci* **13:** 925-936.

Sood, V.D., and Baker, D. 2006. Recapitulation and design of protein binding peptide structures and sequences. *J Mol Biol* **357:** 917-927.

Stocks, M.R. 2004. Intrabodies: production and promise. *Drug Discov Today* **9:** 960-966.

Street, A.G., and Mayo, S.L. 1998. Pairwise calculation of protein solvent-accessible surface areas. *Fold Des* **3:** 253-258.

Tanford, C., Buckley, C.E., 3rd, De, P.K., and Lively, E.P. 1962. Effect of ethylene glycol on the conformation of gama-globulin and beta-lactoglobulin. *J Biol Chem* **237:** 1168-1171.

Vaidehi, N., Floriano, W.B., Trabanino, R., Hall, S.E., Freddolino, P., Choi, E.J., Zamanakos, G., and Goddard, W.A., 3rd. 2002. Prediction of structure and function of G protein-coupled receptors. *Proc Natl Acad Sci U S A* **99:** 12622-12627.

Wollacott, A.M., and Desjarlais, J.R. 2001. Virtual interaction profiles of proteins. *J Mol Biol* **313:** 317-342.

Figure 1. Protein design and protein folding problems.

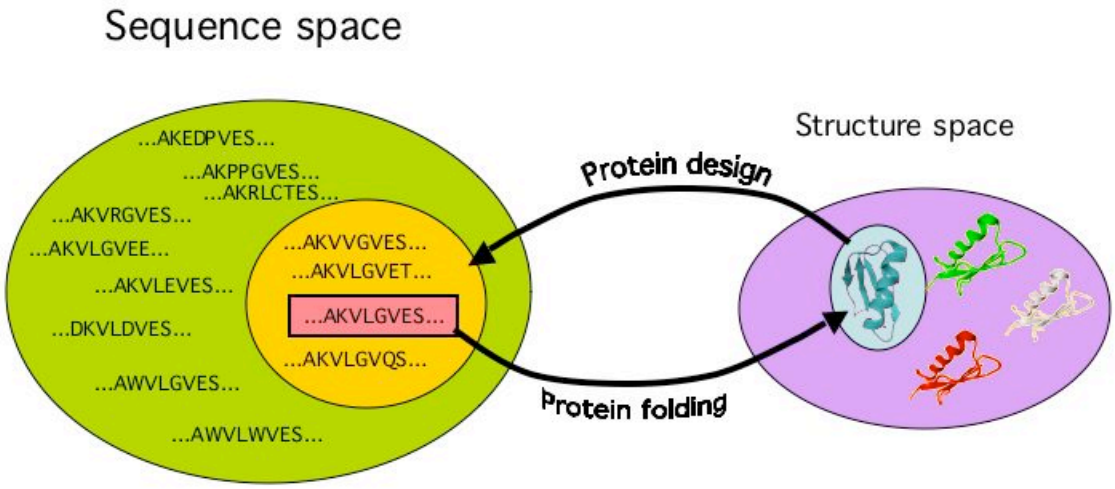

Sequence space
...AKEDPVES...
...AKPPGVES...
...AKRLCTES...
...AKVRGVES...
...AKVLGVEE...
...AKVLEVES...
...DKVLDVES...
...AWVLGVES...
...AWVLWVES...
...AKVVGVES...
...AKVLGVET...
...AKVLGVES...
...AKVLGVQS...
Protein design
Protein folding
Structure space

Figure 2. Potential energy function terms for ORBIT. (A) R: interatomic distance between the two atoms; D_0: geometric mean of the well-depth of the two atoms; R_0: geometric mean of the van der Waals radius of the two atoms; α: van der Waals radius scaling factor (B) R: donor to acceptor distance; D_0: hydrogen bond well-depth (8 kcal/mol); R_0: hydrogen bond equilibrium distance (2.8 Å); F: angle term, which depends on the hybridization state of the donor and acceptor (C) q: charge on atom 1; q': charge on atom 2; r: interaction distance; ε: dielectric constant (D) $A_{np,b}$*: nonpolar buried surface area; $A_{p,b}$*: polar buried surface area; $A_{np,e}$: nonpolar exposed surface area; σ_{np}: nonpolar burial energy; σ_p: polar burial energy; κ: nonpolar exposure multiplication factor; c: polar hydrogen burial energy; N_{bph}: number of buried polar hydrogens not involved in hydrogen bonds.

A. Lennard-Jones 12-6 potential for van der Waals interaction

$$E_{vdw} = D_0\left[\left(\frac{\alpha R_0}{R}\right)^{12} - 2\left(\frac{\alpha R_0}{R}\right)^{6}\right]$$

B. Hydrogen bond 12-10 potential

$$E_{h-bond} = D_0\left[5\left(\frac{R_0}{R}\right)^{12} - 6\left(\frac{R_0}{R}\right)^{10}\right]F(\theta, \phi, \varphi)$$

C. Coulombic function for electrostatic interaction

$$E_{elec} = \frac{qq'}{\varepsilon r}$$

D. Surface area based solvation term

$$E_{as} = -(\kappa + 1)\sigma_{np}A_{np,b} + \kappa\sigma_{np}A_{np,e} + \sigma_{p}A_{p,b} + cN_{bph}$$

Chapter 2

Generation and analysis of proline mutants in protein G

The text of this chapter is adapted from the publication

Choi EJ and Mayo SL. *Protein Eng Des Sel* 2006, in press.

Abstract

The pyrrolidine ring of the amino acid proline reduces the conformational freedom of the protein backbone in its unfolded form and thus, enhances protein stability. The strategy of inserting proline into regions of the protein where it does not perturb the structure has been utilized to stabilize many different proteins including enzymes. However, most of these efforts have been based on trial and error, rather than rational design. Here, we try to understand proline's effect on protein stability by introducing proline mutations into various regions of the B1 domain of Streptococcal protein G. Using two different solvation models, we also applied the ORBIT computational protein design program to determine the extent to which it could predict the stabilizing and destabilizing effects of prolines. The use of a surface area dependent solvation model resulted in a modest correlation between the experimental free energy of folding and computed energies; on the other hand, the use of a Gaussian solvent exclusion model led to significant positive correlation. By including a backbone conformational entropy term to the computational energies, we increased the statistical significance of the correlation between the experimental stabilities and both solvation models.

Introduction

Proline is the only naturally occurring amino acid in which the side chain is bonded to the backbone nitrogen, forming a five-membered pyrrolidine ring. This pyrrolidine ring restricts the rotation of the N-C_α bond, decreasing the backbone conformational entropy of the unfolded form of the protein relative to other naturally occurring amino acids. This allows proline substitution to increase the stability of a protein by decreasing the entropic difference between the unfolded and the folded form, thereby increasing the free energy difference (Nemethy et al. 1966; Matthews et al. 1987). Based on this concept, different residues in various proteins have been mutated to prolines, resulting in increased stability (Matthews et al. 1987; Watanabe et al. 1994).

On the other hand, prolines are also notorious for destabilizing proteins. It is well known that prolines, located internally in α-helices or β-sheets, break these secondary structures, thus destabilizing the protein. Two main factors cause prolines to break secondary structures. One is the absence of the hydrogen on the amide nitrogen, which prohibits prolines from acting as a donor in a hydrogen bond. Another is the steric constraints placed on the proline and the neighboring residues by the pyrrolidine ring, hindering secondary structure formation. The phi and psi angles preferred by prolines are far from the typical range of those for β-sheets and thus, distort the strand significantly. In addition, steric restriction drives the residue preceding proline to prefer the β conformation, thus limiting the occurrence of prolines in α-helices.

Because of the contradictory effects described above, stabilization of proteins by proline incorporation has typically been achieved by trial and error. The general understanding is that proline mutation is stabilizing because of entropic factors, but this

behavior can be masked and even reversed by destabilizing enthalpic changes. Thus, protein stabilization by prolines has been achieved by placing prolines in relatively solvent-exposed locations where they would not disturb the stabilizing interactions of the protein, for example in loops and turns or the first turn of an α-helix (Watanabe and Suzuki 1998). Here we examine the effects of proline mutations in protein G and attempt to "predict" these effects using computational tools with the aim of reasonably incorporating the consideration of proline in computational protein design efforts.

Materials and methods

Mutagenesis and protein purification

Mutants of the B1 domain of protein G (Gβ1) were constructed by inverse polymerase chain reaction in plasmid pET11A, expressed using BL21 (DE3) cells and purified as previously described (Malakauskas and Mayo 1998). Two forms of Gβ1, 56 and 57 residue species, result due to incomplete processing of the N-terminal methionine. In this study, the 56-residue species of Gβ1 mutants and wild type were used. Molecular weights were verified by mass spectrometry.

Circular dichroism analysis

Circular dichroism (CD) data were collected on an Aviv 62DS spectrometer equipped with a thermoelectric unit. Thermal denaturation experiments were monitored at 218 nm from 1°C to 99°C by 1°C increments with an equilibration time of 1.5 minutes using 50 μM protein in 50 mM sodium phosphate at pH 5.5. The midpoint of the thermal unfolding transition (T_m) was determined from a two-state analysis of each denaturation

curve (Minor and Kim 1994). Guanidinium chloride denaturations were performed at 25°C using 5 μM protein in 50 mM sodium phosphate at pH 5.5. Data were collected for 5 minutes and averaged. Free energies of folding (ΔG_f) and error estimates were obtained by fitting the denaturation data to a two-state transition model (Santoro and Bolen 1988) using Kaleidagraph (Synergy Software). Chemical and thermal melting curves for protein G and its variants are presented in Figure 6.

Computational analysis

The crystal structure of wild-type Gβ1 (PDB entry 1pga) was used as the starting template for energy calculations. Explicit hydrogens were added using MolProbity (Lovell et al. 2003) and the structure was energy minimized for fifty steps to remove any steric clashes (Mayo et al. 1990). For each mutant, proline was substituted at the selected position and the protein design program, ORBIT (Optimization of Rotamers By Iterative Techniques) (Dahiyat et al. 1997; Dahiyat and Mayo 1997a; b; Street and Mayo 1998; Pierce et al. 2000), was used to optimize the structure (selecting the optimal rotamer for proline as well as for all the other residues in the protein) and to calculate energies. Solvation energies were calculated using either the method of Street and Mayo (1998) or Lazaridis and Karplus (1999).

Results and discussion

Proline mutants of Gβ1

Prolines tend to have a phi angle of approximately -63°, while the psi angle clusters around two regions in the Ramachandran map, -35° (α region) and 150° (β

region) (MacArthur and Thornton 1991). We selected ten Gβ1 residues with phi and psi angles compatible with proline for mutation: Thr2, Gly9, Lys10, Val21, Ala23, Ala24, Thr25, Val29, Asp36 and Ala48. In order to explore the effect of prolines in different structural environments, these residues were selected from various regions of the protein (Figure 3). Thr2 and Gly9 are located in a β-strand. Ala23, Ala24 and Thr25 are the first three N-terminal residues on the α-helix, Val29 is in the middle of the helix, and Asp36 is the C-terminal residue of the helix. The remaining residues are located in the loops and turns connecting the secondary structural elements. The phi and psi angles of the preceding residue of Thr2, Gly9, Lys10, Val21 and Ala23 are in the β region of the Ramachandran map. Residues preceding proline prefer the β region because their C_β and amide nitrogen sterically conflict with the C_δ of proline (Schimmel and Flory 1968; Matthews et al. 1987; Hurley et al. 1992).

Stability studies and analysis

The stability of each of the mutants was determined by performing thermal and chemical denaturation experiments (Table 1 and Figure 6). The far UV CD spectra before and after thermal denaturation indicate that all mutants except for K10P, T25P and V29P fold reversibly (data not shown). The post-transition region of the melting curves for V21P and V21P/A23P extends beyond the experimental range of 99°C, which leads to large estimated errors.

As expected from the fact that proline is a well-known secondary structure breaker, most of the proline mutants were less stable than the wild-type protein. An exception to this is V21P, which exhibited a ΔG_f enhancement of 0.5 kcal/mol and a T_m

increase of 6°C compared to the wild-type protein. This value agrees well with the expected energy of stabilization generated by the entropic difference between Val and Pro in the unfolded state. According to the method of Nemethy et al. (1966) a Val to Pro mutation should increase stability by 0.5 kcal/mol, while the method of Stites and Pranata (1995) suggests an increase of 0.3 kcal/mol at 25°C. Comparisons between mesophilic and thermophilic proteins and mutational studies indicate that proline residues located in loops help increase the rigidity of the loop and thus, increase the stability of the protein (Vieille and Zeikus 1996). Val21 is the first residue in a two-residue loop, which connects one of the edge strands of the β-sheet and the N-terminus of the α-helix. It is solvent-exposed and does not interact with other residues; thus, mutation to proline does not disturb any energetically favorable interaction. Given these observations, it is not surprising that the increase in stability for V21P is close to the expected value.

Two residues, Lys10 and Ala48, are located in the i+1 position of a β-turn. The fact that proline is the most favored residue for the i+1 position of β-turns has been rationalized by analysis of protein structures. These studies reveal that prolines have phi angles that are favored in that position (Hutchinson and Thornton 1994). In addition, mutational studies report an increase in thermostability with proline substitution in this position (Watanabe and Suzuki 1998). In our case, these two proline mutants are slightly destabilizing compared to the wild-type protein. This may be partly due to the loss of a hydrogen bond between the amide nitrogen of Ala48 and the carboxylate of Asp46, and for Lys10, the loss of electrostatic interactions with negatively charged residues in the vicinity (Asp40, Glu56 and the C terminal carboxyl group).

Prolines occur in the first turn of α-helices with high frequency. It has been suggested that proline residues are not destabilizing in this position because the amide hydrogens in the first turn do not make backbone/backbone hydrogen bonds within the helix (von Heijne 1991). Within the first turn, prolines exist predominantly at the N1 position because steric clashes will result if the preceding residue is in a helical conformation (Yun et al. 1991; Cochran et al. 2001). In our experiments, proline mutation at the N2 position (A24P) was more destabilizing than mutation at the N1 position (A23P), consistent with previous observations. However, both proline mutations were slightly destabilizing. Contrary to some mutational studies, but consistent with our results for A23P and A24P as well as for K10P and A48P, peptide helicity measurements demonstrated that prolines do not stabilize N terminal residues of α-helices or the i+1 residue of β-turns. Instead, it was suggested that prolines occur frequently in certain locations because they are tolerated rather than stabilizing (Cochran et al. 2001). Proline substitution at Thr25, which is located in the N3 position, was highly destabilizing with $\Delta\Delta G_f$ of 2.8 kcal/mol and T_m decrease of 21°C compared to the wild-type protein. The amide nitrogen of residues in the first turn of an α-helix frequently makes hydrogen bonds to a nearby side chain such as the N-cap residue in order to satisfy its hydrogen bond donor potential. This is particularly true for the amide nitrogen of an N3 residue which has frequently been observed to form a hydrogen bond to the N-cap side chain (Penel et al. 1999). In the wild-type Gβ1 structure, the backbone and side chain of Thr25 make extensive hydrogen bonds with the carboxylate group of the N-cap residue, Asp22. This interaction stabilizes and caps the first turn of the helix (Gronenborn et al. 1991).

Substituting proline at this position eliminates this stabilizing interaction, thus destabilizing the protein.

It has been reported that the cost of introducing a proline into an α-helix is about 3.4 kcal/mol (Oneil and Degrado 1990; Yun et al. 1991). Consistent with this, our Val29 to proline mutation destabilized the protein by 3.5 kcal/mol and decreased the T_m by 23°C.

Proline substitution in the β-sheet at Thr2 and Gly9 destabilizes the protein by distorting the secondary structure of the β-sheet. The far UV CD spectra of these two mutants deviate significantly from the wild-type spectrum, suggesting a change in secondary structure content (data not shown).

Introduction of increasing numbers of prolines, up to nine, additively increased the stability of oligo–1,6-glucosidase (Watanabe et al. 1994). To explore whether the stabilizing/destabilizing effects of our proline mutants were additive, we constructed a double mutant, V21P/A23P. To a first approximation, the effect was additive, resulting in a near zero effect on stability, as can be seen in Table 1.

Comparison of computational and experimental energies

The energy of each of the mutants and the wild-type protein was calculated by substituting proline at the respective positions and optimizing all side chains with ORBIT. We used two different methods of calculating solvation energy, a surface area dependent solvation model from Street and Mayo (1998) and a Gaussian solvent exclusion model from Lazaridis and Karplus (1999). The energies are reported in Table 1. Mutants T2P, G9P, V29P and D36P showed significant backbone movement in the

mutated region after minimization of the optimized structure determined by ORBIT and/or exhibited large deviations of their far UV CD spectra compared to the wild type. Since ORBIT utilizes the static backbone of the wild-type crystal structure for its calculations, the ORBIT energy of these mutants is not likely to reflect the energy of the true structure. Thus, they were not considered in the correlation analysis between the calculated ORBIT energy (E_{calc}) and ΔG_f obtained by experiment. However, we would like to point out that ORBIT does predict the destabilizing effects of three (G9P, V29P, D36P) out of the four mutants excluded. These mutants all have very high computed energies, due to large van der Waals clashes between the side chain and the backbone. We propose that this is the reason for these mutants' destabilization and deviation of their CD spectra.

Excluding the mutants mentioned above, the agreement between the ORBIT energy difference between the mutant (P) and the wild type (WT) (ΔE_{calc}(P-WT)), and the experimentally determined free energy difference ($\Delta\Delta G_f$(P-WT)) resulted in an R^2 values of 0.79 using the Street and Mayo (SM) solvation model (Figure 4A), and 0.94 using the Lazaridis and Karplus (LK) solvation model (Figure 5A). Correlation between T_m and ORBIT energy resulted in R^2 values of 0.78 and 0.79, for the SM and LK models, respectively (Figure 4B and Figure 5B). In order to estimate the prediction error of ORBIT energies more accurately and to consider whether the correlation for the data set is dominated by the result for T25P, we used the "leave-one-out" cross validation method on the free energy data set. The cross validation estimate of prediction error was 14.5 for the SM method, while the LK method gave a significantly lower value of 0.12. Thus, the Lazaridis and Karplus excluded volume solvation model based ORBIT energies show

greater correlation to experimental ΔG_u and a lower estimate of prediction error compared to the surface area dependent solvation model based energies, suggesting that the LK model performs better in describing the proline mutants. Overall, despite some false positives, ORBIT is reasonably predictive in ranking the stabilities of the various mutants as indicated by Spearman's rank correlation coefficients of 0.75 ($P < 0.01$) and 0.93 ($P < 0.01$) for the SM and LK free energy correlations, respectively (data not shown).

In our hands, prolines have typically not been included in the set of amino acids that ORBIT considers in protein design calculations because the potential energy function used in ORBIT does not include a conformational entropy term. We tested whether including a backbone conformational entropy term to the computational energy increases the rank correlation between the experimental stabilities and computational energies. Using the backbone entropy scale from Stites and Pranata (1995), the weighting factor for the entropy term was determined by optimizing for the rank correlation between the experimental and computational results. For computational energies calculated with the LK model, a weighting factor in the range of 2.9 to 12.9 gave a rank correlation of 0.96 ($P < 0.01$). For those calculated with the surface area based method, a weighting factor in the range of 15.01 to 15.2 also gave a rank correlation of 0.96 ($P < 0.01$).

As described above, proline stabilizes a protein by decreasing the backbone conformational entropy of the unfolded state. ORBIT predicts the stability of proline mutations reasonably well without an entropic term. This is likely due to the dominance of enthalpic contributions over entropic contributions in protein stability modulation by prolines, which overshadows the missing entropic term in the energy function. Nevertheless, addition of an entropic term with an appropriate weighting factor increases

the correlation between computational energy and experimental energy, especially for the surface area based solvation method.

References

Cochran, D.A., Penel, S., and Doig, A.J. 2001. Effect of the N1 residue on the stability of the alpha-helix for all 20 amino acids. *Protein Sci* **10:** 463-470.

Dahiyat, B.I., Gordon, D.B., and Mayo, S.L. 1997. Automated design of the surface positions of protein helices. *Protein Sci* **6:** 1333-1337.

Dahiyat, B.I., and Mayo, S.L. 1996. Protein design automation. *Protein Sci* **5:** 895-903.

Dahiyat, B.I., and Mayo, S.L. 1997. *De novo* protein design: fully automated sequence selection. *Science* **278:** 82-87.

Dahiyat, B.I., and Mayo, S.L. 1997. Probing the role of packing specificity in protein design. *Proc Natl Acad Sci U S A* **94:** 10172-10177.

Gronenborn, A.M., Filpula, D.R., Essig, N.Z., Achari, A., Whitlow, M., Wingfield, P.T., and Clore, G.M. 1991. A novel, highly stable fold of the immunoglobulin binding domain of streptococcal protein G. *Science* **253:** 657-661.

Humphrey, W., Dalke, A., and Schulten, K. 1996. VMD: Visual molecular dynamics. *J Molec Graphics* **14:** 33-38.

Hurley, J.H., Mason, D.A., and Matthews, B.W. 1992. Flexible-geometry conformational energy maps for the amino acid residue preceding a proline. *Biopolymers* **32:** 1443-1446.

Hutchinson, E.G., and Thornton, J.M. 1994. A revised set of potentials for beta-turn formation in proteins. *Protein Sci* **3:** 2207-2216.

Lazaridis, T., and Karplus, M. 1999. Effective energy function for proteins in solution. *Proteins* **35:** 133-152.

Lovell, S.C., Davis, I.W., Arendall, W.B., 3rd, de Bakker, P.I., Word, J.M., Prisant,

M.G., Richardson, J.S., and Richardson, D.C. 2003. Structure validation by C alpha geometry: phi, psi and C beta deviation. *Proteins* **50:** 437-450.

MacArthur, M.W., and Thornton, J.M. 1991. Influence of proline residues on protein conformation. *J Mol Biol* **218:** 397-412.

Malakauskas, S.M., and Mayo, S.L. 1998. Design, structure and stability of a hyperthermophilic protein variant. *Nat Struct Biol* **5:** 470-475.

Matthews, B.W., Nicholson, H., and Becktel, W.J. 1987. Enhanced protein thermostability from site-directed mutations that decrease the entropy of unfolding. *Proc Natl Acad Sci U S A* **84:** 6663-6667.

Mayo, S.L., Olafson, B.D., and Goddard, W.A. 1990. Dreiding - a generic force-field for molecular simulations. *J Phys Chem-Us* **94:** 8897-8909.

Minor, D.L., Jr., and Kim, P.S. 1994. Measurement of the beta-sheet-forming propensities of amino acids. *Nature* **367:** 660-663.

Nemethy, G., Leach, S., and Scheraga, H. 1966. The influence of amino acid side chains on the free energy of helix-coil transitions. *J Phys Chem* **70:** 998-1004.

Oneil, K.T., and Degrado, W.F. 1990. A Thermodynamic scale for the helix-forming tendencies of the commonly occurring amino-acids. *Science* **250:** 646-651.

