AD_____

GRANT NUMBER DAMD17-94-J-4270

TITLE: A New Generic Method for the Production of Protein-Based Inhibitors of Proteins Involved in Cancer Metastasis

PRINCIPAL INVESTIGATOR: Marshall H. Edgell, Ph.D.

CONTRACTING ORGANIZATION: University of North Carolina
at Chapel Hill
Chapel Hill, North Carolina
27599-4100

REPORT DATE: September 1997

TYPE OF REPORT: Annual

PREPARED FOR: Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	September 1997	Annual (1 Aug 96 - 31 Jul 97)

4. TITLE AND SUBTITLE
A New Generic Method for the Production of Protein-Based Inhibitors of Proteins Involved in Cancer Metastasis

5. FUNDING NUMBERS
DAMD17-94-J-4270

6. AUTHOR(S)
Marshall H. Edgell, Ph.D.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599-4100

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Commander
U.S. Army Medical Research and Materiel Command
Fort Detrick, Frederick, Maryland 21702-5012

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200)*

Our objective is to develop a general method to make protein-based inhibitors against protein targets and as the test case to a proteinase involved in metastasis, stromelysin. The idea is to use a binding epitope display framework which should bind preferentially to the active site pocket of target proteins and which would be screened against targets using phage display. Having spent a lot of effort constructing vectors to generate an expression vector for stromelysin we have now found a source via a collaboration with Dr. Ye at Parke-Davis. Library construction has turned out to more difficult than expected. We have now turned to a construction technology which can be applied to any portion of the framework independent of the presence of restriction sites. We have screened a peptide library against stromelysin and this has yielded binders whose sequences have been determined. This year we have also succeeded in constructing our first library in an aglin based framework. As part of our efforts to charaterize the binding epitope display framework molecule which we use, eglin c, we have also developed and verified a new method for using mutagenesis to study protein structure which we call patterned library analysis.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Breast Cancer	66
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
Unclassified	Unclassified	Unclassified	Unlimited

FOREWORD

Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the U.S. Army.

____ Where copyrighted material is quoted, permission has been obtained to use such material.

____ Where material from documents designated for limited distribution is quoted, permission has been obtained to use the material.

____ Citations of commercial organizations and trade names in this report do not constitute an official Department of Army endorsement or approval of the products or services of these organizations.

NHE In conducting research using animals, the investigator(s) adhered to the "Guide for the Care and Use of Laboratory Animals," prepared by the Committee on Care and Use of Laboratory Animals of the Institute of Laboratory Resources, National Research Council (NIH Publication No. 86-23, Revised 1985).

____ For the protection of human subjects, the investigator(s) adhered to policies of applicable Federal Law 45 CFR 46.

NHE In conducting research utilizing recombinant DNA technology, the investigator(s) adhered to current guidelines promulgated by the National Institutes of Health.

NHE In the conduct of research utilizing recombinant DNA, the investigator(s) adhered to the NIH Guidelines for Research Involving Recombinant DNA Molecules.

NHE In the conduct of research involving hazardous organisms, the investigator(s) adhered to the CDC-NIH Guide for Biosafety in Microbiological and Biomedical Laboratories.

NH Edgell 10-17-97

PI - Signature Date

Table of Contents

INTRODUCTION

Our objective is to learn how to efficiently build proteins which will act as inhibitors to proteins involved in cancer metastasis. It is our expectation that such proteins can be constructed so as to be very specific for the desired target. While such inhibitors might be useful themselves as therapeutic molecules, they will certainly be useful as probes to define the issues associated with inactivating the target proteins; both the primary effects and side-effects.

The approach to be used in this project is to combine molecular genetics and protein biophysics to redirect to the target of interest the activity of a pre-existing protein which will serve as a framework onto which to mount the desired modifications. Molecular genetics will be used to extend the reach of traditional protein engineering. The idea is to make large libraries of structural variants and then use genetic screening and selection strategies to find the best performers. Traditional protein biophysics will then be used to explore the various classes of variants and to make models for what is leading to inhibition. This information will then be used in subsequent cycles of design, construction and screening. We will also employ a cycle of mutation in a mutator strain of bacteria to evolve the inhibitor's affinity and specificity.

The development cycle that we will employ to modify the wild-type eglin c into a new inhibitor is:

1. make our best design guess as to what changes will increase binding to the desired target
2. construct a 'halo' of variants (~10^7) around the design
3. screen the variant library using phage display to find the best binders to the desired target
4. if affinities or specificity are not sufficient pursue two parallel paths to improvement
    - A. Biophysical analysis and redesign
        - i. characterize the binding classes using biophysical techniques (NMR, CD, ANS binding, etc.)
        - ii. use the biophysical information and modeling to build hypotheses concerning binding
        - iii. goto step 2 to carry out another cycle of screening
    - B. Laboratory Evolution
        - i. mutate the population of best binder in bacterial mutator strains
        - ii. goto step 3 to rescreen the population

The simplest inhibitors bind to their targets close to or at the active site and interfere with activity simply by getting in the way and not 'letting go' of the target. That is, such inhibitors lower the energy of the Michaelis complex sufficiently that very few molecules reach the transition state. These inhibitors can be designed to poke a projection into a groove in the target or enfold a pocket over a projection on the target. As our initial protein for protein engineering we wanted a molecule that was small, well mannered and for which we had some reason to think might be structurally compatible with our first set of targets. Our choice was a small proteinase inhibitor, eglin c. This protein is exceptionally stable, has no disulfide bonds, is well characterized and binds very tightly to proteins similar to our first target which is stromelysin, a proteinase implicated in metastasis. Eglin c inhibits its normal targets, serine proteinases, by binding so tightly in the Michaelis complex, that the protein cannot be raised into the transition state. A ten amino acid loop in eglin c binds within the active site groove of the native serine proteinase targets.

High affinity binding requires a sequence (binding epitope) which is compatible with the target and a set of structural constraints on that sequence which prevent it from spending much time in non-productive conformations. The engineering task, which we have set ourselves, is to replace the wild-type binding epitope with one suitable for the target and then to construct a set of new constraints to move the binding epitope into the high affinity domain. Our initial target is stromelysin and hence an appropriate binding epitope is already known, that is, a substrate sequence preferred by the proteinase. Building an inhibitor then reduces to finding a suitable series of structural constraints that can be imposed by the eglin framework on the new binding epitope.

We are pursuing two sub-lines of investigation. One is to expand our information about eglin c as a suitable framework for protein engineering and the other is to start the protein engineering with what we already know.

