DYNAMIC EFFECTS OF MUTATIONS IN EGLIN C

by
Michael Clarkson

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biochemistry and Biophysics.

Chapel Hill
2005

Approved by:

Advisor: Andrew Lee

Reader: Gary Pielak

Reader: Sharon Campbell

Reader: Marshall Edgell

Reader: Dale Ramsden

ABSTRACT

MICHAEL CLARKSON: Dynamic Effects of Mutations in Eglin c

(Under the direction of Andrew Lee)

In order to investigate the mechanism of dynamic propagation in proteins and the role these responses play in energetic communication, a series of novel dynamic perturbation-response experiments using NMR spin-relaxation of ^{15}N and ^{2}H were performed on the serine protease inhibitor eglin c. Although analysis was complicated in some cases by significant structural changes, a considerable number of mutations provoked a substantial response among contiguous side chains while leaving backbone dynamics unaltered. In contrast, some mutations induced dynamic effects that were distributed discontinuously through the protein and associated with very subtle structural changes. Side chains did not always produce reciprocal dynamic responses. Both the "contiguous network responses" and "disperse network responses" were associated with significant free energies of interaction. These results suggest that dynamic networks may play a role in energetic communication.

For my grandmothers, Floreine and Theodora, who were both teachers.

ACKNOWLEDGEMENTS

This work would not have been possible without the operational and intellectual support of several people. I am deeply indebted to my advisor, Andrew Lee, and the members of his lab, particularly Ernesto Fuentes and Steve Gilmore, for their assistance and support. Greg Young, Karl Koshlap, and Fang Yi provided me invaluable aid with the equipment needed for this project. I would also like to thank Jan Hermans, Marshall Edgell, and Gary Pielak for immensely helpful intellectual discussions, and Sharon Campbell and Dale Ramsden for their service on my committee. Barry Lentz contributed by supporting me through the Molecular and Cellular Biophysics Program, for which I am very grateful. Josh Boyer and Eugene DeRose also provided material support and information. Without all of them, this project would not have been possible.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS AND SYMBOLS

AAPF	–	n-Succinyl Ala Ala Pro Phe p-nitroanilide
χ_1	–	First side-chain torsion angle
χ^2	–	Measure of goodness of fit
CI-2	–	Chymotrypsin Inhibitor 2
$\Delta\Delta\Delta G_i$	–	Free energy of interaction
ΔG_u	–	Free energy of unfolding
GuHCl	–	Guanidinium hydrochloride
J_{CC}	–	Three-bond scalar coupling between carbonyl C and Cγ
J_{NC}	–	Three bond scalar coupling between amide N and Cγ
NMR	–	Nuclear Magnetic Resonance
NOE	–	Nuclear Overhauser Effect
PDB	–	Protein Data Bank
RDC	–	Residual Dipolar Coupling
S^2	–	Generalized Lipari-Szabo order parameter
S^2_{axis}	–	Axial Lipari-Szabo order parameter
τ_e	–	Effective internal correlation time
τ_m	–	Global rotational correlation time
TCL	–	Tool Control Language

INTRODUCTION

Proteins use cooperative processes to fold and function. Because these events require many amino acids to work in concert, it is necessary to understand how residues influence each other in order to comprehend these protein behaviors (Luque et al., 2002; Yu and Koshland Jr., 2001). Significant progress has been made in explaining the interactions of adjacent residues, but little has been done on residues that lie far apart within a molecule. Evolutionary studies have indicated that proteins share energy through sparse internal networks of residues in at least some cases, but these experiments do not shed light on the mechanism by which energy is conveyed (Lockless and Ranganathan, 1999; Suel et al., 2003). There are well-known cases, such as that of hemoglobin, in which communication between distal loci can be explained by conformational change (Perutz, 1970). In many instances, however, a structural rationale is not apparent (Mace et al., 1995; McElroy et al., 2002; Meroueh et al., 2002; Ohtaka et al., 2003; Rajagopalan et al., 2002). A deeper understanding of the means by which energy is distributed and relocated within a protein is necessary to advance our understanding of diverse behaviors such as allostery, folding, and ligand specificity.

Experimental and theoretical work has also demonstrated that distal sites within a protein can experience a dynamic response to a perturbation, even when the overall structure of the molecule does not appear to be significantly affected. NMR experiments monitoring dynamic responses to ligand binding have demonstrated long-range effects in many cases, including calmodulin (Lee et al., 2000) and Cdc42 (Loh et al., 2001), as well as isolated PDZ

(Fuentes et al., 2004) and SH2 (Finerty et al., 2002) domains. The reactions have included residues as far as 21 Å from the binding pocket, even though these proteins are not allosteric systems in the classical sense. Mutational experiments have demonstrated similar effects. Dynamics studies in protein L indicated significant changes in slow side-chain dynamics at sites distant from point mutations (Millet et al., 2003). An R→A mutation in chymotrypsin inhibitor 2 (CI-2) caused a dynamic change at a tryptophan 13 Å away (Leatherbarrow and Matthews, 1992). The experimental results are supported by theoretical work; molecular dynamics experiments have also demonstrated distal dynamic effects in response to perturbations (Ceruso et al., 2003; Ichiye and Karplus, 1991). In sum, these observations indicate that long-range dynamic responses should be expected, not surprising, when a protein is perturbed.

In some cases, the phenomena of energetic propagation and dynamic response have similar properties. The energetic networks within proteins are sparse and contiguous (Lockless and Ranganathan, 1999; Suel et al., 2003). This also appears to be true of at least some observed dynamic responses, particularly in the PDZ domain (Fuentes et al., 2004) and also in the case of SH2. These similarities suggest the possibility that the two phenomena are related. *A priori*, dynamic effects reflect the presence of energetic interactions. However, long-range dynamic responses have never been shown to coincide with clear energetic linkages, nor have they been definitively associated with networks within a protein.

Dynamic experiments directed towards the backbone amide groups have almost always shown that the backbone possesses substantial rigidity in folded regions of the protein. The order parameter S^2, which describes the rigidity of a vector and ranges from 0 (isotropic flexibility) to 1 (rigidity) (Lipari and Szabo, 1982), is typically around 0.8 for most backbone

N-H vectors. The main chain alone, therefore, is unlikely to provide an efficient means for transmitting motion within a protein. By contrast, the order parameter of the methyl 3-fold symmetry axis (S^2_{axis}), which has a range and meaning identical to the normal S^2, varies considerably within a given folded protein for alanines, valines, leucines, isoleucines, methionines, and threonines (Constantine et al., 1998; Lee et al., 1999; Mittermaier et al., 1999; Mittermaier et al., 2003). These order parameters have been shown to range from near 0 to more than 0.9, even among residues that are packed into a hydrophobic core (Constantine et al., 1998; Johnson and Handel, 1999; Mittermaier et al., 1999; Wand et al., 1996; Wand, 2001). In order to accommodate this wide range of motion in the comparatively rigid framework of the backbone it is likely that side chain movements are correlated. Molecular dynamics simulations and normal mode analysis support the idea of correlated motions as a means for dynamically linking distal sites (Ceruso et al., 2003; Radkiewicz and Brooks, 2000; Showalter and Hall, 2002; Young et al., 2001). This might provide a mechanical means for energy transmission.

Though the wide range of values suggests that side chains may serve as a conduit for dynamic transmission, the nature of the order parameter makes inferences from a single observation very difficult. The S^2_{axis} value correlates best with distance from the backbone: as the number of bonds between the methyl and the Cα decreases, the order parameter increases (Mittermaier et al., 1999). Other structural parameters, such as accessible surface area and packing density, correlate only weakly to the order parameter. Additionally, because the order parameter does not provide information on the principal axis of a given fluctuation, it cannot be used on its own to identify correlated motions (Lipari and Szabo, 1982). If, however, the protein is disturbed in a localized manner, the distribution of

resulting dynamics changes may reveal the pathways by which motions and energy are dispersed.

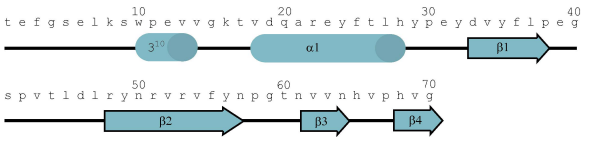

Figure 1: Sequence and secondary structure of WT* eglin c.

In order to investigate the propagation of dynamic effects in proteins, this study employs a novel *dynamic perturbation-response* experiment. Changes in ps-ns side-chain dynamics are monitored in response to a localized perturbation. In this case the perturbation is mutational in nature, but the technique can also make use of ligand binding (Fuentes et al., 2004) or even chemical modification, if appropriate. If the mutation does not provoke a significant change in the structure, one can conclude that the residues experiencing altered dynamics belong to a *dynamic network* involving the mutation site. One can then make use of these networks to test the proposition that dynamic interactions are associated with energetic linkages, using methods already established (Edgell et al., 2003).

The protein investigated in this study was eglin c, a serine protease inhibitor derived from the leech *Hirudo medicinalis* (Seemuller et al., 1977). Like chymotrypsin inhibitor 2, eglin c belongs to the potato inhibitor 1 family, and is atypical for the family in that its sequence contains a large number of aromatic amino acids, more than twice as many as most other members of the family (Figure 1). The majority of methyl-bearing residues in the core are valines; most leucines are at least partially solvent-exposed, and there are no isoleucines or methionines. Nonetheless, due to its small size, the limited number of methyl side chains still samples a majority of the hydrophobic core (Figure 2).

The α-helix of eglin c is packed against a central β-sheet composed of 3 strands (Figure 2). The first and third strands are linked by a highly flexible loop that is bound by eglin c's protease target (Bode et al., 1987). In NMR spectra, the intensities of amide peaks for the residues in this region are attenuated by intermediate chemical exchange processes at 37 ºC; as the temperature is decreased, the peak intensity recovers, probably due to a shift into the slow exchange regime, which may also account for the observation of additional small peaks at this temperature.

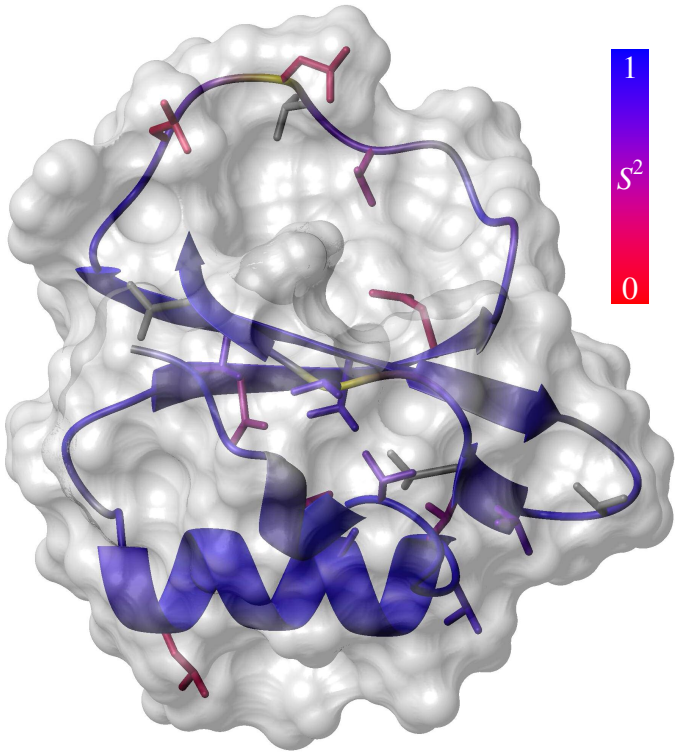

Figure 2: Structure of eglin c (from PDB file 1cse) (Bode et al., 1987) displaying molecular surface and WT* order parameters. S^2 and S^2_{axis} range from 1 (blue) to 0 (red). Yellow backbone segments indicate residues made completely invisible by intermediate exchange; dynamics could not be determined for residues shown in gray. All structure figures were created using MOLMOL (Koradi et al., 1996).

Like most proteins, eglin c has a fairly rigid backbone throughout most of the packed regions (Figure 2). The exposed loop, however, has a great deal of flexibility and possesses complicated dynamics. Despite the widespread rigidity of the backbone, there is a broad range of flexibility for the side chains (Figure 2), although this is again typical for a folded protein. While on average buried residues are more rigid than those that are solvent exposed, some buried residues are more flexible than some exposed residues.

This study primarily addresses the mutational response at 8 sites within eglin c (Figure 3). Each of the residues pictured was mutated to an alanine in order to perform a mutational perturbation-response experiment, except L27, for which only the energetics were analyzed. In addition to an alanine mutation, a less disruptive glutamine mutation was made at R53 in order to investigate the effect of side-chain bulk. The initial mutations (V14A, V54A) were guided by chance and the simple idea of ablating some buried methyls. Subsequent mutations at methylated sites were made on the basis of observations in the initial cases. The mutation to R53 was made on the basis of the previous work in CI-2 (Leatherbarrow and Matthews, 1992).

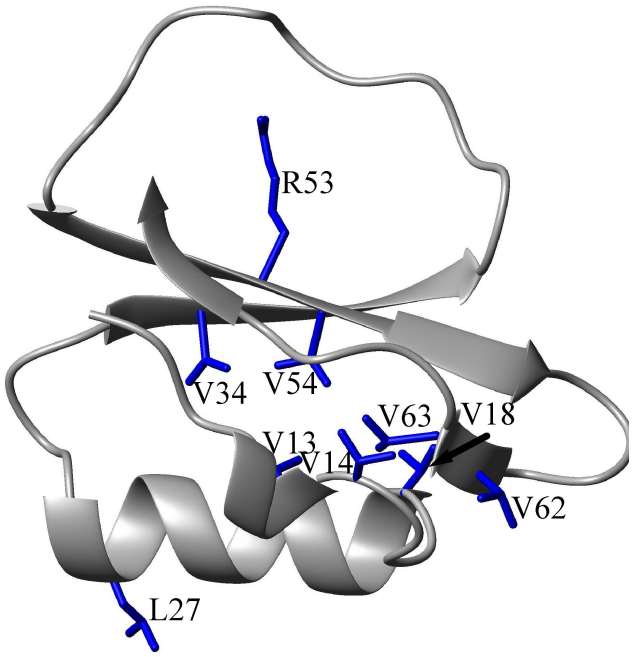

Figure 3: Structure of eglin c with mutation sites highlighted.

In all of these cases the protein bore a F→W mutation at position 10, in order to facilitate detection in chemical denaturation experiments. This mutation preserves aromatic character at residue 10; additionally, all other potato inhibitor I family members thus far identified have a tryptophan at this site. To signify the presence of this mutation, proteins bearing it are designated with an asterisk (*).

The results presented here represent a thorough, but not exhaustive, study of large-to-small mutations among both buried and exposed valines in eglin c, and is the largest study to date of the effects of mutations on side-chain dynamics.

CHAPTER I

CONTIGUOUS NETWORK RESPONSES

Introduction

Comparative studies have frequently found that side-chain dynamics change at sites distant from that of a perturbation. In the case of protein L, for instance, the slow dynamics of some side chains were altered at distances more than 10 Å from a mutated phenylalanine (Millet et al., 2003). Despite this and similar observations, the mechanism by which dynamic responses propagate to distal surfaces is poorly understood. However, a mechanistic explanation of this kind is a critical step towards understanding the significance of these interactions.

One picture that comes easily to mind when considering long-range dynamic responses is that of a *path* of responses going from side chain to side chain, from the locus of perturbation to a distal site. The intuitive appeal of this view is matched by its mechanistic plausibility; side chains can be quite flexible, offering a range of motions that can respond to a change in structure or environment. The coexistence of flexible side chains and a rigid backbone also speaks to the possibility of *correlated motions*, which would provide a convenient means of dispersion for dynamic responses. The simplicity of this view, and its similarities to macroscopic means of force transduction, makes it an attractive first guess about the means of dynamic response.

Several mutational perturbations of eglin c produce responses that at least resemble a pathway. Among these are four – V13A, V14A, V34A, and V54A – that have minimal

effects on structure and backbone dynamics, but significant effects on the side-chain dynamics of the protein. The dynamic responses generally appear to form a sparse network of residues which are in contact with each other. These "contiguous network responses" appear to constitute a significant mechanism for distributing the dynamic effects of mutations through eglin c.

Methods

Expression and Purification

WT* and mutant forms of eglin c were expressed using a pET28a (Novagen) plasmid in *Escherichia coli* BL21(DE3) cells. The bacteria were grown in LB for unlabeled preparations, and in M9 minimal media containing $^{15}NH_4Cl$ and/or [U-^{13}C]D- Glucose (Cambridge) as the sole nitrogen or carbon source to achieve the appropriate labeling. Protein labeled 50% with ^{2}H was prepared using M9 with 60% D_2O.

For all labeling schemes, bacteria were grown at 37 ºC to an OD_{600} of 0.6-0.8 and then induced for 4 hours with 0.5 mM IPTG at the same temperature. The sole exception to this rule was V63A* eglin c, which was induced at 22 ºC for 4 hours. Following the induction period, cells were pelleted by centrifugation and resuspended in 20 mM Tris pH 8.0, then subjected to repeated freeze-thaw cycles. These slurries were then sonicated for a minimum of 3 cycles of 5 minutes in length. The lysates were then centrifuged for 30 minutes at 23,000 *g*. The pellets were discarded, and the supernatants were incubated for 15 minutes on ice with 0.02% polyethyleneimine (Sigma) to precipitate DNA. These suspensions were then centrifuged again with the same speed and duration. The pellets were discarded, and the supernatants were dialyzed overnight against 2 L of 20 mM Tris, pH 8.0.

Following dialysis, the lysates were fractionated over a fast-flow Q sepharose column (Pharmacia), eluting with a linear salt gradient from 0 M to 2 M NaCl (eglin eluted at ~150 mM NaCl). Fractions containing eglin c were then concentrated and further purified using a G-50 Sephadex (Pharmacia) column. The purity of the eluates was assessed by SDS-Page. Pure protein was then concentrated and, if necessary, dialyzed into NMR buffer [20 mM KH_2PO_4, 50 mM KCl, 0.02% NaN_3, 10% D_2O, pH 7.0]. The final concentration of eglin c in samples used for relaxation studies was 2.0 mM unless otherwise noted.

Some double mutants were expressed using an alternate vector containing a His_6 tag. In these cases, the polyethyleneimine precipitation and subsequent steps were skipped, and the raw lysate was purified over a Hi-Trap His column (Pharmacia), eluting with an imidazole gradient. Protein samples were subsequently concentrated and dialyzed into NMR buffer for folding studies.

NMR Data Acquisition and Processing

NMR spectra were acquired using 500 and 600 MHz Varian INOVA spectrometers equipped with triple-resonance probes. Relaxation and assignment spectra were acquired at 37 ºC (temperature calibrated with ethylene glycol) or 25 ºC (temperature calibrated with methanol), as noted. RDC experiments were carried out at 20 ºC, but temperatures were generally not calibrated. Spectra were processed using NMRPipe (Delaglio et al., 1995) and analyzed with the assistance of NMRView (Johnson and Blevins, 1994).

Backbone and methyl side-chain assignments used for this study were determined using gradient-enhanced HNCACB, CBCA(CO)NH, HNCO (Muhandiram and Kay, 1994), and $HCCH_3$-TOCSY (Uhrin et al., 2000) experiments on samples uniformly labeled with ^{15}N and

^{13}C. Stereospecific methyl assignments were determined for most proteins using a constant-time carbon HSQC on a sample fractionally labeled with ^{13}C (Neri et al., 1989). Once these had been performed for numerous mutants, it became apparent that prochiral assignments could be made on the basis of chemical shift relationships alone, and this method was used for subsequent mutants. Assignments of each mutant at 20 ºC were determined by tracing peak motions through a series of ^{15}N HSQCs at different temperatures between 37 ºC and 20 ºC. This method was also used to determine amide and methyl assignments of WT* at 25 ºC.

Chemical shift values for the amide nitrogen and amide proton were compared individually between the WT* and each mutant. The resulting differences were then combined using the formula:

$$\Delta\delta_{vector} = \sqrt{\Delta\delta_H{}^2 + (0.2 * \Delta\delta_N)^2}$$

where $\Delta\delta_H$ is the change in chemical shift (in ppm) of the proton, and $\Delta\delta_N$ is the change in chemical shift of the amide nitrogen. A trimmed mean of these vectors was constructed by excluding the highest 10% of changes. Chemical shift changes from V63A* were excluded from the trimmed mean due to the temperature difference. "Significant" changes in chemical shift were defined as those changes greater than the trimmed average (0.059) plus twice the standard deviation of the trimmed set (0.051). Therefore, only $\Delta\delta_{vector}$ values greater than 0.16 were considered significant.

One bond ^{15}N-^{1}H RDCs (Tjandra and Bax, 1997) were obtained using 1 mM ^{15}N-labeled eglin c in C12E5:hexanol bicelles (Ruckert and Otting, 2000). Concentrated eglin c was added to a solution of C12E5 in water, and hexanol was added dropwise until the solution became transparent and trapped bubbles. Alignment was confirmed via direct observation of quadrupolar splitting of the D_2O signal. Couplings were determined using HSQC-IPAP

experiments (Ottiger et al., 1998) performed at 500 MHz on a sample at 20 ºC in the presence or absence of bicelle material. Splittings were measured using NMRView (Johnson and Blevins, 1994) peaklists. Linear correlations and residuals were calculated using EXCEL.

Rotamer populations were determined on the basis of 3-bond J_{NC} and J_{CC} couplings for valines and threonines. Couplings were measured using a quantitative *J*-correlation method initially described by the Bax group (Bax et al., 1994; Vuister et al., 1993), in which the intensities of two spectra, one of which includes a *J* coupling evolution, are compared. These couplings were then analyzed after the method of Hennig (Hennig et al., 1999):

$$^3J_{exp}(C\gamma N) = p_{180}{}^3J_{trans}(C\gamma N) + (1 - p_{180})^3J_{gauche}(C\gamma N)$$

$$^3J_{exp}(C\gamma CO) = p_{-60}{}^3J_{trans}(C\gamma CO) + (1 - p_{-60})^3J_{gauche}(C\gamma CO)$$

$$p_{60} = 1 - p_{180} - p_{-60}$$

where p_{60}, p_{-60}, and p_{180} correspond to the populations of each rotamer, $^3J_{exp}$ is the experimentally observed coupling, and $^3J_{trans}$ and $^3J_{gauche}$ are the coupling constants expected for fully-occupied rotamers. Following previous work (Schnell et al., 2004), $^3J_{trans}$ was set to 2.1 Hz and 3.6 Hz for NC and CC couplings in valines, and 1.9 Hz and 3.4 Hz for threonines. Similarly, $^3J_{gauche}$ was set to 0.6 Hz, 0.4 Hz, 0.4 Hz, and 0.2 Hz for NC and CC couplings in valines and threonines, respectively. For valines, results from the two methyls were averaged by assigning the γ_2 results to a rotamer shifted 120º from the formula value (i.e. the J_{NC} experiment determined the -60º rotamer). Spectra were analyzed using NMRView (Johnson and Blevins, 1994), and calculations were performed using EXCEL.

NMR Relaxation Studies

Spin-relaxation studies of backbone amides were performed on samples uniformly labeled only with ^{15}N at both spectrometer fields. Standard experiments (Farrow et al., 1994) were used with minor modifications (Hu et al., 2003). ^{15}N T_1 data were collected with relaxation delays of 39, 109, 194, 299, 414, 544, 689, 839, and 1004 ms. ^{15}N T_2 data were collected with relaxation delays of 8, 16, 24, 40, 63, 79, 95, 111, and 127 ms. In some cases, the T_2 experiment was replaced by a $T_{1\rho}$ experiment with relaxation delays of 5, 15, 25, 40, 60, 80, 105, and 135 ms at both fields, and spin-lock frequencies of 2518 Hz at 600 MHz and 2710 Hz at 500 MHz. Following correction for off-resonance T_1 relaxation, the difference between transverse relaxation rates derived from T_2 and $T_{1\rho}$ experiments was negligible. Errors were assessed by comparing the underlined timepoints above with duplicates taken at the end of the experiment. {^{1}H}-^{15}N NOE data were collected with a saturation period of 4.5 s, and errors were assessed from baseline noise. Peak intensities were extracted using NMRView, visually inspected using GNUPLOT, and then fitted to single-exponential decays using a Levinthal-Marquardt algorithm by the in-house program expfit2.

Once relaxation rates were determined, ^{15}N data were fit by the program relxn2.1 (Lee et al., 1999) to a spectral density function, and then to the simple model-free formalism (Lipari and Szabo, 1982):

$$J(\omega) = \frac{2}{5}\left(\frac{S^2 \tau_m}{1+(\omega\tau_m)^2} + \frac{(1-S^2)\tau}{1+(\omega\tau)^2} \right)$$

where $\tau^{-1} = \tau_m^{-1} + \tau_e^{-1}$. In this equation, τ_m represents the global correlation time, S^2 is the order parameter, representing the degree to which the amplitude of an internal fluctuation is restricted, and τ_e is the effective correlation time for the internal fluctuation. The global

correlation time τ_m was fit first by a global minimization for all rigid residues. Flexible residues were identified for exclusion from the global fit using a previously described method (Tjandra et al., 1995). Once τ_m had been determined, the internal parameters S^2 and τ_e were fit individually for each residue. Although a model-selection protocol based on information criteria (Chen et al., 2004; d'Auvergne and Gooley, 2003) identified several residues as experiencing improved fits with alternate formulations of the spectral density, this improvement was marginal, and calculated R_{EX} values were very small except in the case of V13. For this reason, all comparisons use the simple model-free formulation.

In order to determine rotational diffusion tensors for eglin c mutants, a second fit was performed in which τ_m was fit individually for every residue (Lee et al., 1997). These data were then analyzed in concert with solution structures from 1EGL (Hyberts et al., 1992) to determine the angular dependence of the locally-fit global correlation time using the in-house program Qfit. Because no significant anisotropy was noted, no attempt was made to recalculate order parameters on the basis of an alternate diffusion tensor.

Side-chain dynamics were assessed using samples uniformly labeled with ^{13}C and randomly 50% labeled with ^{2}H. ^{2}H spin relaxation data were collected at both 500 and 600 MHz, filtered for CH_2D methyl isotopomers using experiments previously described (Muhandiram et al., 1995). $I_zC_zD_z$ delay times were 3.0, 7.4, 13.1, 19.9, 27.5, 34.0, 45.2, 55.0, and 65.6 ms. $I_zC_zD_y$ delay times were 1.1, 3.2, 5.8, 8.9, 12.4, 16.3, 20.5, 25.1, and 30.0 ms. I_zC_z values were collected with delays of 12, 20, 28, 36, 44, 52, 60, 68, and 76 ms. Uniform uncertainties for all peak intensities were assessed on the basis of duplicate measurements (underlined above). Relaxation times were fit using the same procedure as for backbone NH groups, and the pure D_z and D_y time constants were obtained by subtraction of

I_zC_z rates. Relaxation data were best-fitted to the standard model-free formalism, using the τ_m derived from the analysis of ^{15}N relaxation data. The S^2 values were divided by a factor of 0.111 in order to correct for methyl rotation, yielding S^2_{axis}.

Side-chain methyls for which changes in model-free parameters were greater than twice the standard deviation of the Monte Carlo simulation results (2σ) were identified as having changed dynamics. The application of more stringent standards (such as 3σ, or a 10% change) somewhat diminishes the traceability of responses discussed below, but does not substantially alter most conclusions about long-range responses in these mutants.

Based on previous work (Chou et al., 2003; Hu et al., 2005), it is known that the S^2_{axis} value to some extent reflects the degree to which different rotamers are populated, although it cannot determine specifically *which* rotamer is most highly populated. Therefore, rotamer populations for valines and threonines were calculated from the observed S^2_{axis} values using (Hu et al., 2005):

$$p_{major} = \sqrt{\frac{S^2_{axis} - 0.29}{2.19}} + 0.5$$

where p_{major} is the percentage to which the primary rotamer for the side chain is populated.