Penel, S., Hughes, E., and Doig, A.J. 1999. Side-chain structures in the first turn of the alpha-helix. *J Mol Biol* **287:** 127-143.

Pierce, N.A., Spriet, J.A., Desmet, J., and Mayo, S.L. 2000. Conformational splitting: A more powerful criterion for dead- end elimination. *J. Comput. Chem.* **21:** 999-1009.

Santoro, M.M., and Bolen, D.W. 1988. Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-

chymotrypsin using different denaturants. *Biochemistry* **27:** 8063-8068.

Schimmel, P.R., and Flory, P.J. 1968. Conformational energies and configurational statistics of copolypeptides containing L-proline. *J Mol Biol* **34:** 105-120.

Street, A.G., and Mayo, S.L. 1998. Pairwise calculation of protein solvent-accessible surface areas. *Fold Des* **3:** 253-258.

Vieille, C., and Zeikus, J.G. 1996. Thermozymes: Identifying molecular determinants of protein structural and functional stability. *Trends Biotechnol*. **14:** 183-190.

von Heijne, G. 1991. Proline kinks in transmembrane alpha-helices. *J Mol Biol* **218:** 499-503.

Watanabe, K., Masuda, T., Ohashi, H., Mihara, H., and Suzuki, Y. 1994. Multiple proline substitutions cumulatively thermostabilize Bacillus cereus ATCC7064 oligo-1,6-glucosidase. Irrefragable proof supporting the proline rule. *Eur J Biochem* **226:** 277-283.

Watanabe, K., and Suzuki, Y. 1998. Protein thermostabilization by proline substitutions. *J Mol Catal B* **4:** 167-180.

Yun, R.H., Anderson, A., and Hermans, J. 1991. Proline in alpha-helix - stability and conformation studied by dynamics simulation. *Proteins* **10:** 219-228.

Table 1. Midpoint of thermal unfolding transition (T_m), free energy of folding (ΔG_f) at 25°C, and computed energy for Gβ1 variants (errors determined from non-linear fits)

Name	T_m (°C)	ΔG_f (kcal/mol)	$\Delta\Delta G_f$ (kcal/mol)	E_{calc} (SM)[1] (kcal/mol)	E_{calc} (LK)[2] (kcal/mol)
wild type	89.6 ± 2.6	-5.9 ± 0.4		-90.1	-72.2
T2P	83.0 ± 1.2	-3.2 ± 0.3	2.7	-90.5	-72.9
G9P	72.6 ± 0.7	-3.5 ± 0.3	2.4	120.2	146.9
K10P	81.2 ± 1.1	-5.7 ± 0.4	0.2	-89.0	-71.9
V21P	95.8 ± 14.3	-6.4 ± 0.4	-0.5	-90.5	-73.2
A23P	88.2 ± 1.9	-5.6 ± 0.5	0.3	-93.5	-72.7
A24P	85.0 ± 0.8	-5.4 ± 0.3	0.5	-90.1	-71.2
T25P	68.7 ± 0.4	-3.1 ± 0.2	2.8	-77.5	-60.4
V29P	67.0 ± 0.5	-2.4 ± 0.2	3.5	1464.7	1487.1
D36P	68.6 ± 0.8	-2.8 ± 0.3	3.1	3310.3	3330.0
A48P	82.8 ± 0.7	-5.2 ± 0.3	0.7	-86.4	-70.8
V21P/A23P	96.5 ± 28.4	-5.8 ± 0.4	0.1	-93.9	-73.8

[1]ORBIT energy using Street and Mayo (1998) solvation model.
[2]ORBIT energy using Lazaridis and Karplus (1999) solvation model.

Figure 3. Views of the 10 positions in Gβ1 mutated to proline. These structural figures were generated using VMD (Humphrey et al. 1996).

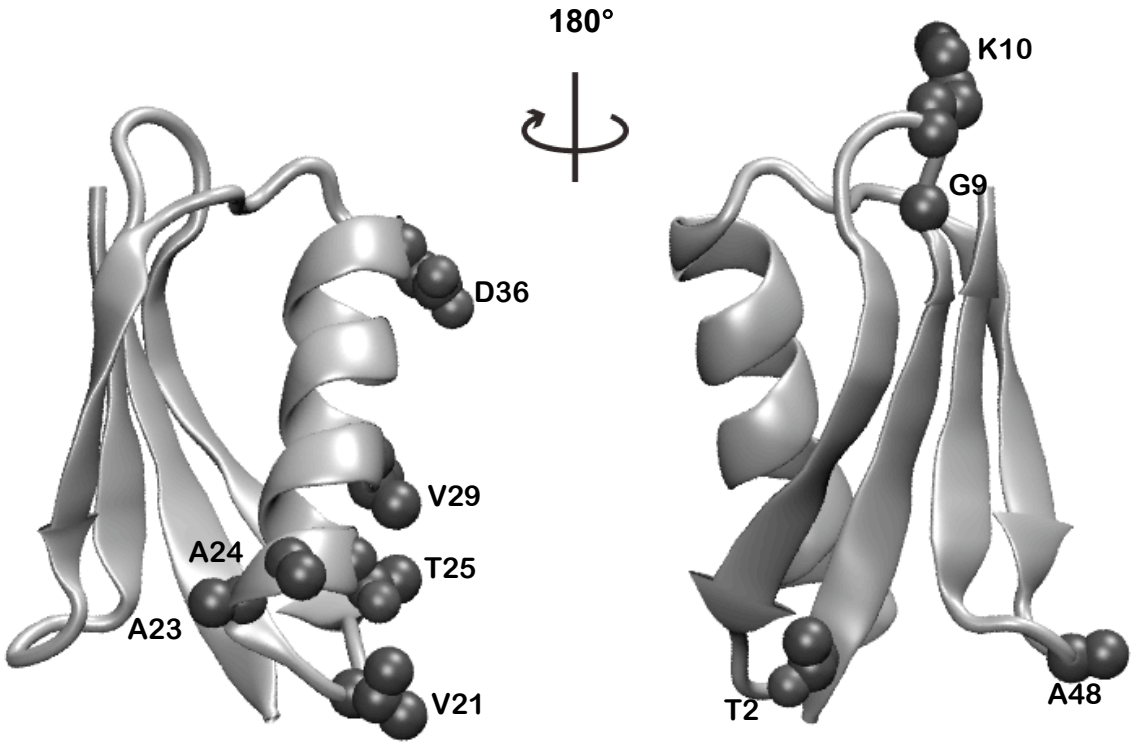
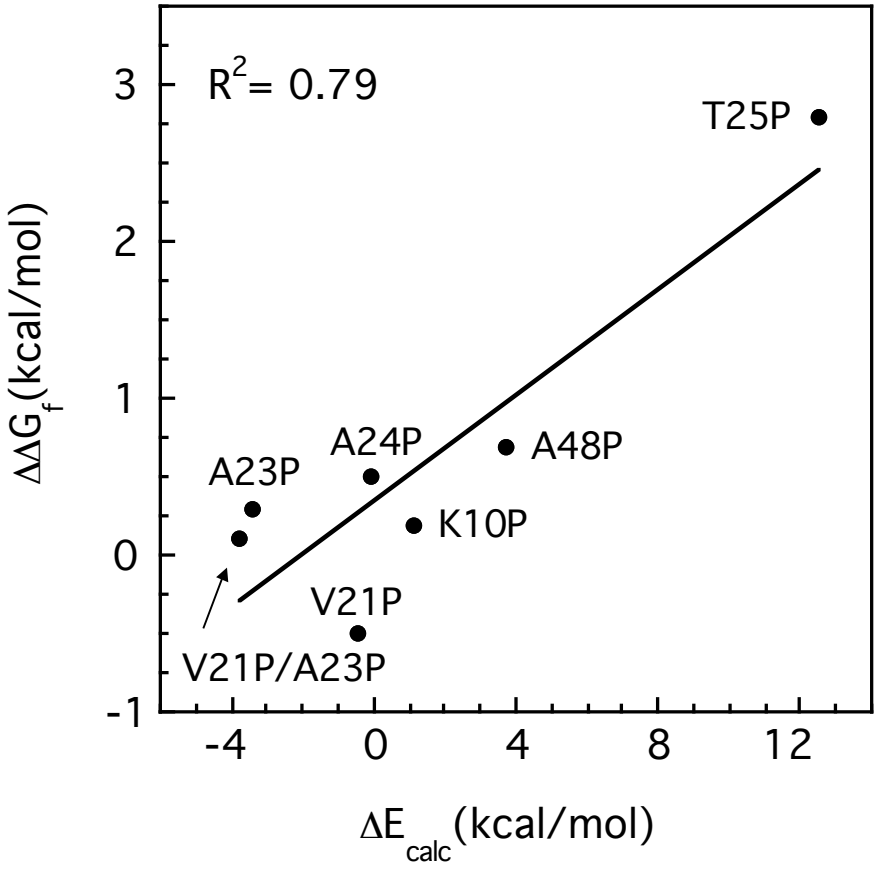

180°
D36
V29
A24
T25
A23
V21
K10
G9
T2
A48

Figure 4. Correlation between calculated energies, using the Street and Mayo (1998) solvation model, and experimental results. (A) Comparison of experimental ($\Delta\Delta G_f$) and calculated (E_{calc}) stability change between mutant and the wild type. (B) Comparison of T_m and calculated energy.

A


R2= 0.79
T25P
A24P
A48P
A23P
K10P
V21P
V21P/A23P
3
2
1
0
-1
-4
0
4
8
12
ΔΔGf(kcal/mol)
ΔEcalc(kcal/mol)

B

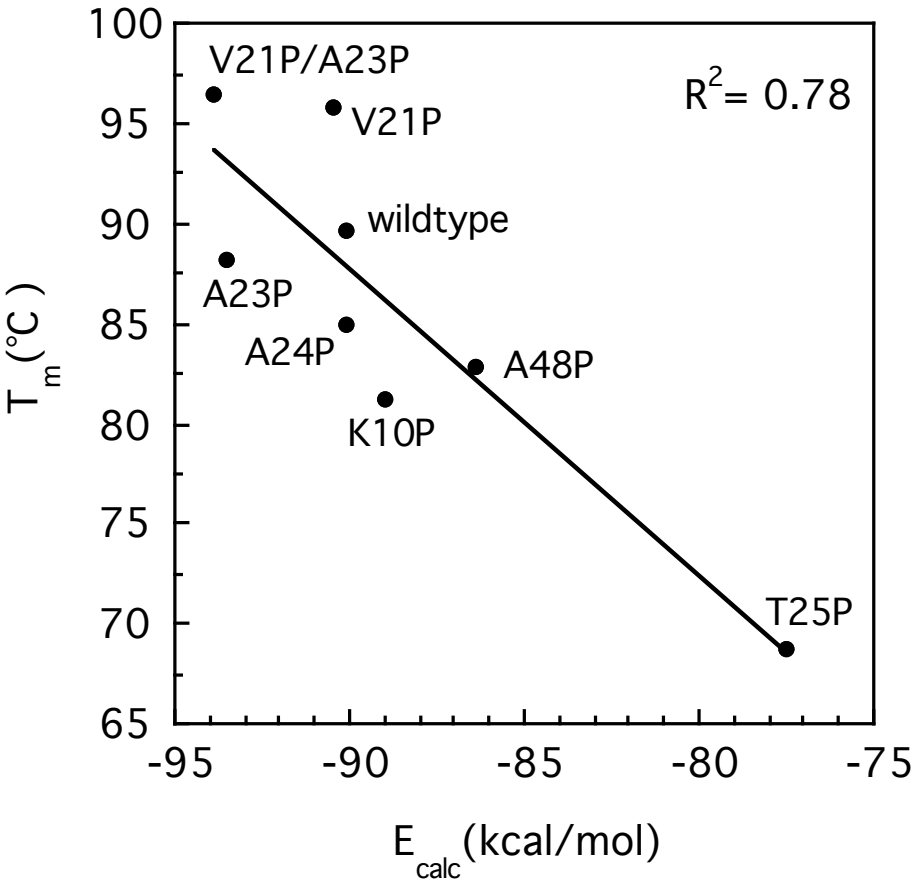
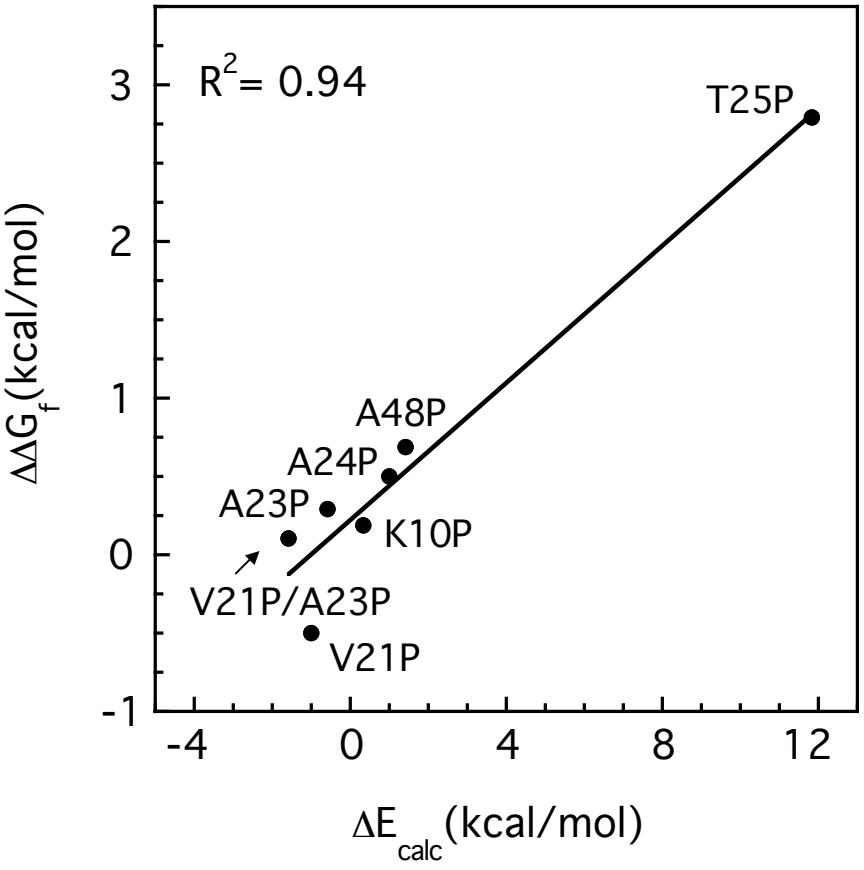

100
95
90
85
80
75
70
65
-95
-90
-85
-80
-75
V21P/A23P
V21P
wildtype
A23P
A24P
A48P
K10P
T25P
R2= 0.78
Tm (°C)
Ecalc (kcal/mol)

Figure 5. Correlation between calculated energies, using the Lazaridis and Karplus (1999) solvation model, and experimental results. (A) Comparison of experimental ($\Delta\Delta G_f$) and calculated (E_{calc}) stability change between mutant and the wild type. (B) Comparison of T_m and calculated energy.

A


R2= 0.94
T25P
A48P
A24P
A23P
K10P
V21P/A23P
V21P
3
2
1
0
-1
-4
0
4
8
12
ΔΔGf(kcal/mol)
ΔEcalc(kcal/mol)

B

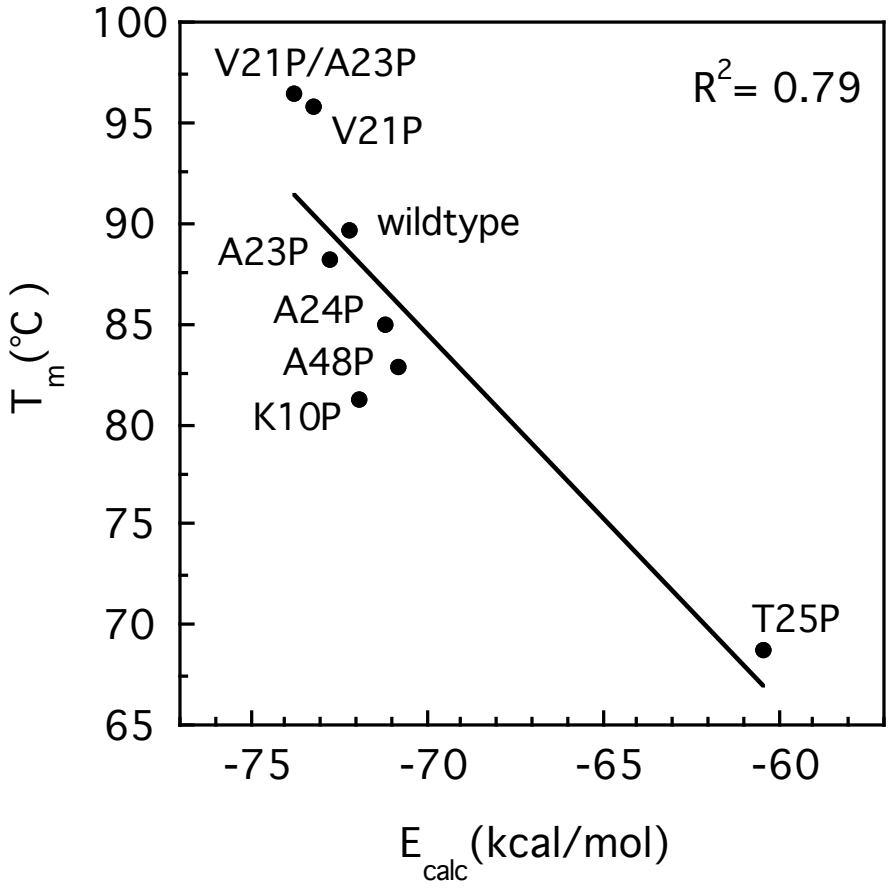
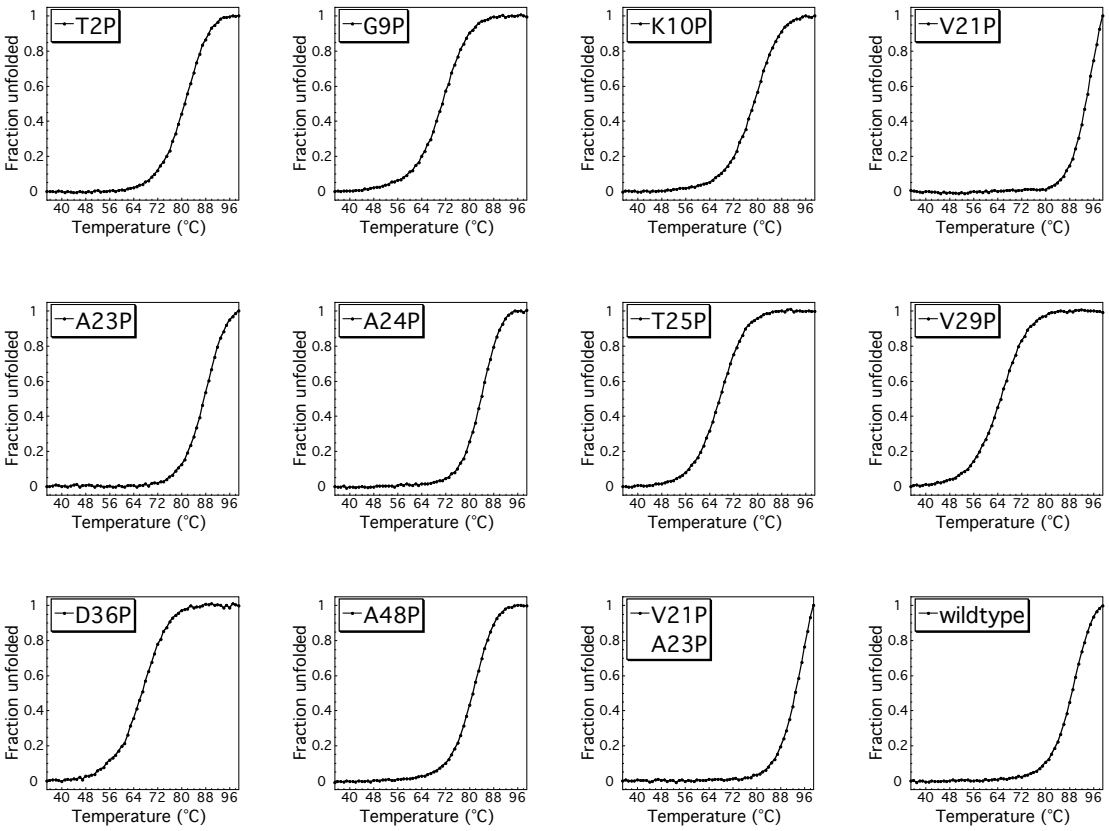
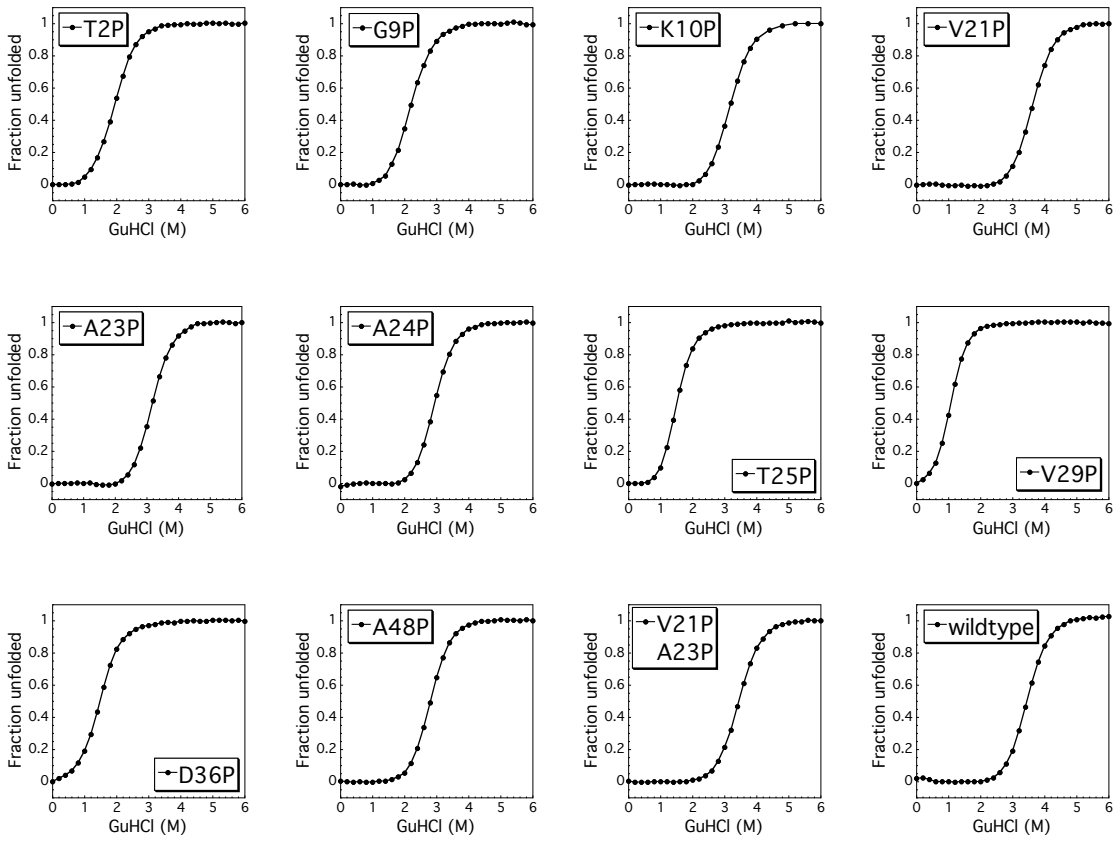

100
95
90
85
80
75
70
65
T_m (°C)
V21P/A23P
V21P
wildtype
A23P
A24P
A48P
K10P
T25P
R^2= 0.79
-75
-70
-65
-60
E_{calc} (kcal/mol)

Figure 6. CD data. (A) Thermal and (B) guanidinium chloride denaturation curves for the various proteins included in this study.