BODY

Summary

During this third year of the project we have spent much of our time on vector constructions. We have worked on stromelysin expression vectors and testing the constructs for active products. Having done that we have recently found a source of stromelysin via a collaboration with Dr. Ye at Parke-Davis. We have also spent much time trying to get the library constructions optimized. This year we have constructed our first eglin-based framework libraries, we have screened a peptide library against stromelysin and found binders, and we have made an exciting methodological discovery in that we have proven that a new method we have devised for analyzing protein structural determinants does indeed work. An outline of our years activities follows:

A. Stromelysin production

1. we have found conditions for high yield recovery of prostromelysin
2. we have been unable to efficiently activate prostromelysin
3. we have constructed several stromelysin catalytic domain expression vectors
4. we have shown that the expression vectors produce active stromelysin
5. we have found a source of stromelysin via a collaboration with Dr. Ye at Parke-Davis

B. Library Construction and Screening

1. random peptide libraries
    a. we have modified our panning protocol to increase diversity
    b. we have showed that random peptide libraries bind to ~50% of random enzymes
    c. we have screened a random peptide library against stromelysin
        i. we have found four stromelysin binders
        ii. analysis of the binders yields two possible binding sequences
2. teglin loop libraries (18 amino acids closed with cysteines)
    a we have tested papain targeted library XFAXXX
        i. we have found many papain binders
        ii. sequence analysis of the binders shows that ALL have lost the cysteines which were supposed to close the loop
    b. we have constructed an eglin-based loop library 8X+2A
3. full framework peglin libraries
    a. we have verified that peglin binds to its native target
    b. we have constructed a 8X+2R library
    c. we have constructing modified vector for library construction

C. Framework Characterization

1. we have continued to determine the thermodynamics of the eglin framework
    a. we have examined the thermal denaturation of eglin c by using circular dichroism spectropolarimetry and differential scanning calorimetry.
    b. we have shown that thermal denaturation is reversible and consistent with a two state model at low protein concentrations.
    c. we have shown that the thermodynamic parameters have an unusual protein concentration dependence due to a thermodynamic non-ideality effect.
2. we have devised a new method, patterned library analysis, to assess hypotheses concerning the structural determinants of proteins
    a. we have worked out resin splitting technology to facilitate the synthesis of oligonucleotides with arbitrary sequence degeneracy
    b. we have carried out proof of principle experiments showing that patterned library analysis can recreate helix amino acid propensity data

A. STROMELYSIN PRODUCTION

In order to screen phage display libraries against stromelysin we need a good source of the proteinase. Last year we reported the construction of an expression vector for prostromelysin. We have worked out growth conditions for this vector and are now getting approximately 13 milligram per liter of culture. Another 10 milligrams or so is found in the non-soluble fraction presumably as inclusion bodies. Unfortunately we have be unable to efficiently convert the prostromelysin to active stromelysin. We have tried both heat (52 C and 58 C) and trypsin activation and have only obtained about 2% conversion from prostromelysin to active stromelysin.

A new literature search indicated that, unlike many other proteinases, a mature form of stromelysin could fold up properly into the active enzyme in the absence of the N-terminal pro-sequence. Hence we then constructed expression vector clones containing only the catalytic domain of stromelysin. Thirteen of fourteen isolates obtained from transformations had stromelysin activity. During this period we had been communicating with a Dr. Quezang Ye at Parke-Davis about his work with catalytic domain clones and purification methods for stromelysin. After some haggling by the lawyers, Parke-Davis agreed to provide stromelysin for this project. Subsequently we were sent 6.7 milligrams of stromelysin.

B. LIBRARY CONSTRUCTION AND SCREENING

We are using three types of binding epitope libraries for our analyses (Figure 1). The first is a traditional randomized peptide library. Our library has twelve randomized amino acids. As effective inhibitors peptides have several disadvantages. There is an upper limit on the binding affinity of an unstructured peptide of around 10^{-8} M. Most peptides bind with affinities which are not useful under physiological conditions. Second, randomized peptides are very promiscuous in terms of locations to which they bind and hence many binders turn out not to be inhibitors. Third, there is a significant limit to the selectivity of peptides to homologous proteins. The premise behind our approach is that epitopes displayed on structured frameworks and in particular on loops on structured frameworks can overcome any of these problems. Hence we desire to test binding epitope libraries on various frameworks to verify our assumptions.

B.1. Random Peptide Libraries

B.1.a. Panning Protocol. Our initial panning protocol involved two panning steps followed by a phage amplification step. This cycle was repeated three times. On titering the phage inputs and outputs we discovered that there could sometimes be only a few hundred phage that get into the first amplification step. To increase the diversity of the screened population we have modified our screening protocol such that now we get at least 10^4 phage into each amplification cycle. We now do one panning step followed by an amplification step. This cycle is repeated six times.

B.1.b. Binding to an assortment of Enzymes. We have chosen eight enzymes on the basis of commercial availability and cost to serve as a test bank of enzymes to characterize the properties and capabilities of the various libraries. The enzymes are: alcohol dehydrogenase, aldolase, alpha amylase, catalase, enolase, hexokinase, l-lactate dehydrogenase, and ribonuclease A. Using our new screening protocol we found that the randomized peptide library binds to three of the eight enzymes strongly and to one weakly (Table 1).

B.1.c. Binders to Stromelysin. We have screened the randomized peptide library against stromelysin. Screening for binders against a proteinase with a peptide has some complications since one might expect weak binding peptides to be cleaved by the enzyme thereby releasing the display phage from the enzyme. Hence we carried out screening against untreated stromelysin, and stromelysin treated with either $MgCl_2$ or $CdCl_2$. The latter reagent should replace the Zn in stromelysin with Cd leaving an inactive enzyme with its structure intact. It is not clear to me whether or not $MgCl_2$ should inactivate the enzyme. We were unable to assay for activity in the presence of either Cd^{++} or Mg^{++} because the ions themselves converted the substrate from colored to colorless. Binders were found to stromelysin under all three conditions (Figures 2,3,4).

We sequenced 30 clones which bound to the Cd treated stromelysin. Four different sequences were found (Figure 5). Only two amino acids were found in all four sequences. Alignments were made

around those common amino acids (Figure 6). These sequences may define a binding motif but further analysis will be required to determine the range of allowed residues in the motif.

B.1. Structured Loop Libraries.