Chemical Denaturation Studies

Thermodynamic energies of unfolding were determined using a protocol described previously (Edgell et al., 2003). Briefly, a solution of ~5 μM eglin c in standard NMR buffer (without D_2O), pH 7.0 was titrated using an Aviv auto-titrating fluorometer with a solution containing an equal concentration of eglin c and 6.8 M guanidinium hydrochloride (Sigma). The temperature in the cell was held at 25 ºC, and the samples were stirred for 1 minute

before exciting fluorescence at W10 with 290 nm light and monitoring emission at 350 nm. Using Sigma Plot, the fluorescence emission curves were fit to (Edgell et al., 2003):

$$F_{obs} = \frac{(f_N + s_N[\text{GuHCl}]) + (f_D + s_D[\text{GuHCl}])\exp\left[\frac{m_{N-D}([\text{GuHCl}] - [\text{GuHCl}]_{1/2})}{RT}\right]}{1 + \exp\left[\frac{m_{N-D}([\text{GuHCl}] - [\text{GuHCl}]_{1/2})}{RT}\right]}$$

where F_{obs} is the observed fluorescence, f_N and f_D represent the fluorescence signal from the native and denatured states, respectively, and s_N and s_D represent the slope of the native and denatured state responses to additional GuHCl. R is the gas constant (1.987 cal/mol) and T is the absolute temperature of the measurements (298 K). The change in free energy with respect to the change in GuHCl concentration is represented by m_{N-D}, while the concentration of GuHCl at which half of the protein molecules are unfolded is represented by $[\text{GuHCl}]_{1/2}$; the ΔG_u for each mutant is calculated by multiplying these two variables. Error was calculated on the basis of repeated measurements; because not all proteins were subjected to the same number of attempts, the standard deviations were averaged over all mutants in order to estimate error.

The $\Delta\Delta\Delta G_i$ values for double mutants were calculated using:

$$\Delta\Delta\Delta G_i = \Delta G_{U,WT*} + \Delta G_{U,AB*} - \Delta G_{U,A*} - \Delta G_{U,B*}$$

where A*, B*, and AB* represent the two single and one double mutant, respectively. The average error in the individual measurements was 0.093 kcal/mol, so the overall error was assumed to be 0.18 kcal/mol for $\Delta\Delta\Delta G_i$. This is essentially the same as the 0.2 kcal/mol error reported in previous mutant-cycle experiments by Stites (Chen and Stites, 2001). In that case, and in a previous paper by Shortle (Green and Shortle, 1993) the standard of 0.3

kcal/mol was applied to determine whether energetic effects were additive; that standard is extended here.

Results and Discussion

Structural Effects

With any mutation it is reasonable to ask whether the structure has been perturbed. In this case, residual dipolar couplings (RDCs) provide a convenient means for assessing the degree of structural perturbation. The dipolar coupling between two nuclei depends on the distance between them and the angle between the line that connects them and the magnetic field vector. In solution, proteins tumble isotropically with respect to the magnetic field vector, and therefore this coupling generally averages to zero (Tjandra and Bax, 1997). In the presence of a liquid crystal medium (in this case polar bicelles) (Ruckert and Otting, 2000) that aligns parallel or perpendicular to the magnetic field, a fraction of this coupling is recovered. These *residual* dipolar couplings provide information about the shape of a molecule and the orientation of a bond within the molecular frame of reference. As such, if the fold of the molecule has changed, the RDCs of the relevant N-H bonds should be affected. This approach has advantages in that the data can be acquired rapidly and the analysis is not computationally intensive.

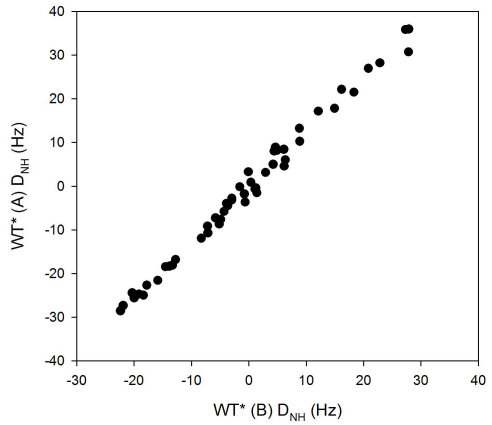

Figure 4: Comparison of RDC values between different preparations of WT* protein and bicelles. R^2=0.99

Because each bicelle solution is different, direct numerical comparisons yield inconsistent results. When two different preps of WT* eglin c are compared, for instance, among RDCs

with a magnitude greater than 5 Hz the average change in magnitude is 30%. Although this numerical comparison seems highly unfavorable, a graphical comparison (Figure 4) is more informative. The two datasets have a tight linear relationship, clearly indicating that the structures are equivalent, at least in terms of backbone N-H vectors. The residuals for the linear relationship are small and random, and average to zero.

Comparisons of RDCs from the four mutants to wild-type are of similarly high quality (Figure 5). Each mutant is shown in comparison to the WT* dataset it most closely matched on the presumption that the bicelle solutions were most similar in that case. Switching the comparison to the other dataset does not cause a large change to the correlation coefficient,

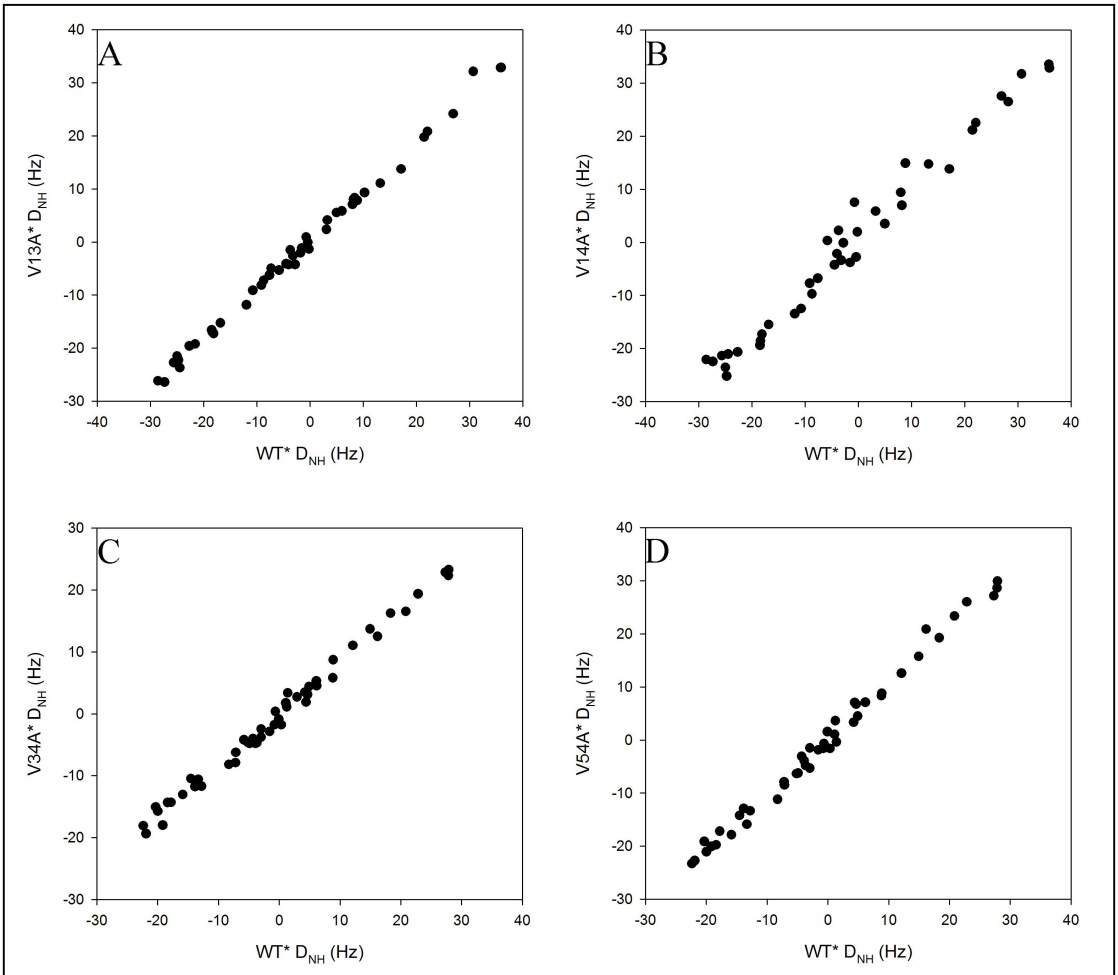

Figure 5: Direct comparisons of RDCs from V13A* (A), V14A* (B), V34A* (C), and V54A* (D) eglin c to WT*. Each linear relationship has an R^2 of 0.99 except V14A*, which has R^2=0.98

or alter the conclusions with respect to the data reported in Chapter 2. The correlation coefficients for plots V13A* (5A), V34A* (5C), and V54A* (5D) RDCs against those of WT* are all 0.99, which is as good as the R^2 when comparing WT* to itself. The exception is V14A* (Figure 5B), which has a correlation coefficient of only 0.98, but even this correlation is excellent.

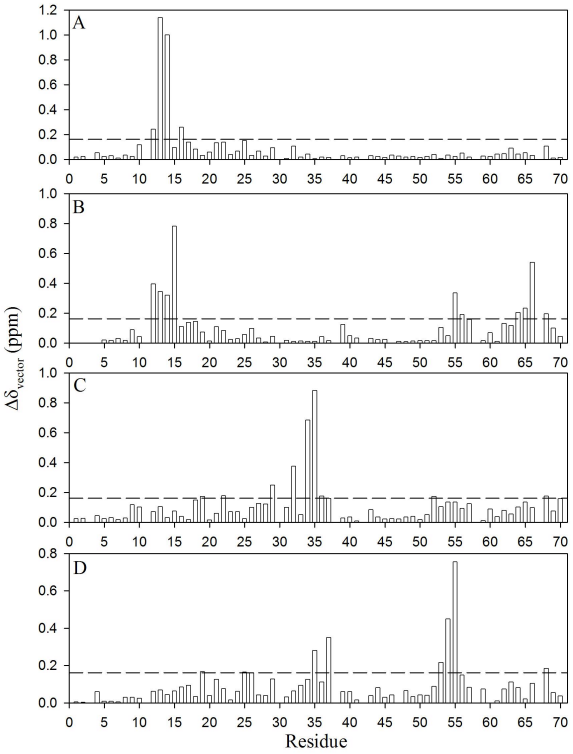

Figure 6: Chemical shift comparisons between WT* and mutant proteins. $\Delta\delta_{vector}$ values are shown for V13A* (A), V14A* (B), V34A* (C), and V54A* (D). Dashed line represents the significance cutoff.

Chemical shift measurements similarly indicate the resilience of the structure; residues with significantly changed chemical shifts are generally restricted to the locale of the mutation (Figure 6). The V14A* and V54A* mutants appear to be exceptions (Figure 6B, 6D), but in these cases the chemical shift changes that appear to be distant in the sequence are actually quite close in the structure. Additionally, none of these mutations alter the secondary structure prediction from chemical shift indexing.

While these data do not rule out the possibility of small, local rearrangements of the side-chains, they firmly establish that the fold has been preserved and that the shape of the molecule is essentially unchanged. This latter contention is also supported by the rotational diffusion anisotropy fits, which indicate near-spherical diffusion tensors for all mutants.

Backbone Dynamics

The small amounts of energy involved in nuclear spin systems make relaxation of these systems by spontaneous emission a very slow process (Cavanagh et al., 1996). Consequently, nuclear spins must return to equilibrium by emitting energy when *stimulated* by nearby fields fluctuating at the appropriate frequency. As a result, the rate of return to equilibrium is sensitive to the motions of a given molecule. In this study, relaxation was primarily studied through two processes: T_1 relaxation, which is the recovery of polarization parallel to the magnetic field vector; and T_2 relaxation, which is the dephasing of polarization in the plane perpendicular to the field vector. Additionally, in the case of the ^{15}N measurements a $\{^1H\}$-^{15}N steady-state NOE was collected, measuring the interdependent relaxation of a proton and nitrogen. The relaxation rates and NOE values acquired for WT* eglin c are shown in Figure 7. Because the relaxation rates are dominated by the overall tumbling of the protein, they are relatively flat across the molecule, but there are notable changes in flexible regions such as the loop and the N-terminus.

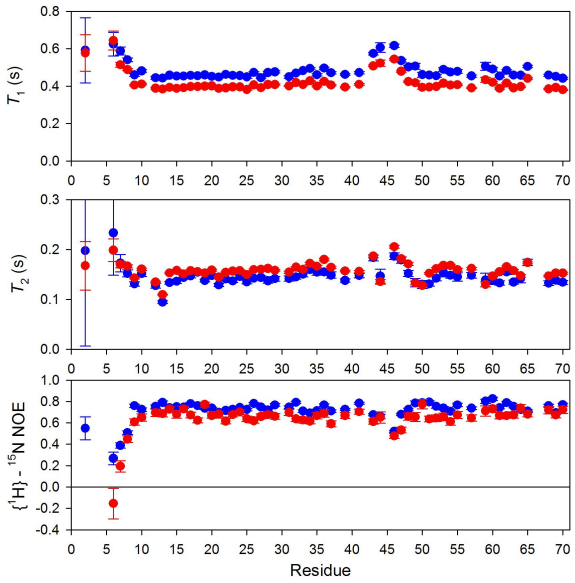

Figure 7: Relaxation data for WT* eglin c. T_1, T_2, and NOE results are shown for 500 (red) and 600 (blue) MHz fields.

All three of these relaxation mechanisms depend on to the *spectral density*, which reflects the power available at a given frequency (Cavanagh et al., 1996; Lipari and Szabo, 1982). The spectral density can in turn be described by three parameters of molecular motion: the global correlation time τ_m, the order parameter S^2, and the effective internal correlation time

τ_e (Lipari and Szabo, 1982). As its name would suggest, the global correlation time describes the rate at which the protein tumbles in solution. This also represents the slowest motion to which actual T_1 and T_2 relaxation rates are sensitive (although the nature of the T_2 experiment means that the measured rate may also reflect slower motions).

The order parameter S^2 can range from 0 to 1 and reflects the degree of restriction of an internal vector. A bond vector for which $S^2 = 1$ is perfectly rigid with respect to the molecular frame of reference; a bond for which $S^2 = 0$ is isotropically flexible with respect to the rest of the molecule (Lipari and Szabo, 1982). Because methyl groups rotate rapidly about the carbon-carbon axis, their C-H bonds naturally have very low order parameters. Dividing the S^2 by 0.111 corrects for this rotation effect and yields the order parameter for the axial bond (S^2_{axis}) (Lipari and Szabo, 1982; Muhandiram et al., 1993).

In WT* eglin c, as in most proteins, the backbone is relatively rigid throughout most of the molecule (Figure 8). The exceptions, as already indicated by the raw relaxation data, are the loop region (residues 39-49) and the N-terminus. In contrast, the side chains appear to possess a wide range of flexibilities, ranging from an S^2_{axis} as low as 0.2 at L27 to a high of 0.8 for A21. In general, solvent-exposed residues possess the lowest order parameters, but

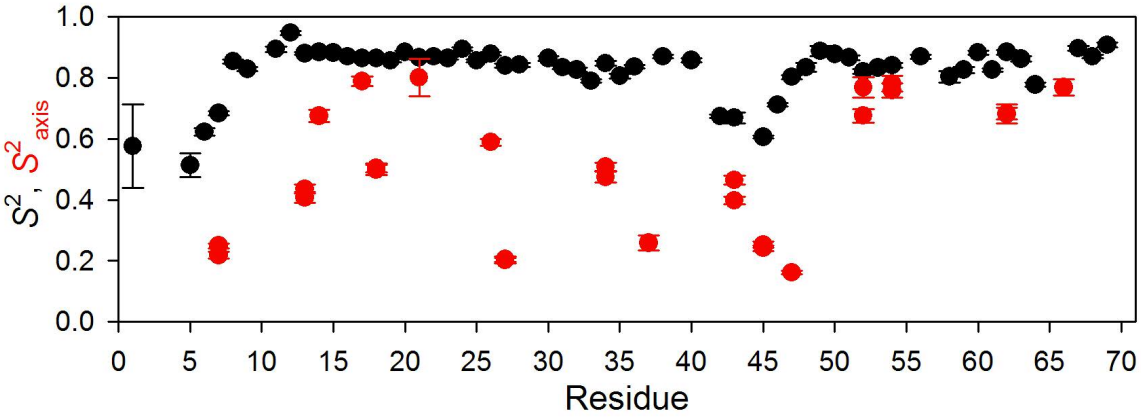

Figure 8: Backbone (black) and side-chain (red) order parameters of WT* eglin c.

this is not true in all cases. T26, for instance, has a higher side-chain order parameter than V18, even though V18 is buried in the core.

The effective internal correlation time τ_e is not fit robustly for backbone amides due to their rigidity and the lower Larmor frequency of ^{15}N; therefore no effort has been made to interpret changes in this parameter for N-H groups. In terms of the generalized order

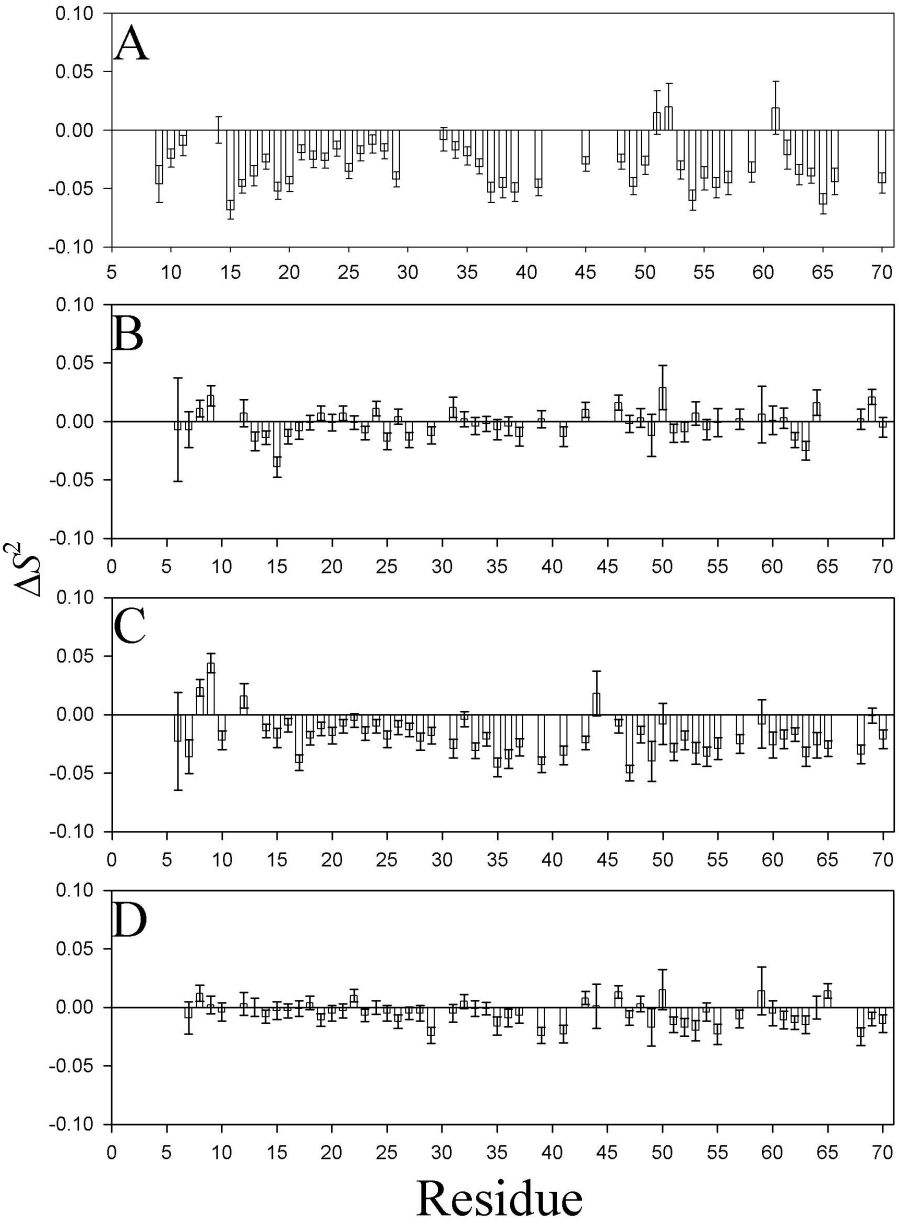

Figure 9: Change in backbone order parameters for eglin mutants. Changes in V13A* (A), V14A* (B), V34A* (C), and V54A* (D) are displayed as (mutant-WT*)

parameter S^2, two kinds of responses to mutation are evident (Figure 9). The first encompasses the V13A* and V34A* mutants (Figure 9A, 9C). In both these cases, it is clear that there is a small, widespread decrease in backbone rigidity on the ps-ns timescale. The average change in S^2 relative to WT* for V13A* is -0.03, and for V34A* it is -0.02, compared to an average order parameter of 0.8 for all residues excluding the N-terminus. The largest changes in S^2 cluster around the mutation site for both proteins. However, none of the order parameters change by 10% or more, suggesting that the perturbation, while global, is not serious. The reduction in order parameters for these mutants is mirrored by the reduction in magnitude of their RDCs relative to wild-type. This latter evidence does not unambiguously prove that the backbone is more flexible, though the similarity of ^{2}H splittings suggests that the media should have produced similar degrees of alignment. However, the reductions in S^2 and RDC magnitude paint a consistent picture of increased backbone flexibility.

In contrast, the backbone S^2 of V14A* and V54A* are very similar to those of WT*, with very few residues worth noting. The average change in order parameter for these proteins is -0.002 (V14A*) and -0.005 (V54A*) relative to averages of about 0.82 for parts of the protein outside the N-terminus. In these cases again the largest changes to the order parameters cluster near the mutation sites. None of the mutations significantly perturbed chemical exchange processes in the loop (residues 39-48) or at V66. A subtle chemical shift exchange process at V13 appears to have been quenched in the V13A mutant, but this is not affected by any other mutations.

For the most part these proteins appear to have similar τ_m values. WT* eglin c has a τ_m of 4.68 ns, which is very similar to 4.61 ns for V34A*, and 4.64 ns for V14A* and V54A*. The

exception here was V13A*, which fit a τ_m of 4.26 ns. It is possible that the decreased τ_m is related to the increase in backbone flexibility, but the changes in S^2 are not an artifact of an erroneously low τ_m fit, as an alternate fit of the data using a longer τ_m showsa *further* decrease in the fitted S^2 values.

The picture that emerges from the structural and dynamic data indicates that by and large the backbone is not significantly perturbed by these mutations. The same can not be said of the side chains. Each of these mutations causes significant changes in side-chain dynamics at sites distant from the mutated residue.

Side-Chain Dynamics

The detection of small changes in side-chain dynamics is almost inevitable in experiments of this kind. In an effort to simplify the analysis, significant changes in side-chain dynamics were defined as those in which ΔS^2_{axis} or $\Delta \tau_e$ was greater than twice the propagated error in the measurements (the error for each measurement is the standard deviation of values from 150 Monte Carlo simulations of the fit). While smaller changes may still be real or important, the 2σ standard identifies residues for which a change can be confidently identified, at least on the basis of the fitted values.

The V13A* mutation causes changes in side-chain dynamics throughout eglin c (Figure 10), and most of the significant changes in S^2_{axis} are concentrated near the mutation site. The order parameter changes significantly for V34 and V54, though the dynamics of V52 are not significantly affected. In addition, the order parameter for T26 changes significantly, as does the τ_e of the adjacent L27. The τ_e effects are generally more widely dispersed throughout the

molecule, encompassing nearly all methyl-bearing groups. Although this response is troublesome, it cannot be dismissed out of hand because τ_e *is* fit robustly for methyl groups.

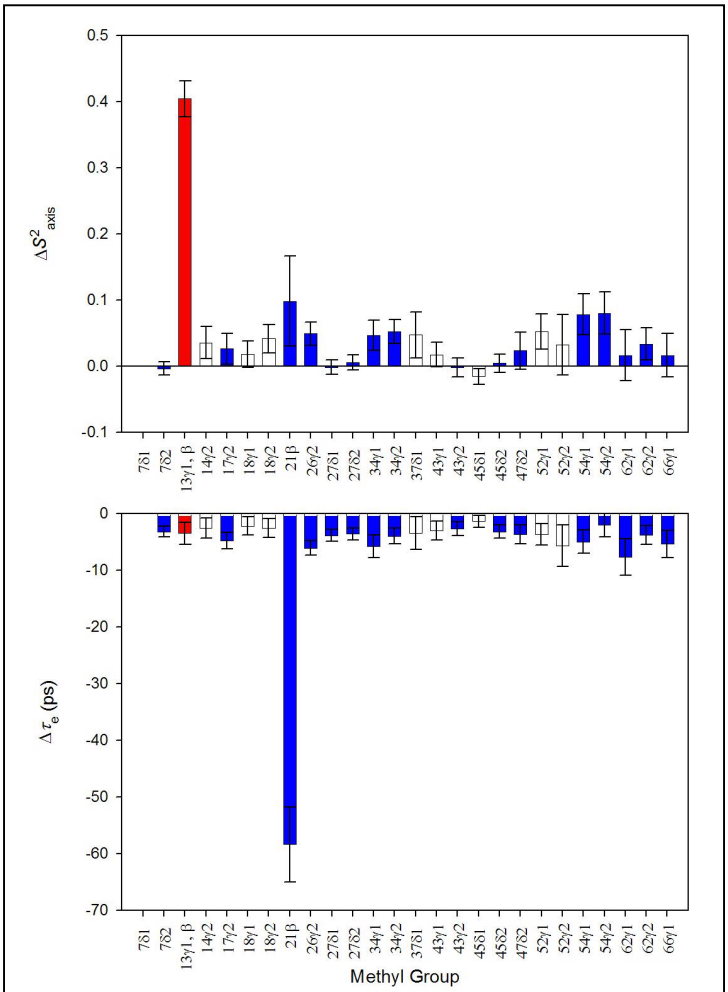

Figure 10: Changes in side-chain order parameters for V13A*.

The τ_e response highlights an important consideration of dynamic experiments, in that one hopes to see a significant number of residues that do not display any response to the mutation (Clarkson and Lee, 2004). When all residues appear to change in a uniform direction, one must be concerned that the fitted τ_m is inaccurate. Even small errors in the correlation time can cause significant changes to the fitted S^2_{axis} values. These changes should be both global and uniform in sign. The absence of this uniform effect suggests but does not guarantee that the fitted τ_m is correct. Similarly, the presence of a global depression of τ_e (as observed in V13A*) does not establish that a τ_m error has been made – it is at least conceivable that all the methyls have accelerated their rotation for some reason. However, taken with the almost uniformly elevated S^2_{axis} values, the τ_e data provide cause for concern.

In terms of the response of the loop residues (V43, L45, L47), L7, and V66, there is additional cause to question the results because these residues are engaged in complicated

dynamic processes. The loop moves fairly freely in solution (Figure 8), and the backbone resonances of numerous residues on it show evidence of intermediate chemical exchange. The backbone amide of V66 is also in chemical exchange, though in this case it's not clear that the exchange is due to any motion on the part of V66. The backbone of L7, in contrast, appears to be in slow exchange, and eglin c does not have a great deal of N-terminal structure. The dynamics at L7 are therefore likely to be just as complex as in the loop. In all of these cases, it is possible that the simple model free dynamics model (S^2, τ_e, τ_m) simply isn't appropriate; the fact that the fits for these residues had χ^2 values above the average of 2.96 (for V13A*) supports this interpretation. In addition, many of these are marginal effects in τ_e, just barely satisfying the condition of significance. Finally, most of these residues are solvent-exposed, which makes changes in methyl rotation rates difficult to explain.

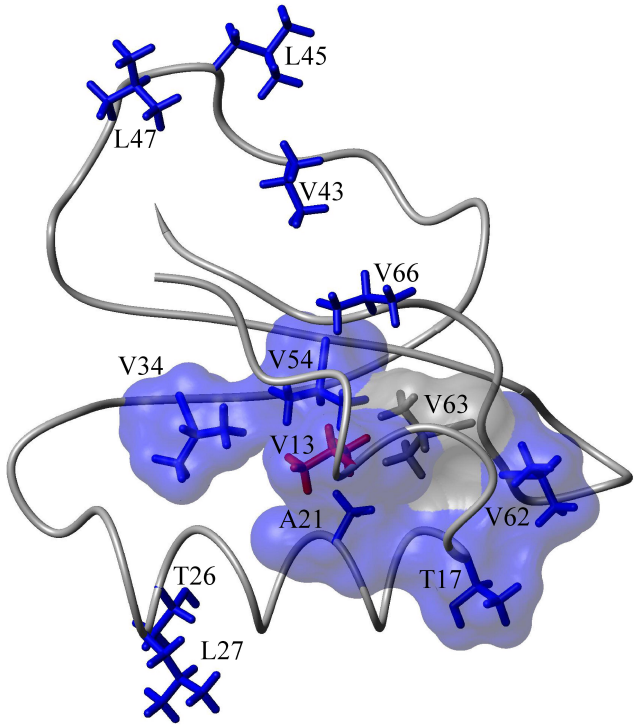

Figure 11: Side chains showing a dynamic response to the V13A mutation. The surface was rendered with a 1.4 Å probe; surfaces associated with V63 are in grey to emphasize the hypothetical nature of its inclusion.