A


T2P
G9P
K10P
V21P
A23P
A24P
T25P
V29P
D36P
A48P
V21P
A23P
wildtype
Fraction unfolded
Temperature (°C)
0
0.2
0.4
0.6
0.8
1
40 48 56 64 72 80 88 96

B


T2P
G9P
K10P
V21P
A23P
A24P
T25P
V29P
D36P
A48P
V21P
A23P
wildtype
Fraction unfolded
GuHCl (M)

Chapter 3

Analysis of computationally designed TRAF6 binding peptides

Abstract

Members of the tumor necrosis factor receptor (TNFR) associated factor (TRAF) family bind to a variety of cell surface receptors and are involved in adaptive and innate immunity, stress response and tissue homeostasis. Starting with the TRAF6-CD40 peptide complex crystal structure, we used the ORBIT computational protein design software to design peptides that would bind to TRAF6 with high affinity. A fluorescence anisotropy binding assay was developed and used to measure binding affinities. All designed peptides had binding affinities similar to the CD40 peptide, with the best showing about a two-fold increase.

Introduction

Tumor necrosis factor receptor (TNFR) associated factors (TRAFs) bind to a variety of cell surface receptors, acting as adaptors in the activation of diverse down stream molecules. They are involved in a wide range of biological functions, including adaptive and innate immunity, stress response and tissue homeostasis (Chung et al. 2002). To date, seven members of the TRAF family have been identified in mammals: TRAF1-7. All TRAF family proteins, excluding TRAF7, contain a TRAF domain at the carboxy terminus, which is further divided into an amino terminal TRAF-N domain and a highly conserved carboxy terminal TRAF-C domain. The TRAF-N domain forms a coiled coil and is involved in oligomerization, while the TRAF-C domain forms an antiparallel β sandwich and interacts with upstream receptors (Figure 7) (Park et al. 1999; Ye et al. 2002a).

TRAF6 mediates signal transduction for members of both the TNFR superfamily and the interleukin-1 receptor (IL-1R)/Toll-like receptor (TLR) superfamily. Among the TNFR superfamily members, the interaction of TRAF6 with CD40 and RANK has been studied extensively. TRAF6 plays an important role in CD40-induced activation of antigen presenting cells and interacts with RANK and thus, affects dendritic cell survival and osteoclast differentiation. TRAF6 differs from the other members of the TRAF family in many ways. It has a unique recognition motif (Pro-X-Glu-X-X-aromatic/acidic residue) and a different mode of ligand binding. Comparison of TRAF6 and TRAF2 crystal structures reveals completely different ligand binding sites. The direction of the bound peptides differs by 40° and TRAF6 ligands assume an extended beta conformation, while TRAF2 ligands are helical (Ye et al. 2002a). TRAF6 is also very

unique in that it activates transcription factors through different downstream components and is the only TRAF known to activate the Src tyrosine kinase family. TRAF6 plays a critical role in (1) adaptive immunity by interacting with CD40 and RANK; (2) bone resorption by interacting with RANK; and (3) innate immunity by interacting with members of the IL-1R/TLR superfamily. TRAF6 has also been implicated in the development of epidermal appendices and central nervous system development (Wu and Arron 2003).

Recently, Wu and co-workers determined the crystal structures of the TRAF-C domain of TRAF6 alone and in complex with receptor peptide ligands. The authors showed that the peptide ligands block TRAF6 mediated signal transduction in RAW264.7 cells and primary mouse derived monocytes. These cells differentiate into multinucleated, tartrate resistant acid phosphatase (TRAP)-positive osteoclasts when stimulated with TRANCE, the ligand for RANK. When co-treated with TRAF6 binding peptide ligands, they observed a dose-dependent decrease of TRAP-positive osteoclasts (Ye et al. 2002a). These results suggest that high affinity peptide ligands of TRAF6 may prove useful in the treatment of osteoporosis.

With this in mind, we used the ORBIT (Optimization of Rotamers By Iterative Techniques) protein design software to design several peptide ligands for TRAF6. The binding affinities of the peptides were determined using anisotropy from the fluorophores attached to the peptides. This method was used instead of isothermal titration calorimetry (ITC), which has been employed by others, because the aggregation-prone characteristics of the protein gave inconsistent results in our ITC experiments. The sensitivity of fluorescence anisotropy allowed us to use nanomolar concentrations of proteins while

ITC experiment required 10-fold excess. All our designed peptides exhibited binding affinities similar to the CD40 peptide, suggesting that the CD40 peptide is highly optimized for the wide binding groove of the TRAF6 protein. Our best design showed a two-fold improvement in TRAF6 binding affinity compared to the CD40 peptide. Since TRAF6 depends on avidity based affinity enhancement (Ye and Wu 2000), we predict that in the context of the full trimer protein receptor, our best design will show over an eight-fold enhancement in TRAF6 binding affinity compared to CD40.

Materials and methods

Protein purification and peptide synthesis

TRAF6 constructs (residues 333-508) containing part of the TRAF-N domain and all of the TRAF-C domain were obtained from Dr. Hao Wu (Cornell Medical College) and was previously described (Ye et al. 2002b). The constructs were expressed in BL21 (DE3) cells and induced overnight at 20°C. The recombinant proteins were purified by Ni^{2+} affinity chromatography and gel filtration using a Superdex 200 column (GE Healthcare). All peptides were chemically synthesized at the Biopolymer Synthesis Center, California Institute of Technology. All peptides used in the fluorescence assay were synthesized with fluorescein attached to the amino terminus and an amide group attached to the carboxy terminus. The peptides used in ITC experiments were acetylated at the amino terminus and amidated at the carboxy terminus to mimic the intact protein. The molecular mass of each peptide was verified by matrix assisted laser desorption ionization time of flight (MALDI-TOF) mass spectrometry. The CD40 peptide used in

this study is a variant from wild-type, with an Asn to Asp mutation at the P_2 position to enhance affinity to TRAF6.

Computational methods

The 1.8 Å resolution crystal structure of TRAF6 in complex with the nine-residue human CD40 peptide (PDB ID: 1LB6) was used as the initial structure for ORBIT. Explicit hydrogens were added using MolProbity (Lovell et al. 2003) and the structures were minimized for 50 steps to remove any steric clashes. For all designs, ORBIT was used to calculate the energies and predict the global minimum energy conformation (GMEC) sequence (Dahiyat et al. 1997; Dahiyat and Mayo 1997a; b; Street and Mayo 1998; Pierce et al. 2000). We used a large backbone-dependent rotamer library, which was expanded about the $\chi 1$ and $\chi 2$ dihedral angles by one standard deviation. On the nine-residue CD40 peptide, three consensus residues (P-X-E-X-X-F) and the first residue, which did not make direct contact with TRAF6, were excluded from the design; the other five residues were allowed to change identity. On the TRAF6 protein, residues within 12 Å of the peptide were allowed to change conformation, and the rest were held fixed. Interaction energies were calculated by using a Monte Carlo search algorithm. 1000 best (lowest energy) sequences were obtained by using the GMEC as the starting structure. The apo-TRAF6 and the peptide ligand were separated in each of these 1000 protein peptide complexes, and their individual energies were calculated and subtracted from the energy of the complex to obtain the interaction energy. The sequences were resorted by interaction energy and the top ranking sequences were inspected visually. ORBIT optimizes for all interactions in the modeled structure, whether they are intramolecular or

intermolecular. To circumvent this problem, we used a modified energy function that biases toward intermolecular interactions (Shifman and Mayo 2003).

Isothermal titration calorimetry (ITC) and fluorescence anisotropy measurement

ITC data were obtained using a micro calorimetry system from MicroCal (Amherst, Ma). All titration experiments were done at 20°C. 5 mM peptide ligand was injected into 1.5 ml of 0.03 ~ 1.1 mM TRAF6 in 3 ~ 5 μl volumes. The titration data were analyzed with the ORIGIN software (MicroCal). The constant heat at the end of the experiment was assumed to be the control background heat and was subtracted out from the raw data before fitting. Nonconstraint fitting was performed for all the samples to extract the binding affinity (K_d).

Fluorescence anisotropy data were obtained using a fluorometer from Photon Technology International (Lawrenceville, NJ). Experiments were done at room temperature or at 4°C. Serial dilutions of TRAF6 were prepared with 12.5 nM of peptide at final volume of 1 ml. All samples were incubated overnight at 4°C and measurements were taken immediately thereafter. Anisotropy measurements were taken for 1 minute and averaged. The average anisotropy and the concentration of TRAF6 were plotted and fit to a binding equation by nonlinear least squares analysis (Lundblad et al. 1996) using Kaleidagraph (Synergy Software).

Results and discussion

Parameterization of ORBIT for the design of high affinity peptides

ORBIT was developed and optimized for stabilization of single isolated proteins with emphasis on hydrophobic packing. We therefore predicted that the default parameters used for protein stabilization would not be optimal for increasing the affinity between a peptide and a protein, especially for TRAF6 and CD40 peptide, where there is little hydrophobic contact. To determine the best combination of ORBIT parameters for this system, we analyzed the results of an experimental colorimetric binding assay in which all the binding site residues were singly substituted to all 20 amino acids (Pullen et al. 1999). Peptide sequences from the assay were divided into two categories: those that bind to TRAF6 (binder), and those that do not (nonbinder). We tried different force field parameter values to calculate the energies of all the peptides used in the experimental assay by ORBIT and looked at how well the energies correlated with the experimental results. We varied the distance dependent dielectric constant (4r or 40r), and used surface area based solvation with either penalties for polar hydrogen burial or polar surface area burial. We calculated two energies: total energy and interaction energy, defined as the sum of all enthalpic interaction energies between the protein and the peptide. The goodness of fit between the experimental data and the computed energies was determined using two scores: T test between the mean energies of the binders and the nonbinders, and the number of false positives. False positives were defined as nonbinders with total energy or interaction energy better than the median of the binders' energies. The results were first sorted by the T score, then the number of false positives were taken into consideration (Table 2). Overall, using total energies gave better correlations than using

interaction energies, and in contrast to our prediction, the default parameter set used in designing isolated proteins ranked very well. We selected four peptides to test experimentally: three were obtained from calculations with parameters which had high-ranking T scores and were selected based on total energies (D1 to D3), and one was predicted using interaction energies (D4) (Table 2).

Binding affinity measurements

Previously, affinities of various peptides to TRAF6 were determined by ITC (Ye et al. 2002a). Unfortunately, in our hands, TRAF6 had limited solubility and aggregated during the ITC run, giving us inconsistent binding results that were, on average, an order of magnitude better than the published values. We therefore developed a fluorescence anisotropy assay to determine the binding affinity of the designed peptides (Figure 8 and Table 3). We synthesized four of the peptides ORBIT predicted using the parameters described above (D1 to D4, Table 3) and measured their binding affinities. Designed peptides D3 and D4 were overly hydrophobic, so additional PDD sequence was added to their C terminus to improve solubility. All affinities reported in Table 3 are relative to the CD40 peptide of the same length as the peptide tested (9-residue or 12-residue). Surprisingly, the K_d values we obtained by fluorescence anisotropy were similar to the K_d values obtained with ITC, and were an order of magnitude better than the K_d value previously reported using ITC (Ye et al. 2002a) (Table 4). Nevertheless, we are confident of the relative affinities from our anisotropy measurements because the standard deviations from multiple experiments are very small (Table 4). Our anisotropy assay was tested for nonspecific binding with two negative control mutant CD40 peptides. The two

negative control peptides used were a mouse CD40 peptide with the Pro at P_{-3} position mutated to Gln (mCD40-P/Q) and human CD40 peptide with the Pro at P_{-3} position mutated to Gly and Glu at P_0 position mutated to Ala (CD40-P/G,E/A) (Table 3). Previously, mCD40-P/Q was shown to have about half the activity of wild-type mouse CD40 peptide in an NF-κB reporter activity assay, and CD40-P/G,E/A was shown *in vitro* to have no affinity for TRAF6 (Pullen et al. 1999; Ye et al. 2002a). In our anisotropy assay, both show little or no binding to TRAF6 (Figure 8).

Other design trials

In the initial designs, our best peptide (D1) exhibited about a two-fold affinity enhancement compared to the CD40 peptide sequence. This improvement encouraged us to try to design a better binder using the knowledge gained from the first four peptides. We decided to use ORBIT's bias function , described in the methods section, to favor the intermolecular interactions between the peptide side chains and the TRAF6 side chains. Using the parameters for D1, we applied a 2-fold bias and a 4-fold bias to the intermolecular optimization, which resulted in the two sequences, D5 and D6 (Table 3). Both are one mutation away from D1, at the P_1 position: instead of an Ile at P_1, D5 has a Leu, and D6 has a Trp. Interestingly, the exact same sequences are obtained when the parameters for D2 are biased 2-fold and 4-fold. We expected D5 to have less affinity for TRAF6 because Ile has higher beta strand propensity than Leu. The ligand peptide binds to TRAF6 in an extended beta strand conformation and a Leu instead of Ile should destabilize the peptide in the beta strand conformation. As predicted, D5 has significantly lower affinity than CD40 peptide. The experimental data of Pullen and co-workers

indicated that D6, however, might bind tighter. This data was used above in determining the ORBIT parameters, in which all 20 amino acids were substituted in all positions of the peptide and binding was observed using membrane binding assays. In their assay, substituting Trp for Ile resulted in 4-fold better binding to TRAF6. In our anisotropy assay, D6 affinity was similar to the wild-type sequence and less than D1. This might be explained by the fact that Pullen and co-workers used a crude colorimetric plate assay while we used purified proteins in a solution state assay. Also, their Trp mutation is in the context of the naturally occurring wild-type sequence, while our designed peptide has two other residues different from the naturally occurring CD40 peptide.

Comparing the D1, D2, D5 and D6 sequences indicates that among the residues tested at P_1, Ile is the most stable. Met, on the other hand, was thought to be destabilizing because of the large entropic loss upon binding. We therefore synthesized and tested another peptide (D7) with the same sequence as D3, but with Ile instead of Met at P_1. Ile turned out to be less optimal than Met in the context of D3, resulting in decreased binding affinity.

Another strategy we tried was to design one of the conserved sequences. When the peptide residues are classified as core, boundary or surface (Dahiyat and Mayo 1997a), only the Pro located in the P_{-2} position is classified as core. This suggested that P_{-2} is the anchoring position in the TRAF6-peptide interaction, and that it makes a large contribution to the binding affinity. It has been shown that hydrophobic interactions are important in affinity while polar interactions are important for specificity (Clackson and Wells 1995). Taking this into consideration, the P_{-2} residue, which is the only completely buried residue at the interface, was thought to be important in determining the binding

affinity of the peptide. To determine whether a larger hydrophobic side chain at P_{-2} would be allowed, we ran three side chain replacement calculations substituting the Pro to Phe, Tyr and Trp. There were no significant steric clashes with all three substitutions when the VDW scale factor for D1 design (0.9) was used. Trp and Phe had high nonpolar exposure, while the hydroxyl group of Tyr made a favorable hydrogen bond with Asp451 of TRAF6. Thus, we decided to test out a peptide with the same sequence as D1 except for a Pro to Tyr substitution at P_{-2} (D8). Unfortunately, the Pro to Tyr mutation decreased the binding affinity to 9% of D1. This implies that Tyr might be sterically clashing with TRAF6, which did not show in our side chain replacement calculation because of the small VDW scale factor used.

Affinity versus avidity

It has been pointed out that the low affinity nature of TRAF-receptor ligand binding ensures that TRAFs bind to their receptors only when the receptors are trimerized and active (Ye and Wu 2000). The structural characteristics of the binding site, which forms a wide and shallow groove, implies a relatively flat binding energy landscape that allows many different sequences to attach, however with low affinity. In addition, the peptide ligand assumes an extended beta conformation when fitting into this binding site, allowing only half of the residues of the peptide to face the protein and be involved in direct contact (Figure 7) The other side chains are facing away from the protein and thus, are highly exposed to the solvent, which may decrease their hydrogen bonding and electrostatic interaction contribution to the binding energy. The fact that the majority of the hydrogen bonding interactions between TRAF6 and the peptide ligand in the crystal

structure are between backbones and that only one interface residue is fully buried implies that the binding affinity between the peptide and protein will be nontrivial to optimize. We have tried to overcome this difficulty by using a biased energy function to optimize specifically for interactions between the side chains of the peptide and TRAF6, but this did not result in a high affinity peptide (D5 and D6). It could be that nature has designed the binding interface of TRAF6 and its receptor to maintain low affinity so that affinity can be achieved through avidity. The fact that the affinities of naturally occurring TRAF6 binding domains are all in the micromolar range (Ye et al. 2002a), even though the sequence diversity is very high in the nonconsensus positions, also supports this hypothesis.

The dynamic nature of proteins is not represented in our design protocol, which employs a fixed backbone. Flexible backbones have been successfully used to design single proteins (Harbury et al. 1998; Larson et al. 2002; Kraemer-Pecore et al. 2003; Kuhlman et al. 2003). Recently, several groups have also designed protein-binding peptides using this method (Wollacott and Desjarlais 2001; Sood and Baker 2006). In these studies, a flexible backbone was necessary to obtain the sequence diversity of experimentally verified ligands. This indicates that a flexible backbone is desirable for successful design and specificity prediction. It would be interesting to see whether a flexible backbone will improve the design effort for the TRAF6 and CD40 peptide system. On the other hand, some results imply that a flexible backbone is not necessary, and that allowing side chain flexibility or decreasing the van der Waals radii scale of the atoms to reduce the effects of using discrete rotamers and a fixed backbone is enough to obtain good results (Su and Mayo 1997; Desjarlais and Handel 1999). If this is the case,

our design, which used a large rotamer library with expansion about both $\chi 1$ and $\chi 2$, and used 0.9 van der Waals scale factor, should give an optimal sequence, even though backbone flexibility was not included. Even with a flexible backbone, the affinity increase for the best designed peptide for the dystrophin protein and Mdm2 were very modest (2.3 fold and 1.3 fold, respectively), comparable to the result we obtained for TRAF6 using a fixed backbone (Sood and Baker 2006). The authors attributed this to the fact that the binding groove is shallow and flat, which is also the case with our protein system. This suggests that a flexible backbone might not benefit designs for our protein system.

We used the ORBIT protein design software to design eight peptides to bind to TRAF6. The relative affinities of our peptides compared to the CD40 peptide ranged from 16% to172%. The modest affinity increase of the best design and the overall similar affinities of all the peptides might be due to the wide, shallow binding site on TRAF6. A high affinity peptide binder might be disallowed because the entropic cost of rigidifying a nine-residue peptide might be much higher than the optimal enthalpic benefit that can be achieved with an extended beta conformation peptide at such a wide and shallow binding site.

References

Chung, J.Y., Park, Y.C., Ye, H., and Wu, H. 2002. All TRAFs are not created equal: common and distinct molecular mechanisms of TRAF-mediated signal transduction. *J Cell Sci* **115:** 679-688.

Clackson, T., and Wells, J.A. 1995. A hot spot of binding energy in a hormone-receptor interface. *Science* **267:** 383-386.

Dahiyat, B.I., Gordon, D.B., and Mayo, S.L. 1997. Automated design of the surface positions of protein helices. *Protein Sci* **6:** 1333-1337.

Dahiyat, B.I., and Mayo, S.L. 1997a. *De novo* protein design: fully automated sequence selection. *Science* **278:** 82-87.

Dahiyat, B.I., and Mayo, S.L. 1997b. Probing the role of packing specificity in protein design. *Proc Natl Acad Sci U S A* **94:** 10172-10177.

Desjarlais, J.R., and Handel, T.M. 1999. Side-chain and backbone flexibility in protein core design. *J Mol Biol* **290:** 305-318.

Harbury, P.B., Plecs, J.J., Tidor, B., Alber, T., and Kim, P.S. 1998. High-resolution protein design with backbone freedom. *Science* **282:** 1462-1467.

Humphrey, W., Dalke, A., and Schulten, K. 1996. VMD: Visual molecular dynamics. *J Molec Graphics* **14:** 33-38.

Kraemer-Pecore, C.M., Lecomte, J.T., and Desjarlais, J.R. 2003. A *de novo* redesign of the WW domain. *Protein Sci* **12:** 2194-2205.

Kuhlman, B., Dantas, G., Ireton, G.C., Varani, G., Stoddard, B.L., and Baker, D. 2003. Design of a novel globular protein fold with atomic-level accuracy. *Science* **302:** 1364-1368.

Larson, S.M., England, J.L., Desjarlais, J.R., and Pande, V.S. 2002. Thoroughly sampling sequence space: large-scale protein design of structural ensembles. *Protein Sci* **11:** 2804-2813.

Lovell, S.C., Davis, I.W., Arendall, W.B., 3rd, de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. 2003. Structure validation by C alpha geometry: phi, psi and C beta deviation. *Proteins* **50:** 437-450.

Lundblad, J.R., Laurance, M., and Goodman, R.H. 1996. Fluorescence polarization analysis of protein-DNA and protein-protein interactions. *Mol Endocrinol* **10:** 607-612.

Park, Y.C., Burkitt, V., Villa, A.R., Tong, L., and Wu, H. 1999. Structural basis for self-association and receptor recognition of human TRAF2. *Nature* **398:** 533-538.

Pierce, N.A., Spriet, J.A., Desmet, J., and Mayo, S.L. 2000. Conformational splitting: A more powerful criterion for dead- end elimination. *J. Comput. Chem.* **21:** 999-1009.

Pullen, S.S., Dang, T.T., Crute, J.J., and Kehry, M.R. 1999. CD40 signaling through tumor necrosis factor receptor-associated factors (TRAFs). Binding site specificity and activation of downstream pathways by distinct TRAFs. *J Biol Chem* **274:** 14246-14254.

Shifman, J.M., and Mayo, S.L. 2003. Exploring the origins of binding specificity through the computational redesign of calmodulin. *Proc Natl Acad Sci U S A* **100:** 13274-13279.

Sood, V.D., and Baker, D. 2006. Recapitulation and design of protein binding peptide structures and sequences. *J Mol Biol* **357:** 917-927.

Street, A.G., and Mayo, S.L. 1998. Pairwise calculation of protein solvent-accessible

surface areas. *Fold Des* **3:** 253-258.