The chief feature which reduces the affinity of unconstrained peptides is the entropy associated with the many structures assessable to the peptide. In effect the configuration suitable for binding is diluted out by all of the other equally probable structures. Constructing peptides which form loops via cysteine bonds reduces the number of accessible configurations. If one suitable for binding is found within the constrained set it should bind with higher affinity than a free peptide. We have constructed various constrained loop libraries. They are all based on the 18 amino acid loop in eglin c which contains the binding epitope of the native inhibitor. We call this truncated constrained loop teglin.

B.1.a. Binders to Papain

Prior to having a source of stromelysin we have used papain as a non-serine proteinase target to model the issues associated with finding binders to enzymes which might cleave the binding epitope. Last year we reported the construction of a teglin library which replaced the serine proteinase binding epitope with XFAXXX, where X is any of the twenty amino acids and F is phenylalanine and A is alanine. FA is a sequence cleaved by papain. Screening this library gave rise to many different binders (Figure 7). However, sequence analysis of these binders showed that most of them had in one devious fashion or another gotten rid of one or more of the cysteines providing the conformational constraints (Figure 8).

B.1.b. New Teglin-based Libraries

We have constructed two new high diversity teglin libraries which contain randomized binding epitopes (Table 2). The ultimate objective is to have dozens of starting libraries and to do screening with them all to find binders. Subsequent rounds of mutagenesis in mutator strains would then be used to fine-tune the binding and specificity of the binders.

LIBRARY	SIZE
8X + 2A	2.3×10^8
8X + 2X	15×10^9

Table 2. Teglin Libraries. The 8X+2A library has eight randomized residues in the center of the eglin loop and has the two arginine residues which in eglin provide a framework-loop interaction via salt brides are replaced by alanines. In the 8X+2X those two arginines are randomized.

B.3. Full Framework Libraries

The full framework protein we have chosen for displaying binding epitopes is eglin c. The expectation is that this small protein will provide the scaffolding for structural constraints and that by presenting the binding epitopes on a loop that they will most likely bind to pockets in the target protein. Most pockets will contain the active site and hence the likelihood of binders being inhibitors is presumed to be large.

B.3.a. Peglin Binds to its Native Target In our first year we showed that eglin c did not in fact function in phage display. Hence we designed a circularity permuted version of eglin called peglin whose construction we reported last year. This year we have shown that peglin does indeed bind to its native target as expected (Figure 9). The binding shown in this figure is weak which retrospectively we realize is due to the presence of inorganic phosphate in the buffer which drastically reduces subtilisin activity. The binding of peglin to subtilisin was verified by measuring the molar effectiveness of peglin as a subtilisin inhibitor as compared to eglin c. We cannot detect the difference in an inhibition assay between peglin and eglin.

B.3.b. Construction of Our First Full Framework Library We have constructed a modest diversity peglin library which we call 8X + 2A. Eight residues are randomized in the center of the peglin loop (Figure 1). In the wild-type protein there are two salt bridges acting as loop constraints which consist of arg to asp and arg to thr interactions. In this library the two arginines have been replaced with alanines (2A). This library was constructed by replacing a Xho/Xba fragment with a degenerate oligonucleotide containing the randomized residues and the arg to ala mutations. The library has a diversity of 3 x 10 5 different phage variants.

B.3.c. An Alternative Library Construction Methodology What we have found is that every time we change enzymes to accommodate the distribution of restriction sites for a library that it takes us a very long time to get the conditions adequate for making the library. To reduce this optimization time we are shifting to a technology which will allow us to use the same enzymes no matter where we need to make changes in the peglin gene. Hopefully, a single optimization will then apply to many obrary constructions.

The approach (Figure 10) makes use of a set of enzymes which cut outside of their recognition sequence. Combining this with PCR technology which allows one to amplify up the entire plasmid makes it possible to replace any portion of the peglin gene using the same set of enzymes. to use this approach we need to remove two Ear I sites in our plasmid. We have gotten rid of one and are in the process of removing the other by site directed mutagenesis.

C. FRAMEWORK CHARACTERIZATION

C.1. Thermodynamic Characterization

We have continued our characterization of the protein which serves as the framework for epitope display. Points C.1. a through c are supported in the manuscript draft "Thermal Denaturation of Eglin c as a Function of Protein Concentration and pH" by Waldner, Edgell and Pielak. We have also made an assessment of whether the his tagged versions of eglin which are produced by some of our expression vectors have the same thermodynamic behavior as the wild-type.

C.1.d. Thermodynamic effect of a poly-histidine-terminal extension on the equilibrium thermodynamics of eglin c denaturation. Poly-histidine-terminal extensions, His-tags, are a common way to purify recombinant proteins. Proteins containing tags bind resin-immobilized divalent cations, most commonly, Ni^{2+}. After contaminates are washed away, the target protein is eluted with imidazole.

Reports of the effects of a His-tag on protein activity structure and stability are inconsistent. Incorporation of a His-tag at the C-terminus of beta-lactamase alters N-terminal processing and decreases T_m, by 1.4 to 3.0 °C.(Ledent et al., 1997). Incorporation of a His-tag into recombinant antibody fragments, however, gives no change in T_m or in the dissociation constant (Deutscher et al., 1996). It seems, therefore, that the effect of a His-tag cannot be predicted. Before starting extensive stability studies of variants from random mutant libraries based on His-tag-containing eglin c, we decided to quantify the tag's thermodynamic effect.

Our His-tag incorporates six histidine residues to eglin c's N-terminus. Denaturation was monitored by both DSC and CD. We have quantified the differences in the thermodynamic parameters, ΔG_D (the free energy of denaturation), ΔH_D (the enthalpy of denaturation), and ΔS_D, (the entropy of denaturation) resulting from incorporation of a His-tag on eglin c. We are able to analyze the data in three different way: using calorimetric enthalpies (cal) van't Hoff (vH) enthalpies from DSC data and vH enthalpies from CD data. As shown in Table 1, the his tag has no effect on eglin c stability.