Even rejecting that those five residues experience real dynamic changes, it is clear that the V13A mutation induces dynamic responses at several residues with which it is not in direct contact (Figure 11). This includes changes in S^2_{axis} and τ_e at V34 and T26, and τ_e changes at

L27, V62, and T17. Of residues in direct contact with V13, only V54 and A21 show a dynamic response.

An additional feature of the dynamic response to the V13A mutation is that most of the residues responding to this mutation can be linked into a contiguous surface (Figure 11), provided one inserts a surface for the V63 atoms. This highlights a practical issue in the interpretation of NMR-based dynamic perturbation-response experiments (Clarkson and Lee, 2004). While a large number of side chains can be analyzed in these experiments, certain types of amino acids or groups will escape analysis. In this instance, the dynamics of methylene groups and aromatic side chains were not analyzed. In addition, features of the spectra may prohibit analysis, as in the case of V63. Both V63 methyls are overlapped in the WT* spectrum, precluding any proof of dynamic change in response to the mutation. Nonetheless, the proximity of V63 to several residues experiencing a dynamic response to the V13A mutation suggests the possibility that it too has responded; it is in the hydrophobic core and lies in close proximity to V13, V62, and V54. Given V63's position, it is at least possible that it responds to the V13A mutation and so its surface has been included, with the caveat that there is no direct evidence to support it. The propagation of dynamic effects to T26 and L27 is more difficult to explain. Although V13 and V34 both directly contact the F25 side chain, it is bent away from T26, and there are no unusual backbone perturbations to indicate a disturbance there. This does not rule out propagation via F25, but these two side chains are not included in the surface nonetheless.

The side-chain dynamic response to the V13A mutation has two notable features. First, residues distal from the mutated site clearly experience altered dynamics. Methyl groups at T17, V62, and V34 are all more than 11 Å away from the V13 β-carbon, and if T26 and L27

are considered the distances are even greater. Second, many of the residues with significant changes in their model-free parameters can be assembled into a contiguous surface. These features are not unique to the V13A response.

The methyl dynamics data of V14A* eglin c also reveal a dispersion stretching away from the mutation site through a contiguous network of interacting side chains (Figure 12 & 13). In contrast to the observations in V13A*, the response involves both increases and decreases in side chain flexibility, and also a large number of residues for which no significant dynamic change is identified. A group of residues including V18 and T17 becomes more flexible in response to this mutation, while V13 and V34 become more rigid. In addition, there appears to be a slight, but significant change in τ_e for the pro-S methyl of L45.

Again in this case there is good cause to question the response at L45, not only because it barely meets the standard of significance but also because its dynamics may be too complex for the simple model-free formalism to accurately handle. The dynamic response at V34 is also subtle, but the well-behaved backbone and traceability

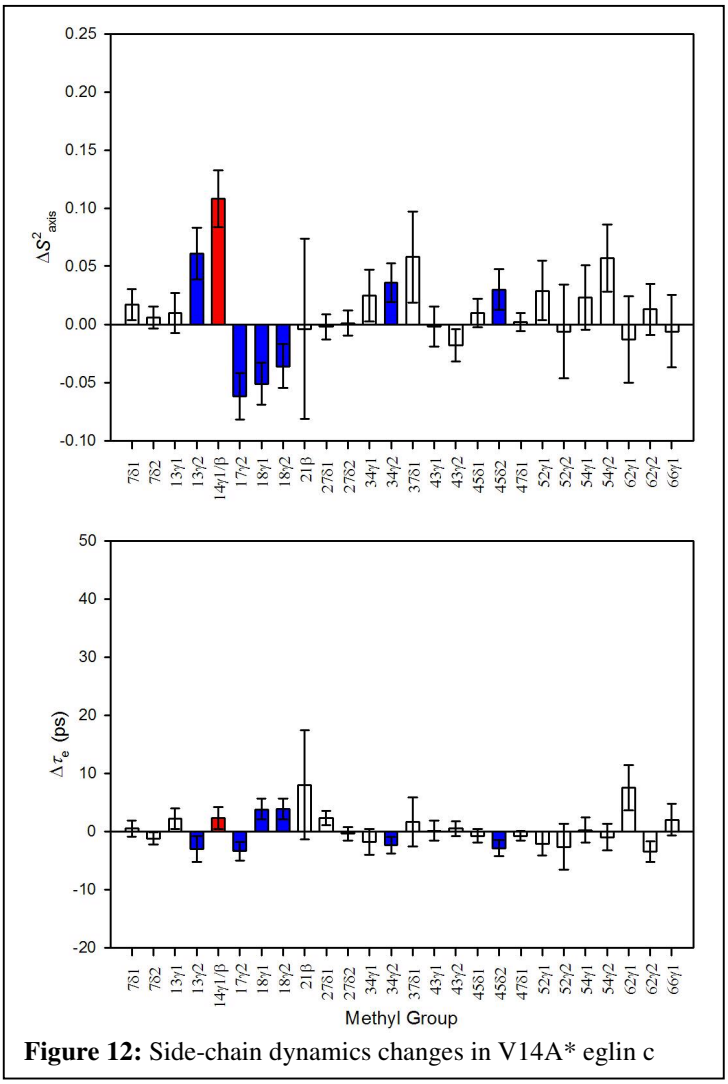

Figure 12: Side-chain dynamics changes in V14A* eglin c

to V14 reinforces the significance of that result.

Regardless of the treatment of L45, the V14A mutation again provokes a clear response at distal residues. The γ-carbons of V18 and T17 lie about 11 Å away from the β-carbon of V14. The methyl carbons of V34 are positioned more than 13 Å away. Additionally, if one takes the liberty of including surfaces for V63 and F25, one can unite every residue with a dynamic response except L45 into a contiguous surface (Figure 13).

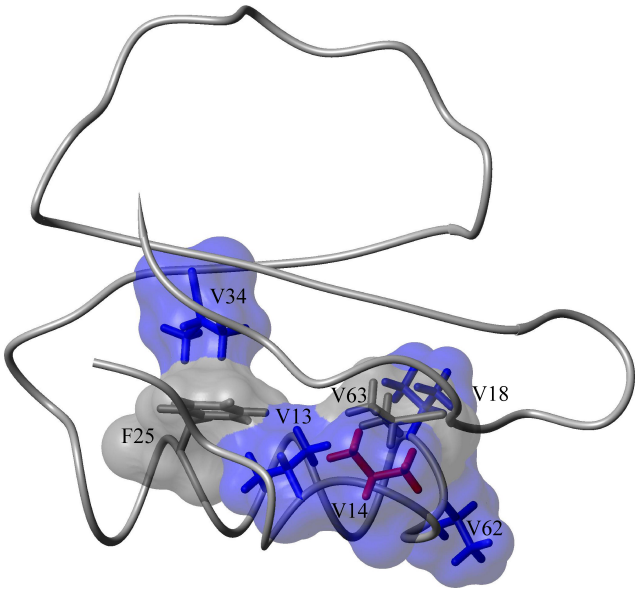

Figure 13: Side chains with a dynamic response to the V14A mutation. Surfaces for F25 and V63 are in gray to emphasize that their membership in the network is hypothetical

The proposed surface appears to have two branches, one of which rigidifies in response to the mutation, the other of which becomes more flexible. The steep drop-off in response between V13 and V34 may be attributable to the fast-timescale inflexibility of F25. Though aromatic rings are known to flip in protein interiors, these motions occur on the sub-millisecond timescale. On the timescale investigated by these experiments – motions of a duration less than τ_m – the bulky, rigid rings probably have low mobility, which may serve to muffle dynamic transmission.

While the V13A and V14A mutations produce clear dynamic responses in distal residues, and though these responses are for the most part larger than those observed on the backbone, one could still argue that they are fairly subtle. For instance, no residue other than the

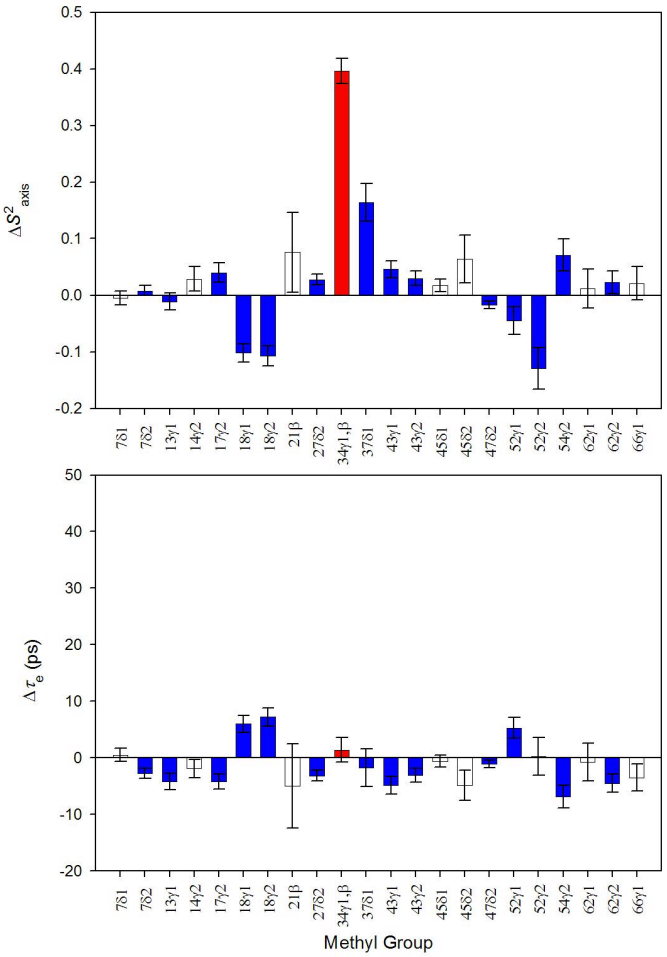

Figure 14: Side-chain dynamics changes in V34A* eglin c. Side chains with significant changes are shown in blue; mutant in red.

mutated site experiences a change greater than 0.1 in S^2_{axis}. In the case of the V34A* mutant, however, a more serious dynamic perturbation appears to take place.

As with V14A, eglin c has a mixed response to the V34A mutation (Figure 14). In this case, however, similar dynamic responses do not co-localize. Areas of the core that become more flexible are separated by stretches of residues that appear to rigidify. The dynamic propagation reaches to a greater distance in this case: the methyls of T17 and V62, for instance, are more than 16 Å away from V34. Additionally, the dynamic perturbation appears to be substantially more severe, not only in terms of S^2_{axis}, but also with respect to side-chain rotamer populations (Table 1).

This large change in rotamer populations for core residues is unique to V34A* eglin c among mutants studied thus far. Though neither methyl's primary rotamer changes as a result of the mutation, the degree to which it is populated shifts dramatically (especially in the case of V52). The population change loosely correlates with a decrease in S^2_{axis}, but for V52 the primary population calculated on the basis of the order parameter is significantly

greater than that calculated on the basis of observed *J* couplings. Because the *J* couplings are sensitive to a much broader timescale than spin-relaxation, this suggests that at least some portion of the observed rotamer averaging is due to motions slower than the overall tumbling of the molecule. In the case of V18, this shift in rotamer populations may represent the effect of a conformational change in F36, which is adjacent to both residues and could shift to fill the new space. In the case of V52 it probably results because that side-chain is directly taking advantage of the hole created by the mutation (see below).

Methyl group	p_{major}, J		p_{major}, S^2	
	WT*	V34A*	WT*	V34A*
V13γ1	0.55	0.51	0.76	0.75
V13γ2	0.55	0.51	0.73	---
V18γ1	0.86	0.68	0.81	0.73
V18γ2	0.86	0.68	0.81	0.71
V52γ1	0.81	0.47	0.92	0.89
V52γ2	0.81	0.47	0.97	0.90
V54γ1	0.91	0.94	0.96	---
V54γ2	0.91	0.94	0.97	1.01
V62γ1	0.83	0.81	0.92	0.93
V62γ2	0.83	0.81	0.92	0.94

Table 1: Major rotamer populations for several side-chains in WT* and V34A* eglin c. Populations were calculated from *J* couplings and S^2_{axis} values.

Curiously, although it contacts both V34 and V52 directly, V54 *rigidifies* in response to the ablation of the V34 methyls. This, however, may be explained by the nature of the population change in V52. In WT* eglin c, the V52 side chain almost exclusively populates the 180º rotamer, putting a single methyl group in contact with both V34 and V54. In the V34A mutant, however, the 60º rotamer is almost equally populated (Figure 15). In this case, the γ_1 methyl appears to swing into the area vacated by the V34 methyls, while the γ_2 methyl takes the place of γ_1. As a result, V54 is now bracketed by 2 methyl groups from the adjacent V52 approximately half the time. This may serve to restrict its available conformational space.

V54 adopts the 180º rotamer 91% of the time in WT*, but the *J* coupling results indicate that this percentage is slightly increased in V34A*. Additionally, the increase in S^2_{axis} for V54 may reflect a narrowing of the well for that particular rotamer.

V34A* eglin c also sees dynamic changes at some residues on the opposite face of the β-sheet, namely L37, and the loop residues V43 and L47. Although the same caveats as before apply to the loop residues, L37 and V43 at least can be linked to the core residues via R53, which lies between the clear responders V52 and V54. As for L47, the effect there is very marginal, as the change in S^2_{axis} is 0.003 above the threshold. It is therefore possible that the response observed there is an artifact.

As with several other mutations, L27 is identified as a responder, though there does not appear to be any direct way to link it back to V34 (Figure 12). The backbone dynamics in that region are also unusually quiescent compared to the rest of the protein, and that end of the helix features some of the best correlations between V34A* and WT* RDCs. L27, in

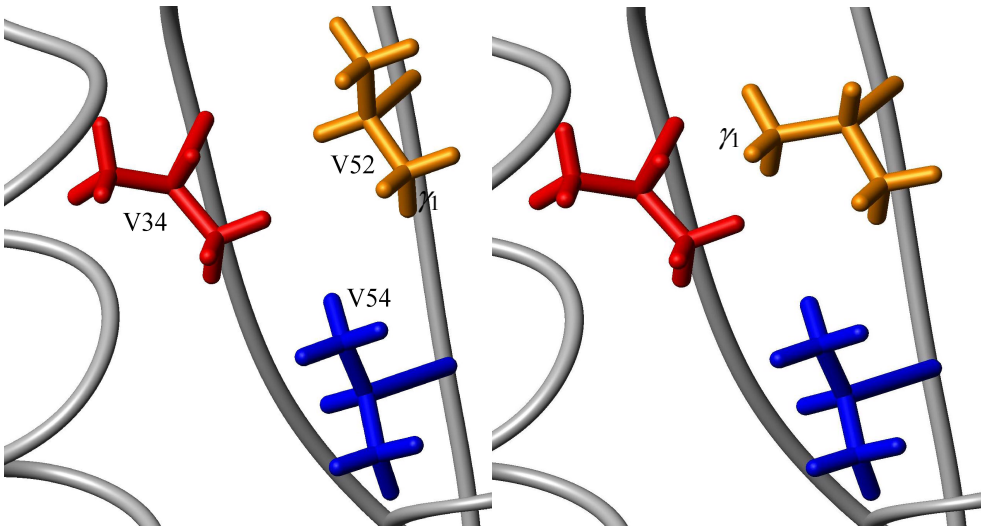

Figure 15: Side-chain dispositions in the core of eglin c. In the WT* protein (left), V52 (orange) primarily populates the 180 rotamer in order to avoid steric clashes with V54 (blue) and V34 (red). In the V34A* protein (right), ablation of the V34 methyls allows V52 to adopt the 60 rotamer, in which the methyls bracket V54γ1.

fact, features the smallest residual observed in the entire protein. However, L27 has some unusual energetic properties (Yi et al., 2003), and its motions may be somehow correlated with those of the loop (J. Fetrow, unpublished data). Thus, there may be some subtle long-range interactions involved. It is also true that the L27 methyls tend to be overlapped slightly in the relaxation spectra, but this occurs in the mutants *and* the wild-type, and is unlikely to give rise to differences between proteins.

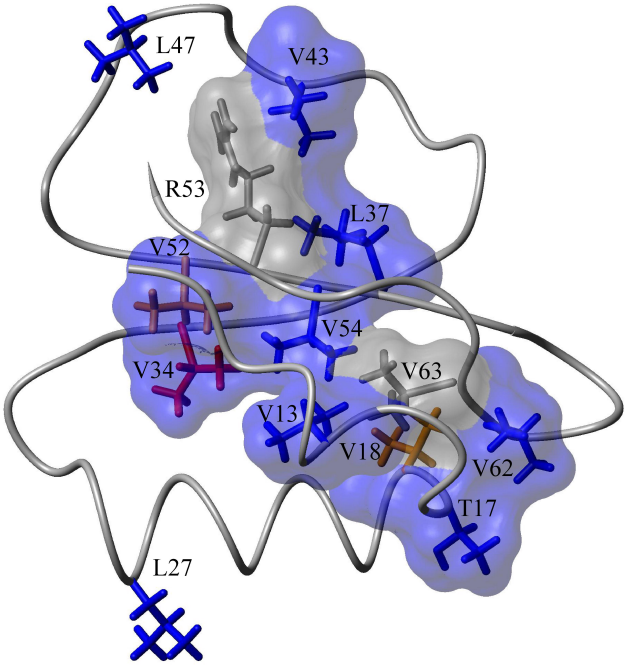

Figure 16: Side chains responding to the V34A mutation. V52 and V18 are shown in orange to emphasize changes in rotamer populations. Surfaces for V63 and R53 are in gray to emphasize the hypothetical nature of their inclusion.

Additionally, it is important to re-emphasize that the fast-timescale dynamics measured in these experiments cover only a fraction of the available range of fluctuations. That a pathway cannot be conveniently drawn encompassing all residues with altered fast-timescale dynamics does not imply that no such pathway could be conceived were the motions of all residues on all timescales known. Dynamic and structural changes below the sensitivity threshold of these experiments may also account for the distribution of effects. The absence of detectable dynamic changes cannot rule out propagation through any particular site.

Nonetheless, with the exception of L27 and L47, the residues experiencing altered dynamics as a result of the V34A mutation can be assembled into a contiguous surface, provided that hypothetical surfaces for R53 and V63 are inserted (Figure 16). Though this

surface does not have the convenient feature that adjacent residues always experience similar changes in dynamics, it at least provides a pathway by which dynamic responders may be linked. Because the example of V52 and V54 showsthat differing dynamic responses in adjacent neighbors can be rationalized in terms of structural principles, this should not be alarming.

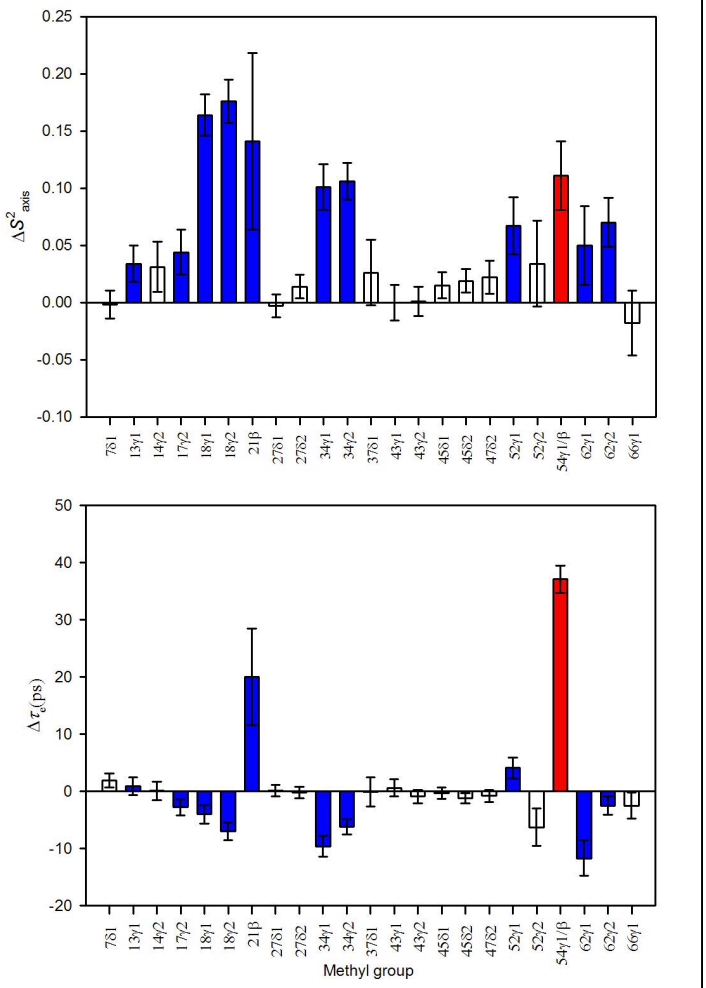

Figure 17: Side-chain dynamics changes in V54A* eglin c. Methyls with significant changes are shown in blue; mutation site in red.

The V54A mutation also provokes a widespread dynamic response involving several large-magnitude changes in S^2_{axis} (Figure 17). Again, many of these responses reach distal residues: the methyl carbons of T17 and V62 are 11 Å away from the β-carbon of V54. Additionally, a contiguous pathway can link all of the response sites, so long as V63 is included in the surface (Figure 18).

Although the magnitude of the response is notable, another key feature of the V54A* mutant is that all the changes appear to involve a rigidification of the methyl groups. In this case, there is no concern for fitting bias as in V13A*, however. The small-magnitude changes in S^2_{axis} that do not meet the 2σ criterion are both positive and negative.

Additionally, the τ_e changes observed include both increases and decreases, suggesting that there is no need to be concerned about τ_m. The rigidification therefore appears to be a real feature of the mutational response, albeit a counterintuitive one. The creation of a hole in the middle of the core would seem to afford more opportunities for conformational flux, not less.

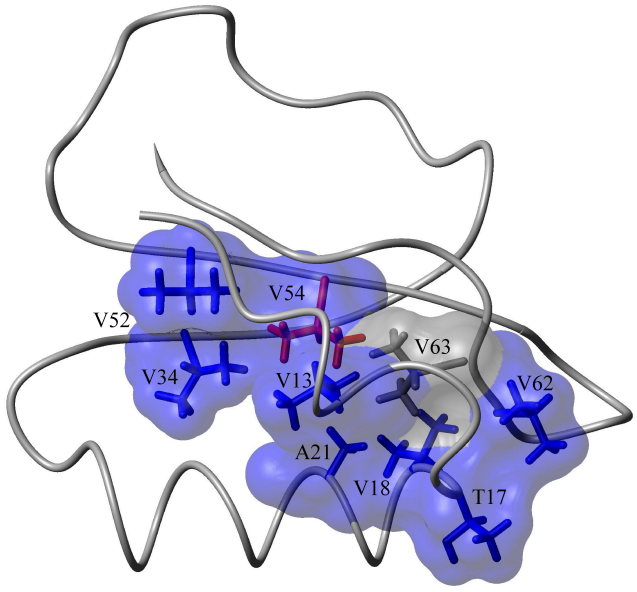

Figure 18: Side chains with a dynamic response to the V54A mutation. The surface for V63 is shown in gray to emphasize the hypothetical nature of its inclusion.

The rigidification may reflect the fact that no fast-moving side chains are appropriately positioned to take advantage of the opportunity created by the ablation of the V54 methyls. In the case of the V34A mutation, V52 can partially fill the new hole simply by adopting an alternate rotamer conformation (Figure 15). This is not possible in the case of the V54A mutation because the methyl groups of V34, V13, and V52 are already as close to the V54 β-carbon as they can be without changing the backbone conformation. Because there is no opportunity to take advantage of the new space, there is no entropic compensation for the loss of enthalpy from van der Waals interactions. This may serve to narrow the energy well and stabilize a single conformation in the core, thus increasing the order parameters.

Clearly, this rigidification does not extend to all timescales; because the stability of the protein is decreased, unfolding is occurring more frequently, implying an increase in backbone flexibility on a timescale of seconds and above. Near the nanosecond timescale, however, it appears that there are no slower compensating motions. Rotamer populations

derived from *J* couplings are nearly identical between WT* and V54A*, showing that there is no motion lasting ms or less that compensates for the increase in rigidity (data not shown). Additionally, because the maximum rotamer population neither increases nor decreases, these data suggest that the increase in order parameter results from decreased fluctuations in particular rotamer states – a narrowing of the energy well for each particular conformation. This is consistent with the picture described above in which side chains sacrifice fluctuations in order to maximize the available hydrophobic contacts.

In a sense, the rotamer transitions suggest that the response to the V54A mutation is more purely dynamic than that to the V34A mutation. In both cases, most side chains in the core rigidify. The only exceptions to this in V34A* eglin c are those residues that are taking advantage of the new pocket directly (V52) or indirectly (V18, via F36). The responses at these residues may be seen as more *structural*, while the rigidifying side chains might be more *dynamic*.

This distinction is somewhat arbitrary. Structure and dynamics are irrevocably intertwined because all proteins possess *dynamic structures*. Although it is simpler to consider a single conformation, the reality for any protein is a temporal or population *ensemble* of structures, of which current techniques identify the lowest-energy (crystallography) or average (NMR) member. To speak of a "purely dynamic" as opposed to a "purely structural" response in some sense misses the point: the structure is what is dynamic. As such, dynamic changes are structural changes. Nonetheless, a distinction *can* be made on the basis of the subtlety of the change. Altered dynamics on the fast timescale, though they unquestionably indicate a conformational shift, reflect differences in environment too small to be reliably detected by

more standard structural means. Dynamics can therefore serve as a sensitive probe for minor, but significant, changes in structure.

An additional feature of interest in the V54A results is that this increase in rigidity does not correlate with an increase in stability. These two concepts are naturally associated (Wray et al., 1999), and it seems rational to suggest that large fluctuations in a structure occur when the interactions of the primary configuration are insufficient to maintain it. Yet in the case of V54A this expectation is not borne out. Though, as mentioned, conformational transitions between the folded and unfolded states have increased in frequency, the fast motions of side chains in the native state have decreased in amplitude. On face this contradiction is perplexing, but a consideration of thermodynamics may serve to explain it. The increase in S^2_{axis} reflects a decrease in conformational entropy, though the possibility of correlated motions makes a straightforward calculation of the magnitude of this effect impossible (Lee et al., 2002; Wand, 2001). Given this loss, the rigidification of V54A* eglin c makes sense as a *cause* of its destabilization. This is consistent with pH experiments on eglin c, in which it was shown that a drop in pH decreased both the global stability and the core methyl flexibility (Hu et al., 2003). In such cases, it may be more appropriate to think of instability in terms of native state fragility rather than structural looseness.

Contiguity of Responses

The four mutants discussed thus far have several unifying features. The first is the preservation of the structure. Although each ablates two methyl groups, RDC and chemical shift analysis do not support any conclusion that the backbone structure has changed significantly from the wild-type. The second feature is dynamic excitation of a network of

side chains that includes residues distant from the mutation site. In each case, dynamic responses were observed beyond the first and second shells of interactions. Additionally, significant changes in both S^2_{axis} and τ_e were observed in each of these mutants. Finally, in each case all, or nearly all, of the side chains where a dynamic response was observed could be assembled into a contiguous surface. These features constitute a specific kind of dynamic reaction to a mutational perturbation: a "contiguous network response".

Several limitations of this analysis are immediately evident. First, while clear changes in model-free parameters create some certainty that a change has occurred, an absence of change in these parameters does not establish that the dynamics are unchanged. Due to their nature, model-free parameters do not include information about the direction of fluctuations. These could conceivably change even when the overall amplitude of fluctuation does not. Additionally, many types of side chains are missed by this analysis, specifically methylenes and aromatic groups. Attempts to measure the dynamics of these groups in the core of eglin c failed in part because their rigidity rendered ^{2}H-based experiments too insensitive, which suggests that their primary flexibility does not lie on the ps-ns timescale. However, this cannot be stated with any certainty. As a result, though the observed networks have the *appearance* of sparseness, they may not actually possess this property.

Secondly, every attempt to create a contiguous surface required the inclusion of hypothetical residues. These consistently included V63, suggesting that it may be a dynamic nexus of this protein. This would seem to indicate that perturbation of this location would cause widespread changes in dynamics.

A V63A mutation, however, does not prompt a significant long-range response (Figure 19). Notable responses occur only at the adjacent residues V14 and V62, and the poorly-modeled V43. Yet it should be noted that the V63A mutation caused a 3.2 kcal/mol decrease in ΔG_u, or more than half of the stability of WT*. The protein could not be studied at the same temperature or concentration as other eglin mutants, and had to be compared to WT* at 25 ºC, 1 mM.