Su, A., and Mayo, S.L. 1997. Coupling backbone flexibility and amino acid sequence selection in protein design. *Protein Sci* **6:** 1701-1707.

Wollacott, A.M., and Desjarlais, J.R. 2001. Virtual interaction profiles of proteins. *J Mol Biol* **313:** 317-342.

Wu, H., and Arron, J.R. 2003. TRAF6, a molecular bridge spanning adaptive immunity, innate immunity and osteoimmunology. *Bioessays* **25:** 1096-1105.

Ye, H., Arron, J.R., Lamothe, B., Cirilli, M., Kobayashi, T., Shevde, N.K., Segal, D., Dzivenu, O.K., Vologodskaia, M., Yim, M., et al. 2002. Distinct molecular mechanism for initiating TRAF6 signalling. *Nature* **418:** 443-447.

Ye, H., Cirilli, M., and Wu, H. 2002. The use of construct variation and diffraction data analysis in the crystallization of the TRAF domain of human tumor necrosis factor receptor associated factor 6. *Acta Crystallogr D Biol Crystallogr* **58:** 1886-1888.

Ye, H., and Wu, H. 2000. Thermodynamic characterization of the interaction between TRAF2 and tumor necrosis factor receptor peptides by isothermal titration calorimetry. *Proc Natl Acad Sci U S A* **97:** 8961-8966.

Table 2. Parameters used to fit experimental binding assay data with computational energies, T test scores and the number of false positives for all the parameters are shown. The parameters in red were used to design peptides used in fluorescence anisotropy assays. The name of the designed peptides are written as a superscript. The underlined parameter set is the default parameter set for ORBIT, which is used to design isolated proteins. polar SA: polar surface area penalty; polar H: polar hydrogen penalty; Total E: total energy; Intxn E: interaction energy.

Parameters			T test	# false positives
Dielectric constant	Solvation penalty	Energy score		
40r	polar SA	Total E^{D3}	0.0012	23
4r	polar SA	Total E	0.0024	14
4r	polar H	Total E^{D2}	0.0054	30
40r	polar H	Total E^{D1}	0.0073	22
4r	polar SA	Intxn E	0.0175	25
40r	polar SA	Intxn E	0.0228	29
40r	polar H	Intxn E^{D4}	0.0394	23
4r	polar H	Intxn E	0.1905	47

Table 3. Sequences and relative affinities of designed peptides.

	P_{-5}	P_{-4}	P_{-3}	P_{-2}	P_{-1}	P_0	P_1	P_2	P_3	P_4	P_5	P_6	Relative affinity*
CD40 peptide	**K**	**Q**	**E**	**P**	**Q**	**E**	**I**	**D**	**F**	**P**	**D**	**D**	**1.00**
D1	**-**	**R**	**-**	**-**	**E**	**-**	**-**	**-**	**-**				**1.72**
D2	**-**	**R**	**-**	**-**	**E**	**-**	**H**	**-**	**-**				**0.50**
D3	**-**	**I**	**H**	**-**	**E**	**-**	**M**	**-**	**-**	**-**	**-**	**-**	**0.43**
D4	**-**	**-**	**Q**	**-**	**E**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**0.83**
D5	**-**	**R**	**-**	**-**	**E**	**-**	**L**	**-**	**-**				**0.36**
D6	**-**	**R**	**-**	**-**	**E**	**-**	**W**	**-**	**-**				**0.89**
D7	**-**	**I**	**H**	**-**	**E**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**0.33**
D8	**-**	**R**	**-**	**Y**	**E**	**-**	**-**	**-**	**-**	**-**	**-**	**-**	**0.16**
mCD40-P/Q	**R**	**-**	**D**	**Q**	**-**	**-**	**M**	**E**	**D**				**~0**
CD40-P/G,E/A	**-**	**-**	**-**	**G**	**-**	**A**	**-**	**N**	**-**	**-**	**-**	**-**	**~0**

Residue position nomenclature derived from Ye et al. (2002a).

* Relative affinity is defined as the K_d of the peptide divided by the K_d of the CD40 peptide that has the same length, either 9 or 12 residues, as the peptide. All affinities were obtained by fluorescence anisotropy assay.

Table 4. Comparison of affinities obtained by fluorescence anisotropy assay and ITC. Previously measured value from Ye et al. (2002a) is also shown for CD40 peptide.

	K_d from fluorescence anisotropy (μM)	K_d from ITC (μM)	K_d from ITC (Ye et al. 2002a) (μM)
9 residue CD40 peptide	7.4 ± 1.1	8.8 ± 1.0	84
9 residue D1	4.3 ± 0.6	7.8 ±3.2	

Figure 7. Ribbon and molecular surface representation of TRAF-C domain of TRAF6 bound to CD40 peptide (PDB ID: 1LB6). These structural figures were generated using VMD (Humphrey et al. 1996).

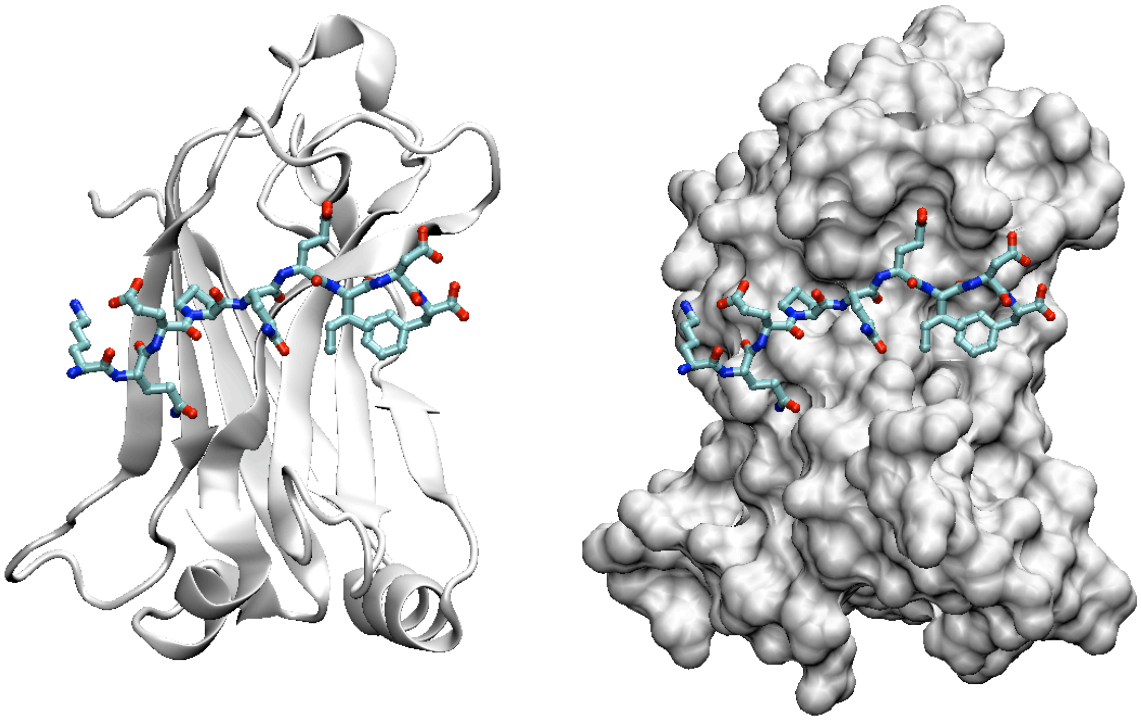
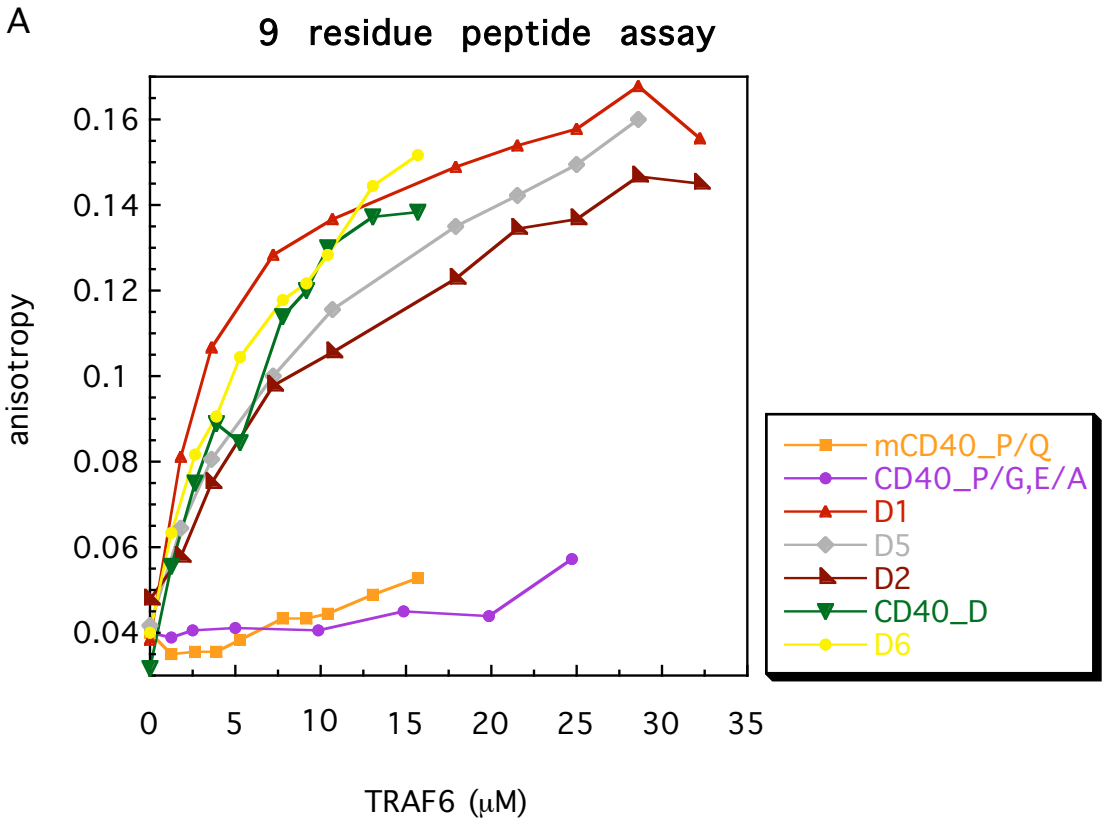
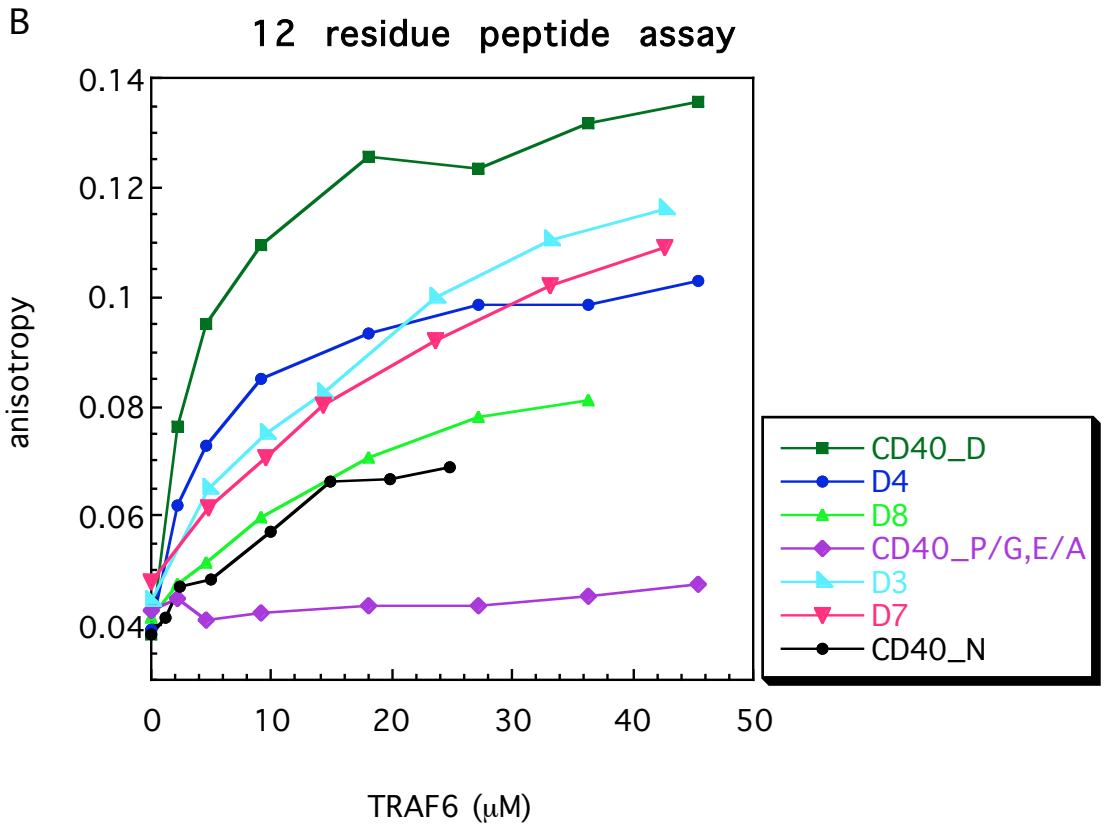


Figure 8. Anisotropy data of all the peptides. (A) Data for all nine residue peptides. (B) Data for peptides with additional three residues, PDD, attached to the carboxy terminus to increase the solubility of the peptide. CD40_N: wild-type CD40 peptide; CD40_D: wild-type CD40 peptide with Asn to Asp mutation at P_2 position.


A
9 residue peptide assay
anisotropy
0.16
0.14
0.12
0.1
0.08
0.06
0.04
0
5
10
15
20
25
30
35
TRAF6 (μM)
mCD40_P/Q
CD40_P/G,E/A
D1
D5
D2
CD40_D
D6


B
12 residue peptide assay
anisotropy
0.14
0.12
0.1
0.08
0.06
0.04
0
10
20
30
40
50
TRAF6 (μM)
CD40_D
D4
D8
CD40_P/G,E/A
D3
D7
CD40_N

Chapter 4

Computational design and biochemical characterization of non-specific lipid transfer protein variants for biosensor applications

The text of this chapter is adapted from the manuscript to be submitted by **Choi EJ**, Mao J and Mayo SL (in preparation).

Abstract

Lipid transfer proteins (LTPs) are a family of proteins that bind and transfer lipids. Utilizing a maize LTP, we have successfully engineered a fluorescent reagentless biosensor for the natural ligand of LTPs; this was achieved by using computational protein design to remove disulfide bridges and attaching a thio-reactive fluorophore. Our LTP variants show affinity to palmitate on the order of the wild-type LTP. These molecules can be used to design biosensors with specificities for various hydrophobic ligands. They can also be used as drug carriers.

Introduction

Biosensors are biological molecules that recognize specific ligands and relay the message to a physical signal, which can be easily detected and measured. Ideally, a biosensor is reagentless and does not change composition as a consequence of the measurement. The specific interaction of a biosensor with its ligand and its reagentless characteristic allow biosensors to be the basis for a wide range of assay methods. Recently, bacterial periplasmic binding proteins (PBPs) with different ligand specificities were tagged with a fluorophore to produce biosensors for their natural ligands, which include sugars, amino acids, dipeptides and ions (de Lorimier et al. 2002). Furthermore, using protein design techniques, these proteins were engineered into biosensors with specificity for novel ligands such as Zn(II), TNT, L-lactate, serotonin, PMPA and dihydroxyacetone phosphate (Dwyer and Hellinga 2004).

The natural ligands of PBPs are hydrophilic and relatively small. The binding site residues are also hydrophilic, thus making it difficult to design in large hydrophobic ligands. Our aim is to expand the chemical scope of target ligands by developing a protein platform capable of specifically binding large and/or hydrophobic ligands. For this purpose, we selected nonspecific lipid transfer protein (LTP) from maize (mLTP). Plant LTPs are a family of proteins known for their ability to bind and transfer lipids. Their biological function is still unclear, but they may be involved in the formation of the cuticle layer, in somatic embryogenesis and in plant responses to pathogenic stress. The two subfamilies of LTPs (LTP1 and LTP2) share eight conserved cysteines that form four disulfide bridges and have nonpolar binding pockets (Capocchi et al. 2005). The larger LTP1 binds various phospholipids, fatty acids, and glycolipids, while the smaller but

more flexible LTP2 can bind to bulkier sterol molecules as well (Samuel et al. 2002; Cheng et al. 2004).

In this paper, we used mLTP to design a biosensor for its natural lipid ligands. mLTP is a 93-residue basic protein comprised of four α-helices (helix 1-4). It is of the LTP1 subfamily and has a tunnel-like hydrophobic cavity that has one wide and one narrow opening on either end. The wide opening is predicted to be the main entrance or exit site for the ligands (Shin et al. 1995; Gomar et al. 1996; Han et al. 2001). The apo-mLTP does not have a tightly packed hydrophobic core because of this hydrophobic cavity, and it is thought that the disulfide bonds are important in maintaining the tertiary structure of the LTPs in the absence of the ligand. Our strategy is to design out the disulfide bonds using ORBIT (Optimization of Rotamers By Iterative Techniques) to allow flexibility in the mLTP molecule. Acrylodan, a thiol-reactive fluorescent probe, will be conjugated to one of the cysteines that are not designed out. We predict that when the ligand is added to the flexible mLTP variant, it will bind to the protein, rigidifying it and changing its conformation, causing a fluorescence change. The small size and the availability of a high-resolution crystal structure of mLTP with its ligand bound makes it a good candidate for computational protein design (Shin et al. 1995).

Recently, there has been an interest in using proteins as carriers for drugs due to their high affinity and selectivity for their targets (De Wolf and Brett 2000). The proteins would not only protect the unstable or harmful molecules from oxidation and degradation, but they would also aid in solubilization and ensure a controlled release of the agents. In a study to determine the suitability of LTPs as drug carriers, various molecules having cosmetic or pharmacological interest were tested for binding to wheat

LTP (wLTP). wLTP was found to bind to a variety of molecules with affinities low enough to allow a slow release (Pato et al. 2001). Our acrylodan-conjugated mLTP variants could be used in a reliable, sensitive, high throughput screening method for binding drug compounds. Also, LTP variants designed to bind a specific ligand could be used as drug carriers.

Materials and methods

Computational protein design

The crystal structure of mLTP with palmitate (PDB ID 1MZM) was briefly energy minimized and its residues were classified as surface, boundary, or core based on solvent accessibility (Dahiyat and Mayo 1997a). Each of the four disulfide bridges was individually reduced by deletion of the S-S bond and addition of hydrogens. The corresponding structures were used in designs for the respective disulfide bridge. The ORBIT protein design program uses an energy function based on the DREIDING force field, which includes a Lennard-Jones 12-6 potential with all van der Waals radii scaled by 0.9, hydrogen bonding and electrostatic terms and a solvation potential (Mayo et al. 1990; Dahiyat et al. 1997; Dahiyat and Mayo 1997b). Both solvent-accessible surface area-based solvation (Street and Mayo 1998) and the implicit solvation model developed by Lazaridis and Karplus (Lazaridis and Karplus 1999) were tried, but better results were obtained with the Lazaridis-Karplus model, and it was used in all final designs. From the fixed composition study on the Engrailed homeodomain protein, polar burial energy was scaled by 0.6 and rotamer probability was scaled by 0.3 (Oscar Alvizo, unpublished data). Parameters from the Charmm19 force field were used. An algorithm based on the

dead-end elimination theorem (DEE) was used to obtain the global minimum energy amino acid sequence and conformation (GMEC) (Pierce et al. 2000). For each design, non-proline, non-glycine residues within 4 Å of the two reduced cysteines were included as the 1st shell of residues and their amino acid identities and conformations were optimized. Residues within 4 Å of the 1st shell were considered the 2nd shell and they were floated; that is, their conformations were allowed to change, but their amino acid identities were held fixed. Finally, the remaining residues were treated as fixed. Based on the results of the initial design calculations, further restricted designs were carried out where only designed positions making stabilizing interactions were included.

Designed protein expression and purification

The *E. coli* expression optimized gene encoding the mLTP amino acid sequence was synthesized and ligated into the pET15b vector (Stratagene) by Blue Heron Biotechnology (www.blueheronbio.com). The pET15b vector includes an N-terminal His-tag. Inverse PCR mutagenesis was used to construct five variants, C4H/C52A/N55E, C4Q/C52A/N55S, C14A/C29S, C30A/C75A and C50A/C89E. The proteins were expressed in BL21(DE3) Gold cells (Stratagene) at 37°C after induction with IPTG (isopropyl-beta-D-thiogalactopyranoside). Cells were resuspended in lysis buffer (50 mM sodium phosphate, 300 mM sodium chloride, 10 mM imidazole, pH 8.0) and lysed by passing through the Emulsiflex at 15,000 psi, and the soluble fraction was isolated by centrifugation at 20,000*g* for 30 minutes. The soluble fraction of the cell lysate was loaded onto a Ni-NTA column and eluted with elution buffer (lysis buffer with 400 mM imidazole). The elutions were further purified by gel filtration chromatography with

phosphate buffer (50 mM sodium phosphate, 150 mM sodium chloride, pH 7.5). The molecular weight and the oxidation state of the purified proteins were verified by SDS-Page and MALDI-TOF. Trypsin digest analysis showed that the N-terminal Met was cleaved. The gel filtration profile for all variants looked similar to that of wild-type mLTP, which we verified to be a monomer by analytical ultracentrifugation (data not shown). Protein concentration was determined using the BCA assay (Pierce) with BSA as the standard.

Circular dichroism

Circular dichroism (CD) data were obtained on an Aviv 62A DS spectropolarimeter equipped with a thermoelectric cell-holder. Wavelength scans and thermal denaturation data were obtained from samples containing 50 μM protein. For wavelength scans, data were collected every 1 nm from 190 to 250 nm with averaging time of 5 seconds. For temperature denaturation, data were collected every 2°C from 1°C to 99°C using an equilibration time of 120 seconds and an averaging time of 30 seconds. Since the thermal denaturations were not reversible, we could not fit the data to a two-state transition. The apparent melting temperatures (T_ms) were obtained from the inflection point of the data. For thermal denaturations of protein with palmitate (Sigma Aldrich), 150 μM palmitate, from stock solution, and 50 μM protein was used.

Acrylodan labeling of mLTP variants

Proteins purified by Ni-NTA column were concentrated to 10-20 μM. 6-acryloyl-2-(dimethylamino)naphthalene (acrylodan) was dissolved in acetonitrile and added to the

purified proteins in 10-fold excess concentration. The mixed sample was incubated at 4°C overnight. All solutions containing acrylodan were protected from light. Precipitated acrylodan and protein were removed by centrifugation and filtering through a 0.2 μm nylon membrane Acrodisc syringe filters (Gelman Laboratory), and the soluble fraction was concentrated. Unreacted acrylodan and protein impurities were removed by gel filtration chromatography with phosphate buffer (50 mM sodium phosphate, 150 mM sodium chloride, pH 7.5). The eluents were simultaneously monitored at 280 nm for protein and 391 nm for acrylodan. Fractions were collected when absorbance was observed at both 280 nm and 391 nm. The conjugation reaction looked to be complete, as both absorbances overlapped for the only protein peak observed. Purified proteins were verified by SDS-Page to be of sufficient purity, and MALDI-TOF analysis showed that they correspond to the oxidized form of the proteins with acrylodan conjugated. Trypsin digest analysis of acrylodan conjugated proteins showed that the acrylodan was attached at the expected cysteine.