Protein	$\Delta G_{D,vH(DSC)}$	$\Delta G_{D,cal(DSC)}$	$\Delta G_{D,vH(CD)}$	$\Delta\Delta G_{D,vH(DSC)}$	$\Delta\Delta G_{D,cal(DSC)}$	$\Delta\Delta G_{D,vH(CD)}$
wt	6.3 " 0.2	6.3 " 0.7	4.4 ± 0.7	---	---	---
H_6	6.2 " 0.2	5.5 " 0.8	4.4 " 0.7	0.1 ± 0.3	0.8 ± 1.1	0.0 ± 1.1

Table 1. Stability of wt and H_6 eglins c from CD and DSC data (pH 3, 300 K units are kcal mol^{-1})

C.2. Determinants of Structure in the Framework Protein Eglin c

One of the most powerful methods to characterize proteins is to utilize mutagenesis to test hypotheses. This approach has yielded considerable insight into the determinants of structure. However, despite these advances in our insights we are still unable to predict the consequences of most changes in a protein or to predict structure from primary sequence. During the course of this project it occurred to us that it might be possible to extend traditional mutagenic approaches to provide a quantitative assessment of hypotheses concerning protein structure. Our approach, which we call patterned library analysis, is conceptually quite straight-forward. Consider a hypothesis about protein structure, e.g. the N-terminal 'cap' of an alpha-helix will consist of [T,S]$\pi\pi$G where π is any hydrophilic amino acid. It is now possible to construct a library of variants all of which are consistent with this hypothesis. If this hypothesis were in fact true then all of the variants in this particular library would have the target structure and hence would be fully active. On the other hand if the hypothesis were only partially true, e.g. perhaps only the thr works and not the ser in the N-terminal position, then not all of the variants would have the target structure and in this example only 50% would be active. That is, the fraction of the library which is active becomes a quantitative measure of the quality of the hypothesis. This simple concept gets complicated quite quickly as one considers the arbitrary nature of activity assays and the fact that proteins do not respond in an all or none way to amino acid replacements. However, we feel that the principle of the approach remains intact particularly when one applies this approach to not assess a single hypothesis but rather compares one to another to decide which is the better.

C.2.a.Library Construction for Patterned Library Analysis To produce libraries of arbitrary complexity at any given codon we employ a technique called resin splitting. Let's say at codon 12 one wanted either a thr or ser. One cannot achieve this degeneracy, which excludes all other amino acids, using some degeneracy in the three nucleotides of the codon. So, one synthesizes the oligonucleotide in the normal fashion until codon 12 at which point the synthesis is stopped. Here we open the column on the oligonucleotide synthesizer with the partially synthesized product and split the resin into two equal portions and put each into a new column. With one we add the thr codon (ACN), to the other the ser codon (TCN). The columns are again opened, the resins removed and combined and synthesis continues. In this fashion one can generate degenerate oligonucleotides which encode an arbitrarily complex set of codons. To accurately split the resin we suspend it in an iso-dense solution (dibromomethane) which we determined empirically since the manufacturers considered the resin density a trade secret. It also turned out that, contrary to a promise from Beckman Instruments, the oligonucleotide synthesizer we purchased from them would not work with columns with removable caps made by Glen Research. The Beckman plastic columns are spot welded shut and hence are considered one-use columns. We were able to reuse the Beckman columns by constructing a brass capping clamp which allows us to hold shut Beckman columns which have been opened and whose seal has been broken in the process.

To determine whether patterned library analysis can indeed assess hypotheses we used the technique to see if it could reconstruct several properties of alpha-helices which have previously been determined from biophysical analyses. From many studies of helix amino acid propensities it has been determined that helices dislike prolines most, glycines next most and like alanines. We constructed four libraries which tested amino acid propensity at four solvent exposed sites in the alpha helix in eglin. The null hypothesis library contained any of six hydrophilic residues (H, Q, N, D, E, K) at the four sites. The three remaining libraries has any of seven residues at the four sites: (H, Q, N, D, E, K, P or H, Q, N, D, E, K, G or H, Q, N, D, E, K, A). When tested at our normal assay conditions virtually all of the variants in the four libraries were active. That is, eglin is so stable that the destablizing effects of even a proline in the helix was not enough to inactivate the protein. So, to increase the sensitivity of the assay we carried out the assay in various concentrations of denaturant to reduce the basal stability of eglin. Fortunately the proteinase target for eglin, subtilisin, is very stable in detergent and hence the assay remained valid in relatively high concentrations of SDS. By using various concentrations of SDS we were able to increase the sensitivity of our assay and distinguish between the four libraries (Figure 11). The four libraries did, in fact, order themselves in exactly the same order as predicted from the conclusions drawn from the biophysical analyses.

Although patterned library analysis represents only a modest extension of prior use of libraries to assess ideas about protein structure it converts a qualitative approach into a quantitative approach. In the

past moving to quantitative approaches has been very productive. In addition, our concept of stability scanning, as implemented by doing the assays in various denaturant concentrations, extracts a new set of information from the experiments. We are very excited about this tool and look forward to finding funding to apply the method to the analysis of protein structure.

CONCLUSIONS

It has taken us much longer to construct libraries than prior experience suggested. While I am hopeful that the new library construction technology outlined above will solve this problem we have also shifted some of our focus from an inhibitor development strategy that is dependent on library construction at each cycle to one which only requires libraries at the first cycle, and generalized libraries at that, i.e. ones which require no target specificity incorporated into their design. This way as we build up an increasingly large collection of libraries. They can all be utilized when we attack a new target. In addition, new libraries could be added to the input population at any stage of working up an inhibitor. If the new libraries do better than the binders being worked on they will fall through the screen and if not will be eliminated.

We have now shown that we can get binders to stromelysin. We have also shown that we can get binders with a randomized library to ~50% of enzymes chosen at 'random'. This data tells us that the concept of being able to design protein-based inhibitors in a generalized fashion for many targets retains viability.

In a less than direct line to inhibitors we have discovered a new approach to assessing hypotheses concerning the structural determinants of proteins. If the method does not develop unexpected problems it will provide another powerful tool for the analysis of proteins. Clearly the acquisition of a capacity to predict structure from amino acid sequences will constitute another revolution in our ability to generate therapeutically useful proteins and to explore biological processes.