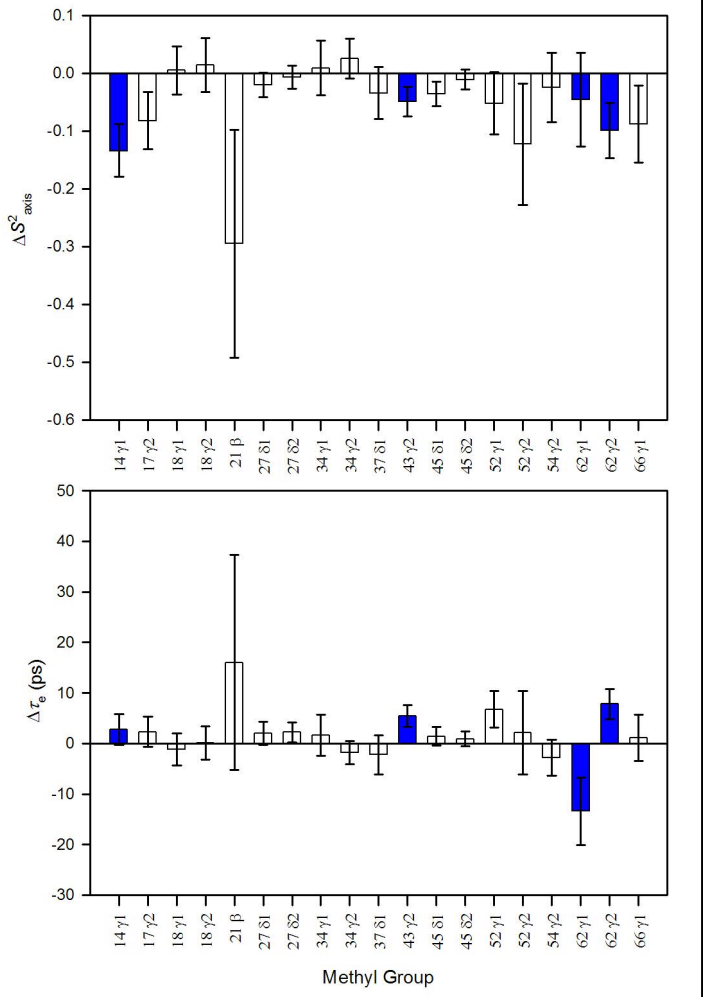

Figure 19: Side-chain dynamics changes in V63A* eglin c. Significant changes are shown in blue.

At this temperature the rigidity of side chains is substantially increased (Figure 20), which may serve to globally dampen dynamic propagation. Although the increase in rigidity is striking, this result is roughly similar to observations for some side chains in calmodulin (Lee et al., 2002). Additionally, V63A* was not entirely stable even at this temperature for long periods of time. As a result, the dynamic data are not as high in quality for this mutant, which may serve to mask real changes in dynamics. Clearly, however, the residue is critical for the stability of eglin c, which serves to reinforce its inclusion in these dynamic networks, though it is not proven that its importance is entirely dynamic in nature.

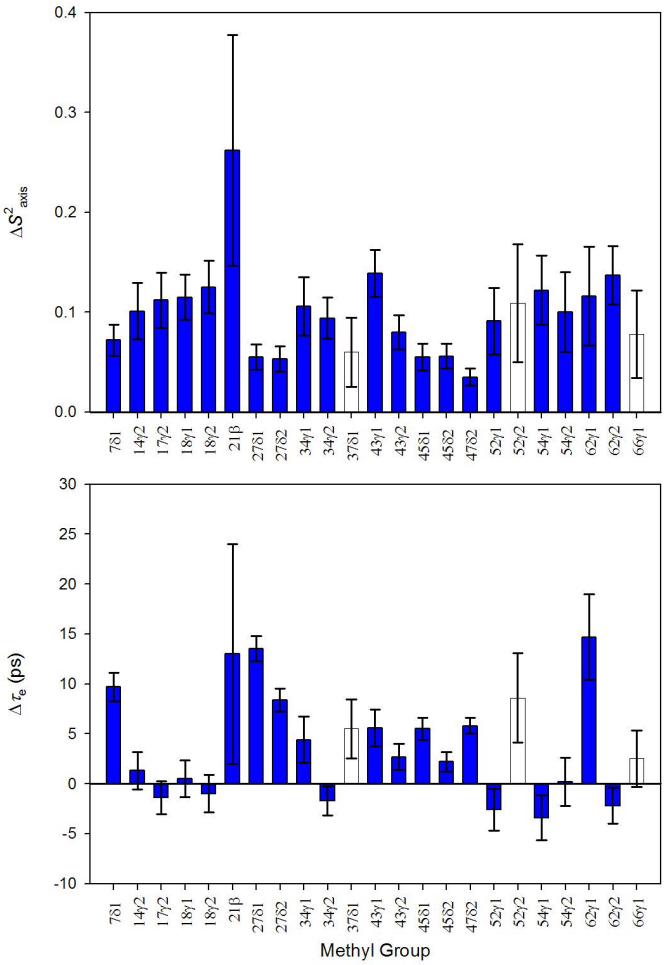

Figure 20: Changes in side-chain dynamics between 25° C and 37° C for WT* eglin c (25°-37°). Residues with significant changes are shown in blue.

Of the other residues which were assumed to be part of these dynamic networks, R53 and F25, both appear to also be critical residues in eglin c, though again it has not been proven that this is due to dynamics. F25L* eglin c forms a protein gel when concentrated to >500 μM at room temperature, and mutations of R53 prompt substantial structural changes in the loop (Chapter 3).

The apparent significance of all these residues suggests that their inclusion in dynamic networks is at least reasonable. Restraint must be exercised, however, in adding such residues to a network. In every case except F25 in the V14A network, each hypothetical inclusion is bracketed by a cluster of responding residues reinforcing its inclusion. The presence of such a cluster should be viewed as a requirement for the insertion of these surfaces in a network.

Assuming that the observation of contiguity is meaningful, what does it imply about the nature of dynamic transmission? In the most trivial sense it would seem to indicate that side chains exert dynamic effects on adjacent side chains, and that the net result of these transmissions is a chain of interactions forming a contiguous surface. Yet this simplistic

interpretation is contradicted by the fact that although all of the observed networks appear to have numerous members in common, they do not have *all* members in common. For instance, although all four mutations appear to affect V63, they do not all alter the dynamics of V18 and V62. Similarly, though V14 causes dynamic changes in the core, it does not affect V54. And though V34 and V13 both appear to be members of the V14A network, V14 does not respond to mutation at either site, nor does it respond to mutation of V54. Finally the V54A mutation causes a rigidification of V18, but the V34A and V14A mutations actually increase V18's flexibility, even though they ostensibly transmit motion to V18 via the same mechanism (V63) as V54A. If adjacency were all that was required for transmission, these observations would be incompatible.

A more powerful explanation of the contiguity could be made using motional modes involving correlated motions of the side chains. Such a possibility cannot be directly established by these experiments because they do not contain information on whether these motions are temporally correlated. Synchronous motions of bond vectors should theoretically give rise to similar τ_e or $\Delta\tau_e$ values for the axial bond. However, the dynamic parameters obtained by ^{2}H spin-relaxation methods are a convolution of the degenerate methyl rotation and the spatial diffusion of the axial bond, and are dominated by the former (Lee et al., 1999). The relative invariance of the bond angle between the hydrogens and the axis makes a deconvolution of the order parameter possible (Muhandiram et al., 1995; Ottiger and Bax, 1999), but the rate of methyl rotation is highly variable, preventing a similar trick for τ_e.

Nonetheless, the contiguity of the residues in each of these networks, and the fact that ΔS^2_{axis} has a similar sign for most of the adjacent members, suggests a mechanism in which

fluctuations that are correlated in the WT protein are collectively affected by a mutation. These fluctuations may assemble into distinct sets of sub-global modes; the existence of many such modes may account for the difference in dynamic response for each of these mutants.

Energetic Connectivity

Motional modes and dynamic interactions both indicate the sharing of energy between distal sites of a protein, and thermodynamic mutant cycle analysis is a classic way of identifying these kinds of interactions. Stabilities of eglin mutants were measured by chemical denaturation and used as a means to detect energetic coupling using mutant cycle analysis. Though numerous dynamic interactions have been noted in the cases of these four mutants, only one coupling (out of five measured) meets the standard of 0.3 kcal/mol used by Shortle and Stites (Table 2). The V18A/V54A cycle has an interaction energy of -0.946 kcal/mol. A cycle measured for a case in which no dynamic interaction was detected – L27A/V54A – has a coupling of -0.221 kcal/mol, similar to the observed values of -0.270 kcal/mol for V14A/V18A, -0.249 kcal/mol for V34A/V62A, and -0.266 kcal/mol for V54A/V62A. The V18A/V34A cycle did not appear to be associated with any interaction energy at all. The energetic results prompt two questions: why the dynamic

Mutant	$\Delta\Delta G_u$	$\Delta\Delta\Delta G_i$
V14A	-1.046	--
V18A	-1.212	--
L27A	-0.170	--
V34A	-1.270	--
V54A	-1.588	--
V62A	-1.022	--
V14A/V18A	-2.528	-0.270
V18A/V34A	-2.504	-0.022
V18A/V54A	-3.745	-0.946
L27A/V54A	-1.979	-0.221
V34A/V62A	-2.541	-0.249
V54A/V62A	-2.876	-0.266

Table 2: Changes in free energy of unfolding due to mutation. All values in kcal/mol.

networks did not generally correlate with detected energetic couplings, and why one of them did.

The near-absence of detectable interaction energies is consistent with the results of previous experiments where the dynamic probe was hydrogen exchange (Spudich et al., 2002). While it is possible that no energetic interaction exists for pairs of residues that did not meet the 0.3 kcal/mol standard, this is difficult to reconcile with the often obvious changes in dynamics. It is equally possible that the free energy of unfolding is simply not an appropriate probe of the interaction. Also, it could be that the energies involved are simply too small to be accurately measured. This does not refute their significance; even if the energetic interaction associated with the dynamic linkage of two particular residues is near zero, the sum of many such energetic linkages throughout a molecule may be substantial.

These findings also point to limitations of the thermodynamic mutant cycle experiment itself. The experiment relies upon the assumption that both mutations ablate the interaction to a similar degree. If the interaction is completely removed (or even gained) by one mutation, but only partially removed or unaffected by the other in the wild-type background, then even a real and large interaction energy will not be detected. The dynamic responses noted above were not generally reciprocal: as an example, V14 did not respond to the V34A mutation, even though V34 responded to V14. This issue will be treated more fully in the next chapter. However, even with the data shown here it is clear that this assumption may be invalid for all or most dynamic responses.

An additional limitation of the experiment is that some elements of a residue cannot be removed. One cannot subtract the backbone atoms of a residue at all, and it is rare that the β-carbon can be ablated without introducing new dynamic complexities to the backbone. In

cases where these portions of the residue serve to mediate the dynamic response, achieving a reciprocal effect is impossible.

Why did a clear interaction appear in the case of the V18A/V54A double mutant? In this case, the explanation may be structural. V18 and V54 have only one layer of side chains between them, and both share substantial surface area with F36. It is possible that slight structural changes can compensate for the loss of contact area on one side of the F36 ring, but that removal of hydrophobic interactions on both faces prompts a major rearrangement at that end of the protein. It is also notable that V18 and V54 both share significant surfaces with V63, a residue that is critical for the stability of eglin c. The loss of hydrophobic surfaces on both sides of this valine may structurally mimic the effect of mutating it to an alanine. V63 does not appear to be structurally perturbed in either single mutant, but no data are yet available for the double mutant. Although the large energetic interaction between V18 and V54 corresponds with the largest S^2_{axis} effect measured in eglin c mutants, it is more plausible that the energetic result reflects structural changes than dynamic ones.

In contrast to the proximity of V18 and V54, the other pairs are positioned too far apart to have any direct interactions. While it is possible that long-range structural effects account for these small for the linkages, it is clear that a gross rearrangement of the kind assumed to explain the V18A/V54A interaction is not likely. In these case, direct enthalpic contributions cannot explain the linkage; entropic contributions (perhaps from shared fluctuations) are a more likely explanation. The fact that the free energy of interaction is negative for all these pairs suggests that the linkage of the two residues is beneficial to the protein: for example, the presence of a valine at position 62 may make new motions (or more motions) available at V34, or vice versa. It is also possible that these non-additive interactions are the product of

behavior in the unfolded state (Green and Shortle, 1993). However, because the NMR data are unambiguously derived from the native state, it is far more likely that the interactions occur there. This result therefore lends experimental weight to Cooper and Dryden's suggestion that entropy alone can give rise to allostery (Cooper and Dryden, 1984).

Conclusions

Because of the ambiguity in these experiments it is not certain that thermodynamic mutant cycle analysis will accurately determine the amount of energy bound up in dynamic interactions. Nonetheless, it is clear from the ^{2}H spin-relaxation studies that side chains can experience significant changes in dynamics at positions distant from a mutational perturbation, even in the absence of obvious structural change. The contiguous network responses described here suggest a mechanism – correlated motions – whereby dynamic effects can propagate through the protein. The existing thermodynamic mutant cycle data indicate that in at least some cases these dynamic interactions are associated with measurable energetic interactions. In sum, these data show that perturbations such as mutations can have energetic consequences for regions of the protein beyond the first shell of interactions, and represent the first demonstration of coincident dynamic and energetic linkage. These facts may have consequences for the interpretation of mutational data in proteins, and may prove particularly helpful in understanding how mutations exert energetic influence on distal binding and catalytic sites.

CHAPTER II

DISPERSE NETWORK RESPONSES

Introduction

Although contiguous responses possess a certain conceptual neatness, there's no particular reason to expect dynamic traceability on the fast timescale. The range of motions quantified by the Lipari-Szabo formalism is actually rather small, covering only fluctuations with characteristic times less than that of the protein's rotational diffusion. In eglin c, this limits the slowest motions measured by spin relaxation to around 4 ns, but in practice most of the observed motions last just several hundred ps. Even if a contiguous dynamic response is actually necessary, there's no particular reason to expect that response to be expressed in terms of fast dynamics for all residues in the pathway. Slower fluctuations, or even subtle structural changes, might also cause dynamic propagation. With this in mind it is actually somewhat surprising that dynamic effects in the eglin c mutants already discussed appear to propagate through discrete, contiguous pathways.

The contiguous response networks described in Chapter 1, however, are not the only kind of dynamic effect resulting from mutations to eglin c. Using the same methods described previously, eglin c mutants V18A* and V62A* were examined for dynamic responses. Although both mutations induce long-range dynamic effects, the affected residues cannot be linked by side chains experiencing altered methyl dynamics unless numerous dynamically unobservable side chains are included in the network. Additionally, changes to τ_e predominate among affected residues. These "disperse response networks" are associated

with slightly larger structural perturbations than the contiguous responses, which in some instances may explain the observed dynamic reactions.

Methods

Nearly all aspects of data collection were as described in Chapter 1. However, partially deuterated forms of both V18A* and V62A* eglin c were observed to aggregate at 2 mM and 37 ºC. As a result, dynamic measurements on V18A* and V62A* were carried out using 1 mM protein, though the temperature was kept the same as in previous studies.

Results and Discussion

Dynamic Effects

The V62A mutation causes only minor dynamic perturbations to eglin c (Figure 21A). Backbone dynamics of V62A* are remarkably similar to WT* – as is typical for most other mutants, the observed changes to backbone dynamics are only marginally greater than the 2σ cutoff. The exceptions to this are A21 and E39, though because the latter is positioned on the dynamically complex loop its significance is difficult to assess. In terms of the side-chain dynamics (Figure 21B, C), only a few residues experience altered dynamics, and these are all differences in τ_e. Additionally, most of the measured responses are only barely greater than 2σ. Overall, the dynamic response to the V62A* mutation appears to be very subtle compared to those of previously discussed mutations.

An additional feature of the V62A perturbation is that the affected residues do not appear to be adjacent in structure. Residues that experience a dynamic response are spread throughout the protein (Figure 22), and include several that are quite distant from V62.

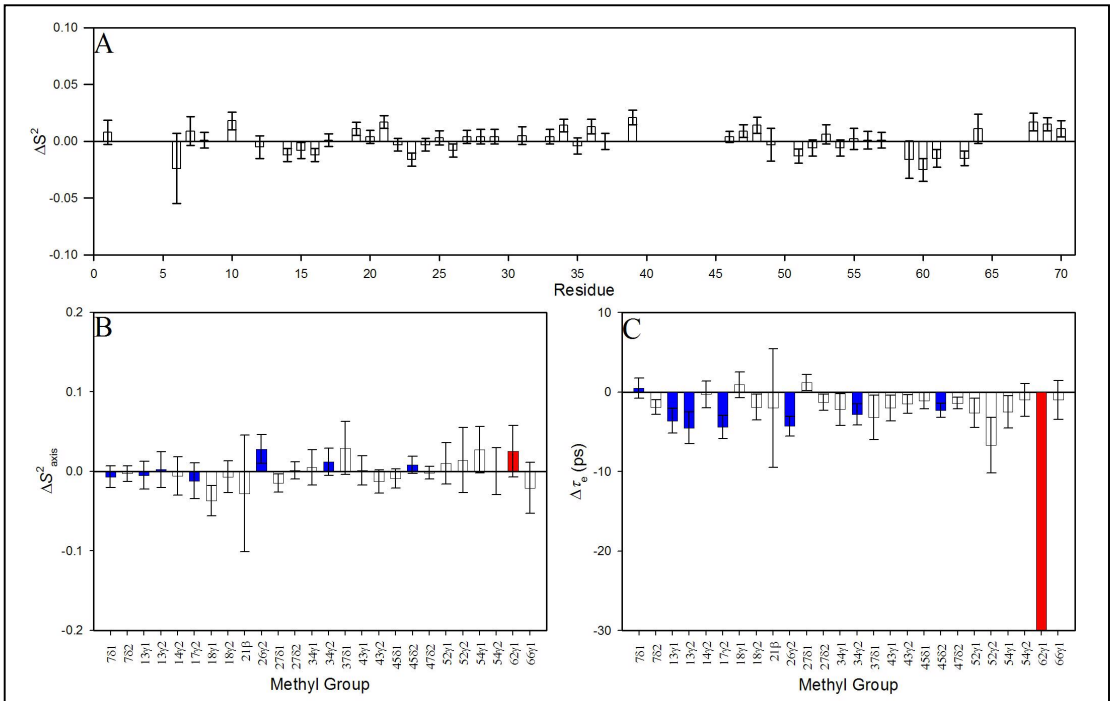

Figure 21: Changes in backbone (A) and side-chain (B,C) dynamics parameters for V62A* eglin c. Side chains with significantly changed dynamics are shown in blue, mutation in red.

This is consistent with observations of distal responses in the previous chapter. However, in the case of V62, it is not possible to construct a contiguous surface that encompasses all residues without including either a large number of unobservable residues, or some methyl-bearing residues for which dynamics could be calculated but that did not display altered dynamics. For instance, V13 and V34 could be linked to each other and to V62 if surfaces were included for V63 and V54. As before, dynamics could not be measured for V63 in WT*, so it is at least possible

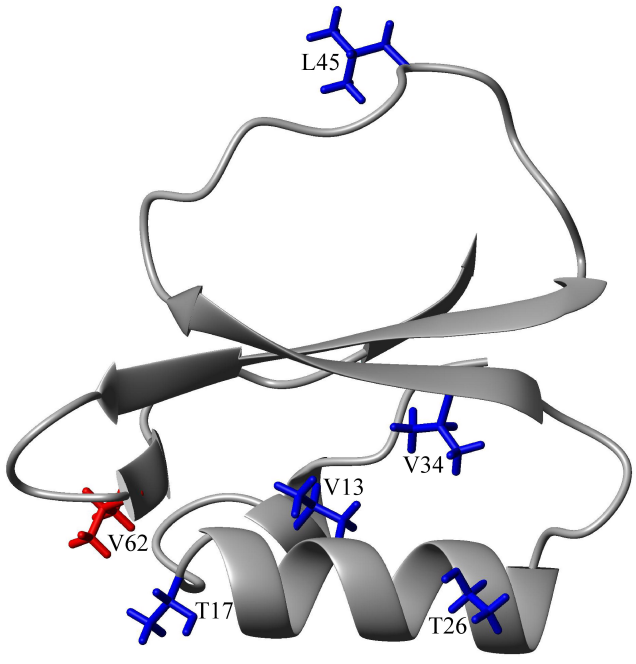

Figure 22: Side chains responding to the V62A mutation.

that its dynamics changed. However, the V54 methyls were clearly observed in both the WT* and V62A* spectra, and just as clearly did not experience any dynamic change. Its inclusion therefore cannot be justified *on the basis of fast dynamics*. V13 and V34 could conceivably be linked by F25, but this puts multiple hypothetical steps into the pathway, in the absence of any supporting evidence. In contrast with the dynamic responses of Chapter 1, the residues experiencing altered dynamics in the V62A* mutant are dispersed throughout the protein, without readily apparent linkage.

The V18A mutation produced a dynamic response that was similar in its distribution to that of the V62A mutation. As before, the backbone dynamics are not greatly perturbed by the mutation (Figure 23A), though a subtle increase in flexibility seems to have occurred near V18, and a similar change may have occurred about halfway down the helix. A fairly significant change is also noticeable at H65, which may indicate a small structural effect.

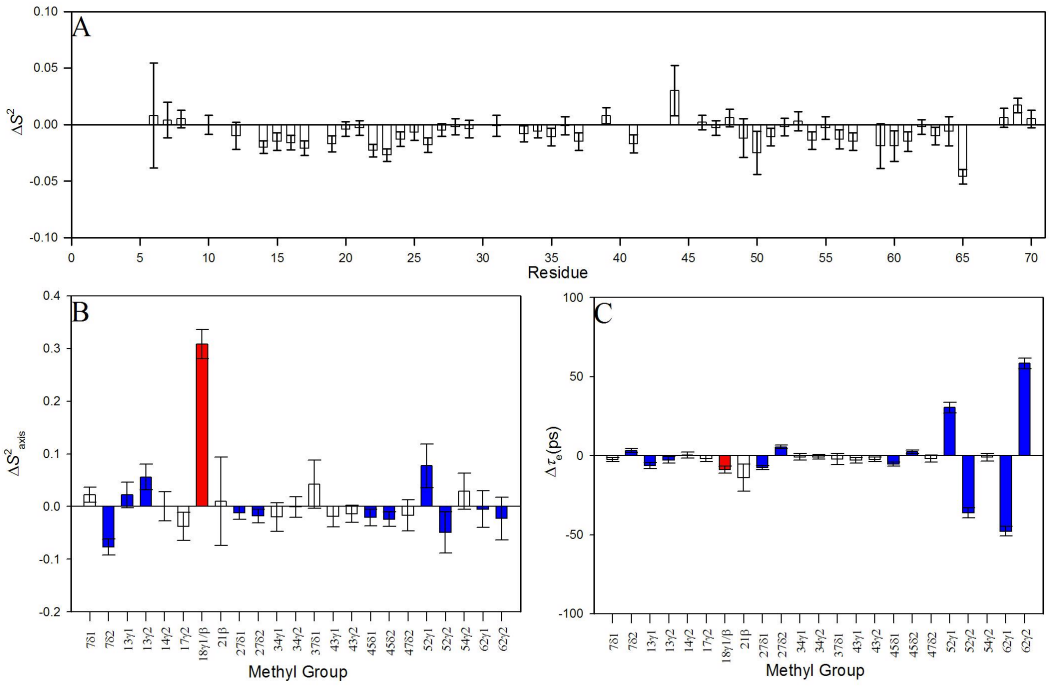

Figure 23: Backbone (A) and side-chain (B, C) dynamics changes in V18A* eglin c. Side chains with significant changes are shown in blue.

Again, however, in most cases the altered dynamics barely exceed the 2σ cutoff.

As regards the side chains, however, the results of the mutation are much more dramatic (Figure 23B, C). As with V62A*, most residues do not experience a change in S^2_{axis}, though small effects are evident at V13 and L7. Instead, the V18A mutation dramatically alters τ_e for several residues, most notably V62 and V52. The changes in V62 and V13 can perhaps be rationalized as transfer through V63. The more distant residues experiencing dynamic changes (V52, L27), however, cannot be explained structurally without incorporating a number of unobserved side chains, or methyl-bearing side chains that did not have altered dynamics. Again, at least a portion of the clear dynamic changes appear to be dispersed through the protein, though in this case they are much more dramatic than in V62A (Figure 24). In neither of these cases were any changes in rotamer populations detected.

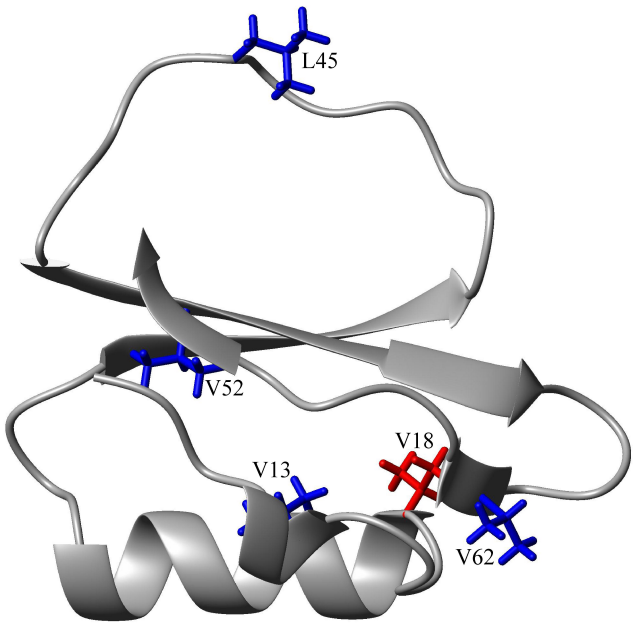

Figure 24: Side chains experiencing dynamic changes in response to the V18A mutation

Thus far, these responses have two features in common: differences in τ_e as the primary dynamic effect, and difficulty in connecting parts of the network without numerous hypothetical inclusions. An additional commonality of these responses becomes apparent when RDCs are compared between these mutants and WT* eglin c (Figure 25).

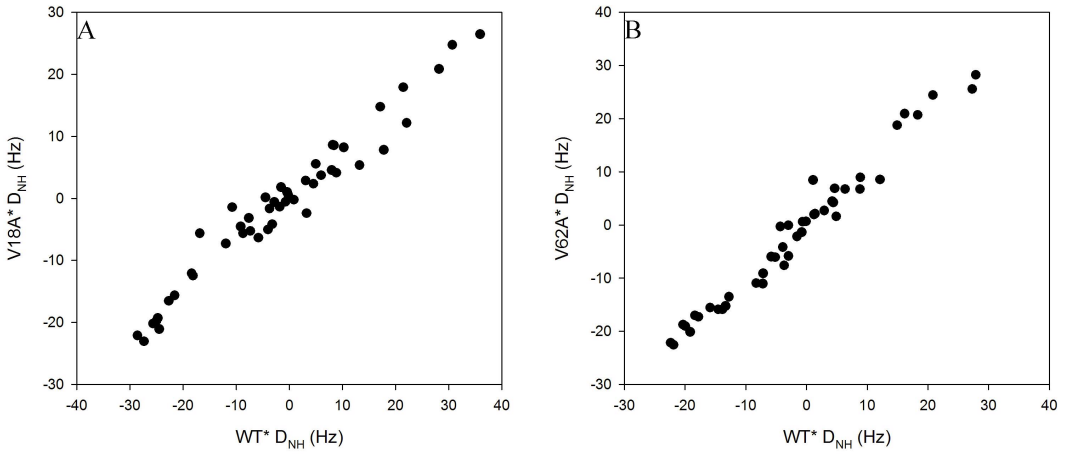

Figure 25: Direct comparisons of V18A* (A) and V62A* (B) RDCs to WT* RDCs. R^2 was 0.96 and 0.97 for V18A* and V62A* respectively.

Structural Effects

When comparisons are made between RDCs from these mutants and WT*, the correlation coefficients are still quite good (R^2 = 0.96 for V18A*, 0.97 for V62A*), but slightly less than the excellent correlations of the previous chapter. Residuals were determined for linear relationships between WT* and mutant RDCs for every mutant discussed thus far. The residuals for each mutant average to zero and appear to be randomly distributed (Figures 26A, 27A). Across all studied relationships (including a comparison of the two WT* datasets collected), the standard deviation of the residuals is 1.91 Hz. A cutoff of twice this value was used to identify residues with altered structure, though this does not imply a major change or even a change that would be noticeable by visual comparison of solution structures.