Fluorescence emission scan and ligand binding assay

Ligand binding was monitored by measuring the fluorescence emission of protein-acrylodan conjugates with the addition of palmitate. Fluorescence measurements were performed on a Photon Technology International Fluorometer equipped with a stirrer at room temperature. Excitation wavelength was set to the excitation maximum for each variant and measurements were taken at the wavelength of the emission maximum for 30 – 60 sec at 2 nm intervals and 0.5 second integration time. Stock solution of palmitate dissolved in ethanol was titrated into 2 ml of 500 – 1000 nM protein-acrylodan

conjugate. The total amount of palmitate in ethanol was never allowed to exceed 1% of the sample volume. The dissociation constants (K_d) were determined by fitting the decrease in fluorescence to an equation for a non-cooperative binding model (Dubreil et al. 1997).

Results and discussion

mLTP designs

mLTP contains four disulfide bridges: C4-C52, C14-C29, C30-C75, and C50-C89. We used the ORBIT protein design software to design variants without each disulfide bridge. Calculations were evaluated and five variants, C4H/C52A/N55E, C4Q/C52A/N55S, C14A/C29S, C30A/C75A and C50A/C89E, were chosen (Figure 9). The disulfide bridge C4-C52 anchors two helices to each other, with C52 more buried than C4. In the final designs C4H/C52A/N55E and C4Q/C52A/N55S, the disulfide bridge is lost, but residue 4 and 55 form an interhelical hydrogen bond, H4-E55 and Q4-S55, with heavy atom distances of 2.8 Å. C14A/C29S gains a hydrogen bond between S29 and S26. The C30-C75 disulfide bond is surrounded by nonpolar residues and both residues are mutated to alanine in C30A/C75A mutant. C50-C89 anchors the C-terminal loop to helix 3. The mutation C89E introduces hydrogen bonds with R47, S90 and K54.

Experimental validation

The circular dichroism wavelength scans of mLTP and the variants (Figure 10) show that three of the five variants (C4H/C52A/N55E, C4Q/C52A/N55S and C50A/C89E) have similar spectra to the wild type, with minima at 208 nm and 222 nm,

characteristic of helical proteins. C14A/C29S and C30A/C75A, on the other hand, are not folded properly, with wavelength spectra resembling those of ns-LTP with scrambled disulfide bonds (Lin et al. 2004). In the mLTP crystal structure, C14-C29 and C30-C75 both link the end of helix 1 (residues 4-18) and helix 4 (residues 65-75), respectively, to helix 2 (residues 27-39) near the narrow opening of the protein. The narrow opening was suggested to have only limited flexibility by the low B-factor of the residues around it (Shin et al. 1995). These observations plus our result that C14A/C29S and C30A/C75A variants are unfolded imply the importance of these two disulfide bridges in anchoring the molecule together. On the other hand, the two disulfides that were designed out without a major change in the protein's fold, C4-C52 and C50-C89, are connecting the N-terminal and C-terminal region of the protein to helix 3 (residues 43-58) near the wide opening of the protein. This opening is very flexible and is speculated to be a ligand entry and exit site; thus, eliminating C4-C52 and C50-C89 disulfide bonds should be tolerated better.

We determined the thermal stability of the variants in the absence and presence of palmitate and compared it to wild-type mLTP (Figure 11). The removal of the disulfide bridge C4-C52 significantly destabilizes the protein relative to the wild type, lowering the T_m by as much as 28°C (Table 5). The disruption of C50-C89 leads to a T_m only 10°C lower than that of the wild type. The variants are still able to bind palmitate, as evidenced by the increase in the apparent melting temperatures when palmitate is present, as is observed for the wild-type protein. The C4H/C52A/N55E and C4Q/C52A/N55S mutants' thermodynamic behaviors were similar, since each variant supplied one potential hydrogen bond to replace the disulfide bond. There was a much larger gain in stability

upon binding palmitate for the two mutants above (~20°C) than was observed for the wild-type protein (8°C), suggesting that these mutants might undergo a large conformational change upon ligand binding. Therefore, they were predicted to be good candidates for biosensor design. The difference in T_ms between the palmitate-bound mutants and wild type was 18°C, 10°C lower than the 28°C difference observed between the unbound proteins. The deletion of disulfide bonds in the mutants decreased the stability of the apo-protein far more than the complexed protein, confirming the idea that it is highly likely these mutants show large conformational differences between the apo and the complexed form, an important criterion in our biosensor design.

It is interesting that C50A/C89E is 15°C more stable than the C4-C52 variants. The disulfide C50-C89 anchors the long C-terminal loop to helix 3. Disruption of this disulfide only lowered the T_m by 10°C. This could be due to the three introduced hydrogen bonds that were a direct result of the C89E mutation. The stability gained by palmitate binding only raises the T_m by 6°C, similar to the 8°C observed for wild-type mLTP. For wild-type mLTP, the crystal and solution structures show little change in conformation upon ligand binding (Shin et al. 1995; Gomar et al. 1996), and we suspect this to be the case for C50A/C89E.

The C4-C52 mutated variants look promising as the basis for the development of a reagentless biosensor. Fluorescent sensors are extremely sensitive to their environment. By conjugating a sensor molecule to the site of conformational change, these mutants could be used as a reporter for ligand binding.

Protein-acrylodan conjugates

We chose two of the variants, C4H/C52A/N55E and C50A/C89E, and mutated one of the original cysteine residues in each variant back. This did not disturb any of the interactions we designed into the former disulfide bridge because all designed interactions were not between the two residues constituting the disulfide bond. This gave us four new variants: C52A, C4H/N55E, C50A, and C89E. We conjugated acrylodan (Ac), an environment sensitive thiol-reactive fluorophore (Prendergast et al. 1983), to the resulting free cysteine in each protein. Trypsin digest and tandem mass spectrometry of C52A/C4-Ac confirmed that the acrylodan was conjugated to the correct cysteine. Analytical ultracentrifugation of C52A/C4-Ac showed that it is a monomer (data not shown), and all gel filtration profiles for all acrylodan-conjugated variants showed a single dominant peak similar to C52A/C4-Ac, implying that all variants are monomers.

We obtained circular dichroism wavelength scans of the acrylodan-conjugated variants to ensure they were properly folded (Figure 12). While all four conjugates appeared folded with characteristic helical protein minima near 208 nm and 222 nm, the C52A/C4-Ac spectra was closest to the wild-type mLTP spectra. Thermal denaturation experiments (Table 5) show that when acrylodan is conjugated to C4 or C52, the apo-protein stability is increased while the stability of the ligand bound protein is decreased, compared to the C4H/C52A/N55E variant. For the mutants with the acrylodan conjugated to C50 or C89, we see a different phenomenon in which the T_m of both the apo and ligand bound form of the protein decrease or stay the same. This and additional data mentioned below points to the possibility that acrylodan conjugated to C4 or C52 may fit into the ligand-binding pocket of mLTP, stabilizing the apo-protein. When ligand is added, the

acrylodan is displaced by the ligand; thus, the acrylodan will not induce any stabilizing effect on the ligand-binding form of the protein. As mentioned below, the displacement of the acrylodan by the ligand could also cause the fluorescence change observed with ligand titration. This same mechanism has been suggested for the fluorescence change that occurs upon ligand binding to acrylodan-conjugated intestinal fatty acid binding protein (Richieri et al. 1992).

Ligand binding assays

We performed titrations of the protein-acrylodan conjugates with palmitate to test the ability of the engineered mLTPs to act as biosensors. As predicted, of the four protein-acrylodan conjugates, C52A/C4-Ac and C4H/N55E/C52-Ac show the most marked difference in signal when palmitate is added (Figure 13A). For each titration, the average of the fluorescence intensity time-based measurement at the emission maxima (480nm) was used to fit a non-cooperative binding equation (Dubreil et al. 1997). We determined the K_d of palmitate to be 0.6 μM for C52A/C4-Ac (Figure 13B) and 4 μM for C4H/N55E/C52-Ac (Figure 13C). These values were close to the 3 μM K_d of palmitate to wild-type LTP we measured using tyrosine fluorescence (Douliez et al. 2000).

We successfully engineered mLTP into a fluorescent reagentless biosensor for nonpolar ligands. This was achieved using computational protein design to remove one of the disulfide bridges and then attaching a thio-reactive fluorophore to the free cysteine. We believe the observed change in acrylodan signal is a measure of the local conformational change the protein variants undergo upon ligand binding, which causes the displacement of the fluorophore from the hydrophobic binding pocket. The removal

of the C4-C52 disulfide bridge provides the N-terminal helix more flexibility and allows acrylodan to insert into the binding pocket. Upon ligand binding, however, acrylodan is displaced and shifts from an ordered nonpolar environment to a disordered polar environment. The observed decrease in fluorescence emission and the red shift of the emission maxima as palmitate is added is consistent with this hypothesis.

References

Capocchi, A., Fontanini, D., Muccilli, V., Cunsolo, V., Saviozzi, F., Saletti, R., Lorenzi, R., Foti, S., and Galleschi, L. 2005. NsLTP1 and NsLTP2 isoforms in soft wheat (Triticum aestivum Cv. Centauro) and farro (Triticum dicoccon Schrank) bran. *J Agric Food Chem* **53:** 7976-7984.

Cheng, C.S., Samuel, D., Liu, Y.J., Shyu, J.C., Lai, S.M., Lin, K.F., and Lyu, P.C. 2004. Binding mechanism of nonspecific lipid transfer proteins and their role in plant defense. *Biochemistry* **43:** 13628-13636.

Dahiyat, B.I., Gordon, D.B., and Mayo, S.L. 1997. Automated design of the surface positions of protein helices. *Protein Sci* **6:** 1333-1337.

Dahiyat, B.I., and Mayo, S.L. 1997a. *De novo* protein design: fully automated sequence selection. *Science* **278:** 82-87.

Dahiyat, B.I., and Mayo, S.L. 1997b. Probing the role of packing specificity in protein design. *PNAS* **94:** 10172-10177.

de Lorimier, R.M., Smith, J.J., Dwyer, M.A., Looger, L.L., Sali, K.M., Paavola, C.D., Rizk, S.S., Sadigov, S., Conrad, D.W., Loew, L., et al. 2002. Construction of a fluorescent biosensor family. *Protein Sci* **11:** 2655-2675.

De Wolf, F.A., and Brett, G.M. 2000. Ligand-binding proteins: their potential for application in systems for controlled delivery and uptake of ligands. *Pharmacol Rev* **52:** 207-236.

Dubreil, L., Compoint, J.P., and Marion, D. 1997. Interaction of puroindolines with wheat flour polar lipids determines their foaming properties. *J Agr Food Chem* **45:** 108-116.

Dwyer, M.A., and Hellinga, H.W. 2004. Periplasmic binding proteins: a versatile superfamily for protein engineering. *Curr Opin Struct Biol* **14:** 495-504.

Gomar, J., Petit, M.C., Sodano, P., Sy, D., Marion, D., Kader, J.C., Vovelle, F., and Ptak, M. 1996. Solution structure and lipid binding of a nonspecific lipid transfer protein extracted from maize seeds. *Protein Sci* **5:** 565-577.

Han, G.W., Lee, J.Y., Song, H.K., Chang, C., Min, K., Moon, J., Shin, D.H., Kopka, M.L., Sawaya, M.R., and Yuan, H.S. 2001. Structural basis of non-specific lipid binding in maize lipid-transfer protein complexes revealed by high-resolution X-ray crystallography. *Journal of Molecular Biology* **308:** 263-278.

Lazaridis, T., and Karplus, M. 1999. Discrimination of the native from misfolded protein models with an energy function including implicit solvation. *Journal of Molecular Biology* **288:** 477-487.

Lin, C.-H., Li, L., Lyu, P.-C., and Chang, J.-Y. 2004. Distinct unfolding and refolding pathways of lipid transfer proteins LTP1 and LTP2. *The Protein Journal* **23:** 553-566.

Mayo, S.L., Olafson, B.D., and Goddard, W.A. 1990. Dreiding - a generic force-field for molecular simulations. *Journal of Physical Chemistry* **94:** 8897-8909.

Pato, C., Le Borgne, M., Le Baut, G., Le Pape, P., Marion, D., and Douliez, J.-P. 2001. Potential application of plant lipid transfer proteins for drug delivery. *Biochemical Pharmacology* **62:** 555-560.

Pierce, N.A., Spriet, J.A., Desmet, J., and Mayo, S.L. 2000. Conformational splitting: A more powerful criterion for dead- end elimination. *J. Comput. Chem.* **21:** 999-1009.

Prendergast, F.G., Meyer, M., Carlson, G.L., Iida, S., and Potter, J.D. 1983. Synthesis, spectral properties, and use of 6-acryloyl-2- dimethylaminonaphthalene (Acrylodan).

A thiol-selective, polarity- sensitive fluorescent probe. *J. Biol. Chem.* **258:** 7541-7544.

Richieri, G.V., Ogata, R.T., and Kleinfeld, A.M. 1992. A fluorescently labeled intestinal fatty acid binding protein. Interactions with fatty acids and its use in monitoring free fatty acids. *J Biol Chem* **267:** 23495-23501.

Samuel, D., Liu, Y.-J., Cheng, C.-S., and Lyu, P.-C. 2002. Solution structure of plant nonspecific lipid transfer protein-2 from rice (Oryza sativa). *J. Biol. Chem.* **277:** 35267-35273.

Shin, D.H., Lee, J.Y., Hwang, K.Y., Kyu Kim, K., and Suh, S.W. 1995. High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. *Structure* **3:** 189-199.

Street, A.G., and Mayo, S.L. 1998. Pairwise calculation of protein solvent-accessible surface areas. *Folding & Design* **3:** 253-258.

Table 5. Apparent T_ms of mLTP and designed variants.

	Apparent T_m		ΔT_m
	Protein alone	**Protein + Palmitate**	
mLTP	84	92	8
C4H/C52A/N55E	56	76	20
C4Q/C52A/N55S	56	74	18
C50A/C89E	74	80	6
C4H/ N55E/C52-Ac	62	71	9
C52A/C4-Ac	59	66	7
C50A/C89-Ac	64	73	9
C89E/ C50-Ac	71	80	9

Figure 9. Ribbon diagram of mLTP and the designed variants of each disulfide. The palmitate-bound mLTP (cyan) is superimposed on the unbound protein (green). Palmitate is shown in spheres with carbon in magenta and oxygen in red. Disulfides are in orange. In panels, mutated residues and the residues they form hydrogen bonds with are shown in stick with CPK-inspired colors, and the modeled hydrogen bonds are shown with yellow dashed lines, with measured heavy atom distances between 2.8 and 3.0 Å.

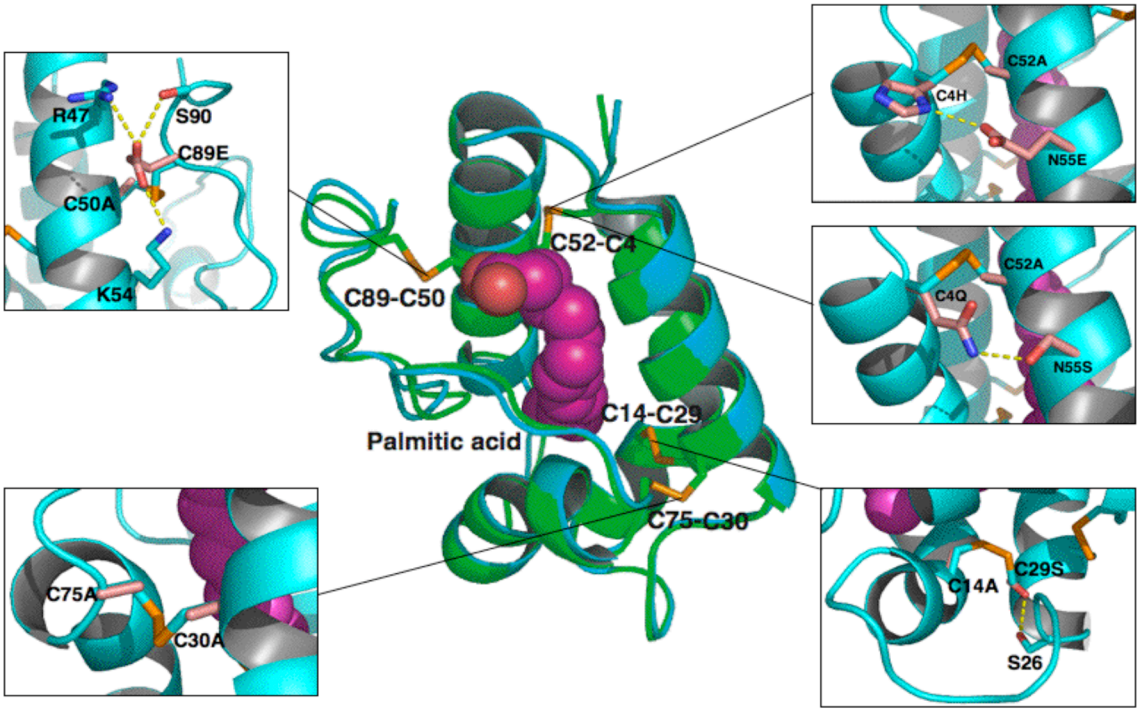

R47
S90
C89E
C50A
K54
C52A
C4H
N55E
C52-C4
C89-C50
C52A
C4Q
N55S
C14-C29
Palmitic acid
C75-C30
C75A
C30A
C14A
C29S
S26

Figure 10. Wavelength scans of mLTP and designed variants. Variants C4H/C52A/N55E and C4Q/C52A/N55S and C50A/C89E are folded similar to wild-type mLTP, with minima at 208nm and 222nm, but C14A/C29S and C30A/C75A are misfolded.

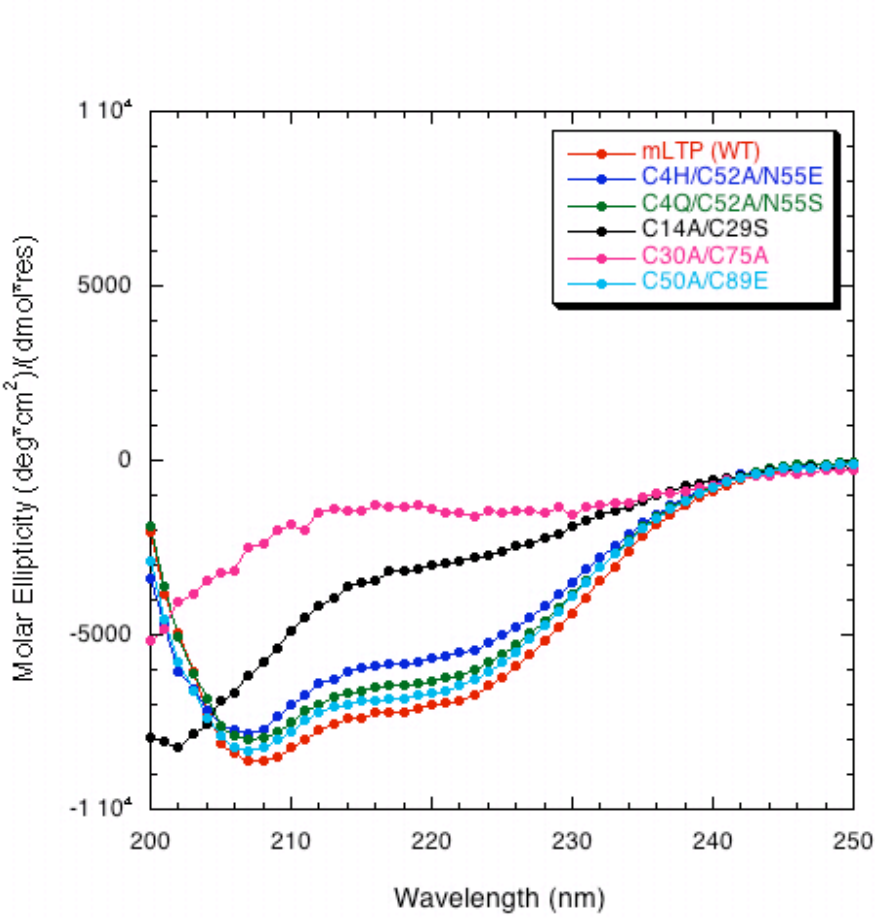

1 10⁴
5000
0
-5000
-1 10⁴
200
210
220
230
240
250
Molar Ellipticity (deg*cm²)/(dmol*res)
Wavelength (nm)
mLTP (WT)
C4H/C52A/N55E
C4Q/C52A/N55S
C14A/C29S
C30A/C75A
C50A/C89E

Figure 11. Thermal denaturations of mLTP and designed variants: mLTP (red), C4H/C52A/N55E (blue), C4Q/C52A/N55S (green), and C50A/C89E (cyan). Solid lines are protein alone, dashed lines are protein with palmitate added. Removal of disulfide bridges significantly destabilized the protein, but the variants still bound palmitate.

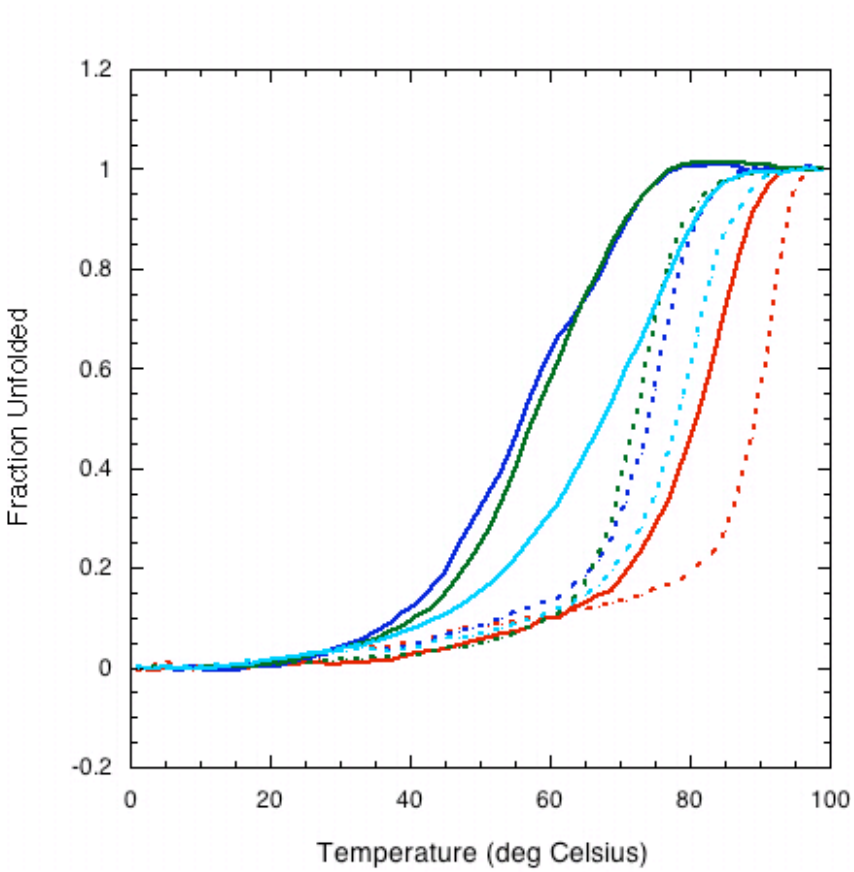

Fraction Unfolded
Temperature (deg Celsius)
1.2
1
0.8
0.6
0.4
0.2
0
-0.2
0
20
40
60
80
100

Figure 12. Circular dichroism wavelength scans of the four protein-acrylodan conjugates. Each conjugate shows the characteristic minima near 208nm and 222nm for helical proteins. C52A/4C-Ac is most like wild-type mLTP.