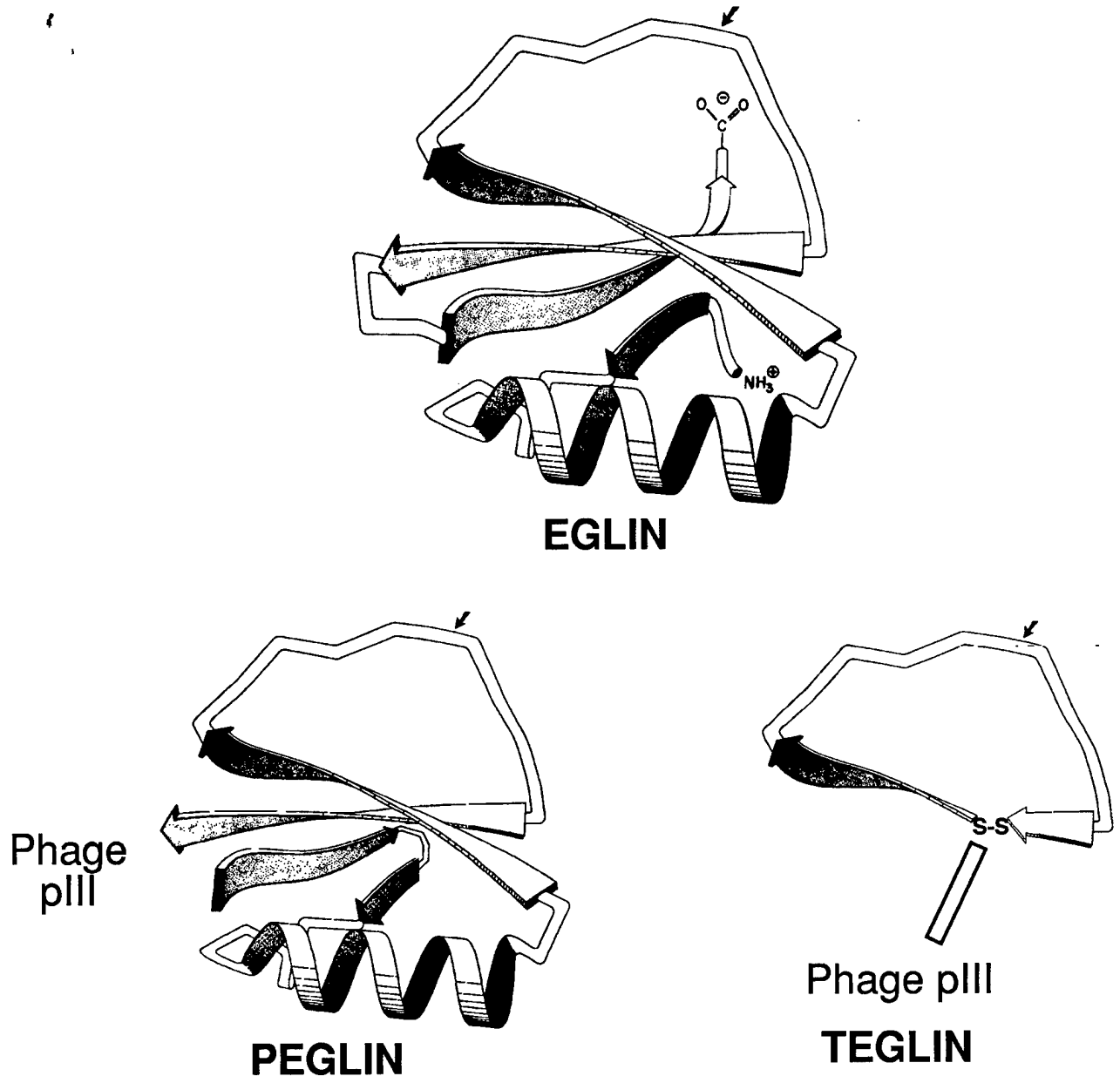


Figure 1. Three Libraries Used For Screening. Our randomized peptide library has the twelve N-terminal residues randomized and is fused to the phage gene pIII protein via a GGGG linker. The TEGLIN libraries are embedded in an 18 amino acid loop constrained via a disulfide bond. The non-randomized residues in teglin are derived from eglin. The loop sequences are tethered to the pIII protein via a QGGGG linker. The PEGLIN libraries have the epitope sequences embedded in the same loop but now constrained by the entire framework and framework-loop interactions.

Individual Binders from Pan6 (unamplified) vs Background
Scheme:Pan/amplify x6

	Background	Tested	≥2x	3x	4x	5x	6x	≥10x
Alcohol dehydrogenase	.069	88	25	39	10	2	0	0
Aldolase	067	92	62	0	0	0	0	0
Alpha Amylase	.1635	94	3	0	2	1	0	0
	.07	96	39	12	6	1	1	3
Catalase	.079	92	8	1	3	0	0	53
Enolase	.065	90	33	5	0	1	0	0
Hexokinase	.106	90	2	0	0	0	0	0
L-lactate dehydrogenase	.169	3	0	0	0	0	0	0
Ribonuclease-A	.07	92	15	61	15	0	0	1

Table 1. Binders to the General Enzyme Panel. Individual clones from a phage population after six rounds of panning were tested for binding to the eight enzymes in our panel using an ELISA assay to detect phage binding. Listed are the number of clones which showed binding at the indicated levels above background.