The RDC residuals for V18A* eglin c clearly identified the central β-strand as an altered element of the structure (Figure 26). This strand spans residues 50-57 and includes 4 residues identified as undergoing a structural change. Tellingly, this strand also includes one the most strongly responding side chains in this mutant: V52. These changes in RDCs also correlate with notable amide chemical shift changes at Y56 and N57, and with the previously

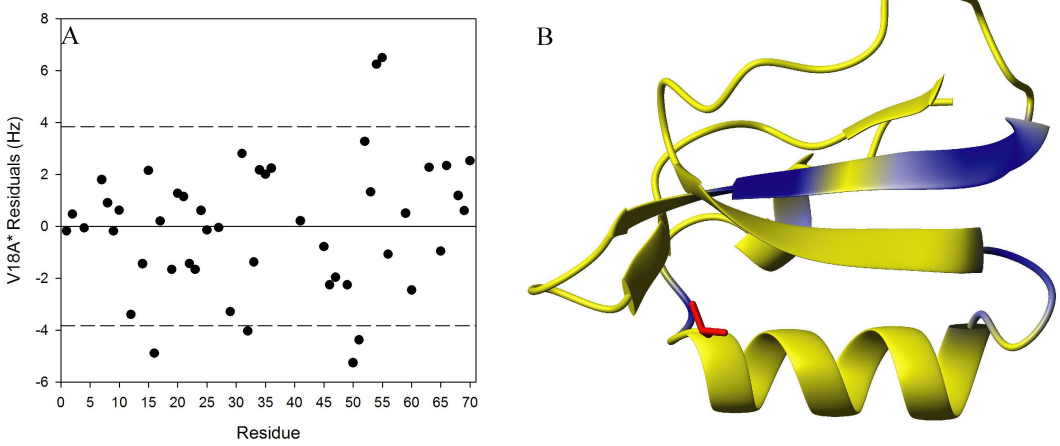

Figure 26: Residuals plotted (A) and mapped onto the structure (B) for V18A*. Residues with residuals above the threshold (see text) are shown in dark blue; those with residuals near the threshold are shown in light blue. Other residues are in yellow. V18 side chain shown for context.

noted backbone S^2 change at H65 (which may form a hydrogen bond with F55). Collectively, these data point to a minor structural shift in this strand, possibly induced by a change in the side-chain disposition of Y56, which contacts the V18 side chain in the WT structure. Because RDCs contain both structural and dynamic data, it is difficult to state exactly what causes the observed differences. However, it is worth noting that for every residue on the strand, the observed RDC had a magnitude less than expected on the basis of the linear relationship. This is consistent with an increase in flexibility relative to the WT and the rest of the structure. That such a change does not manifest in the S^2 values of the backbone NH vectors suggests that this would represent a slow fluctuation of the sheet. It is also possible that these results signify that the twist of the sheet has changed, although it seems that this would cause even more widespread changes in RDCs as other elements shifted to compensate.

Similar to its dynamic results, the structural effects of the V62A* mutation are subtler and more difficult to interpret. Although several residues clearly have notably large residuals, they localize not to a single element of structure, but rather to the N-termini of both major strands (Figure 27). More curiously, this region lies at the opposite end of the protein from

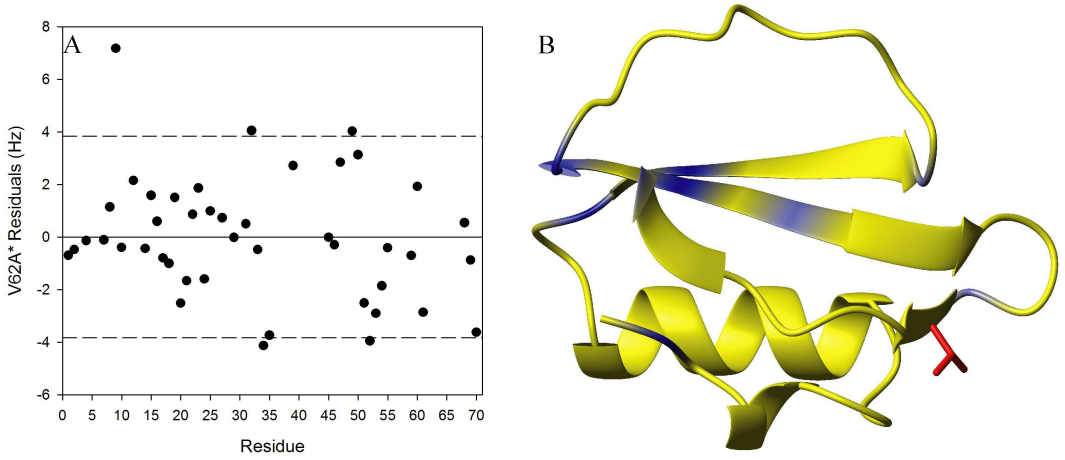

Figure 27: Residuals plotted (A) and mapped onto the structure (B) for V62A*. Residues with residuals above the threshold (see text) are shown in dark blue; those with residuals near the threshold are shown in light blue. Other residues are in yellow. V62 side chain is shown for context. The view is reversed relative to Figure 26 for added clarity.

V62. This distribution of structurally-altered residues does not correlate with any unusual change in chemical shift or backbone S^2. Though they co-localize, the affected residues do not experience a monotonic effect on RDC magnitude, suggesting that in this case dynamics is not to account for the observed changes in RDCs. Because the V62 side chain is pointed towards the solvent, it is difficult to understand how these structural changes might occur. However, because these structural changes are likely to be very subtle, it is possible that they result from slight rearrangements of side chains in the core caused by a repositioning of V63.

Significantly, V34, which had a dynamic response in terms of τ_e, is among the residues identified as experiencing a structural change. Again, changes in structure appear to correlate with observations of dynamic change that are otherwise difficult to understand.

The V18A and V62A mutations appear to cause dynamic responses in a number of residues that are distant in structure, much like networks noted in Chapter 1. In contrast, however, these responses are limited almost exclusively to changes in τ_e, and are difficult to assemble into a contiguous body of residues. However, at least some of the distal residues lie in regions of the structure that have experienced subtle conformational shifts as a result of the

mutation. This suggests that the contiguous response networks described in Chapter 1 are supplemented by a second class of dynamic perturbations, "disperse response networks". These are identified by 1) difficulty in linking responding residues, 2) preference for τ_e changes, and 3) subtle structural or dynamic shifts of the backbone.

Unifying Features

Although they are close to each other in the structure (Figure 28), V18 and V62 appear to have little in common. While V62 is pointed out of the protein and partially exposed, V18 is buried in the core, with less than 8 $Å^2$ of solvent-accessible surface area. V18 is buried, but its immediate neighbors in the hydrophobic core are Y56, F36, and V63 (Figure 28). Because of their bulk, the large aromatic side chains are probably rigid on the dynamic timescale being measured here; fast dynamic motions are therefore unlikely to propagate through them. This means that the primary conduit for transmitting motional changes from V18 to the rest of the mobile core residues is probably V63. Propagation of dynamic effects from V62 to the core is also stymied, in this case by isolation instead of rigid surroundings. However, if dynamics are transmitted through the backbone, then V62 has a conduit for this kind of transmission, namely V63.

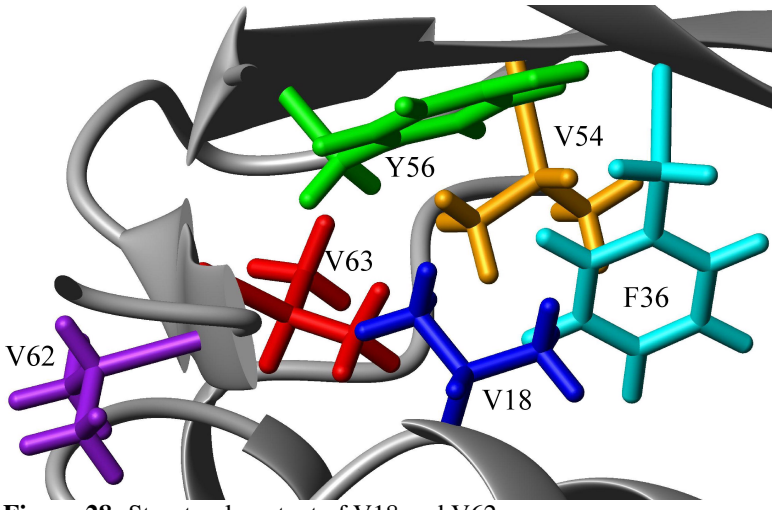

Figure 28: Structural context of V18 and V62

As noted in Chapter 1, mutation of V63 to alanine has a profound effect on the stability of eglin c, and so the residue probably plays a key role in maintaining the fold. The proximity of this critical residue to V18 and V62 and the requirement that it serve as a conduit for transmitting motion provide important commonalities which may explain why these two mutations with very different local structures nonetheless have similar dynamic responses.

In the case of eglin c, at least, residues that induce these "disperse network responses" appear to have two characteristic features. The first is dynamic isolation – the position of V62 and the highly aromatic environment of V18 both work to counteract dynamic propagation into the core. In addition, their proximity to V63 may be important to explaining both the mode of propagation and the slight differences from the wild-type structure. While mutations made to other residues that contacted V63 excited contiguous response networks, these residues also had the feature of possessing alternate routes for dynamic propagation. The absence of such avenues for dynamic dispersion may force the structure to compensate for the mutation. It is therefore likely that both dynamic isolation and proximity to a critical stabilizing residue are necessary to produce a disperse network response. As such, these responses are more likely to occur in solvent-exposed residues in proteins without a strong aromatic presence in the core.

Reciprocity in Dynamics

The set of mutants thus far considered possesses an additional interesting feature: the occasional lack of reciprocity in dynamic response. In the most general sense, reciprocity occurs when a mutation at site A causes a dynamic response at site B, and a mutation at site B causes a dynamic response at site A. A more specific form of reciprocity, response-in-

kind, occurs when the dynamic disturbance of B by A and the dynamic disturbance of A by B have a similar dynamic nature. Of the 18 instances in which a mutation to eglin c produced a dynamic response in a residue that was also the subject of a mutation experiment, 10 of them formed a reciprocal relationship, producing five pairs of linked residues. These were V13A/V34A, V13A/V54A, V13A/V62A, V34A/V54A, and V34A/V62A. Given the 6 methyl-bearing sites under consideration, a discovery of 5 reciprocal linkages (out of a possible 15 mutant pairs) is a respectable percentage. The question remains, however, whether the dynamic effects of each mutation are similar in nature.

A simplistic test for response-in-kind is to ask whether the parameters that change at the disturbed sites are similar for both mutations, without any regard for the *degree* of change. In the case of V34A, the τ_e of V62 changes but the S^2_{axis} does not. Similarly, in the V62A mutant only the τ_e of V34 changes. Because the same parameter changes, this qualifies as a response-in-kind. This simplistic standard is also met by the V34A/V54A (S^2_{axis} and τ_e) and V13A/V62A (τ_e only) reciprocal responses.

Of course, it is possible that reciprocal responses reflect local structural changes. This is a matter of particular concern for the V13A/V54A and V34A/V54A pairs, because V13 and V34 both contact V54. This is also a factor in the consideration of the V13A/V34A pair because the two residues involved are spatially separated by a single aromatic residue, and their connection might also be mediated by V54, with which both are in contact. It is worth noting, however, that V18 and V54 have similar structural entanglements (separated by F36, possibly linked by V63) and do not show a reciprocal dynamic response. Furthermore, direct structural contact cannot account for the reciprocal nature of the V13A/V62A and V34A/V62A pairs (both of which also feature a response-in-kind relationship). While it is

possible that some reciprocal dynamic interactions are the result of local rearrangements, this is clearly not necessarily the case.

Local rearrangements may also explain why reciprocity is not a universal feature of dynamically-linked sites. Although it is simplest to assume that the results of a mutation do not stem from changes to the structure, the above discussions of V18A and V62A indicate that this is clearly not so. Local (or even long-range) rearrangements of side-chain structure might compensate for perturbations to dynamic networks. Sub-global motional modes may also account for the lack of reciprocity – the motional mode affected by the ablation of methyls at position B may not be the same mode affected by methyl subtraction at A.

The absence of reciprocity suggests a weakness of the mutational perturbation-response experiment, in that a single mutation may not identify all the side-chains sharing a network with the mutation site. Indeed, there may not be any mutation that identifies the network if, for instance, the connection is mediated by the backbone. This is, however, a common weakness of mutation experiments, and its influence here, while disappointing, is hardly a surprise. The possibility that a mutation will not excite a given network is simply another caveat to the experiment.

Reciprocity would appear to be a necessary feature of dynamic systems expected to show coupling by thermodynamic mutant cycle analysis. As mentioned in the previous chapter, finding a coupling using thermodynamic mutant cycle analysis relies upon the assumption that whatever interaction exists between the pair of residues under investigation is disrupted to a similar extent by each mutation. The absence of reciprocal dynamic response between some pairs of residues would seem to indicate that in several cases it is possible that a dynamic linkage between two residues is not necessarily perturbed by a mutation to one of

them. The expectation following from this – that mutant cycles that do not feature reciprocal response will not reveal an energetic coupling – has largely been borne out. The V14A/V18A, V18A/V34A, and V54A/V62A double-mutant cycles did not reveal any significant individual interaction energy (Table 2), nor did they include a reciprocal dynamic response. The V34A/V62A cycle also did not appear to show a linkage even though the dynamic states of these mutants suggested a reciprocal dynamic response-in-kind. The V18A/V54A cycle, which is not accompanied by dynamic reciprocity, nonetheless features a substantial non-additivity. Overall, too few double-mutant cycles have been made to draw any firm conclusions about the necessity of dynamic reciprocity for the observation of energetic non-additivity. Further energetic experiments, focusing on cases where long-range reciprocity of S^2_{axis} changes has been observed, may do more to indicate the role of reciprocal responses.

CHAPTER III

STRUCTURAL RESPONSES

Introduction

Most of the point mutations discussed thus far did not have profound effects on the structure of eglin c. Even the mutation that produces the greatest apparent deviation from the WT* structure (V18A*) has RDCs that compare favorably, with a correlation coefficient near one. This tight structural correspondence is not limited to the global fold, either, for the side-chain rotamer populations of buried valines are also remarkably consistent across all mutants (with the exception of V34A*). For many of these targets, such as V13, V34, and V54, tolerance to mutation would be expected on the basis of fairly low sequence conservation. However, the largest structural response, that of the V18A* mutation, results from the perturbation of a site that is very poorly conserved. Additionally, mutations at highly conserved sites such as V14 and V62 did not produce major structural changes. While the set of mutations studied here is limited, the structural data are of high quality and suggest that sequence conservation is not an absolute indicator of the potential structural effects of a point mutation. This result is consistent with experiments on T4 lysozyme that demonstrated that the structure was maintained even when multiple residues in its α-helical or β-sheet domains were replaced by alanines (He et al., 2004; Zhang et al., 1992). Even at highly conserved sites, the self-reinforcing nature of tertiary interactions preserves the fold and most other structural features.

Structural resiliency is not absolute, however; the structural effects of single mutations are known to cause pathologies such as sickle-cell anemia (Hunt and Ingram, 1959). Eglin c reproduces this phenomenon in that although most mutations do not cause major alterations to the structure, there are a few, at R53 and V63, that produce significant changes. These strong structural perturbations are associated with functional effects. However, the dynamic changes resulting from these mutations do not resemble each other. Examination of these results suggests that significant structural changes are not necessarily associated with extended dynamic responses. Additionally, when these mutations *do* excite dynamic networks, the dynamic changes are not limited to the locus of structural change.

Methods

Assignment and Conditions

Unless otherwise noted, experimental techniques match those of previous chapters. NMR conditions for V63A* eglin c were as described in Chapter 1; NMR conditions for R53A* and R53Q* were 2 mM at 37 ºC. Assignments for R53A* eglin c were a gift from J. Boyer, but acquired in the same general fashion as described in Chapter 1.

Alignment Tensors

RDCs were obtained as described in Chapter 1. Once obtained, the RDCs were fit to alignment tensors using the fortran program MALIGN and the crystal structure of eglin c from the PDB file 1CSE. The alignment tensor roughly describes the orientation preference of the protein with respect to the alignment medium, and can be easily calculated when a

structure is known. The measured residual dipolar coupling D_{NH} for a given N, H pair is given by (Tjandra and Bax, 1997):

$$D_{NH}(\theta,\phi) = S\frac{\mu_0}{4\pi}\gamma_N\gamma_H h\left[A_a\left(3\cos^2\theta - 1\right) + \frac{3}{2}A_r\sin^2\theta\cos 2\phi\right] / 4\pi^2 r_{NH}^3$$

Where γ_N and γ_H are the gyromagnetic ratios of the nuclei, r_{NH} is the distance between them, h is Planck's constant, and μ_0 is the vacuum permittivity constant. **S** is an order parameter (but not directly related to the Lipari-Szabo S^2) for the bond, and A_a and A_r are the axial and rhombic components of the alignment tensor, which depend on the nature and magnitude of the alignment. When (as here) the alignment depends on steric interactions, the shape of the molecule also contributes to these parameters. The angles θ and ϕ are cylindrical coordinates that describe the orientation of the vector in the axis system of the alignment tensor. These in turn depend on the Euler angles α, β, and γ, which indicate the rotations necessary to transform coordinates from the structure frame into the alignment tensor frame. Where the interaction is steric in nature, these will also likely be influenced by the shape of the molecule. Only RDCs from rigid regions of the protein were used in the calculation; this excluded the loop and N-terminus.

Activity Assays

Eglin c inhibitory activity was assessed in a manner similar to that used previously for CI-2 (Jackson and Fersht, 1993). Subtilisin (Sigma) was dissolved at 20 μM in a solution of 200 mM Tris pH 8.0 and 10% glycerol and flash-frozen in liquid nitrogen. WT* and mutant forms of eglin c were then diluted to a concentration of 25 nM in assay buffer (200mM Tris pH 8.0, 5 mM $CaCl_2$, 0.01% Tween-20). The frozen stocks of subtilisin were thawed on ice

and diluted in ice-cold assay buffer to a concentration 1 nM, then warmed to room temperature and combined with the eglin c solutions for a concentration of 250 pM subtilisin and 1 nM eglin c in 200 μL. These solutions were incubated overnight at room temperature. Following an average incubation of 18 hours, 800 μL of 1.25 mM AAPF dissolved in assay buffer was added for a final concentration of 50 pM subtilisin, 200 pM eglin c, and 1 mM AAPF. These solutions were then incubated at 25 ºC for 3 hours, with the absorbance at 412 nm continually monitored by a Shimadzu spectrophotometer equipped with a temperature-controlled 6-cell changer.

Chemical denaturation of eglin c by GuHCl, acquisition of backbone and side-chain dynamics data, and determination of rotamer populations were performed as described in Chapter 1.

Results and Discussion

V63A Structure and Dynamics*

The trimmed mean chemical shift change across all mutants except V63A* was a $\Delta\delta_{vector}$ value of 0.059 ± 0.051 ppm. Changes in chemical shift vectors greater than or equal to 0.162 (trimmed mean + twice its error) were generally restricted to the immediate vicinity of the mutation for most mutants of eglin c. This proved to be true also for V63A* (Figure 29). Most of the residues experiencing non-trivial chemical shift changes clustered around V63, though some were dispersed through the core of the protein (Figure 29B). Though the chemical shift of E2 also changes, that residue is highly flexible and in slow exchange among multiple conformers; the difference in chemical shift probably results from settling into a single conformer at the lower temperature and not from the V63A mutation. The change in

temperature almost certainly accounts for the high background in Figure 25A as well. However, the chemical shift changes at V13, K16, V63, and V64 are clearly not trivial, as they fall in the upper 5th percentile of all observed chemical shift changes. These results suggest that the structural changes are localized.

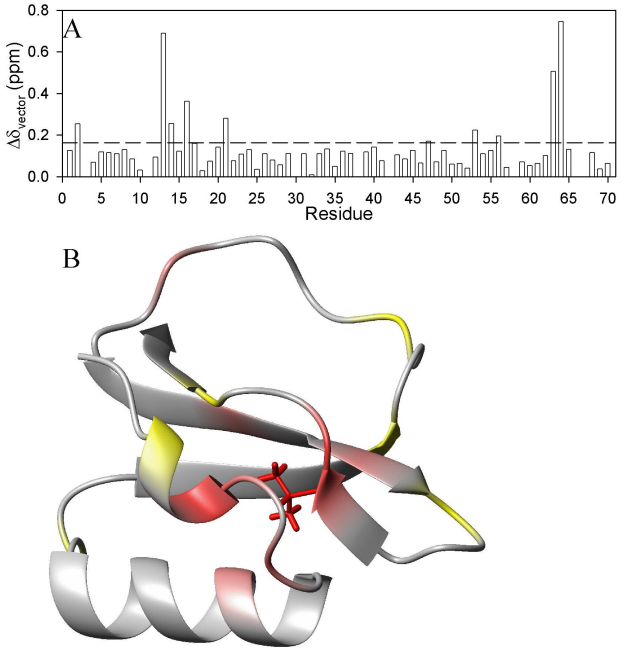

Figure 29: Chemical shift changes plotted (A) and mapped onto the structure (B) for V63A*. Residues are colored red based on intensity of change. Prolines are shown in yellow. The V63 side chain is shown for context. Because complete chemical shift assignments were not obtained for WT* at 25 ºC, V63A* chemical shifts were compared to WT* chemical shifts obtained at 37 ºC.

Dynamic data support the contention of a localized disturbance. Order parameters for the backbone (Figure 30) indicate a clear change in dynamics only for K16. Similarly, the methyl dynamics data indicate significant changes only in V14 and V62 (Figure 19), though overlap prevented comparisons of order parameters for the V13 methyls. Dynamic changes at the V43 methyls are difficult to interpret due to the complex dynamics of the loop, but may be related to the chemical shift changes at R53.

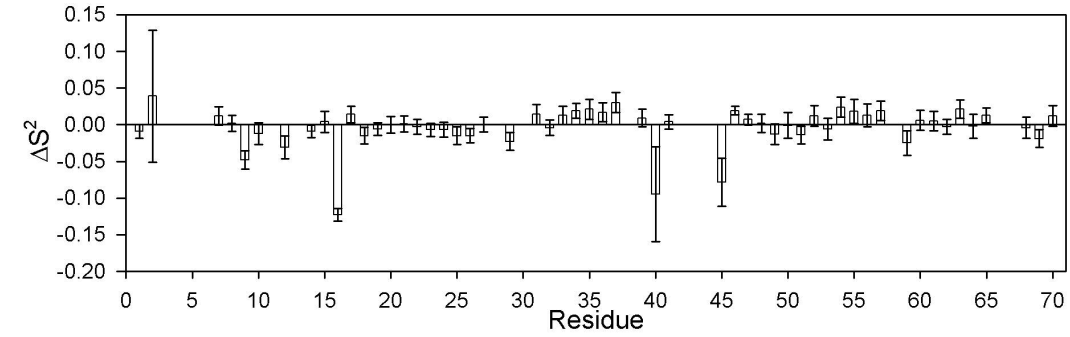

Figure 30: Changes in backbone order parameter for V63A* eglin c

Direct structural data also support the interpretation of a localized structural change. Comparisons of RDCs indicate a clear change in the orientation of the backbone NH vector for K16 (Figure 31). Excluding this single point brings the correlation coefficient for the linear fit close to 0.99, and when the residuals are examined the difference at that site is obvious. Strangely, other residues in the vicinity do not seem to have highly altered structures. However, K16 is positioned in a loop, adjacent to G15, and may be taking advantage of the conformational freedom thus available. In this position, K16 may be able to move around quite easily, and because glycines are highly flexible around the Cα, conformational changes C-terminal to it might not be transmitted to the G15 amide. Additionally, the increased flexibility of the K16 amide as observed in the backbone order parameters may be contributing to the observed RDC.

Structural features of K16 provide a way to unify several of the observations. Importantly, the amide of K16 forms a hydrogen bond with the carbonyl oxygen of V13 in the crystal structure. In addition, the carbonyl oxygen of K16 forms a hydrogen bond with the amide of V63. This interaction encompasses three of the residues with the largest chemical shift changes. Given with the large increase in flexibility at K16 and the evidence from RDCs that the angle of its N-H bond has changed, these data suggest that one or both of the hydrogen bonds has broken as a result of the mutation. The effect may be a "fraying" of the structure at that end of the

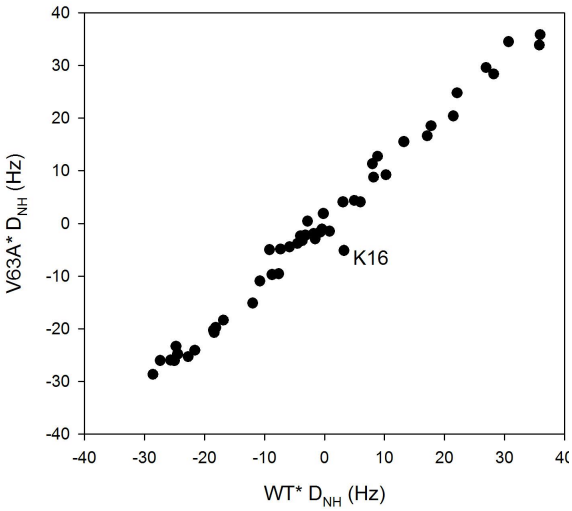

Figure 31: Direct comparison of V63A* RDCs with WT*. R^2=0.98

protein, moving that loop away from the small β3 strand and allowing water into the core.

The data for V63A* eglin c indicate that the mutation has produced a highly localized dynamic and structural disturbance. This produces an effect which is unlike the contiguous and dispersive responses, because it lacks any significant non-local component. In this case there is a small but clear structural change that is highly localized.

R53 Mutation Structural Effects

In contrast, mutations to R53 produced numerous effects that spread beyond the first shell of interactions. Two mutants of this kind were studied; R53A* and R53Q*, both of them designed to disrupt a hydrogen bond between this arginine and T44. This bond is known to form in the bound complex (as demonstrated in PDB entry 1CSE (Bode et al., 1987)), and based on the evidence presented here, probably appears in the unbound state as well. Chemical shift diagrams (Figure 32) reveal striking changes in the structure resulting from both of these mutations.

The R53Q mutation has a significant effect on the chemical shift of several residues, but notably, they include neither residue 53 nor its nearest neighbors in the sequence. Instead, the most significant changes in chemical shift take place

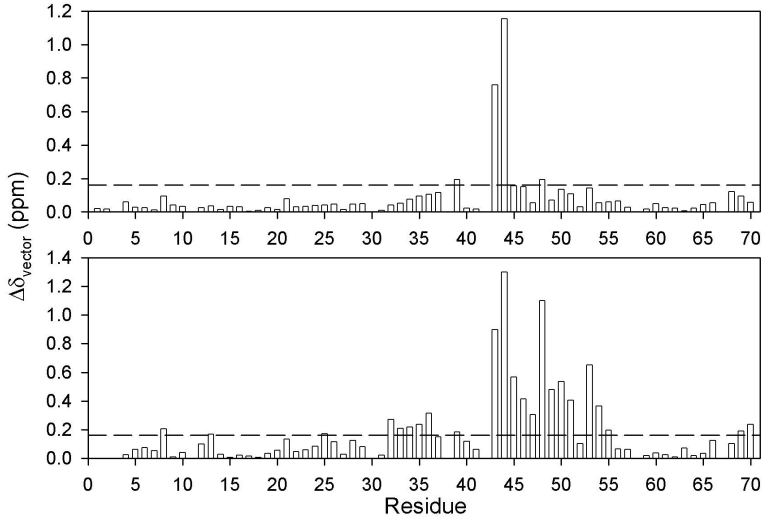

Figure 32: Changes in chemical shift for R53Q* (top) and R53A* (bottom). Dashed line represents the significance cutoff.

on the loop, with the most strongly affected residues being V43 and T44. In an additional indication that the loop is the primary locus of structural change for this mutant, intermediate-timescale chemical exchange processes at L45 were quenched in this mutant. Furthermore, J_{CC}- and J_{NC}- derived rotamer populations for V43 indicate that the 180 and -60 χ_1 rotamers are equally populated at 42% in R53Q* eglin c, as opposed to an 80% population in -60 for all non-R53 mutants. This suggests that an obstruction to the population of the 180 rotamer has been removed.