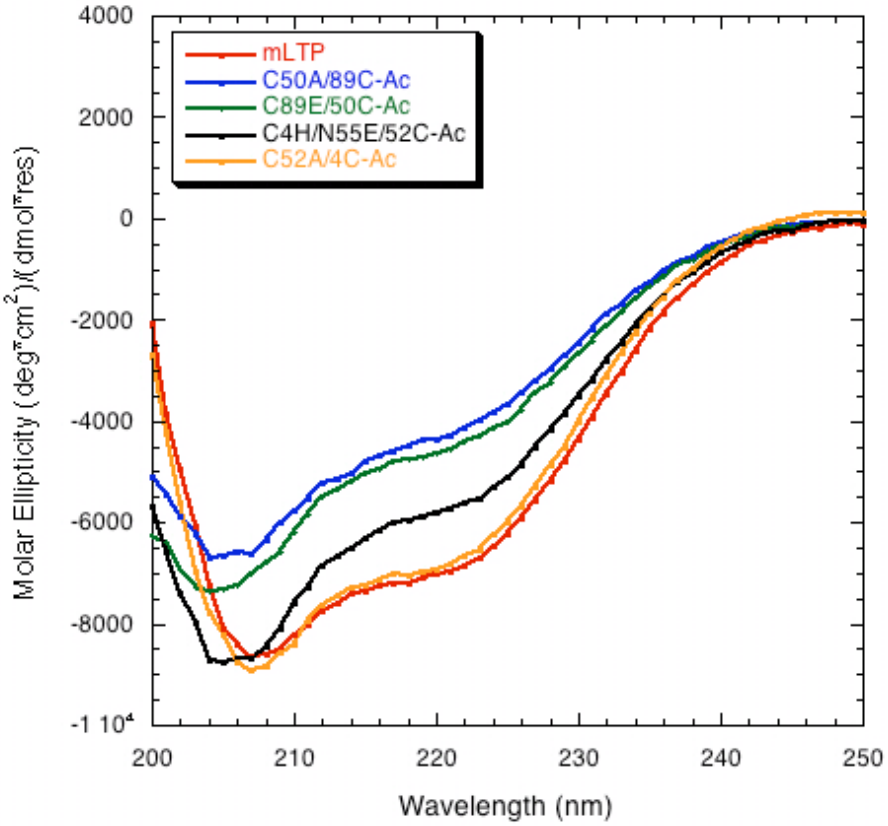
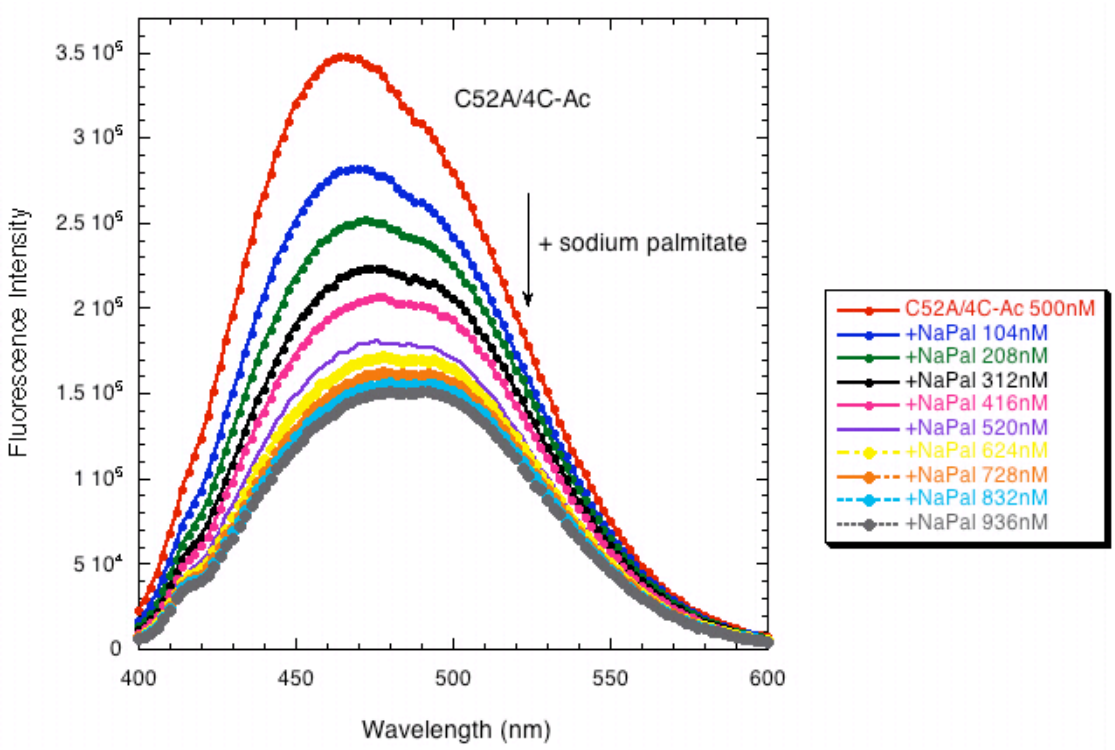
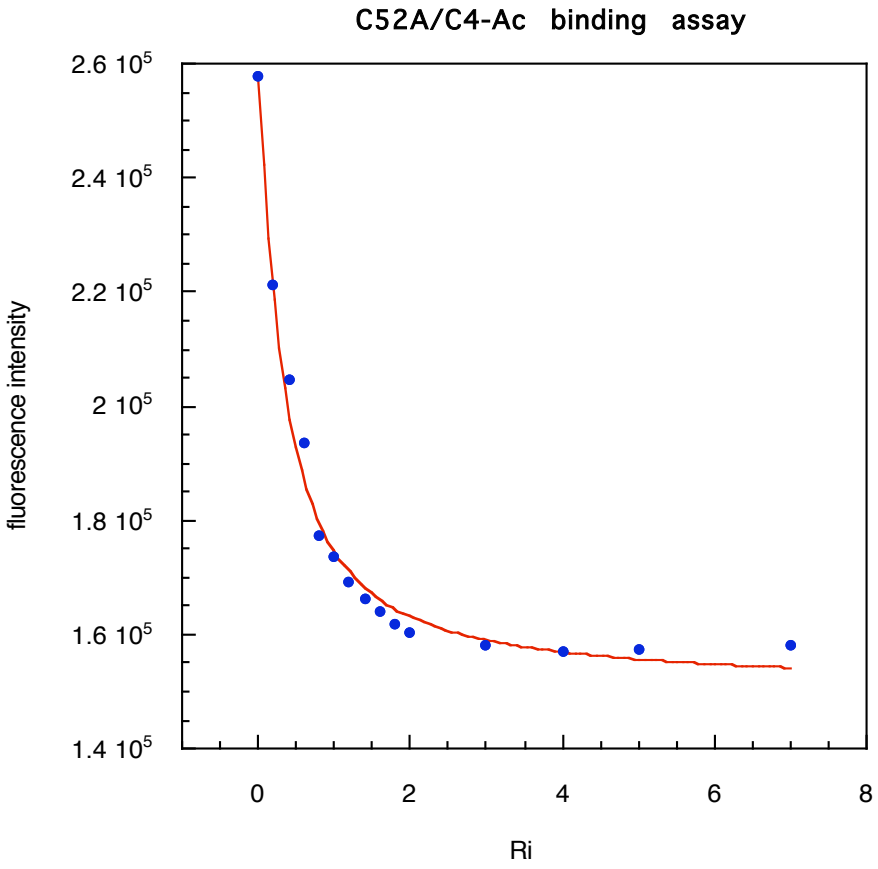

mLTP
C50A/89C-Ac
C89E/50C-Ac
C4H/N55E/52C-Ac
C52A/4C-Ac
Molar Ellipticity (deg*cm²)/(dmol*res)
Wavelength (nm)
4000
2000
0
-2000
-4000
-6000
-8000
-1 10⁴
200
210
220
230
240
250

Figure 13. Titration of C52A/C4-Ac with palmitate monitored by fluorescence emission. (A) Fluorescence emission scans of C52A/4C-Ac (red) show decreased fluorescence with addition of increasing concentrations of sodium palmitate. Only a subset of experimental data is shown. Excitation wavelength is 363nm. Fluorescence monitored at 466nm was used to determine the K_d. (B) C52A/C4-Ac K_d is 0.6 μM and (C) C4H/N55E/C52-Ac K_d is 4 μM.

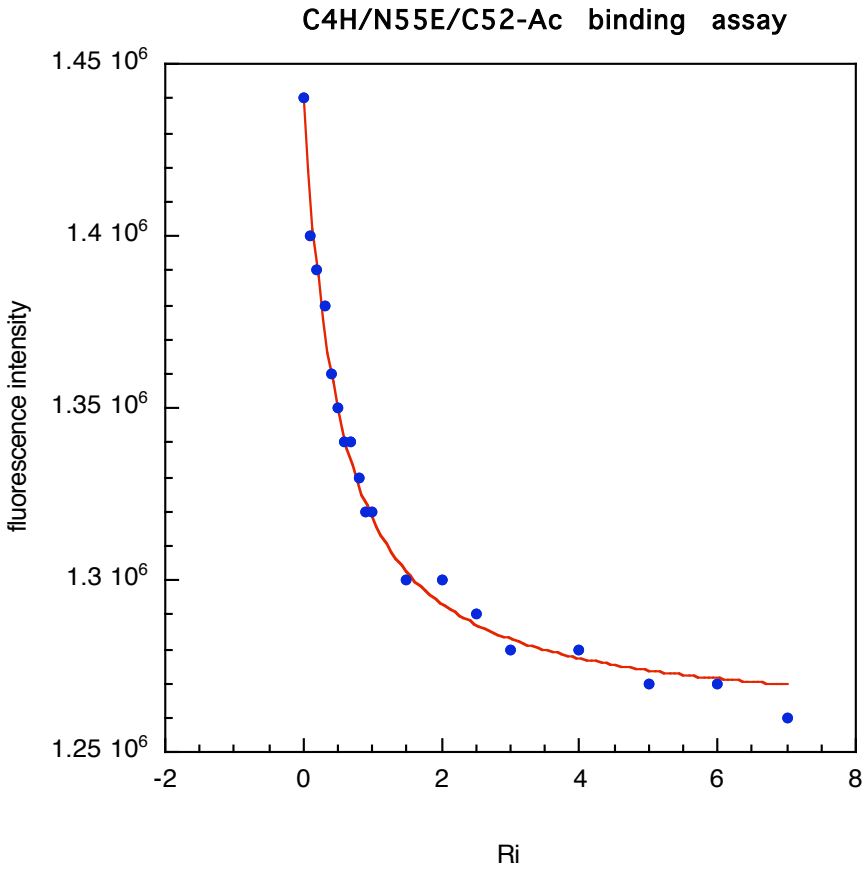
A


C52A/4C-Ac
+ sodium palmitate
Fluorescence Intensity
Wavelength (nm)
3.5 10⁵
3 10⁵
2.5 10⁵
2 10⁵
1.5 10⁵
1 10⁵
5 10⁴
0
400
450
500
550
600
C52A/4C-Ac 500nM
+NaPal 104nM
+NaPal 208nM
+NaPal 312nM
+NaPal 416nM
+NaPal 520nM
+NaPal 624nM
+NaPal 728nM
+NaPal 832nM
+NaPal 936nM

B


C52A/C4-Ac binding assay
fluorescence intensity
2.6 10^5
2.4 10^5
2.2 10^5
2 10^5
1.8 10^5
1.6 10^5
1.4 10^5
0
2
4
6
8
Ri

C


C4H/N55E/C52-Ac binding assay
fluorescence intensity
1.45 10^6
1.4 10^6
1.35 10^6
1.3 10^6
1.25 10^6
-2
0
2
4
6
8
Ri

Chapter 5

Computational design and biochemical characterization of a solubility enhanced MW7 V_H variant, an antibody variable domain fragment for huntingtin protein

Introduction

Antibodies often play a key role in biological studies and have been widely used in the treatment of disease. By binding to specific antigens, antibodies can block interactions, tag target molecules, lock proteins into specific conformation, and even act as catalysts (Stocks 2004). However, due to their large size and the fact that antibodies naturally function in the extracellular space, their utility as therapeutics for intracellular targets has been limited. Recently, however, two major developments have allowed us to overcome these limitations. One is the application of recombinant DNA technology to engineer antibodies into smaller fragments that still retain their binding properties. Although all antibody domains serve a purpose, only the variable domain is necessary if binding is all that is desired. Fragments of whole antibodies have been engineered including the antigen binding fragment (Fab), the variable region fragment (Fv), single chain Fvs (scFv), single domains (V_H or V_L), and single loops. These smaller fragments also have the benefit of being able to be expressed in functional form in *E. coli*, making simple and large-scale production possible. Another major development is the discovery of antibodies that fold correctly and maintain their fold inside cells without the formation of disulfide bonds. Because the cytoplasm is strongly reductive, disulfide bonds cannot form in this environment. These intracellular antibodies, called intrabodies, are functional inside cells and are usually scFvs, but single antibody domains have also been successfully utilized. Many intrabodies related to human diseases, including AIDS, cancer, and various neurological disorders, have been isolated and studied (Lobato and Rabbitts 2004). The ability of intrabodies to modify the interior of cells provides a new

approach and affords them tremendous potential for use as therapeutics and biological reagents.

Several intrabodies have been engineered that bind to exon 1 of huntingtin protein (htt), the mutant form of which is believed to cause neural toxicity in the striatum and cortex and result in Huntington's disease (Lecerf et al. 2001; Khoshnan et al. 2002; Colby et al. 2004). One such intrabody is a scFv of the MW7 antibody, which recognizes the poly proline (polyP) domain of htt. When co-expressed with htt in 293 cells, MW7 scFv interferes with htt aggregation and decreases cell toxicity (Khoshnan et al. 2002). A similar response is also observed when only the V_H of MW7 is expressed (P.H. Patterson, personal communication).

We propose to design a highly soluble and stable mutant of MW7 V_H using ORBIT (Optimization of Rotamers By Iterative Techniques), a computational protein design program developed by Stephen Mayo and co-workers. ORBIT predicts the optimal protein sequence for a given three-dimensional fold. It employs an unbiased, quantitative design algorithm based on a force field that includes van der Waals interactions, hydrogen bonding, solvation and electrostatic forces, and a discrete set of amino acid conformers called rotamers. Since MW7 V_H is a small protein and we will only be considering a few positions for mutation, this problem is amenable to both computational design and extensive experimental manipulation and characterization. MW7 V_H is functional when expressed in mammalian cells, but when expressed in *E. coli,* it aggregates and is not prone to refolding. Our objective is to use computational methods to predict mutations that will maintain binding to htt but will improve the solubility of the protein, allowing *E. coli* or baculovirus to be used for expression. After

expression and purification of the MW7 V_H mutant molecules, we intend to assess functionality, biochemically characterize the active variants, and use X-ray crystallography to solve the structure of selected variants, alone and in complex with the polyP domain plus adjacent amino acid residues of htt.

Currently, only three structures of engineered human V_H exist in the PDB database, and no structures of antibodies that bind to htt are available, although one is to be published (P.J. Bjorkman, personal communication). An X-ray crystal structure of an MW7 V_H mutant, especially with a ligand bound, will help elucidate the mechanism this antibody uses in preventing intracellular aggregation of htt. The crystal structure would also be useful in the computational design of an even more stable molecule. ORBIT employs the backbone of a crystal structure to predict mutations; since the crystal structure of MW7 V_H is not available, we will initially utilize the backbone structure of the V_H domain of an anti-breast tumor antibody, SM3 (PDB ID: 1SM3). Although the sequence identity of the V_H domain of SM3 with MW7 V_H is very high (86%), the backbone structures of the two probably differ. Using the actual structure should give more accurate results. Moreover, the crystal structure of MW7 V_H with ligand bound can be used for affinity design. Thus, determining the ligand-bound structure of a MW7 V_H variant will enhance the accuracy of our calculations and give us further opportunities to improve the stability and the binding for better treatments for Huntington's disease.

Materials and methods

Expression of MW7 V_H and variants in E. coli and baculovirus

The MW7 V_H genes, with and without cysteines, with an N terminal flag tag cloned into pcDNA3.1 (Invitrogen) plasmid were obtained from Dr. Paul Patterson (California Institute of Technology). These genes were subcloned into pET-22b(+) (Novagen), which added a 6x-His tag on the C terminus, and pET-11a (Novagen) plasmids for *E. coli* expression. For baculovirus expression, the genes were cloned into the pAcGP67A vector (BD Biosciences) with a C terminal His tag to be used in purification. All variants were constructed by site-directed mutagenesis using inverse PCR. All *E. coli* expression was done using BL21 (DE3) cells (Stratagene). Cells were induced with isopropyl-β-D-thiogalactoside (IPTG) for three hours at 37°C. Inclusion bodies were solubilized in 7 M guanidine-HCl, 10 mM reduced glutathione and 1 mM oxidized glutathione. The proteins were refolded by a rapid dilution method in refolding buffer (100 mM Tris, 400 mM L-arginine pH 8.0). Refolded proteins were concentrated and purified by gel filtration chromatography using a Superdex 200 column (GE Healthcare) and Ni^{2+} affinity chromatography where applicable. All baculovirus expression was done by the Protein Expression Center, California Institute of Technology.

Computational methods

All variants were designed with the ORBIT protein design software (Dahiyat and Mayo 1996; Dahiyat et al. 1997; Dahiyat and Mayo 1997; Dahiyat and Mayo 1997; Street and Mayo 1998; Pierce et al. 2000). ORBIT predicts the optimal protein sequence for a given three-dimensional fold. Since a crystal structure of MW7 V_H was not available, an alternative structure, 1SM3, was used for design. Details of the selection of

1SM3 will be explained in the **Results and discussion** section. Explicit hydrogens were added using MolProbity (Lovell et al. 2003) and the structure was energy minimized for fifty steps to remove any steric clashes. Solvation energies were calculated using the Lazaridis and Karplus method (Lazaridis and Karplus 1999).

Coexpression of MW7 V_H variants with 103-polyglutamine (polyQ) htt in cultured cells

293 cells were grown in DMEM supplemented with 10% heat-inactivated bovine serum and 2mM glutamine. Cells were grown to 70% confluence and co-transfected with 12 μg of MW7 V_H or variants cloned into pcDNA3.1 and 3 μg of htt exon 1, containing a 103-residue Gln domain (polyQ) that was fused to EGFP using a calcium phosphate transfection method. The transfected cells were examined under the microscope after 48 hours of incubation.

Results and discussion

Selection of backbone structure and design of MW7 V_H variants

The backbone structure to be used as the initial structure for ORBIT design, was chosen by searching the PDB database for antibody structures with very high sequence identity to MW7 V_H. Other criteria were high resolution, preferably better than 2 Å, and the least number of gaps in the sequence when aligned with MW7 V_H. A breast tumor specific antibody (SM3) structure (PDB ID: 1SM3) was the best candidate found (Figure 14). Its V_H domain sequence has 86% sequence identity to MW7 V_H, its resolution is 1.95 Å, and it has only one, single amino acid gap when the two sequences are aligned. Additionally, the epitope recognized by SM3 is the core repeating motif of epithelia

mucin, MUC1, which is rich in prolines (Dokurno et al. 1998). The fact that MW7 V_H recognizes the polyP domain and SM3 binds to a proline rich motif indicated that the 1SM3 structure would be a good candidate for MW7 V_H modeling.

Our design strategy was to design out the hydrophobic residues on the former V_L interface of MW7 V_H. Without a V_L domain, these hydrophobic residues will be exposed to solvent and might destabilize MW7 V_H. This is a common strategy used to stabilize single domains of antibodies. Modifying the V_H sequence so that it mimics camel V_H domains, which occur naturally without light chains, or mass screening of antibody single domain libraries to find a molecule that is stable are the usual approaches (Ward et al. 1989; Davies and Riechmann 1996; Wirtz and Steipe 1999; Tanaka et al. 2003; Colby et al. 2004). Our approach is to use computational protein design to replace these hydrophobic residues located in the interface region. Two methods were used to determine which hydrophobic residues to consider in the design. The first one employed ORBIT's RESCLASS classification protocol in which the residue positions in the structure were assigned a core, boundary or surface classification based on the distance between the solvent accessible surface and the residue's Cα and Cβ atoms (Dahiyat and Mayo 1997). The hydrophobic residues whose classification changed (from core to boundary, core to surface, or boundary to surface) when the V_L domain of 1SM3 was removed were considered for the design. The second method included all hydrophobic residues on the V_H domain within 5 Å of the V_L domain. Since residues on the CDR are involved in antigen binding, they were excluded from the design. The first method selected Leu45, Trp47, Ile95, Tyr97 and Trp107 (Trp106 in MW7 V_H) to be designed. The second method selected all the residues from the first method, plus Val37. Val37 was

classified as core by RESCLASS whether the V_L domain was present or not. Nevertheless, it is located in the middle of the interface and is surrounded by the other residues considered in the design. When mutated to hydrophilic residues, the designed residues might not appreciate a hydrophobic residue poking out among them. Visual inspection of preliminary design calculation results showed that when Tyr97 is included in the design, it retains its current identity (because of an optimal hydrogen bond with Gln39), or is mutated into a residue that has poor contacts with neighboring residues. Thus, the identity of Tyr97 was retained in all subsequent designs. None of the residues selected for design was within 6 Å of the peptide antigen in the 1SM3 structure. Non-selected residues that were within 10 Å of V_L domain and classified as surface or boundary were allowed to change conformation, but not identity. All other residues were fixed.