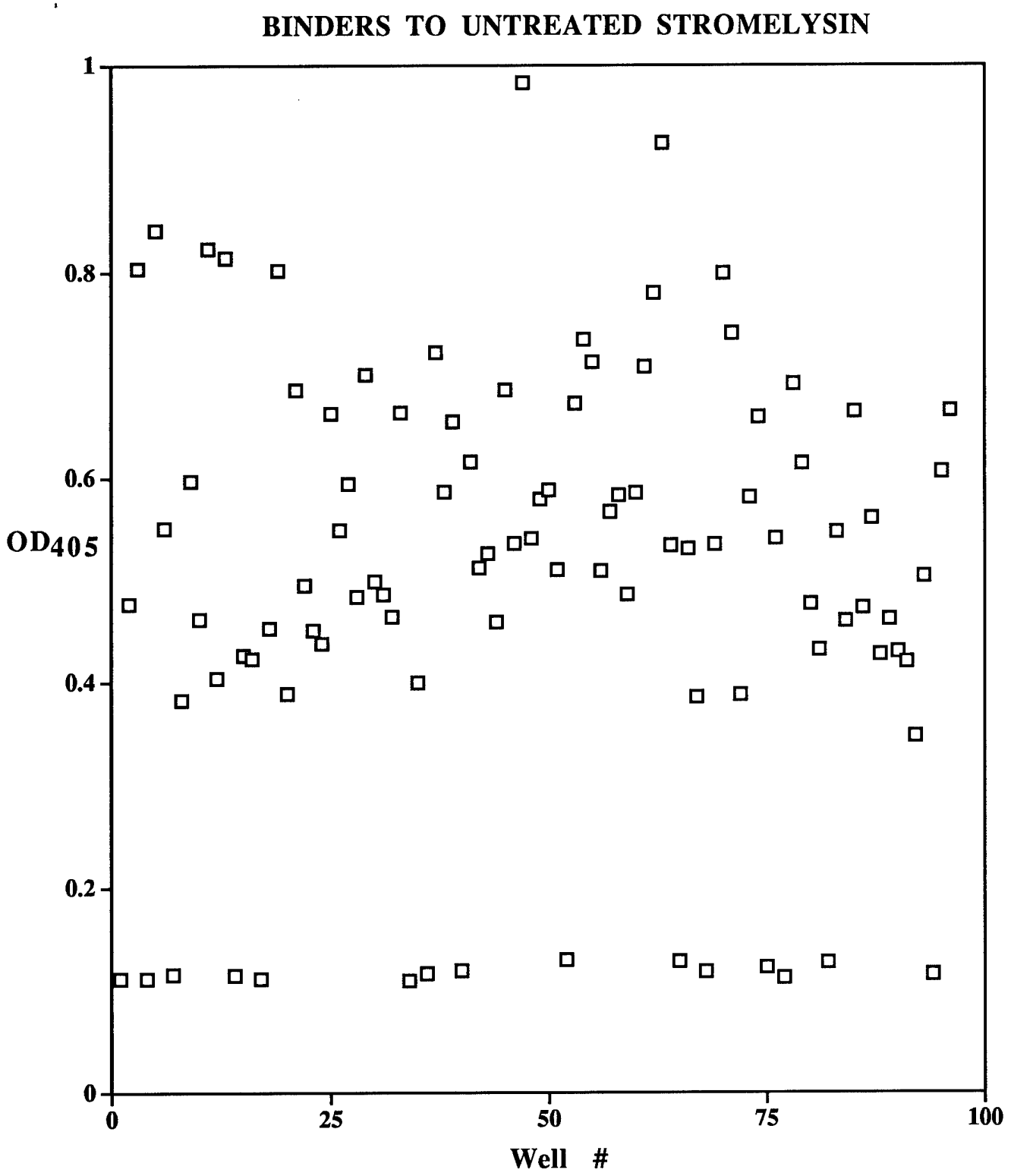


Figure 2. ELISA Values for Individual Clones After Six Rounds of Panning Against Stromelysin. Individual clones are isolated at random from the population which was present after six rounds of panning against stromelysin and bound to a well in a 96-well plate coated with stromelysin. The amount of phage that binds is determined by an ELISA assay using a primary antibody against wild-type M13 bacteriophage.

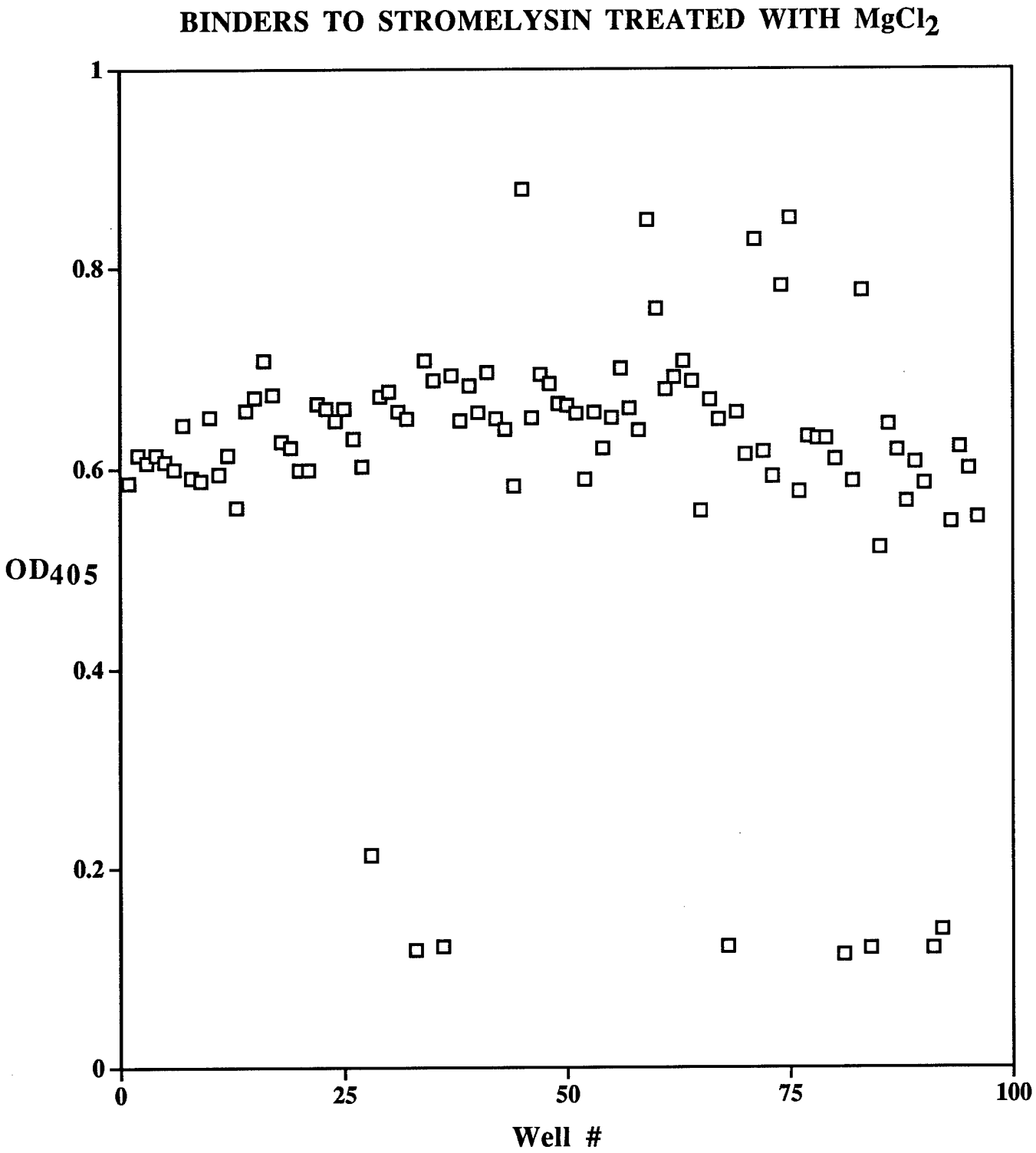


Figure 3. ELISA Values for Individual Clones After Six Rounds of Panning Against Stromelysin Treated with $MgCl_2$. Individual clones are isolated at random from the population which was present after six rounds of panning against $MgCl_2$ treated stromelysin and bound to a well in a 96-well plate coated with stromelysin. The amount of phage that binds is determined by an ELISA assay using a primary antibody against wild-type M13 bacteriophage.