The R53A mutation has even more widespread chemical shift changes; indeed, the backbone chemical shifts of R53A* eglin c at 37 ºC (average $\Delta\delta_{vector}$ = 0.194) differ more greatly from WT* than the V63A* shifts at a temperature 12 degrees different (average $\Delta\delta_{vector}$ = 0.135). Here again, the largest chemical shift changes are observed in the flexible loop residues, but in this case significant chemical shift differences are also observed for the β-strands encompassing residues 32-36 and 49-55. It is nonetheless clear that the whole molecule has not been perturbed. The loop and short β-strand around V63 do not experience significant changes, and only small changes appear in the α-helix and N-terminus. When the changes are mapped onto the structure (Figure 33), it becomes clear that proximity to the loop is the primary factor in chemical shift change. This produces a picture consistent with that

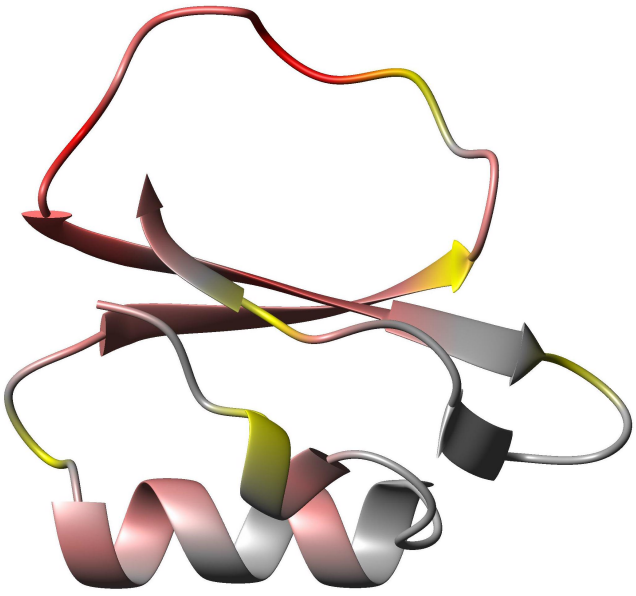

Figure 33: Chemical shift changes in eglin c due to the R53A mutation. Intensity of red color reflects magnitude of change; prolines are shown in yellow.

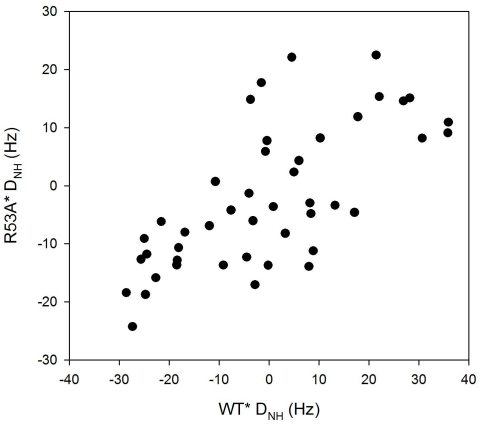

Figure 34: Direct comparison of R53A* and WT* RDC values

from R53Q*, in that the loop feels the structural effects of R53 mutations most strongly, and in some cases these effects are spread to nearby regions of the protein.

A significant perturbation to the loop is also consistent with the RDC data for R53A* (Figure 34). Even a cursory inspection of the comparison plot against WT* indicates that significant changes have taken place, and a more rigorous analysis confirms it, as the correlation coefficient for a linear fit is 0.55. However, the nature of RDC measurements allows for two possibilities that may account for the observed changes. While a significant alteration to the fold *is* a possibility, it is equally possible that these results merely reflect a large change in the alignment tensor.

In order to distinguish between these alternatives, alignment tensors were fit to all the RDC datasets of eglin c mutants that had been acquired so far (including two datasets for WT* eglin c), using the crystal structure of WT eglin c from 1CSE. The results of the analysis are shown in Table 3. Note that there is a fair degree of variation in the parameters of the alignment tensor even among the mutants that experienced minimal structural change (and even between the

Mutant	A_a	A_r	α	β	γ
WT* (A)	1.35E-03	5.65E-04	1.379	0.762	2.899
WT* (B)	1.76E-03	5.41E-04	1.430	0.802	2.838
V13A*	1.57E-03	5.68E-04	1.432	0.858	2.842
V14A*	1.48E-03	6.29E-04	1.584	0.827	2.962
V18A*	1.05E-03	6.48E-04	1.466	0.803	2.795
V34A*	1.08E-03	4.62E-04	1.399	0.914	2.940
R53A*	1.23E-03	4.75E-04	1.260	1.586	0.074
V54A*	1.39E-03	4.86E-04	1.443	0.919	2.810
V62A*	1.18E-03	5.30E-04	1.446	1.000	2.938
V63A*	1.69E-03	6.79E-04	1.528	0.847	2.777

Table 3: Parameters defining the alignment tensors for eglin mutants

two acquisitions of WT* RDCs). This is most likely the result of differences between the batches of bicelles prepared for each of the experiments.

Despite the degree of variation, two significant pieces of information can be gleaned from these results. The first of these is that the alignment tensor for R53A* is significantly different from that of all other mutants. The value for the β parameter is 1.586 radians for R53A*, compared to an average of 0.859 ± 0.074 radians for the other mutants. The α parameter is also somewhat lower for R53A* than for the other mutants, and the α/β ratio for R53A* is 0.794, significantly different from the ratio of 1.706 ± 0.159 for all other mutants. The γ angle is also very clearly different in R53A* compared to all other mutants. The axial and rhombic components of the R53A* alignment tensor, however, are largely similar to those of the overall set. This suggests that the interactions with the liquid crystalline medium are largely the same, but the translations needed to get from the molecular frame to the alignment frame have changed, possibly as a result of an altered molecular shape.

An additional point of significance is that the RDCs determined for the rigid residues of R53A* *can* be fit to an alignment tensor using the WT structure. This suggests that the core of the protein has not suffered a serious deformation as a result of the mutation. The dipolar couplings predicted by MALIGN compare favorably with the experimental values (Figure 35). Although the correlation between prediction and experiment is not as tight

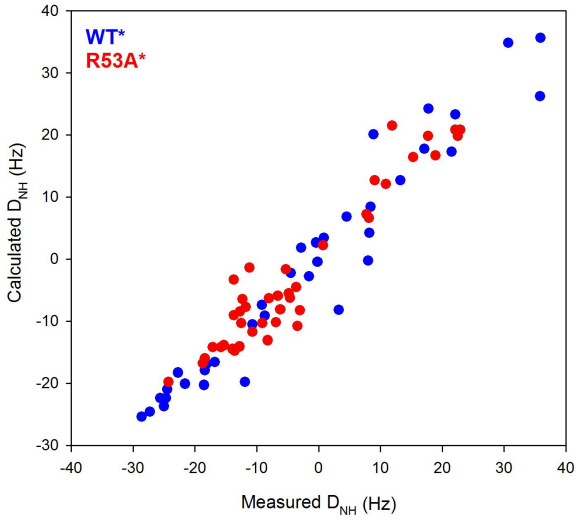

Figure 35: Comparison of predicted and observed RDCs for WT* (blue) and R53A* (red) eglin c.

for R53A* (R^2 = 0.91) as for WT* (R^2 = 0.94), the spread for each is not vastly dissimilar. While there may be some small changes in structure, the overall fold clearly is preserved.

From the dynamic effects on the loop, the clear change in the alignment tensor, and the distribution of chemical shift changes, it appears that there has been a shift in the position of the loop. It is unlikely that the loop has frozen out its motion completely. However, there is a strong possibility that the dynamic distribution of its positions has moved closer to the core of the protein in order to take advantage of potential hydrophobic interactions between loop residues and the β-strands. Given that there is a large amount of structural entropy to be gained from loop flexibility, it is probable that any such interactions are transient.

A condensation of the structure is also suggested by the enhanced rotational diffusion of the R53 mutants. Global τ_m fits of rigid residues produced values of 4.00 ns and 4.18 ns for R53A* and R53Q* respectively, compared to 4.68 ns for WT*. This suggests that the molecule has become smaller on average. Local τ_m values fit for these mutants were used to produce a rotational diffusion tensor, which was spherical for both R53 mutants and the WT. This reinforces the transient nature of the loop collapse, because if the loop formed long-lasting hydrophobic interactions with the β-strands, the diffusion tensor would probably be more cylindrical.

A similar "collapse" of the loop was proposed as a partial explanation of eglin c rigidification in reponse to a pH shift from 7 to 3 (Hu et al., 2003). Unfortunately, RDCs have not been taken at that pH, and the differences in chemical shifts are difficult to interpret because of the widespread changes in charge.

The ability of R53 to form hydrogen bonds is important for the inhibitory activity of eglin c. When incubated with subtilisin for ~18 hours before the addition of a colorigenic

substrate, WT* eglin c causes a significant inhibition of activity, but this inhibition does not occur when R53Q* eglin c is used in the incubation. The difference in inhibitory activity cannot be attributed to a change in the stability of the protein, because $\Delta\Delta G_U$ is minimal for this mutation (-0.61 ± 0.093 kcal/mol). The change in stability is somewhat larger for R53A* (-0.87 ± 0.093 kcal/mol), but this still does not represent a major global loss of stability. Furthermore, even the most destabilizing mutant with respect to folding (V63A*) does not cause a significant change in inhibitory activity. The loss of activity is therefore due not to a change in the intrinsic stability of the mutant, but rather to structural interactions in the active site involving R53, which are altered by these mutations. This result is consistent with previous work in CI-2 (Radisky et al., 2003) that indicates the core's primary role is to correctly position the ends of the binding loop and properly orient the R65 and R67 side chains.

Dynamic Effects in R53 Mutants

Both R53 mutants produce strong, but highly localized effects on the backbone of eglin c. Backbone perturbations are generally mild in the case of the R53Q mutation (Figure 36A), but T44 shows a significant decrease in the order parameter. Because the transient hydrogen bond no longer forms, a significant stabilizing factor for this residue is removed. An increase in flexibility is also consistent with the observation that intermediate exchange behavior at L45 is quenched. In the absence of the transient bond, the loop motions accelerate into the fast exchange regime.

The changes to backbone dynamics in R53A* are more subtle, but also more widespread (Figure 36B) than in R53Q*. Most loop residues experience a significant increase in

flexibility, which is consistent with the observations in R53Q*. V43 appears to rigidify, but this may reflect the appearance of new R_{EX} at that site affecting the T_2 value. R_{EX} models are among the best-fit for this residue, but in general the χ^2 for any model at V43 is very high, suggesting that the model-free model may not be applicable. In both cases, the largest dynamic effects occur in the loop region of eglin c, again suggesting that it is the primary locus of structural change.

For all that the backbone dynamics of the two mutants are similar, the side-chain effects are very different. The R53Q mutation provokes only subtle effects in the core of the protein (Figure 37A, B). The τ_e values of methyls at V53, V34, and V13 change in response to the mutation, but beyond this the mutation appears to have had little effect. In contrast, the R53A* mutation appears to have had a large, generalized effect on side chain motions across the molecule (Figure 37 C, D). Because the changes include both increases and decreases in

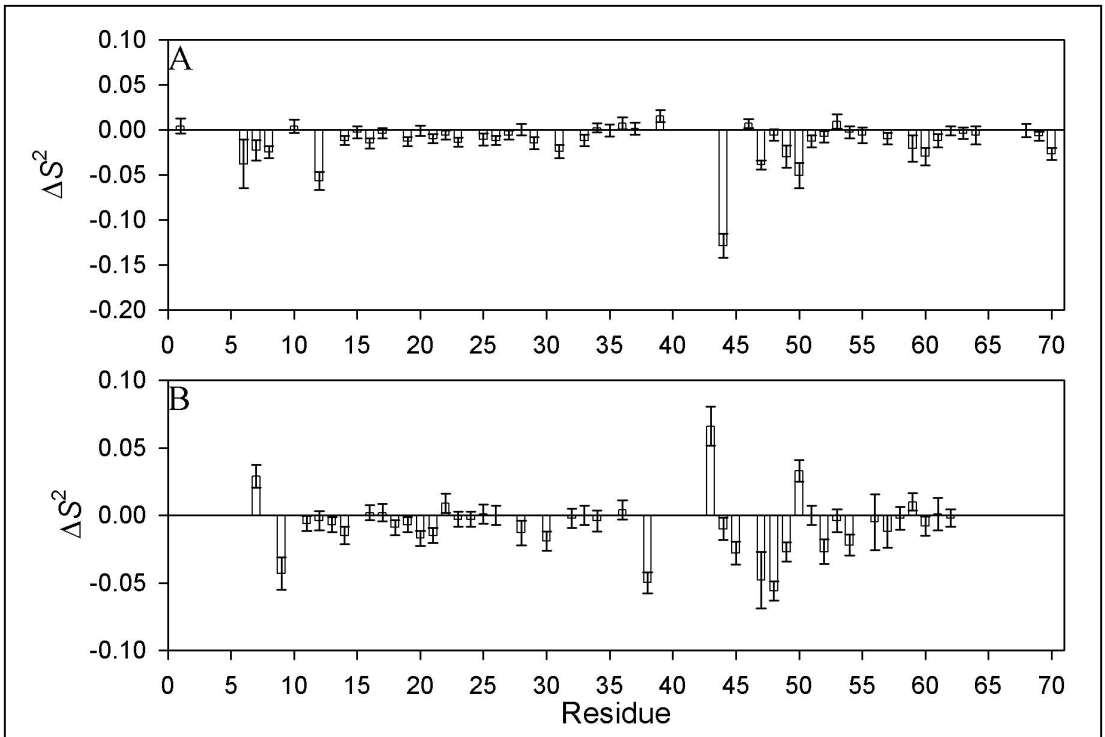

Figure 36: Backbone dynamics changes in R53Q* (A) and R53A* (B)

flexibility and τ_e, it is unlikely that this large perturbation is the result of a fitting or temperature error (as these would produce uniformity). Rather, it is likely that a dual effect is occurring: residues on the loop have rearranged themselves into new environments, causing diverse changes in flexibility, while the core has almost uniformly rigidified. The opposing τ_e effects at V52 in this case are reminiscent of the effect of the V18A mutation. It may be that the central strand (which contains R53) is undergoing a similar fluctuation or conformational shift, although in this case it appears that the dynamic response is more severe.

The R53A* response is also reminiscent of eglin c's side-chain response to a drop in pH. In that case, side chains rigidified uniformly in a response that was associated with a collapse of the loop. The change in S^2_{axis} for these methyl groups in R53A* is not as large as that

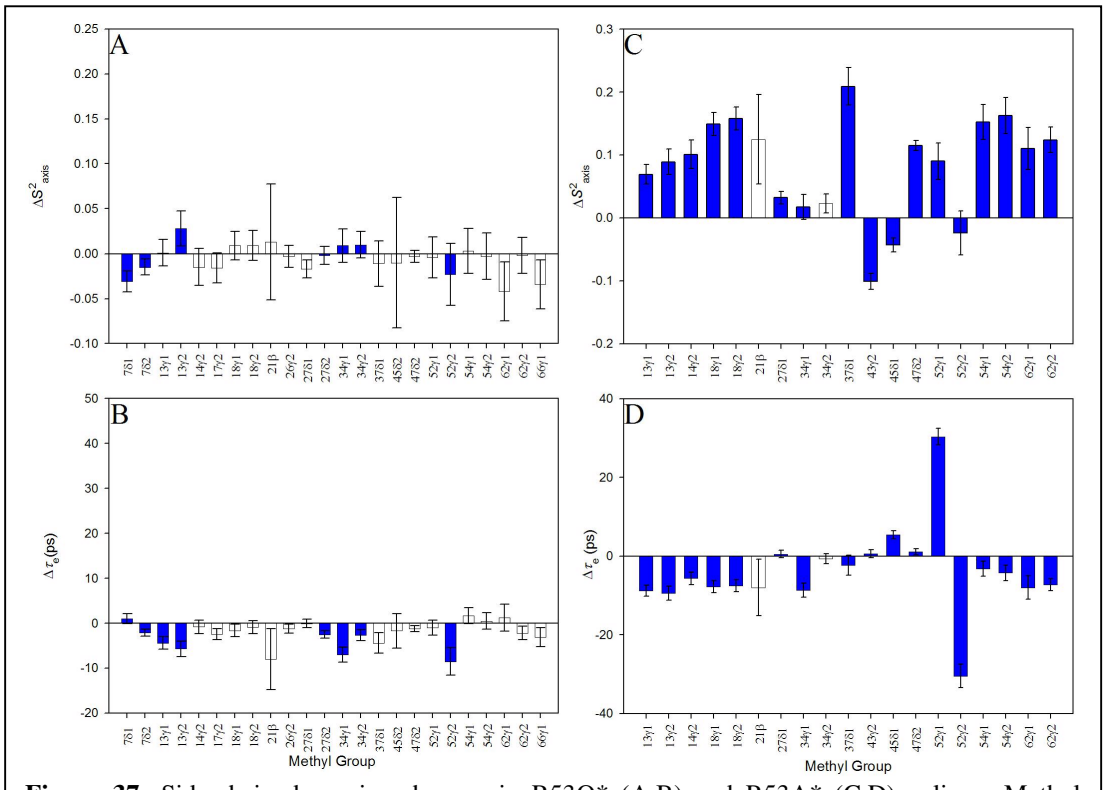

Figure 37: Side-chain dynamics changes in R53Q* (A,B) and R53A* (C,D) eglin c. Methyl groups with significant changes in dynamics are shown in blue.

observed for the pH change. However, in most cases the order parameters for side chains lie between the WT* pH 7 and pH 3 values, suggesting that R53A* is in an intermediary state. It is also possible that these values reflect an average of two states, but there is no evidence of interconversion on the slow or intermediate timescales. The differences between R53A* and pH 3 WT* probably arise from the fact that the pH drop causes charge effects that do not occur in the mutant, and the fact that the mutation has removed a very large side-chain from the loop region. The bulk of the R53 side chain may also account for the differences between R53Q* and R53A*. Even absent the hydrogen bond, the loop may still stay "out" in R53Q* due to the size of the glutamine, or the formation of alternative hydrogen bonds. In R53A*, the removal of that surface, and all hydrogen-bonding potential, may prompt a side-chain reorganization that results in a more notable loop collapse.

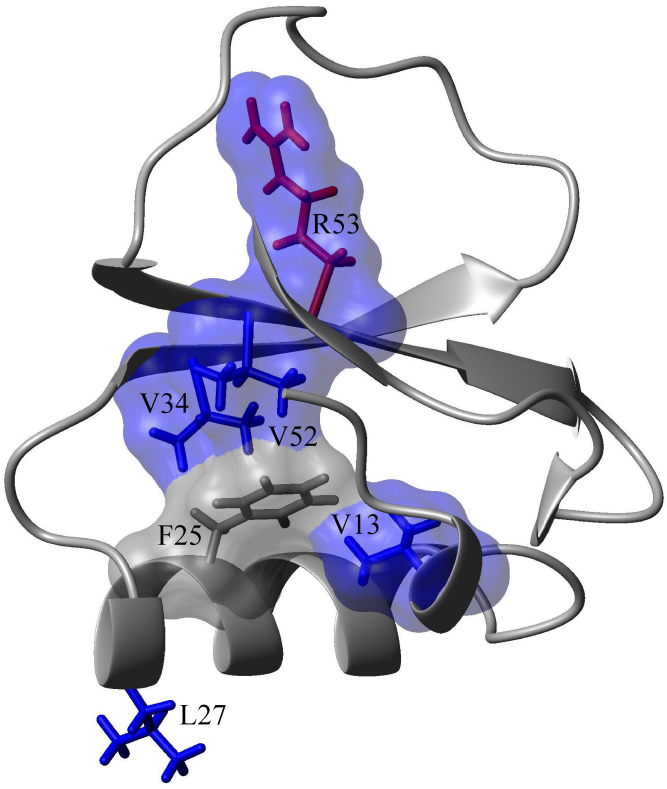

Figure 38: Surface diagram for R53Q* eglin c

In both these cases the observed responses can be rationalized structurally. The R53Q* response, subtle as it is, actually recapitulates some elements of the dynamic pathway observed in the V14A mutant (Figure 38). The dynamic response follows a contiguous pathway from R53 to V52 to V34, then across F25 to V13. Previous experiments on CI-2

produced a similar phenomenon (Leatherbarrow and Matthews, 1992). In those cases, a mutation to R67 of CI-2 (homologous to R53 of eglin c) caused an increase in the flexibility of the indole ring of W24 (homologous to W10 of WT* eglin c). V13, V34, and V52 are all close to W10 in the structure, suggesting that a similar dynamic propagation occurs in these two related proteins. Further experiments to determine the methyl dynamics of CI-2 and its mutants should provide a firmer indication of whether proteins that share sequence and structure also share similar dynamic response pathways. However, it should be noted that the R67A mutation in CI-2 caused a decrease in rigidity, while the similar mutation in eglin c caused an increase. Additionally, the R67A mutation appears to destabilize CI-2 more than the R53A mutation destabilizes eglin c.

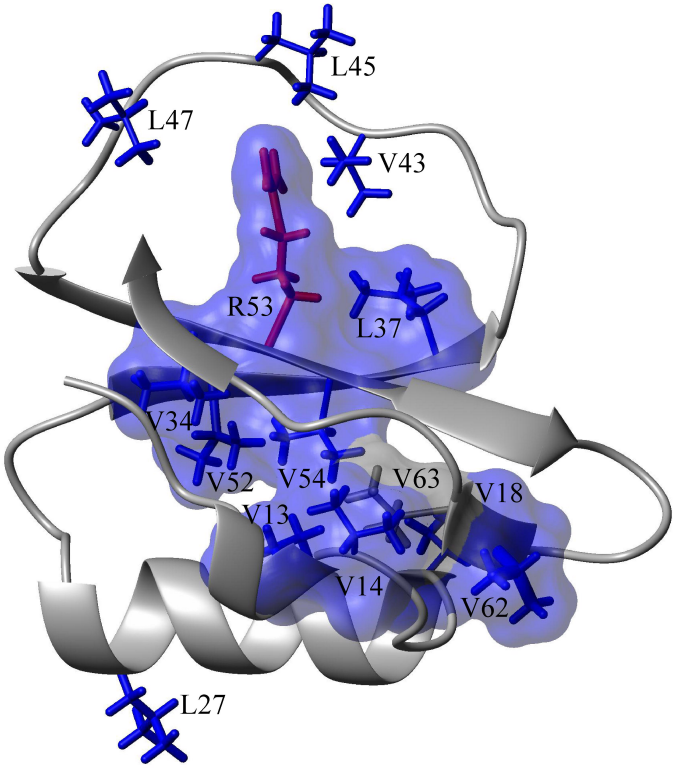

Figure 39: Surface diagram for R53A* eglin c

Although the change in dynamics at L27 cannot be traced through a contiguous network easily, it is consistent with the information that L27 and L45 motions are correlated. Whatever mechanism links them may also be affecting L27 in this case.

The R53A response also reaches a contiguous pathway, though in this case the changes are much more significant and widespread (Figure 39). No effort was made to link residues on

the loop to R53, though L45 and V43 contact it directly; the serious perturbation there makes altered side-chain dynamics inevitable. The change in L37 dynamics, however, may be attributable to direct interactions between its side chain and that of R53. The rest of the internal residues are linked to R53 via V52 and V54. The resulting network looks very much like the V54A* response.

Note that the effect may actually be larger than represented here. The assignments for R53A* did not include information for threonine methyl groups. As a result, the effect of this mutation on dynamics at T17 and T26 is not known.

Notably, several solvent-exposed residues have changes in S^2_{axis}, and several of these effects are as large as those in the interior of the protein. It is not clear why this should be the case, but the sign of the response is structurally coherent, reinforcing its significance.

Conclusions

None of these three mutations produces a clear structural change in eglin c. Chemical shifts and RDCs suggest that V63A* and the R53 mutants have conformations that are substantially similar to that of WT*. Despite its significant effect on protein stability, the V63A mutation causes only a local disturbance to the structure (K16) and the dynamics. The structural response to R53 mutations is less localized, but largely does not perturb the core components of eglin c. Instead, the R53Q* and R53A* mutations appear to change the "structure" of the disordered loop, probably by shifting its average distribution closer to the core of the protein. These mutations therefore do not counter the presumption that a point mutation generally preserves a protein's conformation. However, the relatively small structural changes induced by these mutations have profound functional effects in terms of

both inhibition and stability. Additionally, they reinforce the difficulty in predicting dynamic propagation resulting from mutation.

Although V63 is predicted to be central to several dynamic pathways (Chapter 1), mutation of this residue did not produce a widespread dynamic response. In contrast, mutation of R53 produces two distinct kinds of dynamic response. One alters the motions of the loop in a manner dominated by changes in the motions of the backbone. The other is directed *away* from the loop, towards the interior of the protein, and appears to be largely mediated by the side chains. In both cases, at least some dynamic effects overlap with the locus of structural change. However, in the case of R53A*, the dynamic response also extends beyond this region to reach the opposite side of the protein, in what appears to be a contiguous response pathway, or a generalized loss of flexibility. Even when the structural perturbation is localized, one cannot assume that the dynamic response will be.

Conclusions from Mutational Experiments

The mutations to eglin c that have been discussed induce three separate kinds of responses. In "contiguous network" responses, the mutation produces widespread changes in S^2_{axis} and τ_e that are associated with minimal changes to the overall structure. In these cases, most of the affected side chains can be assembled into contiguous surfaces with the addition of only a few residues that have no supporting evidence. The "disperse network" responses of the V18A and V62A mutations also feature long-range dynamic effects, but in these cases it is not possible to build a contiguous network without numerous hypothetical inclusions. The distal responses in these cases, however, are accompanied by subtle structural changes that may account for the side-chain effects. Finally, structural responses such as those seen in

V63A and the R53 mutants feature clear local disruptions of structure and dynamics, with an extended response in some cases. In one instance – V18A and V54A – the dynamic interactions were clearly associated with energetic connectivity.

In cases where the overall structure is nearly unaffected by a mutation, therefore, it appears that the dynamic effects propagate away from the perturbation site via networks of connecting side-chains. This behavior can be recapitulated in some cases where structural change occurs (namely the R53 mutants), but in those cases it appears that propagation into the core is not associated with structural changes to it. These observations support the idea that correlated motions in the native state are collectively affected by a mutational perturbation. The fact that these systems, though they have many residues in common, do not have *all* residues in common also supports this interpretation. Though these measurements by their nature cannot prove the existence of sub-global motional modes in the native state, the features of these mutants taken together strongly suggest that this is the case.

In mutants where the structure of rigid portions of eglin c changes (V18A*, V62A*, V63A*), dynamic propagation either does not occur or appears to be mediated to a large extent by the altered backbone structure. In the case of V18A* the mechanism of propagation is fairly obvious, but this was not the case for the V62A* mutation. In both these cases the residues inducing disperse responses appear to be dynamically isolated from the core and connected to V63, possibly explaining the similarity of their responses.

In all but one of the cases discussed here, dynamic or structural effects propagate to regions of the protein more than 10 Å away, and beyond the first shell of side-chain interactions. Aside from conclusions about the means for dynamic propagation, these experiments collectively demonstrate that one cannot safely assume that a point mutation

only affects its immediate surroundings. Although the energetic effects at distal locations appear to be minor in many cases, the demonstration of clear energetic interactions between pairs of residues sharing a dynamic response indicates that these long-range responses cannot be disregarded. Even when a mutation does not appear to alter a protein's conformation it is likely to have dynamic and energetic consequences even at residues distant in the structure.

Perspective

The study related here had two purposes: to explore the effects of mutations on a protein from a strictly dynamic point of view, and to determine if there was any correlation between dynamic and energetic linkages in proteins. The approach used – an array of point mutations in the context of a single protein, checking energetic effects after dynamic linkages were found – while acceptably serving both ends, was unfortunately not ideal for either. This work represents an essential first step on the path to a greater understanding of the interplay between dynamics and energetics, but does not on its own provide solid answers to the essential questions it sought to address.

An important point with respect to understanding the dynamic effects of mutations is that these responses are likely to be idiosyncratic and highly depend on the structural context in which the mutations take place. This can be seen even in eglin c itself, where similar mutations in some contexts produce long-range responses in contiguous networks, in others cause more disperse responses, and in still others appear to cause structural changes. Similarly, the very nature of these responses may be a particular property of eglin c. In most cases the core residues of eglin c respond to mutation by rigidifying. Does this signify a global trait of dynamic responses to mutation, or is it simply an effect of the highly

constrained (by virtue of the small valine side chains and large aromatic residues) disposition of eglin c's core? It may be that the V34A mutation provides a paradigm for understanding what occurs in the case of a large-to-small mutation, in that nearby side chains that *can* take advantage of the new conformational space do so on a *slow* timescale, while all other side chains rigidify at the fast timescale. Or perhaps this, too, is a feature unique to eglin c. The only way to truly begin to understand the nature of dynamic effects is to explore them in a variety of proteins with dissimilar folds, sizes, and compositions. A larger body of experience is the only way to clearly separate the universal from the idiosyncratic.

At the same time, even in this context it is possible that more could have been done. Although the results produced were highly interesting, it is possible that alternate mutations would have proved more interesting. Valine-to-alanine mutations, after all, are intended primarily to ablate existing interactions. Valine-to-leucine mutations, on the other hand, might have the effect of creating new ones; providing a mobile link to connect two residues that did not previously interact. Using point mutations that minimally perturb the DNA sequence might also provide important information about the ways in which evolution produces proteins that resist disruption by single mutations. A greater emphasis on mutating aromatic residues might also have benefited this study.