ORBIT design results are as follows. Method 1 produced D1, which has L45R, W47K, I95T and W107E mutations. Method 2 produced D2, which has V37Q, L45R, W47K, I95T and W107E mutations. All the mutated residues showed extensive interactions with neighboring residues and with each other. Lys47 makes electrostatic contacts with Glu50, and to a lesser degree, with Asn35 (Figure 14B). Thr95 hydrogen bonds with Thr112 (Figure 14C) and Gln37, Arg45 and Glu107 make a salt bridge/hydrogen bond network with each other (Figure 14D). Previously, residues Leu45 and Trp47 (with Gly44) on a V_H molecule of human origin were mutated to Arg and Gly, respectively, successfully producing a soluble V_H molecule (Davies and Riechmann 1996). The fact that our methods selected both of these residues for our design and that one of our predicted mutations is the same (L45R) is an encouraging result.

Expression of MW7 V_H and variants in E. coli

MW7 V_H proteins, with and without cysteines, were expressed in a variety of *E. coli* strains including BL21 (DE3), BL21 (DE3) Lys, and Rosetta Blue. The amount of soluble protein expression did not differ between strains. MW7 V_H expressed in inclusion bodies with no soluble proteins, and MW7 V_H without cysteines expressed in soluble form in very small amounts; all subsequent expressions were therefore done in BL21 (DE3) cells. Different plasmids, with and without His tag, were also tried with no significant difference. The refolding protocol described in the Materials and Methods section was used to refold MW7 V_H and its variants. There were very few or no properly folded proteins after refolding. Using hydrogen deuterium exchange experiments by NMR, Pluckthun and co-workers showed that Ile95 on the heavy chain was highly protected during folding of an anti-phosphorylcholine scFv, implying that Ile95 might be important in initiating folding (Freund et al. 1996). Thus, a variant of D1 was made that did not include the I95T mutation. This molecule also expressed in inclusion bodies and was not prone to refolding. It is well known that V_H domains often tend to form insoluble aggregates in the periplasm of *E. coli*. This might be because no general periplasmic chaperones are present to help the molecule with folding. Taking this and the fact that MW7 V_H is functional in 293 cell culture into account, we decided to express MW7 V_H and D1 in baculovirus, predicting that the chaperones in eukaryotic cells would assist in the proper folding of these molecules. This experiment is currently in progress; preliminary data indicates that both MW7 V_H and D1 express in soluble form.

Expression of MW7 V_H variants in 293 cells

To test for the functionality of MW7 V_H variants in cell culture, 293 cells were co-transfected with MW7 V_H with and without cysteines and its variants, with 103-polyQ htt fused to EGFP. It has been shown previously that MW7 V_H, but not MW7 V_H without cysteines, decreased htt aggregation and cell toxicity (P.H. Patterson, personal communication). Although the transfection data alone is insufficient to support any conclusions, it shows that the amount of transfected cells expressing htt is similar for MW7 scFv and D2 with and without cysteines. Our future work will include confirming whether the variants have functionality by analyzing the lysates of the transfected cells for aggregated htt with western blot using antibodies for htt and GFP. Staining the cells with flag and GFP antibody to observe the co-localization of MW7 V_H variants with htt and TUNEL staining to show the degree of cell toxicity also need to be shown.

Future directions

We have designed MW7 V_H variants that are predicted to be more soluble and stable than MW7 V_H. Preliminary data indicate that MW7 V_H and its variant, D1, express solubly in baculovirus. These molecules will be purified using a Ni^{2+} affinity column and their stability will be measured by thermal and chemical denaturation using circular dichroism spectroscopy. The variant's melting temperature and the free energy of folding will be compared to those of MW7 V_H. The binding of the variants and MW7 V_H will be initially determined by mixing the purified proteins with thioredoxin or GST fusion proteins of huntingtin exon 1 (TRX-HTT) containing normal length or expanded glutamines (Scherzinger et al. 1997; Bennett et al. 2002) then running a native gel or

using gel filtration chromatography. A more detailed binding study will be done with a BIACORE instrument to obtain the binding affinities. This system measures the surface plasmon resonance change that occurs when flowing analytes bind to ligands that are attached to a biosensor chip. TRX-HTT will be attached to CM5 biosensor chips using primary amine chemistry, and MW7 V_H and its variants will be flowed over the sensor chip as previously described (P. Li, manuscript in preparation). Our final biochemical goal will be to crystallize MW7 V_H or its variants with and without the polyP ligand and solve the structure by X-ray crystallography. Prior trials to crystallize anti-htt antibodies (MW1 Fab and Fv) with the polyQ ligand, a 10 – 15 residue peptide, were not successful (P. Li, manuscript in preparation). We will be attempting the same goal, but with a different system. PolyQ is notorious for being insoluble and aggregation prone, and the increased length of the polyQ domain is the direct cause for htt aggregation and leads to Huntington's disease. On the other hand, polyP region is thought to confer structural stability to the N-terminal region of htt (Qin et al. 2004). Although polyP peptides are also insoluble, many proline rich peptides have been crystallized with their counterpart proteins. Also, the proline rich motif located between the polyP domains is soluble. Including a minimal polyP stretch with the proline rich motif might create a good candidate for co-crystallization with MW7 V_H. Because the polyP domain does not directly cause htt aggregation, a variety of different peptides and fusion proteins that might improve its solubility can be tried as co-crystallization agents with MW7 V_H. A pre-crystallization step is required to check if the peptides or fusion proteins for co-crystallization actually bind to MW7 V_H. The BIACORE assay mentioned above will be a good and simple validation method to determine binding. The polyP domain has the

potential to interact with proteins that have Src homology 3 (SH3) and WW domains. Recent research showed that the deletion of the polyP domain from htt reduced the number of cells with htt-labeled autophagic bodies that are known to contribute to cell death, and also reduced or eliminated amounts of vesicle trafficking proteins in htt bodies (Qin et al. 2004). This demonstrates the importance of the polyP domain in recruiting and sequestering cellular proteins in both normal and pathological cells. It also hints at the mechanism that MW7 uses to prevent cell toxicity in cell culture. Thus, solving the structure of the polyP domain bound to MW7 V_H will be as important as, if not more important than solving the structure of MW7 V_H bound to polyQ. The crystal structure of MW7 V_H with its ligand will show the detailed molecular interaction between the two, giving us a more definitive picture of the mechanism.

The functionality of the variants in cell culture should also be corroborated. Although the BIACORE binding assay will show us whether our designs bind to htt or not, the affinities of intrabodies measured *in vitro* do not necessarily correlate with their function *in vivo* (Tanaka and Rabbitts 2003). We have already observed that when the variants are tranfected into 293 cells, they do not cause any major toxicity to the cell or influence the expression of htt co-transfected with them (data not shown). As mentioned above, experiments that confirm the expression of the variants and determine whether they decrease htt aggregation should be performed; this can be done with western blot analysis. Immunohistochemical methods will be used to visualize the co-localization of htt and the variants, and TUNEL staining will be done to quantitate the reduction in cell toxicity.

The inability of our designed MW7 V_H variants, D1 and D2, to be expressed as soluble protein in *E. coli* and to refold was disappointing. One explanation is that designing out the 4 to 5 hydrophobic residues we considered was not enough to solubilize the protein. Several hydrophobic residues located on the CDR at the interface were not considered for design, since we were concerned that altering these might disrupt the interaction between the antibody and the antigen. An alanine scanning analysis could be done to determine which hydrophobic residues on the CDR are not necessary for antigen binding. These residues could be additionally considered in future designs. Furthermore, Gly44 is another potential position to be considered. Previous attempts to camelise human V_H were successful when residues 45, 47 and 44 were mutated to other residues (Davies and Riechmann 1996). One of the criteria we used to select residues for design was to require that they be hydrophobic. Even though residue 44 was chosen using both selection methods mentioned above, it was eliminated from the pool of residues considered because it was not hydrophobic.

Another reason for the insolubility of the designed variants might be due to differences in the crystal structure used to model MW7 V_H and the actual structure. Although SM3 has high sequence identity with MW7 V_H and thus, was considered to be a good model, 1SM3 structure is of a Fab fragment, which includes not only V_H, but V_L, C_H and C_L domains. Comparison of the structure of a camelised V_H with the structure of V_H domain of Fv fragments from the same human V_H family shows differences in certain parts of the backbone, especially at the interface (Riechmann 1996). By using the actual crystal structure of MW7 V_H, this problem can be solved. Thus, the crystal structure of

MW7 V_H can also be utilized in a second round of designs to produce variants with higher solubility and stability.

Although all MW7 V_H molecules mentioned were tested both with and without cysteines, the molecules without cysteines were not emphasized because they were not functional in cell culture (P.H. Patterson, personal communication). The reductive environment in the cytoplasm does not allow disulfide bonds to form; thus, we would expect similar functionality from both forms of MW7 V_H. Removal of intra-domain disulphide bonds from intrabodies has been shown to have no substantial effect on *in vivo* expression or function (Tanaka et al. 2003). The contradictory results we observed with MW7 V_H might be because the cysteine to alanine mutation used to delete cysteines did not result in optimal packing, which could destabilize the molecule. Packing is especially important in this case because the disulfide bond is deeply buried in the protein. In this sense, using ORBIT to design out the disulfide bonds should result in an optimal substitution for cysteines. ORBIT has been successfully used in the past to design out disulfide bonds (unpublished result) and repack protein cores (Dahiyat and Mayo 1997).

Our ultimate goal in this project is to design a highly soluble and stable anti-htt polyP MW7 V_H that could be used for a variety of applications. These include use in yeast surface screening for antibodies with improved functionality (Colby et al. 2004), as bio reagents to detect htt, and as therapeutics for Huntington's disease.

References

Bennett, M.J., Huey-Tubman, K.E., Herr, A.B., West, A.P., Jr., Ross, S.A., and Bjorkman, P.J. 2002. Inaugural Article: A linear lattice model for polyglutamine in CAG-expansion diseases. *Proc Natl Acad Sci U S A* **99:** 11634-11639.

Colby, D.W., Garg, P., Holden, T., Chao, G., Webster, J.M., Messer, A., Ingram, V.M., and Wittrup, K.D. 2004. Development of a human light chain variable domain (V(L)) intracellular antibody specific for the amino terminus of huntingtin via yeast surface display. *J Mol Biol* **342:** 901-912.

Colby, D.W., Kellogg, B.A., Graff, C.P., Yeung, Y.A., Swers, J.S., and Wittrup, K.D. 2004. Engineering antibody affinity by yeast surface display. *Methods Enzymol* **388:** 348-358.

Dahiyat, B.I., Gordon, D.B., and Mayo, S.L. 1997. Automated design of the surface positions of protein helices. *Protein Sci* **6:** 1333-1337.

Dahiyat, B.I., and Mayo, S.L. 1996. Protein design automation. *Protein Sci* **5:** 895-903.

Dahiyat, B.I., and Mayo, S.L. 1997. *De novo* protein design: fully automated sequence selection. *Science* **278:** 82-87.

Dahiyat, B.I., and Mayo, S.L. 1997. Probing the role of packing specificity in protein design. *Proc Natl Acad Sci U S A* **94:** 10172-10177.

Davies, J., and Riechmann, L. 1996. Single antibody domains as small recognition units: design and *in vitro* antigen selection of camelized, human VH domains with improved protein stability. *Protein Eng* **9:** 531-537.

Dokurno, P., Bates, P.A., Band, H.A., Stewart, L.M., Lally, J.M., Burchell, J.M., Taylor-Papadimitriou, J., Snary, D., Sternberg, M.J., and Freemont, P.S. 1998. Crystal

structure at 1.95 A resolution of the breast tumour-specific antibody SM3 complexed with its peptide epitope reveals novel hypervariable loop recognition. *J Mol Biol* **284:** 713-728.

Freund, C., Honegger, A., Hunziker, P., Holak, T.A., and Pluckthun, A. 1996. Folding nuclei of the scFv fragment of an antibody. *Biochemistry* **35:** 8457-8464.

Khoshnan, A., Ko, J., and Patterson, P.H. 2002. Effects of intracellular expression of anti-huntingtin antibodies of various specificities on mutant huntingtin aggregation and toxicity. *Proc Natl Acad Sci U S A* **99:** 1002-1007.

Lazaridis, T., and Karplus, M. 1999. Effective energy function for proteins in solution. *Proteins* **35:** 133-152.

Lecerf, J.M., Shirley, T.L., Zhu, Q., Kazantsev, A., Amersdorfer, P., Housman, D.E., Messer, A., and Huston, J.S. 2001. Human single-chain Fv intrabodies counteract *in situ* huntingtin aggregation in cellular models of Huntington's disease. *Proc Natl Acad Sci U S A* **98:** 4764-4769.

Lobato, M.N., and Rabbitts, T.H. 2004. Intracellular antibodies as specific reagents for functional ablation: future therapeutic molecules. *Curr Mol Med* **4:** 519-528.

Lovell, S.C., Davis, I.W., Arendall, W.B., 3rd, de Bakker, P.I., Word, J.M., Prisant, M.G., Richardson, J.S., and Richardson, D.C. 2003. Structure validation by C alpha geometry: phi, psi and C beta deviation. *Proteins* **50:** 437-450.

Pierce, N.A., Spriet, J.A., Desmet, J., and Mayo, S.L. 2000. Conformational splitting: A more powerful criterion for dead- end elimination. *J. Comput. Chem.* **21:** 999-1009.

Qin, Z.H., Wang, Y., Sapp, E., Cuiffo, B., Wanker, E., Hayden, M.R., Kegel, K.B., Aronin, N., and DiFiglia, M. 2004. Huntingtin bodies sequester vesicle-associated

proteins by a polyproline-dependent interaction. *J Neurosci* **24:** 269-281.

Riechmann, L. 1996. Rearrangement of the former VL interface in the solution structure of a camelised, single antibody VH domain. *J Mol Biol* **259:** 957-969.

Scherzinger, E., Lurz, R., Turmaine, M., Mangiarini, L., Hollenbach, B., Hasenbank, R., Bates, G.P., Davies, S.W., Lehrach, H., and Wanker, E.E. 1997. Huntingtin-encoded polyglutamine expansions form amyloid-like protein aggregates *in vitro* and *in vivo*. *Cell* **90:** 549-558.

Stocks, M.R. 2004. Intrabodies: production and promise. *Drug Discov Today* **9:** 960-966.

Street, A.G., and Mayo, S.L. 1998. Pairwise calculation of protein solvent-accessible surface areas. *Fold Des* **3:** 253-258.

Tanaka, T., Lobato, M.N., and Rabbitts, T.H. 2003. Single domain intracellular antibodies: a minimal fragment for direct *in vivo* selection of antigen-specific intrabodies. *J Mol Biol* **331:** 1109-1120.

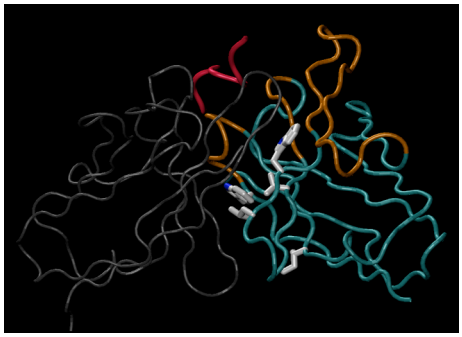
Tanaka, T., and Rabbitts, T.H. 2003. Intrabodies based on intracellular capture frameworks that bind the RAS protein with high affinity and impair oncogenic transformation. *Embo J* **22:** 1025-1035.

Ward, E.S., Gussow, D., Griffiths, A.D., Jones, P.T., and Winter, G. 1989. Binding activities of a repertoire of single immunoglobulin variable domains secreted from Escherichia coli. *Nature* **341:** 544-546.

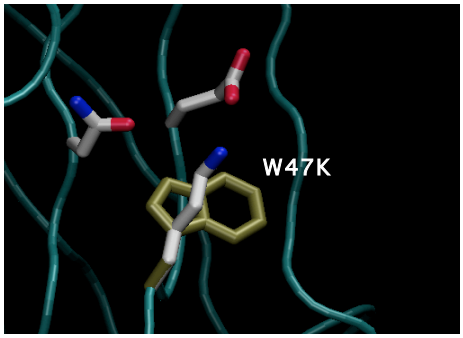
Wirtz, P., and Steipe, B. 1999. Intrabody construction and expression III: engineering hyperstable V(H) domains. *Protein Sci* **8:** 2245-2250.

Figure 14. Structures of 1SM3 and its mutants showing the mutated residues (A) Structure of 1SM3 Fv. V_H, V_L, peptide ligand and CDR of V_H is shown in cyan, gray, red and orange, respectively. The side chains of the residues mutated are shown. The wild-type residues are shown in gold. (B) W47K (C) I95T (D) V37Q, L45R, W107E mutations in variant MW7 V_H are shown in CPK-inspired colors.

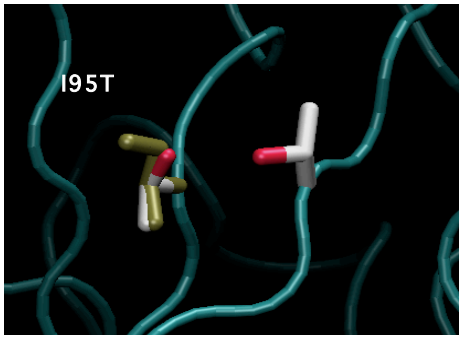
A



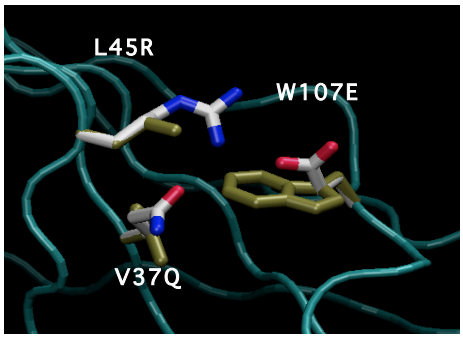
B



C



D



Appendix A

Ab initio modeling and ligand binding study of Edg4 and Edg6

The text of this chapter is adapted from the publication

Vaidehi N, Floriano WB, Trabanino R, Hall SE, Freddolino P, **Choi EJ**, Zamanakos G, and Goddard WA 3rd. *Proc Natl Acad Sci USA* 2002 99, 12622-7.

Abstract

We performed computational *ab initio* modeling and ligand docking studies of the endothelial differentiation gene (Edg) G protein-coupled receptors Edg4 and Edg6. By using a well-established molecular dynamics method, we derived atomic models for the two receptors. The two receptors show different ligand specificities, with Edg4 binding to lysophosphatidic acid and Edg6 binding to sphingosine-1-phosphate. Using these molecules as positive agonists, we docked each of them to the predicted structure of each of the receptors and found the most favorable binding site. Next, agonists and antagonists were docked to the putative binding sites and the binding energies were calculated. The results are consistent with experimental data.

Introduction

Different extracellular stimuli transfer signals to the interior of the cell via receptors on the cell surface. Many hormones, lipid mediators, oligopeptides, odorants and taste molecules bind to G protein-coupled receptors (GPCRs) and induce their conformational change, which initiates signal transduction in the cytoplasm. The GPCR family is the largest known receptor family, comprising approximately 1% of the proteins encoded by the humane genome (Sautel and Milligan 2000). The cellular mechanisms that these diverse GPCRs are involved in are extremely important; hence, they are frequently used as drug targets. The fact that over 50% of approved drugs are agonists or antagonists of these receptors emphasizes their therapeutic significance (Gudermann et al. 1995).

The use of molecular biology and comprehensive high throughput drug screening has resulted in the discovery of many GPCR ligands, especially small molecule ligands of peptide-binding GPCRs (Muller 2000). Nevertheless, ligand identification is difficult due to the enormous number of possible ligands to be tested. Complications arising from the elusive nature of certain ligands and cell assay systems with endogenous GPCRs make the process more difficult (Marchese et al. 1999). In addition, even with the ligand identified, a detailed understanding of ligand-receptor interactions is needed for drug design and to gain insight into diseases caused by malfunction of the receptors. For these reasons, structural and functional analysis of GPCRs is critical. Unfortunately, structural information on GPCRs is insufficient. This problem can be overcome by using atomic level models.

The endothelial differentiation gene (Edg) receptor family is a subfamily of rhodopsin-like GPCRs that has been implicated in diverse biological processes such as cell proliferation, differentiation and migration (Lynch and Im 1999; Pyne and Pyne 2000). The most significant ligands for these receptors are sphingosine-1-phosphate (S1P), a phosphorylated derivative of sphingosine that binds to Edg1, Edg3, Edg5, Edg6 and Edg8, and lysophosphatidic acid (LPA) that binds to Edg2, Edg4 and Edg7 (An et al. 1997a; An et al. 1997b; An et al. 1998; Lee et al. 1998; Im et al. 2000a; Im et al. 2000b; Van Brocklyn et al. 2000). Edg receptors are widely distributed in many tissues and tend to interact with multiple intracellular signal transduction pathways including MAPK, phospholipase C/D, adenylate cyclase, JNK and small GTPases (Fukushima et al. 2001; Takuwa et al. 2001). However, the biological functions and molecular mechanisms of these receptors are poorly understood at the present time.

We modeled Edg6 and Edg4 using a computational GPCR tertiary structure prediction method called MembStruk. Starting with these predicted structures, we then used the HierDock ligand docking procedure to analyze receptor-ligand interactions (Floriano et al. 2000). Edg6 was modeled because, in contrast to the universal distribution of the majority of the Edg receptors, Edg6 expression is restricted to the tissues of the immune system and thus, it is assigned important cell signaling roles in various immune responses (Graler et al. 1998). Edg4 has ligand specificity opposite to Edg6 and was selected because it has been implicated in ovarian cancer (Contos and Chun 2000; Fang et al. 2002). The MembStruk/HierDock approach to modeling could be applied to find ligands of receptors and to design ligands to be used as drugs. It could also be used to assist in functional studies such as finding the site for mutagenesis studies, to understand

the mechanism of ligand-receptor interactions, and to study the overall mechanism of the receptor and its involvement in diseases.

Materials and methods

Prediction of Edg receptor structures by MembStruk

17 Edg receptor gene sequences were obtained by BLAST homology search (Altschul et al. 1997) and used to obtain the sequence alignment profile by ClustalX (Altschul et al. 1990). The transmembrane helices were identified on the basis of hydrophobicity by the multi-sequence profile method (Donnelly 1993) implemented in PERSCAN. A window size of 20 residues was used. The sequence for the Edg6 receptor was used to build canonical right-handed helices with extended side chains. The structures of these helices were optimized using the Newton-Euler Inverse Mass Operator torsional MD method that scales linearly with the number of torsional degrees of freedom (Jain et al. 1993; Mathiowetz et al. 1994; Vaidehi et al. 1996). This procedure fixes the bonds and angles, optimizing the side chain conformations.

The orientation of each helical axis was determined using the bovine rhodopsin 7.5 Å electron density map (Schertler 1998). In addition, the hydrophobic moment of each helix was directed outwards. The helices of Edg4 were rotated by running a coarse grain rotation dynamics program called CoarseRot (Vaidehi et al. 2002). The helix bundle was placed in the middle of a membrane bilayer composed of dilauroylphosphatidyl choline, and the packing of the helix bundle was optimized using rigid body dynamics. The DREIDING force field (Mayo et al. 1990) and CHARMM22

(MacKerell et al. 1998) charges were utilized and the rigid body dynamics protocol was run for 150 ps to attain equilibration.