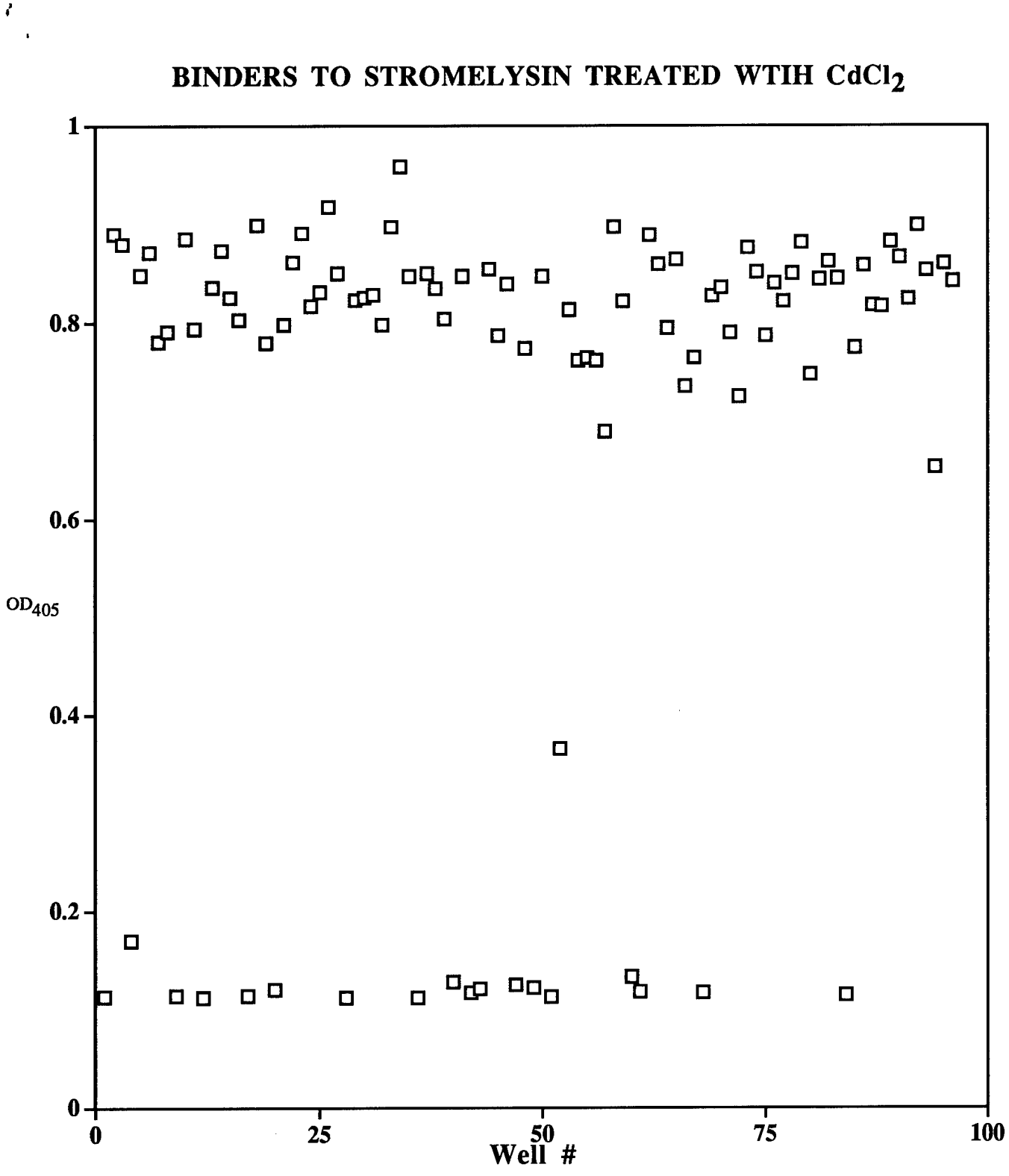


Figure 4. ELISA Values for Individual Clones After Six Rounds of Panning Against Stromelysin Treated with $CdCl_2$. Individual clones are isolated at random from the population which was present after six rounds of panning against $CdCl_2$ treated stromelysin and bound to a well in a 96-well plate coated with stromelysin. The amount of phage that binds is determined by an ELISA assay using a primary antibody against wild-type M13 bacteriophage.

Design	TCC TCG AGT NNK NNK NNK NNK NNK NNK NNK NNK NNK NNK NNK TCT AGA CCT
(27)	TCC TCG AGT *CCG CTT GAG AGG TTG ATG GCG CGG ATG GCT ACT CCT* TCT AGA CCT pro leu glu arg leu met ala arg met ala thr pro
(1)	TCC TCG AGT *CGG TCT GGG TTG GAG TCT TAT TGG AGG AGT GCG GAG* TCT AGA CCT arg ser gly leu glu ser tyr trp arg ser ala glu
(1)	TCC TCG AGT *TTG GAT GCG TGG CCG GAT GGT CCG AAG CGG ATT GCG* TCT AGA CCT leu asp ala trp pro asp gly pro lys arg ile ala
(1)	TCC TCG AGT *GGT AGG TCG GCT TGG ACG ATT GAT GGG ACT GTT GTG* TCT AGA CCT gly arg ser ala trp thr ile asp gly thr val val

Figure 5. Sequences of Clones Which Bound to $CdCl_2$ Treated Stromelysin. Most of the isolates had a single sequence. The number of isolates with a given sequence are indicated in parenthesis.

met **ala** arg met	glu **arg** leu met
met **ala** thr pro	ala **arg** met ala
ser **ala** glu ser	ser **arg** ser gly
asp **ala** trp pro	trp **arg** ser ala
ile **ala** ser arg	lys **arg** ile ala
ser **ala** trp thr	gly **arg** ser ala

Figure 6. Alignments of Sequences Around the Two Amino Acids Present in all of the Variants. Only alanine and arginine are present in the variant portion of all of the isolates. Alignments are presented around each of these two residues.