Dynamic studies of double mutants are also a tempting possibility. Such experiments would at least have the advantage of possibly answering some questions about whether particular residues mediate particular pathways. One could, for instance, remove V63 and ask whether a V34A mutation still produces a dynamic effect at V62. At the same time, such experiments are difficult and complicated to interpret. Two mutations, particularly two *interesting* mutations, may be enough to bring eglin c (or another protein for that matter) to

the verge of total instability, making relaxation analysis difficult or at least highly error-prone. Furthermore, the two mutations bring with them an increased risk of structural change, and the difficulty of separating overlapping effects. The researcher's ability to avoid these risks will depend on his ability to select a set of mutations that is *just right*. The single mutations should excite significantly different networks (or the same network in different ways), one should be on a propagation pathway from another, and the double mutation should preserve both the structure and its stability. In the case of eglin c, it is unlikely that any two mutations thus far identified would fulfill these criteria. In this regard, either of the two steps mentioned above (more proteins or more mutations) should come first. An expanded library of single mutants with known effects will make possible an interpretable double-mutant experiment.

There are two conflicting directions here. The first is to study point mutations in a greater variety of protein contexts. The second is to more intensively study a single protein (or even a single site) through an array of possible substitutions. Both are likely to be essential, but one is clearly more practical. Certainly the research program required would be less intense if NMR or other experiments could simply answer the question of whether motions are correlated in the native state and to what degree this has an influence on protein behavior. However, the available means for such investigations (MD primarily) are not yet up to the task of replicating NMR data, much less reliably going beyond it. As such, we are forced to rely upon what we can imply about motions from simple experiments, which means we must have the broadest possible set of data in order to minimize the chance of overgeneralization. Expanding the eglin studies to CI-2 should be only the first step in a broader program

involving many more (and larger) proteins. As time goes on these proteins should not be left behind; rather, the number of mutations should be expanded.

What would justify such an effort? The failure, in most cases, of this study to demonstrate significant long-range energetic connectivity coinciding with dynamic linkage suggests that in many or all cases where residues are dynamically interdependent there will be no apparent energetic consequence. Although the preponderance of negative values in these energetic experiments is interesting, no amount of peering at the numbers would serve to differentiate the cycle in which no dynamic linkage was observed from the others. Even in the case in which the dynamic linkage coincided with an energetic connection, the explanation appeared to be structural in nature. As such, one could reasonably conclude (with the caveats already mentioned) that dynamic connectivity and significant energetic linkage do not regularly coincide. Given the limited set of experiments, one could just as reasonably conclude the opposite, which implies that more work is needed. Addressing the energetic question, however, needs a different approach.

The primary deficiency of the energetic experiments described here lies in the fact that there is no way to know whether they are actually probing an activity related to the dynamic linkage. There are two ways to address this problem. The first is to identify cases in which long-range energetic connectivity appears, and *then* attempt to determine if there is a dynamic linkage that coincides. The energetic experiment in this case retains all its deficiencies; however, the experimental approach compensates by only looking at cases where thermodynamic mutant cycle analysis actually succeeds in identifying a coupling. A set of experiments in this vein would have the advantage of unambiguously indicating whether non-additive cycles coincide with dynamic linkages, though it would neither answer

the question of whether those linkages *caused* the energetic connection, nor determine the energy bound up in dynamic interactions *generally*. The other approach is not available in eglin c, but relies using an alternate energetic probe, perhaps one more directly related to motion. In this regard, ligand binding (not necessarily catalysis, though that might also serve) particularly recommends itself. BIACORE or a similar approach could be used to achieve large-scale, high-quality measurements of affinity that could potentially be sensitive enough (particularly in a high-affinity system) to provide a reliable estimate of the energetic consequences of dynamic linkages.

A thorough understanding of the nature of side-chain motions in proteins and an answer to whether those motions are energetically relevant are both noble pursuits. This work suggests but cannot prove that there are extensive correlated motions in proteins, and indicates that some but not all dynamic interactions coincide with energetic linkages. It thus provides a starting point for addressing both goals. Going forward, from the perspective of funding and impact, an effort should be made to address the second concern first. Long-range dynamic responses are no longer a secret. At one time, nobody cared about these responses simply because they did not know about them. Now nobody cares about long-range dynamic responses because they appear to be energetically and functionally irrelevant. A demonstration of a pattern of coincidence between dynamic responses and energetic linkages will be critical in focusing research and funding interest to support studies to address the more fundamental phenomenology of correlated motions.

CHAPTER IV

POLYMERASE μ BRCT DOMAIN

Introduction

DNA polymerase μ is a member of the Pol X family of DNA polymerases (Aoufouchi et al., 2000) that is involved in the non-homologous end joining (McElhinny et al., 2005) pathway for the repair of double-strand breaks (Mahajan et al., 2002). Like other Pol X family members it possesses an N-terminal BRCT domain (Aoufouchi et al., 2000), which has been shown to be important in the recruitment of the polymerase to the end-joining complex (Ma et al., 2004). This interaction also depends on the presence of Ku, and XRCC4/DNA Ligase IV. The precise mechanism whereby Ku functions to recruit Pol μ remains unclear, however.

Because structural and dynamic information may provide important context for further biological work, the structure and dynamics of the BRCT domain of Pol μ were determined using NMR. The BRCT domain is a common module with loose sequence conservation originally identified as a tandem domain at the C-terminus of the tumor suppressor BRCA1 (Altschul et al., 1997; Callebaut and Mornon, 1997). BRCT domains frequently appear in this tandem setup, and are generally involved in protein-protein interactions. Some tandem BRCT repeats and isolated BRCT domains have been shown to bind phosphopeptides (Manke et al., 2003; Yu et al., 2003); this is believed to be a general feature of their functionality. However, it is not known if Pol μ specifically possesses the ability to bind

phosphopeptides or if this is the means by which it is recruited to double-strand breaks by Ku.

Methods

Pol μ BRCT was expressed using a pET28a vector (Novagen) in *E. Coli* BL21(DE3) cells; the DNA was a gift from D. Ramsden and the vector was prepared by S. Gilmore. The construct begins at position 20, with an S to G mutation at this site; residue numbers for the features described here identify this site as residue 1. The construct ends at glutamate 124 of the native protein (here E105). Bacteria were grown in M9 minimal media containing $^{15}NH_4Cl$ and/or [U-^{13}C]D- Glucose (Cambridge) as the sole nitrogen or carbon source to achieve the appropriate labeling. Cells transformed with the pET28a – BRCT plasmid were grown to an OD_{600} of 0.6-0.8 while shaking at 37 ºC. The cultures were then cooled with ice to below room temperature and induced with 0.4 mM IPTG. Induction lasted 5 hours, shaking at 22 ºC, after which cells were pelleted by centrifugation, then resuspended in 25 mM Tris pH 8.0 and then frozen. After several rounds of freezing and thawing, the cell slurry was lysed by sonication for at least 4 rounds of 3 minutes at a 50% duty cycle.

Crude lysates were centrifuged for 30 minutes at 23,000 *g* and the pellets discarded. The supernatant was dialyzed for 4-16 hours against 2L of 25 mM Tris, 5 mM DTT, pH 8.0 at 4 ºC. The dialyzed lysate was then fractionated over a column containing S-Sepharose Fast Flow resin (Pharmacia) at 4 ºC, and eluted via a step gradient using a buffer with 25 mM Tris pH 8.0, 150 mM NaCl, and 2 mM DTT. The protease inhibitor AEBSF (Roche) was then added to the fraction presumed to contain BRCT at the maximum recommended concentration, because these buffer conditions are unfavorable for AEBSF.

Following incubation with AEBSF, the BRCT fraction was concentrated at 4 ºC and further purified over a G-75 Sephadex column (Pharmacia), with a buffer of 25 mM Tris, 150 mM NaCl, 2 mM DTT, and 0.02% NaN_3. Purity was assessed via SDS-Page, and fractions containing BRCT were then further concentrated to < 1 mL and dialyzed against 500 mL BRCT NMR buffer [25 mM deuterated Tris, 50 mM KCl, 5 mM DTT, 0.02% NaN3, 10% D_2O, pH 7.9].

Spectra of BRCT were acquired at 10 ºC, but acquisition and analysis of NMR relaxation data were in all other respects identical to the description in Chapter 1 for eglin c. Assignments and structures were a gift from E. DeRose of the National Institute of Environmental Health Sciences.

Alignments with TdT were performed using BLAST (Altschul et al., 1997).

Results and Discussion

The BRCT domain of Pol μ has an α/β construction with helices packed around a central sheet (Figures 40, 42). BRCT domains have highly diverse sequences (altschul), and the Pol μ domain is most closely related to that of TdT (Figure 40). High-quality relaxation data were collected at 500 and 600 MHz fields (Figure 41), but at the temperature used, the insensitivity of the NOE experiment made the noise a more significant consideration. These

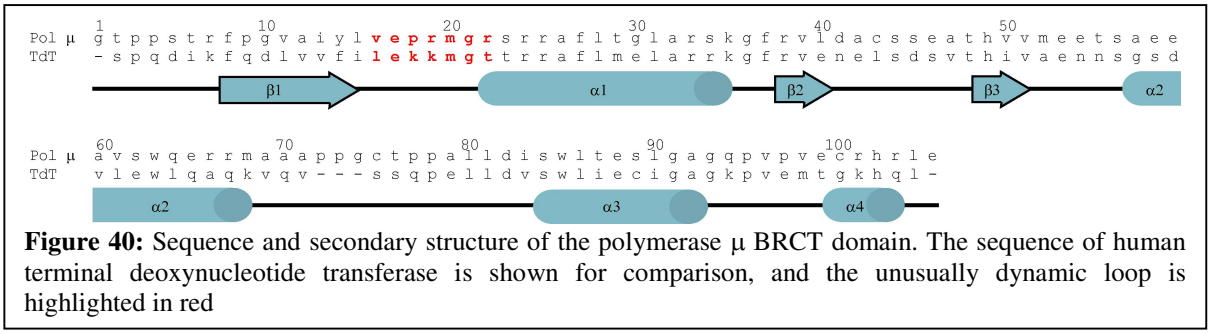

Figure 40: Sequence and secondary structure of the polymerase μ BRCT domain. The sequence of human terminal deoxynucleotide transferase is shown for comparison, and the unusually dynamic loop is highlighted in red

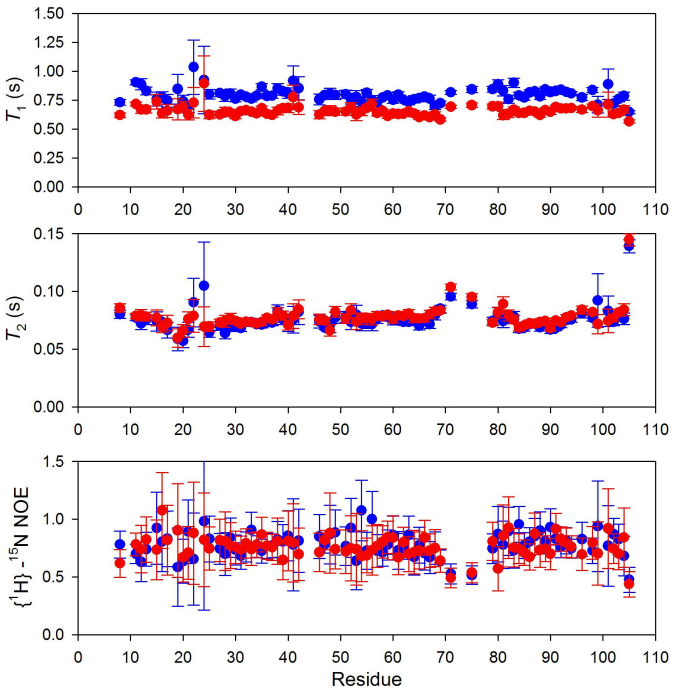

Figure 41: Relaxation data for BRCT at 500 (red) and 600 (blue) MHz.

data, however, were automatically given less weight in the fitting algorithm.

Fits of order parameters assumed isotropic rotation. Anisotropy fits indicated that the rotational diffusion tensor was not perfectly spherical ($D_{\parallel}/D_{\perp} \approx 1.1$), but the use of an anisotropic tensor did not improve the χ^2 values associated with fitting to any model. The model-free parameters produced by the isotropic rotational fit are, as is typical for folded proteins, high and fairly uniform across the backbone (Figure 42). Order parameters decrease slightly near solvent-exposed loops, and the loops themselves often suffer from intermediate exchange or other processes that attenuate peak height and make fitting of order parameters difficult.

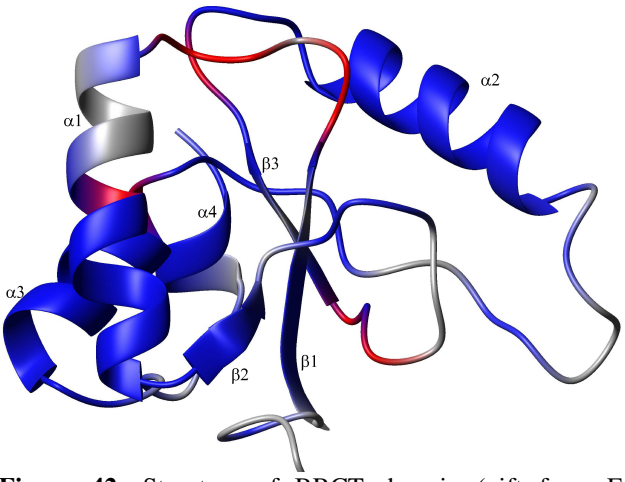

Figure 42: Structure of BRCT domain (gift from E. deRose) with S^2 mapped onto it. Intensity of blue indicates magnitude of S^2. Residues for which relaxation data were collected but which could not be fit are shown in red. The 16-22 loop is at the top of the molecule.

Even considering well-defined peaks, however, some residues were not amenable to being fit to the

standard formulation. Most notably, each member of a stretch of residues from V16 to S22 causes the fitting procedure of relxn2.2 to go out of bounds when trying to fit a model that contains S^2 and τ_e. In most of these cases, alternative formulations of the spectral density proved effective in producing a fit. Nonetheless, the dynamics in this region may still be too complex to describe using any model-free model.

The primary evidence for this proposition is that the peaks in this region are less intense than their neighbors, and have unusually short T_2 values. This indicates that some chemical exchange process is occurring in the area of these residues. However, a model-selection algorithm based on the Akaike and Bayesian information criteria (Chen et al., 2004; d'Auvergne and Gooley, 2003) selects an R_{EX} parameter only for M20, and even then the apparent exchange rate (at 500 MHz) is not well-determined: $4x10^4 \pm 2x10^4$ s^{-1}.

The other residues in this stretch can be fit if τ_e is fixed near zero while S^2 is allowed to vary, but this may be an artifact of the fitting procedure. The fitting program relxn2.2 uses a comprehensive grid search for fits in which S^2 alone is allowed to vary, while a Powell minimization is used for all other fits. The comprehensive grid search has hard boundaries, while the Powell minimization can escape its initial bounding conditions. Because the order parameters fit using S^2 only are very high and have large errors, it is likely that they are pushing towards the edge of the bounding conditions. Were the "S^2 only" fits to proceed using a Powell minimization, they would probably become as unbounded and ill-defined as the two-parameter fits.

Attempts to fit the order parameters using a Powell minimization only succeeded when τ_m was allowed to vary locally. Even in this case the errors in S^2 were large, but the values were moderate. The average local τ_m for the stretch from 16-22 was 13.7 ns, compared to 10.4 ns

for the global fit. The low T_2 values for these residues give rise to the elevated rotational correlation times. Again, this suggests a chemical exchange process, though models including R_{EX} parameters were not selected. Additionally, the ratio of T_2 values at 600 MHz to those at 500 MHz was unusually low for several of these residues, namely E17, M20, and G21, which is consistent with a chemical exchange process because these effects scale with field strength. Because the NOE values in this region are also depressed, it is reasonable to conclude that the observed peak attenuation and chemical exchange are due to conformational flexibility of the backbone, and not due to motions of adjacent regions of the protein.

This finding of conformational flexibility is somewhat surprising because it appears that the loop in question has some contacts with residues in the adjacent loop from M52 to S56 that should stabilize it. Additionally, the side chains for V16 and M20 point towards the core of the protein, which should provide additional anchor points to hold the loop in place. However, E53 of the second loop also experiences complicated dynamics, suggesting the possibility of a shared motion. Additionally, the conformational plasticity in this region has correlations with observations in other BRCT domains (Figure 43). As shown here, a structural alignment of the Pol μ BRCT domain with the 2nd BRCT domain of

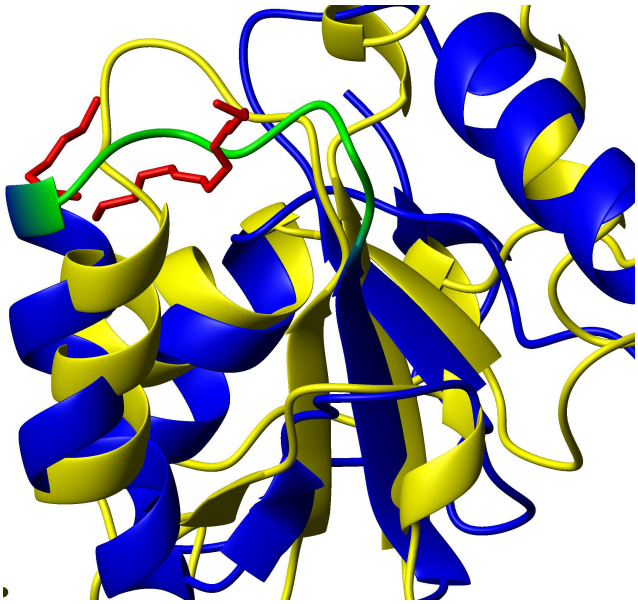

Figure 43: Overlay of BRCA1 BRCT2 (yellow) and Pol μ BRCT (blue). The BACH1 phosphopeptide is shown in red and 16-22 loop in green. Structure from Clapperton et al., 2004

BRCA1 indicates that the dynamically complex 16-22 loop of Pol μ is structurally equivalent to a loop of BRCA1 BRCT2 that directly contacts the BACH peptide (Clapperton et al., 2004; Williams et al., 2004). It may therefore be that the observations in Pol μ speak to the existence of a binding activity that has not yet been defined.

The possibility of phosphopeptide binding capability in Pol μ would seem to be countered by the fact that this activity has mostly been found in BRCT domains with tandem repeats (Manke et al., 2003; Yu et al., 2003). In addition, the BRCT2 domain of BRCA1 primarily serves to define the binding pocket and provide hydrophobic surfaces. However, the 16-22 loop of Pol μ BRCT contains two arginines that could stabilize the charge of a phosphate. Additionally, the single BRCT domain of TdT was also found to possess phosphopeptide binding activity (Yu et al., 2003), and the Pol μ BRCT domain is closely related in terms of sequence (Figure 40). Several residues in this loop are conserved between these two domains, and positive charge is also conserved at one of the arginines. It is therefore possible that Pol μ possesses a similar ability to bind phosphopeptides on its own. Future experiments should clarify what role this stretch of residues plays in targeting Pol μ to the end-joining complex and whether recognition is phosphorylation-dependent.

APPENDIX

TCL TOOLS FOR RELAXATION DATA ANALYSIS

Introduction

The principal steps in the analysis of NMR relaxation data are the extraction of peak intensities, the fitting of those intensities to monoexponential decays, the determination of a global correlation time, the assessment of tumbling anisotropy, and the fitting of relaxation data to the model-free dynamics formalism. More steps may be added – the inclusion of a modeł s election algorithm, or close inspection of the decay curves – but these represent the fundamental and necessary steps. Early computational protocols required the execution of at least one TCL or perl script for every step, and additional scripts to collate data so that it could be interpreted by the in-house programs expfit2, relxn2.1 (and eventually 2.2) (Lee et al., 1999), and qfit. Additionally, visual inspection of the data was difficult at many points due to its separation into separate files for each residue.

This prompted the creation of 2 scripts to streamline data processing and analysis. The dependent TCL script “relax_extract.tcl” is used to extract relaxation rates from pseudo-3D NMR datasets, compare duplicate points to assess errors, create necessary adjunct files, and execute expfit2 to fit the extracted intensities to monoexponential decays. It also provides a convenient way to visually inspect relaxation data. The independent TCL application rvi provides a graphical front end to collate relaxation data, select and exclude individual pieces of data, interpret the possibility of backbone flexibility, transparently control fitting options, and run both relxn2.2 and qfit. Additionally, rvi provides a model-selection algorithm for the fitting of relaxation data. These programs were created to minimize the amount of time spent

on script creation and editing so as to focus more laboratory time on data interpretation and analysis.

Methods

These programs are both written in the Tool Control Language and its Toolkit (TCL/Tk), initially developed by John Ousterhout (Ousterhout, 1994). Because it makes use of commands and variables that are unique to the NMR data analysis program NMRView, relax-extract.tcl was developed to be run using NMRView's internal TCL console, and makes use of the native NMRView TCL libraries. As a result, relax_extract.tcl *cannot* be executed independently of NMRView.

In contrast, rvi uses more advanced functions of TCL/Tk, it *cannot* be executed within NMRView. To use rvi on a given system, TCL/Tk libraries must be downloaded and installed independently. These libraries are freely available on the web; download locations are listed at the TCL developers' site (www.tcl.tk). The program works best with libraries from the ActiveTCL distribution available from www.activestate.com. In addition, rvi makes use of the "expect" package by Don Libes (which comes with ActiveTCL), and the "scrolledframe" package by ulis (available at the developers' site, comes packaged with rvi). Additionally, rvi requires that the user be able to execute relxn2.2 or qfit by typing the program name on the command line – this requires that these programs or a link to them be placed in a directory that is in the users PATH. It is not enough to create an alias.

Protocol

Figure 44: Main window of relax_extract.tcl

A pseudo-3D relaxation dataset and peaklist are required to use relax_extract.tcl. Once the dataset and peaklist have been loaded into NMRView, the script is called from the TkCon command line. This brings up the main window (Figure 44). The dataset and peaklist to be used in the analysis are selected from the pulldown menus, and the number of planes to be analyzed, kind of relaxation, and initial guess for relaxation time are put into the first row of entry boxes. The destination directory for the files created by relax_extract.tcl is typed into the large entry box, and the filename root and names of the log file, summary file, and name file are typed in the remaining boxes. The delays associated with the planes can be typed into a special window or read in from a file which begins with a single comment line and in which each data line has the relaxation delay as the final element separated by whitespace.

Once these fields are set up to the user's satisfaction, clicking on the "Extract Data" button will begin the extraction algorithm (which is based on earlier incarnations of this script by A. Lee). This algorithm makes use of the NMRView 'jitter' variable to obtain the greatest intensity from the inner 50% of the peak box; this result is compared to the 'max' (the greatest intensity in the entire peak box) in order to identify overlapped or misplaced peaks. The intensities are then read into an

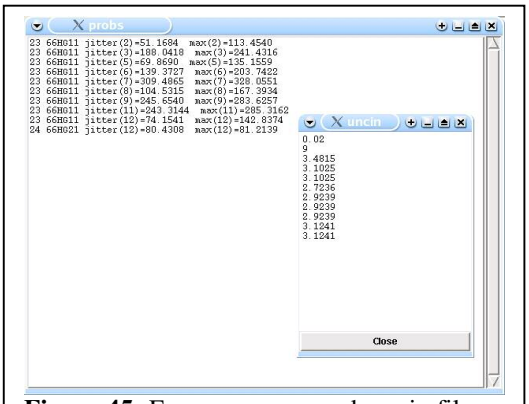

Figure 45: Error messages and unc.in file

internal array and rearranged based on the entered relaxation times. Duplicate points are automatically identified and compared; all the information produced in these steps is written to a log file, while the net RMS error is written to a file named "unc.in" which contains uncertainty information for use by expfit. Onscreen, the program displays a text box containing any error messages from the extraction, and a smaller box showing the contents of the "unc.in" file (Figure 45). The error messages and the raw data used to determine peak uncertainties for the unc.in file are stored in a log file. Residue names are simultaneously extracted to a “namefile” for use with rvi.

The onscreen display also includes the "Residue Selection Window" (Figure 46). Using this window the user can exclude residues from the fitting procedure. Additionally, if the user double-clicks on a residue name, relax-extract.tcl creates an instance of gnuplot displaying the decay of peak intensity. This allows a visual inspection of the data before proceeding to exponential fitting and can help the user to assess the quality of the data independent of the χ^2 values from the fitting procedure. However, due to quirks of the gnuplot program, the graph window should never be closed manually; relax_extract.tcl will close it properly once the user chooses to either proceed or quit.

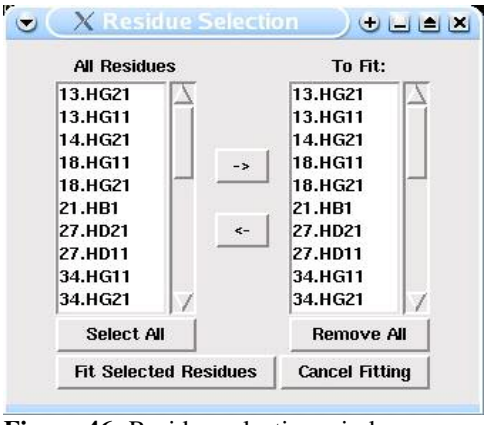

Figure 46: Residue selection window

Once the user is satisfied with the group of residues designated for fitting, clicking on the "Fit Selected Residues" button will call expfit to produce a monoexponential fit of each residue's dataset. The output from each of these procedures is saved to a ".out" file and the

time constant (the reciprocal of the fitted rate) is extracted to a summary file and displayed in a text window. Depending on the number of residues and the amount of time the researcher wishes to spend inspecting the data, a complete analysis of a pseudo-3D can take as little as 10 minutes.

It should be noted that relax_extract.tcl can only handle datasets that involve monoexponential decays. As such, it is not an appropriate tool for analyzing $\{^{1}H\}$-^{15}N NOE spectra.

Once relaxation data have been pulled from NMRView, rvi can be run from the command line. All data files produced by relax_extract.tcl can be read by rvi, but data files from any source can be made readable by making note of the following format expectations. For all data files, a "column" should be understood to be a piece of data separated by whitespace. Any initial whitespace is disregarded; the first column is the first piece of actual information. For all file types rvi expects this column to contain the residue name. For ^{15}N and ^{13}C T_1 and T_2 data, rvi expects to find the time constant in the fourth column and the error in the sixth column. For NOE data, rvi expects the value in the second column and the error in the third column. For D_z and D_y data, rvi expects the time

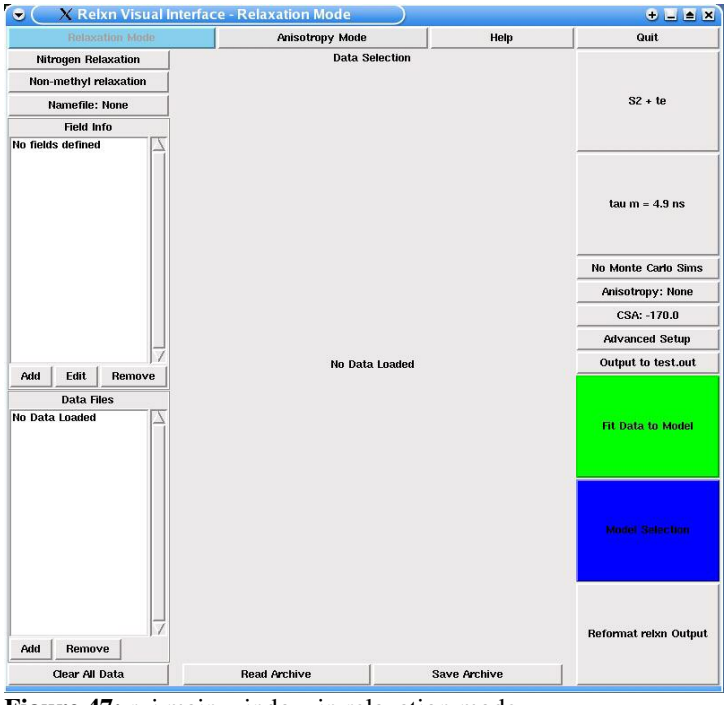

Figure 47: rvi main window in relaxation mode

constant to be in the fifth column and the error to be in the sixth column. Once a data file has been read into rvi there is no need to do it again: rvi has the ability to store data archives that contain all the necessary information from a dataset. These archives are a convenient way to store collated data.

The operation of rvi proceeds in a natural way suggested by its layout (Figure 47). In either relaxation mode or anisotropy mode the correct thing to do is work through each column from top to bottom, handling the columns from left to right. For a relaxation analysis, one first selects the nucleus and indicates if methyl groups are the subject of the study. The user must then load a *namefile* containing one unique entry for every vector intended for consideration. In the "Field Info" box, the user then clicks "Add" to enter the Larmor frequencies of the nuclei for that particular magnet. Once all magnets have been entered, the user clicks "Add" under "Data Files" to bring up a data selection window (Figure 48). Clicking on a data file will bring up a secondary window where the user can indicate what kind of relaxation data is being loaded and at what field it was acquired. If significant errors are made, the user can click "Clear All Data" to start over.