Loops were then added to the helices with WHATIF software (Altenberg-Greulich and Vriend 2001), and a molecular dynamics optimization was performed using the MASSIVELY PARALLEL SIMULATION program (MPSim). (Ding et al. 1992a; b; Lim et al. 1997) The procedure first optimizes loops with the helices fixed, then the entire receptor is optimized.

Docking studies with HierDock

Each ligand was built in the extended conformation. The starting conformations were optimized by potential energy minimization with the conjugate gradient method employing the DREIDING force field (Mayo et al. 1990) and Gasteiger charges (Gasteiger and Marsili 1980). The minimized conformations were used as starting conformations for docking. The HierDock protocol (Floriano et al. 2000) and the molecular surface of the protein model were used to locate the binding site using the positive ligand. First, a coarse grain docking procedure using DOCK 4.0 (Ewing and Kuntz 1997) was carried out. Void regions within the receptor are divided into several boxes and multiple ligand conformations are sampled in them. Then, 1000 ligand configurations are generated for each box and the 100 best scoring structures are kept. Next, these best 100 ligand structures are minimized with the protein fixed using the DREIDING force field and QEq charges. The structures are then ranked using both energy and solvation. The binding site is defined as the site where most of the best scoring ligands cluster. After defining the binding site, a library of ligands is docked to

this binding region. The library docking procedure uses the protocol mentioned above, plus another step in which the best structures (the top 10%) are further minimized with all atoms movable. Then, each structure is scored and ranked using the binding energies, which are calculated as the difference between the energy of the ligand in the receptor and the energy of the free ligand.

Results and discussion

Structural model of Edg6 and Edg4 receptors

The seven helical transmembrane GPCRs of the Edg family are divided into two subgroups according to the extent of amino acid sequence homology and ligand specificity. The first group is selective for S1P and is comprised of Edg1, Edg3, Edg5, Edg6 and Edg8 receptors. The second is selective for LPA, and includes Edg2, Edg4 and Edg7. Using MembStruk and a proper description of the differential solvent environment, we determined atomic level models of Edg6 and Edg4. No prior structural information was used, except the receptor sequence and the 7.5 Å rhodopsin crystal structure, which was used to align the helix axis (Figure 15).

Identification of the most probable binding site

Without any previous knowledge of the binding site, S1P and LPA were used as positive ligands to find the putative binding sites for Edg6 and Edg4, respectively. The binding site predicted by the HierDock protocol using the modeled structures was the same for both receptors: between helix 3,5,6 and 7. This correlates well with the binding region of the retinal in rhodopsin as well as the ligand-binding site in the cannabinoid

receptor, which has the highest homology to the Edg receptors outside the Edg family (Graler et al. 1998).

Binding affinity ordering

Having located the putative binding site, we docked S1P, LPA, sphingosylphosphorylcholine (SPC) and 2-amino-2-[2-(4-octylphenyl)ethyl]-propane-1,3-diol (FTY720) into the putative binding site and calculated the binding energies. SPC has been reported to be a low affinity ligand for the S1P binding receptors with approximately 10- to 100-fold lower activation. But this is controversial because contradictory results have also been reported. Recently, it has been shown that activation of Edg receptors by SPC might have been the result of contaminants or the indirect effect of other SPC-specific GPCR (e.g., OGR1, G2A) activation (Van Brocklyn et al. 2000; Fukushima et al. 2001; Kluk and Hla 2002; Xu 2002). FTY720 is an immunosuppressant known to sequester lymphocytes in secondary lymphoid organs, thus decreasing the number of patrolling lymphocytes, especially T lymphocytes (Brinkmann et al. 2000; Brinkmann et al. 2001). Recently, the phosphate ester metabolite of FTY720 was found to displace S1P binding with pico to nanomolar inhibitory concentration 50% (IC50) values on S1P binding receptors (Mandala et al. 2002).

The binding energies from the docking results are listed in Table 6. The binding energies of S1P and LPA for both receptors are consistent with the experimental results. Edg6 prefers S1P to LPA, while Edg4 prefers LPA to S1P. The calculated binding energy for Edg6 and SPC is worse than that for Edg6 and LPA, which does not correlate with experimental results reported previously. But as mentioned above, this result is

controversial. Our result supports the suggestion that SPC is at best a very weak ligand of Edg receptors. Nevertheless, we cannot eliminate the possibility that the crudeness of the model is reflected by this inconsistency with experimental data. Another explanation for the bad binding energy obtained for SPC could be that SPC binds to Edg6, but at a different binding site from S1P. Thus, docking SPC in the putative binding site located by using S1P as a positive ligand might give a bad binding energy because the SPC binding site is somewhere else. The calculation results for FTY720 show that it is a good ligand for Edg6, with a binding energy close to the positive ligand S1P, while with Edg4, it has a bad binding energy, close to that of S1P. This corroborates the assumption made in prior publications in which FTY720 was hypothesized to be an Edg6 ligand because of its immunosuppressant activity and its structural similarity with S1P,. This is further substantiated by a recent report that the phosphate ester metabolite of FTY720 is a high affinity agonist of at least four of the five S1P-specific Edg receptors.

Residues predicted as directly involved in binding the ligands

Figure 16 shows the predicted recognition site for Edg6 and Edg4 with the residues involved in ligand binding displayed. The phosphate of S1P makes hydrogen bonds with the Thr127 and Trp291 of Edg6, and the nitrogen of the ammonium group makes a hydrogen bond with Glu284. On the other hand, Ser114 and Lys278 interact with LPA in our model of Edg4. These residues are on helix 3 and 7, which correlates with the helices in rhodopsin and the cannabinoid receptor interacting with each of their own ligands.

Not many site-directed mutagenesis studies have been done on the Edg receptors. Homology modeling of Edg1 and Edg2 followed by experimental validation was published recently (Parrill et al. 2000; Wang et al. 2001). These studies suggest that Edg1 uses Arg120, Arg292 and Glu121 to interact with S1P, while Edg2 uses Arg124 and Lys304 to interact with LPA. The three residues our model predicts to interact with the ligand in Edg6 differ from those reported for Edg1. Parrill and co-workers showed that their three predicted ligand recognition residues are conserved in all S1P-binding Edg receptors including Edg6. Our predicted residues are more than 5 Å away from their prediction. This disagreement could be due to different S1P binding sites in different Edg receptors. Different S1P-binding Edg receptors have different affinities and activity responses to different agonists and antagonists; thus, it has been surmised that the binding sites for each of the receptors may be different (Spiegel and Milstien 2000; Kluk and Hla 2002). Also, Edg6 is more likely to differ from the other S1P-binding Edg receptors because it has relatively low sequence homology with them, and higher homology with the LPA-binding Edg receptors. Thus, Edg6 is included in the S1P-binding Edg receptor subfamily by its ligand specificity, but when considering only sequence homology, it cannot be clearly grouped into either the S1P-binding or the LPA-binding subfamily (Lynch and Im 1999; Van Brocklyn et al. 2000; Fukushima et al. 2001). Alignment of the sequence for the human and the mouse Edg6 gene shows that Glu284 and the Trp291 residues are conserved, indicating that these residues are important and could indeed be the residues utilized in ligand recognition.

In contrast to the results obtained with Edg6, the residues we predicted with our Edg4 model correlate well with the results from the Parrill and co-workers. Lys278

predicted by our model correlates with Lys304 of Edg2 when the two sequences are aligned, and the Ser114 is conserved in all the LPA-binding receptors.

Conclusion

We used a modeling and docking protocol to predict the structures of two Edg receptors and to dock predetermined ligands onto them. The docking study enabled us to validate the modeling technique by allowing us to compare the computational results with experimental results from the literature. It also allowed us to predict the receptor binding site and putative ligands. During the procedure, the structural information used included only the receptor sequences and the bovine rhodopsin crystal structure (for helix alignment). Although the results achieved using these theoretical methods must be validated experimentally, they are in good agreement with current published reports on these receptors. This is certainly encouraging and accumulation of similar studies will lay the foundation for theoretical modeling studies of proteins in the future.

References

Altenberg-Greulich, B., and Vriend, G. 2001. Where to attach dye molecules to a protein: lessons from the computer program WHAT IF. *J. Mol. Struct.* **598:** 1-8.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215:** 403-410.

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25:** 3389-3402.

An, S., Bleu, T., Hallmark, O.G., and Goetzl, E.J. 1998. Characterization of a novel subtype of human G protein-coupled receptor for lysophosphatidic acid. *J Biol Chem* **273:** 7906-7910.

An, S., Bleu, T., Huang, W., Hallmark, O.G., Coughlin, S.R., and Goetzl, E.J. 1997a. Identification of cDNAs encoding two G protein-coupled receptors for lysosphingolipids. *FEBS Lett* **417:** 279-282.

An, S., Dickens, M.A., Bleu, T., Hallmark, O.G., and Goetzl, E.J. 1997b. Molecular cloning of the human Edg2 protein and its identification as a functional cellular receptor for lysophosphatidic acid. *Biochem Biophys Res Commun* **231:** 619-622.

Brinkmann, V., Chen, S., Feng, L., Pinschewer, D., Nikolova, Z., and Hof, R. 2001. FTY720 alters lymphocyte homing and protects allografts without inducing general immunosuppression. *Transplant Proc* **33:** 530-531.

Brinkmann, V., Pinschewer, D., Chiba, K., and Feng, L. 2000. FTY720: a novel transplantation drug that modulates lymphocyte traffic rather than activation. *Trends Pharmacol Sci* **21:** 49-52.

Contos, J.J., and Chun, J. 2000. Genomic characterization of the lysophosphatidic acid receptor gene, lp(A2)/Edg4, and identification of a frameshift mutation in a previously characterized cDNA. *Genomics* **64:** 155-169.

Ding, H.Q., Karasawa, N., and Goddard, W.A. 1992a. Atomic level simulations on a million particles - the cell multipole method for Coulomb and London nonbond interactions. *J. Chem. Phys.* **97:** 4309-4315.

Ding, H.Q., Karasawa, N., and Goddard, W.A. 1992b. The reduced cell multipole method for Coulomb interactions in periodic-systems with million-atom unit cells. *Chem. Phys. Lett.* **196:** 6-10.

Donnelly, D. 1993. Modelling alpha-helical transmembrane domains. *Biochem Soc Trans* **21:** 36-39.

Ewing, T.J.A., and Kuntz, I.D. 1997. Critical evaluation of search algorithms for automated molecular docking and database screening. *J. Comput. Chem.* **18:** 1175-1189.

Fang, X., Schummer, M., Mao, M., Yu, S., Tabassam, F.H., Swaby, R., Hasegawa, Y., Tanyi, J.L., LaPushin, R., Eder, A., et al. 2002. Lysophosphatidic acid is a bioactive mediator in ovarian cancer. *Biochim Biophys Acta* **1582:** 257-264.

Floriano, W.B., Vaidehi, N., Goddard, W.A., 3rd, Singer, M.S., and Shepherd, G.M. 2000. Molecular mechanisms underlying differential odor responses of a mouse olfactory receptor. *Proc Natl Acad Sci U S A* **97:** 10712-10716.

Fukushima, N., Ishii, I., Contos, J.J., Weiner, J.A., and Chun, J. 2001. Lysophospholipid receptors. *Annu Rev Pharmacol Toxicol* **41:** 507-534.

Gasteiger, J., and Marsili, M. 1980. Iterative partial equalization of orbital electronegativity - a rapid access to atomic charges. *Tetrahedron* **36:** 3219-3228.

Graler, M.H., Bernhardt, G., and Lipp, M. 1998. EDG6, a novel G-protein-coupled receptor related to receptors for bioactive lysophospholipids, is specifically expressed in lymphoid tissue. *Genomics* **53:** 164-169.

Gudermann, T., Nurnberg, B., and Schultz, G. 1995. Receptors and G proteins as primary components of transmembrane signal transduction. Part 1. G-protein-coupled receptors: structure and function. *J Mol Med* **73:** 51-63.

Im, D.S., Heise, C.E., Ancellin, N., O'Dowd, B.F., Shei, G.J., Heavens, R.P., Rigby, M.R., Hla, T., Mandala, S., McAllister, G., et al. 2000a. Characterization of a novel sphingosine 1-phosphate receptor, Edg-8. *J Biol Chem* **275:** 14281-14286.

Im, D.S., Heise, C.E., Harding, M.A., George, S.R., O'Dowd, B.F., Theodorescu, D., and Lynch, K.R. 2000b. Molecular cloning and characterization of a lysophosphatidic acid receptor, Edg-7, expressed in prostate. *Mol Pharmacol* **57:** 753-759.

Jain, A., Vaidehi, N., and Rodriguez, G. 1993. A fast recursive algorithm for molecular-dynamics simulation. *J. Comput. Phys.* **106:** 258-268.

Kluk, M.J., and Hla, T. 2002. Signaling of sphingosine-1-phosphate via the S1P/EDG-family of G-protein-coupled receptors. *Biochim Biophys Acta* **1582:** 72-80.

Lee, M.J., Van Brocklyn, J.R., Thangada, S., Liu, C.H., Hand, A.R., Menzeleev, R., Spiegel, S., and Hla, T. 1998. Sphingosine-1-phosphate as a ligand for the G protein-coupled receptor EDG-1. *Science* **279:** 1552-1555.

Lim, K.T., Brunett, S., Iotov, M., McClurg, R.B., Vaidehi, N., Dasgupta, S., Taylor, S., and Goddard, W.A. 1997. Molecular dynamics for very large systems on massively parallel computers: The MPSim program. *J. Comput. Chem.* **18:** 501-521.

Lynch, K.R., and Im, D.S. 1999. Life on the edg. *Trends Pharmacol Sci* **20:** 473-475.

MacKerell, A.D., Bashford, D., Bellott, M., Dunbrack, R.L., Evanseck, J.D., Field, M.J., Fischer, S., Gao, J., Guo, H., Ha, S., et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **102:** 3586-3616.

Mandala, S., Hajdu, R., Bergstrom, J., Quackenbush, E., Xie, J., Milligan, J., Thornton, R., Shei, G.J., Card, D., Keohane, C., et al. 2002. Alteration of lymphocyte trafficking by sphingosine-1-phosphate receptor agonists. *Science* **296:** 346-349.

Marchese, A., George, S.R., Kolakowski, L.F., Jr., Lynch, K.R., and O'Dowd, B.F. 1999. Novel GPCRs and their endogenous ligands: expanding the boundaries of physiology and pharmacology. *Trends Pharmacol Sci* **20:** 370-375.

Mathiowetz, A.M., Jain, A., Karasawa, N., and Goddard, W.A. 1994. Protein simulations using techniques suitable for very large systems - the cell multipole method for nonbond interactions and the Newton-Euler Inverse Mass Operator method for internal coordinate dynamics. *Proteins* **20:** 227-247.

Mayo, S.L., Olafson, B.D., and Goddard, W.A. 1990. Dreiding - a generic force-field for molecular simulations. *J. Phys. Chem.* **94:** 8897-8909.

Muller, G. 2000. Towards 3D structures of G protein-coupled receptors: a multidisciplinary approach. *Curr Med Chem* **7:** 861-888.

Parrill, A.L., Wang, D., Bautista, D.L., Van Brocklyn, J.R., Lorincz, Z., Fischer, D.J., Baker, D.L., Liliom, K., Spiegel, S., and Tigyi, G. 2000. Identification of Edg1 receptor residues that recognize sphingosine 1-phosphate. *J Biol Chem* **275:** 39379-39384.

Pyne, S., and Pyne, N.J. 2000. Sphingosine 1-phosphate signalling in mammalian cells. *Biochem J* **349:** 385-402.

Sautel, M., and Milligan, G. 2000. Molecular manipulation of G-protein-coupled receptors: a new avenue into drug discovery. *Curr Med Chem* **7:** 889-896.

Schertler, G.F. 1998. Structure of rhodopsin. *Eye* **12 (Pt 3b):** 504-510.

Spiegel, S., and Milstien, S. 2000. Functions of a new family of sphingosine-1-phosphate receptors. *Biochim Biophys Acta* **1484:** 107-116.

Takuwa, Y., Okamoto, H., Takuwa, N., Gonda, K., Sugimoto, N., and Sakurada, S. 2001. Subtype-specific, differential activities of the EDG family receptors for sphingosine-1-phosphate, a novel lysophospholipid mediator. *Mol Cell Endocrinol* **177:** 3-11.

Vaidehi, N., Floriano, W.B., Trabanino, R., Hall, S.E., Freddolino, P., Choi, E.J., Zamanakos, G., and Goddard, W.A., 3rd. 2002. Prediction of structure and function of G protein-coupled receptors. *Proc Natl Acad Sci U S A* **99:** 12622-12627.

Vaidehi, N., Jain, A., and Goddard, W.A. 1996. Constant temperature constrained molecular dynamics: The Newton-Euler inverse mass operator method. *J. Phys. Chem.* **100:** 10508-10517.

Van Brocklyn, J.R., Graler, M.H., Bernhardt, G., Hobson, J.P., Lipp, M., and Spiegel, S. 2000. Sphingosine-1-phosphate is a ligand for the G protein-coupled receptor EDG-6. *Blood* **95:** 2624-2629.

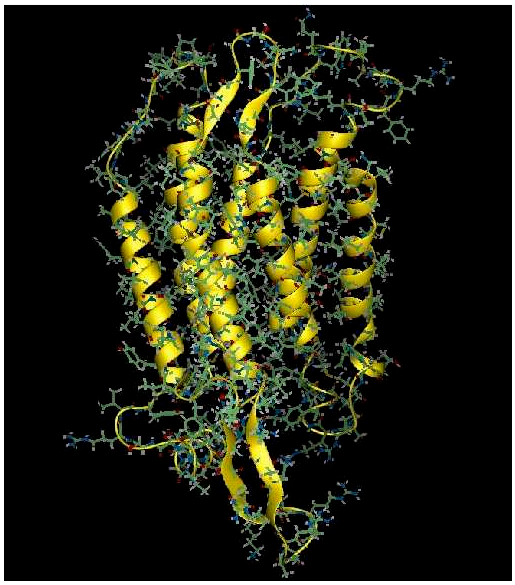
Wang, D.A., Lorincz, Z., Bautista, D.L., Liliom, K., Tigyi, G., and Parrill, A.L. 2001. A single amino acid determines lysophospholipid specificity of the S1P1 (EDG1) and LPA1 (EDG2) phospholipid growth factor receptors. *J Biol Chem* **276:** 49213-49220.

Xu, Y. 2002. Sphingosylphosphorylcholine and lysophosphatidylcholine: G protein-coupled receptors and receptor-mediated signal transduction. *Biochim Biophys Acta* **1582:** 81-88.

Table 6. Computational binding energies of Edg6 and Edg4 to S1P, LPA, SPC and FTY720.

	Edg6	**Edg4**
S1P	-41.7	-33.6
LPA	-34.1	-40.6
SPC	33.4	39.2
FTY720	-40.3	-34.9

Figure 15. Modeled structures of Edg6 (left) and Edg4 (right).



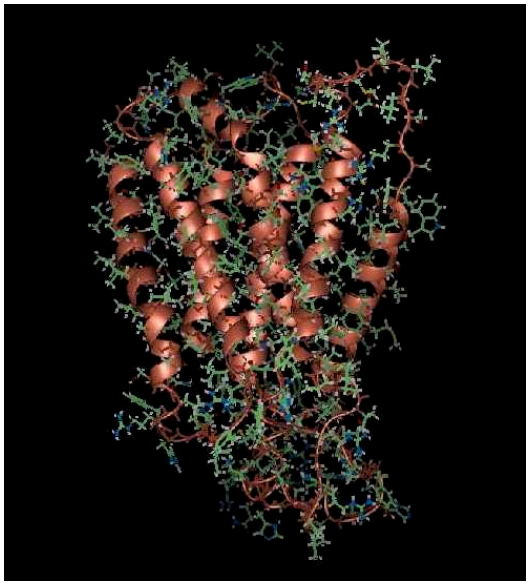
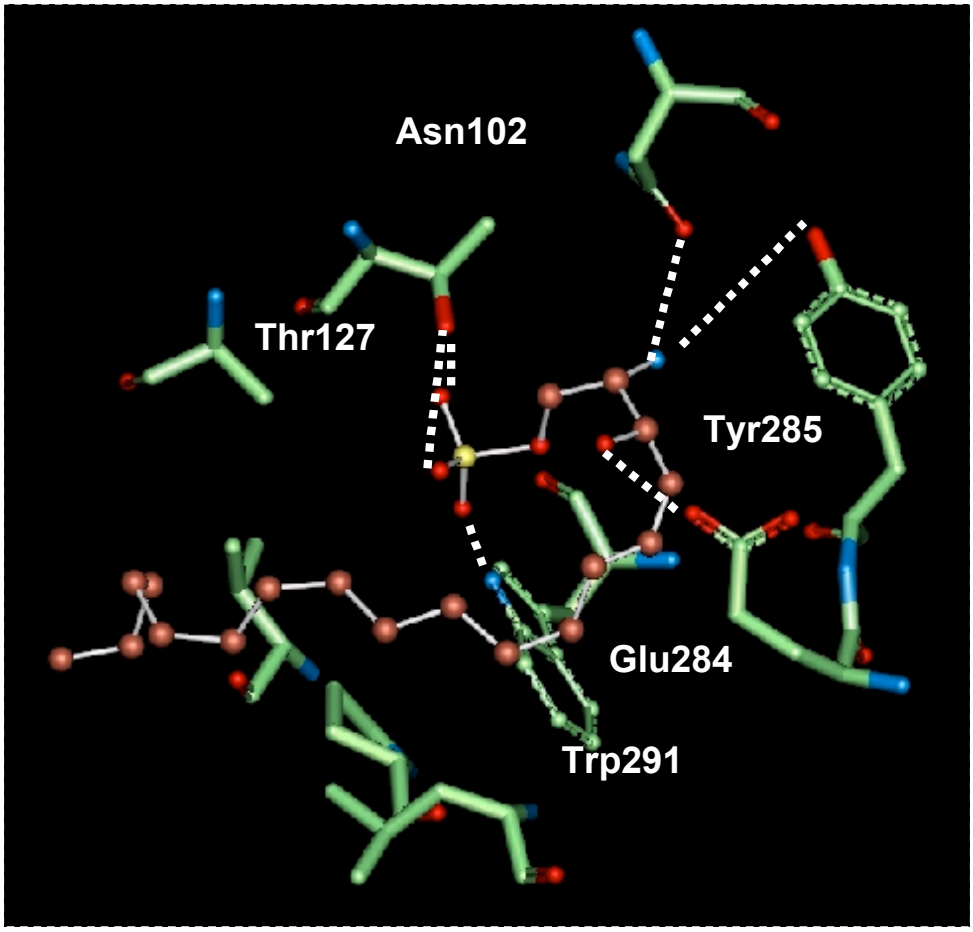


Figure 16. Predicted recognition site for Edg6 (A) and Edg4 (B) with the residues involved in ligand binding displayed.

A



B