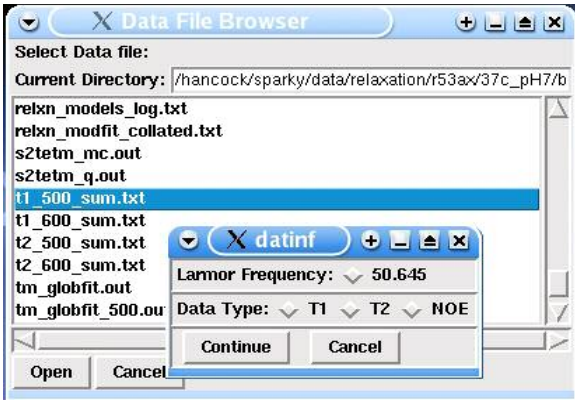

Figure 48: Primary and secondary data selection windows

Once all the desired data has been loaded, the user should proceed to the central column for data selection. At the bottom of this column, a user can choose to either load an existing data archive, or save a new one based on the data already loaded. In the example shown here, a ^{15}N relaxation dataset has been loaded (Figure 49). The variable colors indicate the

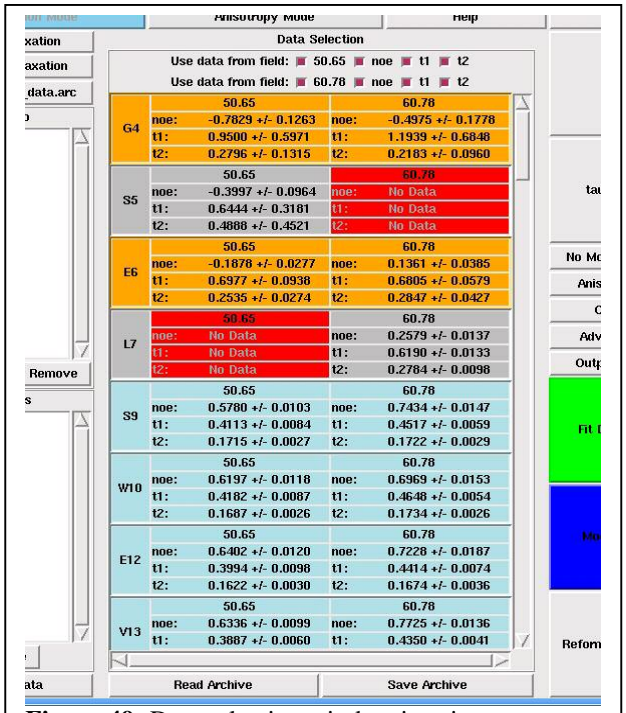

Figure 49: Data selection window in rvi

suitability of each residue to be used in a global analysis of τ_m. Residues in Carolina blue may be safely used, residues in orange have $\{^1H\}$-^{15}N NOE values that are too low, and residues in yellow have acceptable NOE values but have a relationship between T_1 and T_2 that suggests the possibility of R_{EX} or other flexibility (Tjandra et al., 1995). A red background for a piece of data indicates that it will not be used in the analysis because the user has excluded it (or it has not been loaded). Residues that have gray backgrounds do not have sufficient data to determine whether they are appropriate for a global τ_m fit or not. If the data are for nuclei other than ^{15}N, they will all appear in Carolina blue.

The user can choose to exclude any single piece of data by clicking on it. To exclude all data from a single field for a single residue, click on the Larmor frequency of that field next to the residue. To exclude all data for a given residue, the user clicks the residue name. At the top of the selection window are global checkboxes, with which the user can exclude all the data for a given field, or all of a particular type of data for that field.

Once the user has chosen what data to exclude, he should press the model selection button, which always bears the identifier of the model currently selected. This initiates a popup containing all the available models. Nine models are available, listed by their parameters

being fitted for each: Global τ_m; S^2 only; S^2 & τ_e; S^2, τ_e, & τ_m; S^2, τ_e, & CSA; S^2, τ_e, τ_m, & CSA; S^2, τ_f, & τ_s; S^2_f, S^2_s, & τ_s; S^2 & R_{EX}; and S^2, τ_e, & R_{EX}. Of these, only the "Global τ_m" model minimizes a parameter using all residues; every other model fits all parameters locally. Once a model has been chosen, the τ_m button should be pressed. This brings up a window to enter a new τ_m value, or in the case of the "Global τ_m" model, a window with upper and lower limits and a stepsize for the τ_m fit. By pressing the "Monte Carlo Sims" button a user can choose whether to use Monte Carlo simulations, and if so, how many. The Anisotropy button will bring up a menu to input corrections based on anisotropic tumbling; this approach requires a separate file containing direction cosines for all vectors. That file *must* use the same names as the relaxation data and must be in the same order. The CSA button will allow the user to alter the CSA value in use. The "Advanced Setup" button should be avoided unless the user knows precisely how relxn2.2 works, and contains fields for setting grid conditions, search types, and data weighting.

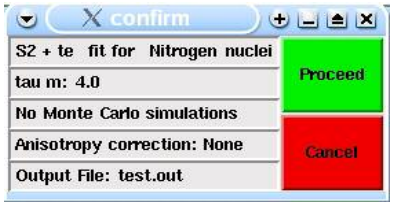

Figure 50: Confirmation window

Once the user has set up the data and fitting parameters to his satisfaction, he can define a new filename for the relxn2.2 output by pressing the "Output to:" button. This accomplished, he can press the "Fit Data to Model" button, which will bring up a confirmation window (Figure 50), and then a monitoring window that contains the screen output from relxn2.2. The interplay between rvi and relxn2.2 is handled by the Expect package, and allows the user to cancel and monitor a fit in progress, and also to interact with relxn2.2 if something goes wrong.

Once the fit is completed, the "Reformat relxn Output" button takes the user to a screen where raw relxn2.2 output files can be converted into easier-to-read text files.

Figure 51: Overall and residue-specific model selection windows

If the user wishes, most of the third column can be skipped by pressing the "Model Selection" button, though this should only be done after the user has fit a satisfactory τ_m. The model selection algorithm of rvi initially performs a quick run of five possible model-free models without implementing Monte Carlo simulations. Once this run is complete, rvi calculates an AIC and BIC value for each fit, and opens a model selection panel (Figure 51) containing the parameters of the model with the best AIC value. If the AIC and BIC disagree, the button is yellow; otherwise it is Carolina blue. When a button is clicked, it opens a window containing the χ^2, AIC, and BIC values for each model, along with buttons that allow the user to select which one will be applied (Figure 51).

Once the user is satisfied with the model selected for each residue, rvi will proceed to fit each residue to its selected model using Monte Carlo simulations (the user does not need to press the "Monte Carlo Sims" button). When the run is complete, rvi will automatically collate the output into a single file, saving it to disk and displaying it onscreen.

The Anisotropy Mode follows a similarly logical protocol. Running an anisotropy fit requires that the user have performed a three-parameter (S^2, τ_e, τ_m) fit with Monte Carlo simulations. The raw output of this fit from relxn2.2 may be loaded in Anistropy Mode by pressing "Format Input File" (Figure 52). Alternately, if the user has already created an input file for qfit (automatically saved by rvi as qfit_data.inp) then that can be loaded using the

"Read Data from File" button. One or more PDB files can be selected for use with the "Add" button at the bottom of the column. Because PDB files are not sufficiently standardized, it is highly recommended that only PDB files saved from MOLMOL be used, as these are sure to have a format qfit can handle.

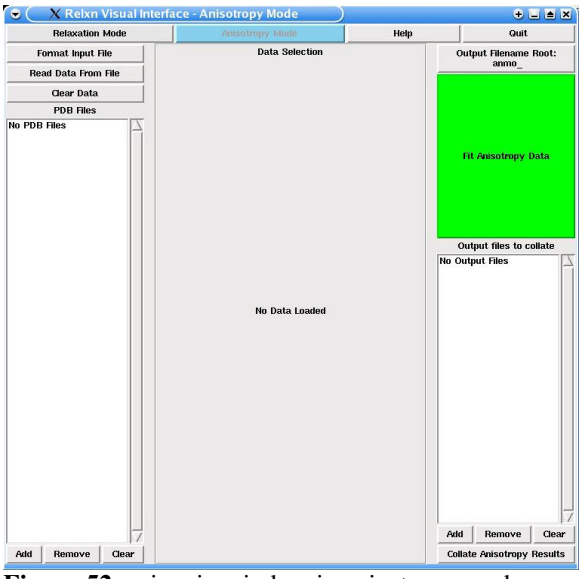

Figure 52: rvi main window in anisotropy mode

Once the data have been loaded, the individually-fit τ_m values and their associated errors will appear in the central "data selection" column. Values that appear in yellow lie more than a standard deviation away from the average, and may cause qfit to fail. They can be de-selected by clicking the adjacent box. Once the user has excluded data to his satisfaction, he can define a new output root: this is a root name for the output file, to which a number will be appended to indicate which PDB file from the list it used. The large green button initiates qfit, again using Expect, and produces interaction screens similar to those for relxn2.2. Once the fit is complete, the output files can be selected using the "Add" button at lower right, and then collated using the bottom button in the third column. The output from this collation includes averages from the Monte Carlo simulations for each PDB file, and averages across all simulations.

The standalone rvi program greatly streamlines data selection and manipulation for relxn2.2 and qfit, which due to their inflexible input formats can be otherwise difficult to work with.

REFERENCES

1. 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.

2. Aoufouchi, S., Flatter, E., Dahan, A., Faili, A., Bertocci, B., Storck, S., Delbos, F., Cocea, L., Gupta, N., Weill, J. C., and Reynaud, C. A. (2000). Two novel human and mouse DNA polymerases of the polX family. *Nucleic Acids Res.* **28**, 3684-3693.

3. Bax, A., Vuister, G. W., Grzesiek, S., Delaglio, F., Wang, A. C., Tschudin, R., and Zhu, G. (1994). Measurement Of Homonuclear And Heteronuclear J-Couplings From Quantitative J-Correlation, In Nuclear Magnetic Resonance, Pt C, pp. 79-105.

4. Bode, W., Papamokos, E., and Musil, D. (1987). The high-resolution X-ray crystal structure of the complex formed between subtilisin Carlsberg and eglin c, an elastase inhibitor from the leech Hirudo medicinalis. Structural analysis, subtilisin structure and interface geometry. *Eur. J. Biochem.* **166**, 673-692.

5. Callebaut, I., and Mornon, J. P. (1997). From BRCA1 to RAP1: A widespread BRCT module closely associated with DNA repair. *FEBS Lett.* **400**, 25-30.

6. Cavanagh, J., Fairbrother, W. J., Palmer, A. G., and Skelton, N. J. (1996). *Protein NMR Spectroscopy: Principles and Practice* (New York: Academic Press).

7. Ceruso, M. A., Grottesi, A., and DiNola, A. (2003). Dynamic effects of mutations within two loops of cytochrome c_{551} from *Pseudomonas aeruginosa*. *Proteins* **50**, 222-229.

8. Chen, J., and Stites, W. E. (2001). Energetics of side chain packing in staphylococcal nuclease assessed by systematic double mutant cycles. *Biochemistry* **40**, 14004-14011.

9. Chen, J. H., Brooks, C. L., and Wright, P. E. (2004). Model-free analysis of protein dynamics: assessment of accuracy and model selection protocols based on molecular dynamics simulation. *J. Biomol. NMR* **29**, 243-257.

10. Chou, J. J., Case, D. A., and Bax, A. (2003). Insights into the Mobility of Methyl-Bearing Side Chains in Proteins from $^{3}J_{CC}$ and $^{3}J_{CN}$ Couplings. *J. Am. Chem. Soc.* **125**, 8959-8966.

11. Clapperton, J. A., Manke, I. A., Lowery, D. M., Ho, T., Haire, L. F., Yaffe, M. B., and Smerdon, S. J. (2004). Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer. *Nature Struct. Mol. Biol.* **11**, 512-518.

12. Clarkson, M. W., and Lee, A. L. (2004). Long-range dynamic effects of point mutations propagate through side chains in the serine protease inhibitor eglin c. *Biochemistry* **43**, 12448-12458.

13. Constantine, K. L., Friedrichs, M. S., Wittekind, M., Jamil, H., Chu, C. H., Parker, R. A., Goldfarb, V., Mueller, L., and Farmer, B. T. (1998). Backbone and side chain dynamics of uncomplexed human adipocyte and muscle fatty acid-binding proteins. *Biochemistry* **37**, 7965-7980.

14. Cooper, A., and Dryden, D. T. F. (1984). Allostery without conformational change - a plausible model. *Eur. Biophys. J.* **11**, 103-109.

15. d'Auvergne, E. J., and Gooley, P. R. (2003). The use of model selection in the model-free analysis of protein dynamics. *J. Biomol. NMR* **25**, 25-39.

16. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe - a multidimensional spectral processing system based on unix pipes. *J. Biomol. NMR* **6**, 277-293.

17. Edgell, M. H., Sims, D. A., Pielak, G. J., and Yi, F. (2003). High-precision, high-throughput stability determinations facilitated by robotics and a semiautomated titrating fluorometer. *Biochemistry* **42**, 7587-7593.

18. Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D., and Kay, L. E. (1994). Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. *Biochemistry* **33**, 5984-6003.

19. Finerty, P. J., Muhandiram, R., and Forman-Kay, J. D. (2002). Side-chain dynamics of the SAP SH2 domain correlate with a binding hot spot and a region with conformational plasticity. *J. Mol. Biol.* **322**, 605-620.

20. Fuentes, E. J., Der, C. J., and Lee, A. L. (2004). Ligand-dependent dynamics and intramolecular signaling in a PDZ domain. *J. Mol. Biol.* **335**, 1105-1115.

21. Green, S. M., and Shortle, D. (1993). Patterns of Nonadditivity between Pairs of Stability Mutations in Staphylococcal Nuclease. *Biochemistry* **32**, 10131-10139.

22. He, M. M., Wood, Z. A., Baase, W. A., Xiao, H., and Matthews, B. W. (2004). Alanine-scanning mutagenesis of the beta-sheet region of phage T4 lysozyme suggests that tertiary context has a dominant effect on beta-sheet formation. *Protein Sci.* **13**, 2716-2724.

23. Hennig, M., Bermel, W., Spencer, A., Dobson, C. M., Smith, L. J., and Schwalbe, H. (1999). Side-chain conformations in an unfolded protein: chi(1) distributions in

denatured hen lysozyme determined by heteronuclear C-13, N-15 NMR spectroscopy. *J. Mol. Biol.* **288**, 705-723.

24. Hu, H., Clarkson, M. W., Hermans, J., and Lee, A. L. (2003). Increased rigidity of eglin c at acidic pH: Evidence from NMR spin relaxation and MD simulations. *Biochemistry* **42**, 13856-13868.

25. Hu, H., Hermans, J., and Lee, A. L. (2005). Relating side-chain mobility in proteins to rotameric transitions: Insights from molecular dynamics simulations and NMR. *J. Biomol. NMR* **32**, 151-162.

26. Hunt, J. A., and Ingram, V. M. (1959). A terminal peptide sequence of human haemoglobin? *Nature* **184**, 640-641.

27. Hyberts, S. G., Goldberg, M. S., Havel, T. F., and Wagner, G. (1992). The solution structure of eglin-c based on measurements of many NOEs and coupling-constants and its comparison with x-ray structures. *Protein Sci.* **1**, 736-751.

28. Ichiye, T., and Karplus, M. (1991). Collective motions in proteins: a covariance analysis of atomic fluctuations in molecular dynamics and normal mode simulations. *Proteins* **11**, 205-217.

29. Jackson, S. E., and Fersht, A. R. (1993). Contribution of long-range electrostatic interactions to the stabilization of the catalytic transition state of the serine protease subtilisin BPN'. *Biochemistry* **32**, 13909-13916.

30. Johnson, B. A., and Blevins, R. A. (1994). NMRView: A computer program for the visualization and analysis of NMR data. *J. Biomol. NMR* **4**, 603-614.

31. Johnson, E. C., and Handel, T. M. (1999). Effect of hydrophobic core packing on sidechain dynamics. *J. Biomol. NMR* **15**, 135-143.

32. Koradi, R., Billeter, M., and Wuthrich, K. (1996). MOLMOL: a program for the visualization and analysis of macromolecular structures. *Journal of Molecular Graphics* **14**, 51-55.

33. Leatherbarrow, R. J., and Matthews, S. J. (1992). Carbon-13 nuclear magnetic resonance relaxation study of chymotrypsin inhibitor-2 (CI-2). *Magn. Reson. Chem.* **30**, 1255-1260.

34. Lee, A. L., Flynn, P. F., and Wand, A. J. (1999). Comparison of ^{2}H and ^{13}C NMR relaxation techniques for the study of protein methyl group dynamics in solution. *J. Am. Chem. Soc.* **121**, 2891-2902.

35. Lee, A. L., Kinnear, S. A., and Wand, A. J. (2000). Redistribution and loss of side chain entropy upon formation of a calmodulin-peptide complex. *Nat. Struct. Biol.* **7**, 72-77.

36. Lee, A. L., Sharp, K. A., Kranz, J. K., Song, X. J., and Wand, A. J. (2002). Temperature dependence of the internal dynamics of a calmodulin-peptide complex. *Biochemistry* **41**, 13814-13825.

37. Lee, L. K., Rance, M., Chazin, W. J., and Palmer, A. G. (1997). Rotational diffusion anisotropy of proteins from simultaneous analysis of ^{15}N and $^{13}C^{\alpha}$ nuclear spin relaxation. *J. Biomol. NMR* **9**, 287-298.

38. Lipari, G., and Szabo, A. (1982). Model-Free Approach to the Interpretation of Nuclear Magnetic Resonance Relaxation in Macromolecules: 1. Theory and Range of Validity. *J. Am. Chem. Soc.* **104**, 4546-4559.

39. Lockless, S. W., and Ranganathan, R. (1999). Evolutionarily conserved pathways of energetic connectivity in protein families. *Science* **286**, 295-299.

40. Loh, A. P., Pawley, N., Nicholson, L. K., and Oswald, R. E. (2001). An increase in side chain entropy facilitates effector binding: NMR characterization of the side chain methyl group dynamics is Cdc42Hs. *Biochemistry* **40**, 4590-4600.

41. Luque, I., Leavitt, S. A., and Freire, E. (2002). The linkage between protein folding and functional cooperativity: Two sides of the same coin? *Annu. Rev. Biophys. Biomolec. Struct.* **31**, 235-256.

42. Ma, Y. M., Lu, H. H., Tippin, B., Goodman, M. F., Shimazaki, N., Koiwai, O., Hsieh, C. L., Schwarz, K., and Lieber, M. R. (2004). A biochemically defined system for mammalian nonhomologous DNA end joining. *Molecular Cell* **16**, 701-713.

43. Mace, J. E., Wilk, B. J., and Agard, D. A. (1995). Functional linkage between the active-site of alpha-lytic protease and distant regions of structure--scanning alanine mutagenesis of a surface loop affects activity and substrate-specificity. *J. Mol. Biol.* **251**, 116-134.

44. Mahajan, K. N., McElhinny, S. A. N., Mitchell, B. S., and Ramsden, D. A. (2002). Association of DNA polymerase mu (pol mu) with Ku and ligase IV: Role for pol mu in end-joining double-strand break repair. *Mol. Cell. Biol.* **22**, 5194-5202.

45. Manke, I. A., Lowery, D. M., Nguyen, A., and Yaffe, M. B. (2003). BRCT Repeats as Phosphopeptide-Binding Modules Involved in Protein Signaling. *Science* **302**, 636-639.

46. McElhinny, S. A. N., Havener, J. M., Garcia-Diaz, M., Juarez, R., Bebenek, K., Kee, B. L., Blanco, L., Kunkel, T. A., and Ramsden, D. A. (2005). A gradient of template dependence defines distinct biological roles for family X polymerases in nonhomologous end joining. *Molecular Cell* **19**, 357-366.

47. McElroy, C., Manfredo, A., Wendt, A., Gollnick, P., and Foster, M. (2002). TROSY-NMR studies of the 91 kDa TRAP protein reveal allosteric control of a gene regulatory protein by ligand-altered flexibility. *J. Mol. Biol.* **323**, 463-573.

48. Meroueh, S. O., Roblin, P., Golemi, D., Maveyraud, L., Vakulenko, S. B., Zhang, Y., Samama, J. P., and Mobashery, S. (2002). Molecular dynamics at the root of expansion of function in the M69L inhibitor-resistant TEM beta-lactamase from *Escherichia coli*. *J. Am. Chem. Soc.* **124**, 9422-9430.

49. Millet, O., Mittermaier, A., Baker, D., and Kay, L. E. (2003). The effects of mutations on motions of side-chains in protein L studied by ^{2}H NMR dynamics and scalar couplings. *J. Mol. Biol.* **329**, 551-563.

50. Mittermaier, A., Kay, L. E., and Forman-Kay, J. D. (1999). Analysis of deuterium relaxation-derived methyl axis order parameters and correlation with local structure. *J. Biomol. NMR* **13**, 181-185.

51. Mittermaier, A., Davidson, A. R., and Kay, L. E. (2003). Correlation between ^{2}H NMR side-chain order parameters and sequence conservation in globular proteins. *J. Am. Chem. Soc.* **125**, 9004-9005.

52. Muhandiram, D. R., Farrow, N. A., Xu, G. Y., Smallcombe, S. H., and Kay, L. E. (1993). A gradient ^{13}C NOESY-HSQC experiment for recording NOESY spectra of ^{13}C-labeled proteins dissolved in H2O. *J. Magn. Reson. Ser. B* **102**, 317-321.

53. Muhandiram, D. R., and Kay, L. E. (1994). Gradient-enhanced triple-resonance 3-dimensional NMR experiments with improved sensitivity. *J. Magn. Reson. Ser. B* **103**, 203-216.

54. Muhandiram, D. R., Yamazaki, T., Sykes, B. D., and Kay, L. E. (1995). Measurement of ^{2}H T_1 and $T_{1\rho}$ relaxation times in uniformly ^{13}C-labeled and fractionally ^{2}H-labeled proteins in solution. *J. Am. Chem. Soc.* **117**, 11536-11544.

55. Neri, D., Szyperski, T., Otting, G., Senn, H., and Wuthrich, K. (1989). Stereospecific nuclear magnetic-resonance assignments of the methyl-groups of valine and leucine in the DNA-binding domain of the 434-repressor by biosynthetically directed fractional C-13 Labeling. *Biochemistry* **28**, 7510-7516.

56. Ohtaka, H., Schon, A., and Freire, E. (2003). Multidrug resistance to HIV-1 protease inhibition requires cooperative coupling between distal mutations. *Biochemistry* **42**, 13659-13666.

57. Ottiger, M., Delaglio, F., and Bax, A. (1998). Measurement of *J* and Dipolar Couplings from Simplified Two Dimensional NMR Spectra. *J. Magn. Reson.* **131**, 373-378.

58. Ottiger, M., and Bax, A. (1999). How tetrahedral are methyl groups in proteins? A liquid crystal NMR study. *J. Am. Chem. Soc.* **121**, 4690-4695.

59. Ousterhout, J. K. (1994). *Tcl and the Tk Toolkit*: Addison-Wesley).

60. Perutz, M. F. (1970). Stereochemistry of cooperative effects in haemoglobin. *Nature* **228**, 726-739.

61. Radisky, E. S., King, D. S., Kwan, G., and Koshland Jr., D. E. (2003). The Role of the Protein Core in the Inhibitory Power of the Classic Serine Protease Inhibitor, Chymotrypsin Inhibitor 2. *Biochemistry* **42**, 6484-6492.

62. Radkiewicz, J. L., and Brooks, C. L. (2000). Protein dynamics in enzymatic catalysis: Exploration of dihydrofolate reductase. *J. Am. Chem. Soc.* **122**, 225-231.

63. Rajagopalan, P. T. R., Lutz, S., and Benkovic, S. J. (2002). Coupling interactions of distal residues enhance dihydrofolate reductase catalysis: Mutational effects on hydride transfer rates. *Biochemistry* **41**, 12618-12628.

64. Ruckert, M., and Otting, G. (2000). Alignment of Biological Macromolecules in Novel Nonionic Liquid Crystalline Media for NMR Experiments. *J. Am. Chem. Soc.* **122**, 7793-7797.

65. Schnell, J. R., Dyson, H. J., and Wright, P. E. (2004). Effect of cofactor binding and loop conformation on side chain methyl dynamics in dihydrofolate reductase. *Biochemistry* **43**, 374-383.

66. Seemuller, U., Meier, M., Ohlsson, K., Muller, H. P., and Fritz, H. (1977). Isolation and characterisation of a low molecular weight inhibitor (of chymotrypsin and human granulocytic elastase and cathepsin G) from leeches. *Hoppe-Seyler's Z. Physiol. Chem.* **358**, 1105-1117.

67. Showalter, S. A., and Hall, K. B. (2002). A functional role for correlated motion in the N-terminal RNA-binding domain of human U1A protein. *J. Mol. Biol.* **322**, 533-542.

68. Spudich, G., Lorenz, S., and Marqusee, S. (2002). Propagation of a single destabilizing mutation throughout the *Escherichia coli* ribonuclease HI native state. *Protein Sci.* **11**, 522-528.

69. Suel, G. M., Lockless, S. W., Wall, M. A., and Ranganathan, R. (2003). Evolutionarily conserved networks of residues mediate allosteric communication in protein. *Nat. Struct. Biol.* **10**, 59-69.

70. Tjandra, N., Feller, S. E., Pastor, R. W., and Bax, A. (1995). Rotational diffusion anisotropy of human ubiquitin from N-15 NMR relaxation. *J. Am. Chem. Soc.* **117**, 12562-12566.

71. Tjandra, N., and Bax, A. (1997). Direct Measurement of Distances and Angles in Biomolecules by NMR in a Dilute Liquid Crystalline Medium. *Science* **278**, 1111-1114.

72. Uhrin, D., Uhrinova, S., Leadbeater, C., Nairn, J., Price, N. C., and Barlow, P. N. (2000). 3D $HCCH^3$-TOCSY for resonance assignment of methyl-containing side chains in ^{13}C-labeled proteins. *J. Magn. Reson.* **142**, 288-293.

73. Vuister, G. W., Wang, A. C., and Bax, A. (1993). Measurement Of 3-Bond Nitrogen Carbon-J Couplings In Proteins Uniformly Enriched In N-15 And C-13. *J. Am. Chem. Soc.* **115**, 5334-5335.

74. Wand, A. J., Urbauer, J. L., McEvoy, R. P., and Bieber, R. J. (1996). Internal dynamics of human ubiquitin revealed by C-13- relaxation studies of randomly fractionally labeled protein. *Biochemistry* **35**, 6116-6125.

75. Wand, A. J. (2001). Dynamic activation of protein function: a view emerging from NMR spectroscopy. *Nat. Struct. Biol.* **8**, 926-931.

76. Williams, R. S., Lee, M. S., Hau, D. D., and Glover, J. N. M. (2004). Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. *Nature Struct. Mol. Biol.* **11**, 519-525.

77. Wray, J. W., Baase, W. A., Lindstrom, J. D., Weaver, L. H., Poteete, A. R., and Matthews, B. W. (1999). Structural analysis of a non-contiguous second-site revertant in T4 lysozyme shows that increasing the rigidity of a protein can enhance its stability. *J. Mol. Biol.* **292**, 1111-1120.

78. Yi, F., Sims, D. A., Pielak, G. J., and Edgell, M. H. (2003). Testing Hypotheses about Determinants of Protein Structure with High-Precision, High-Throughput Stability Measurements and Statistical Modeling. *Biochemistry* **42**, 7594-7603.

79. Young, M. A., Gonfloni, S., Superti-Furga, G., Roux, B., and Kuriyan, J. (2001). Dynamic coupling between the SH2 and SH3 domains of c-Src and Hck underlies their inactivation by C-terminal tyrosine phosphorylation. *Cell* **105**, 115-126.

80. Yu, E. W., and Koshland Jr., D. E. (2001). Propagating conformational changes over long (and short) distances in proteins. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 9517-9520.

81. Yu, X., Chini, C. C. S., He, M., Mer, G., and Chen, J. (2003). The BRCT Domain is a Phospho-Protein Binding Domain. *Science* **302**, 639-642.

82. Zhang, X. J., Baase, W. A., and Matthews, B. W. (1992). Multiple Alanine Replacements Within Alpha-Helix-126-134 Of T4 Lysozyme Have Independent, Additive Effects On Both Structure And Stability. *Protein Sci.* **1**, 761-776.