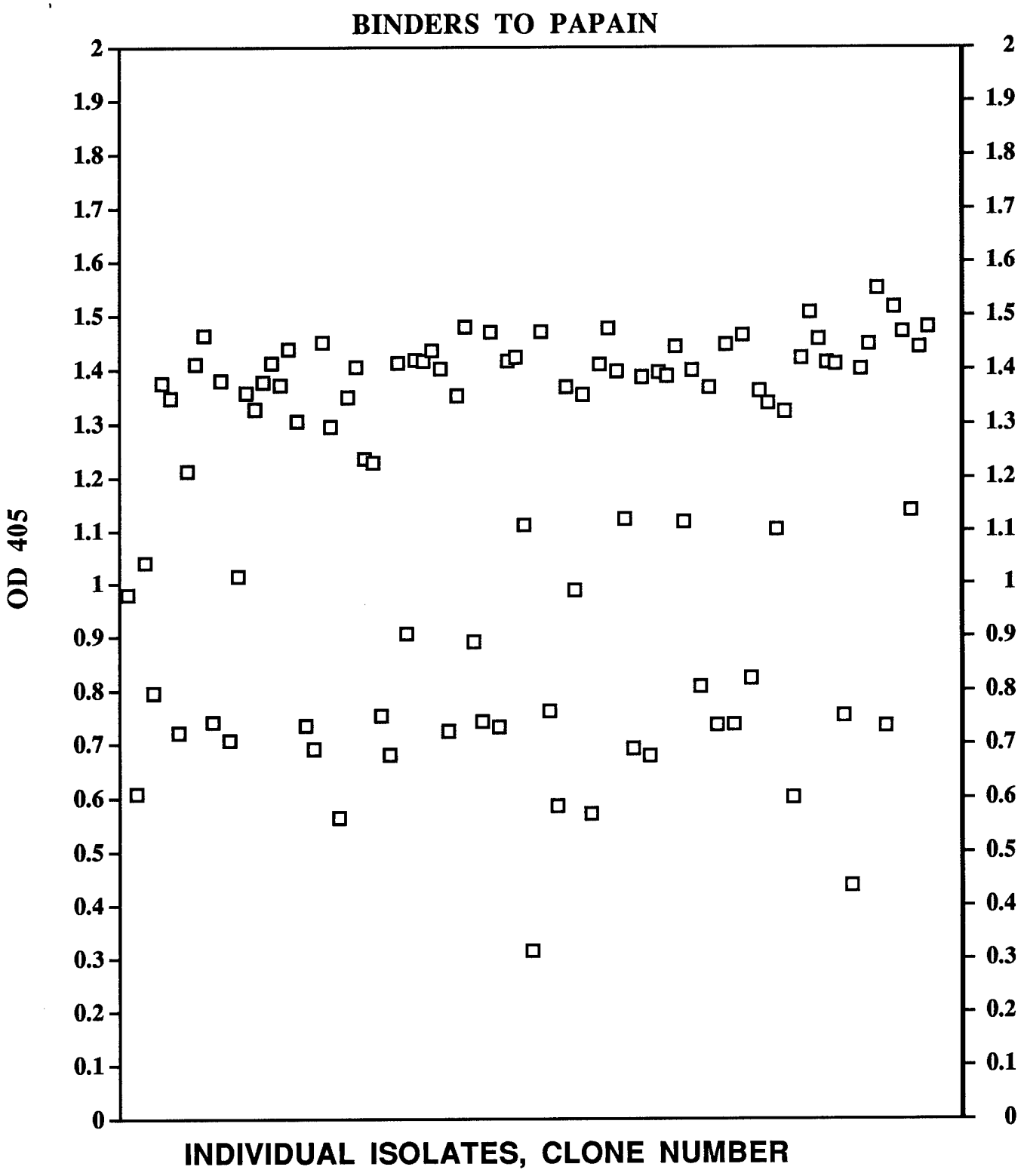


Figure 7. ELISA Values for Individual Clones After Six Rounds of Panning Against Papain. Individual clones are isolated at random from the population which was present after six rounds of panning against papain and bound to a well in a 96-well plate coated with stromelysin. The amount of phage that binds is determined by an ELISA assay using a primary antibody against wild-type M13 bacteriophage.

Design

```
TGC GGT ACC ATC NNS TTC GCT NNS NNS NNS ATC GAC CGC ACC CGT TCC TTC TGT
cys gly thr ile xxx phe ala xxx xxx xxx ile asp arg thr arg ser phe cys
```

High Papain

```
TGG GTA CCA TCG GGT TCG CTG GGA CGC GGG ATC GAC CGC ACT TGT TCC TTC TGT
trp val pro ser gly ser leu gly arg gly ile asp arg thr cys ser phe cys

TGG CGG TAC CAT CAA GTT CGC TCG GAG GCT GAT CGA CCG CAC CCG TTC CTT CTT
trp arg tyr his gln val arg ser glu ala asp arg pro his pro phe leu leu

TGC GGT ACC ATC GGG TTC GCT CCG AGG CTG ATC GAC CGC ACC CAT TCC TTC TTT
cys gly thr ile gly phe ala pro arg leu ile asp arg thr his ser phe phe

TGG CGG TAC CAT CAC GTT CGC TCC GAG CCC GAT CGA CCG CAC CCG TTC CTT CTG
trp arg tyr his his val arg ser glu pro asp arg pro his pro phe leu leu

TGC GGT ACC ATC GAC TTC GCT AAG AGG ACG ATC TAC CGC ACC CAT TCC TTC TGG
cys gly thr ile asp phe ala lys arg thr ile tyr arg thr his ser phe trp
```

Intermediate Papain

```
TAC CAT CTA GTT CGC TGG GGG GAG GAT CGA CCG CAC CCG TTC CTT CTG TCG GGT
tyr his leu val arg trp gly glu asp arg pro his pro phe leu ser gly gly

TGC GGT ACC ATC GAG TTC GCT GGG GGC GGG ATC GAC CGC ACC CGT TCC TTC TGT
cys gly thr ile glu phe ala gly gly gly ile asp arg thr arg ser phe cys

TGC GGT ACC ATC TGG TTC GCT GGG GGG GAT ATC GAC CGC ACC CGT TCC TTC TGT
cys gly thr ile trp phe ala gly gly asp ile asp arg thr arg ser phe cys
```

Figure 8. Sequences of Binders to Papain. All but two of the sequences have undergone some mutation relative to the library design which has removed the possbility of forming a constrained loop through a disulphide bond.

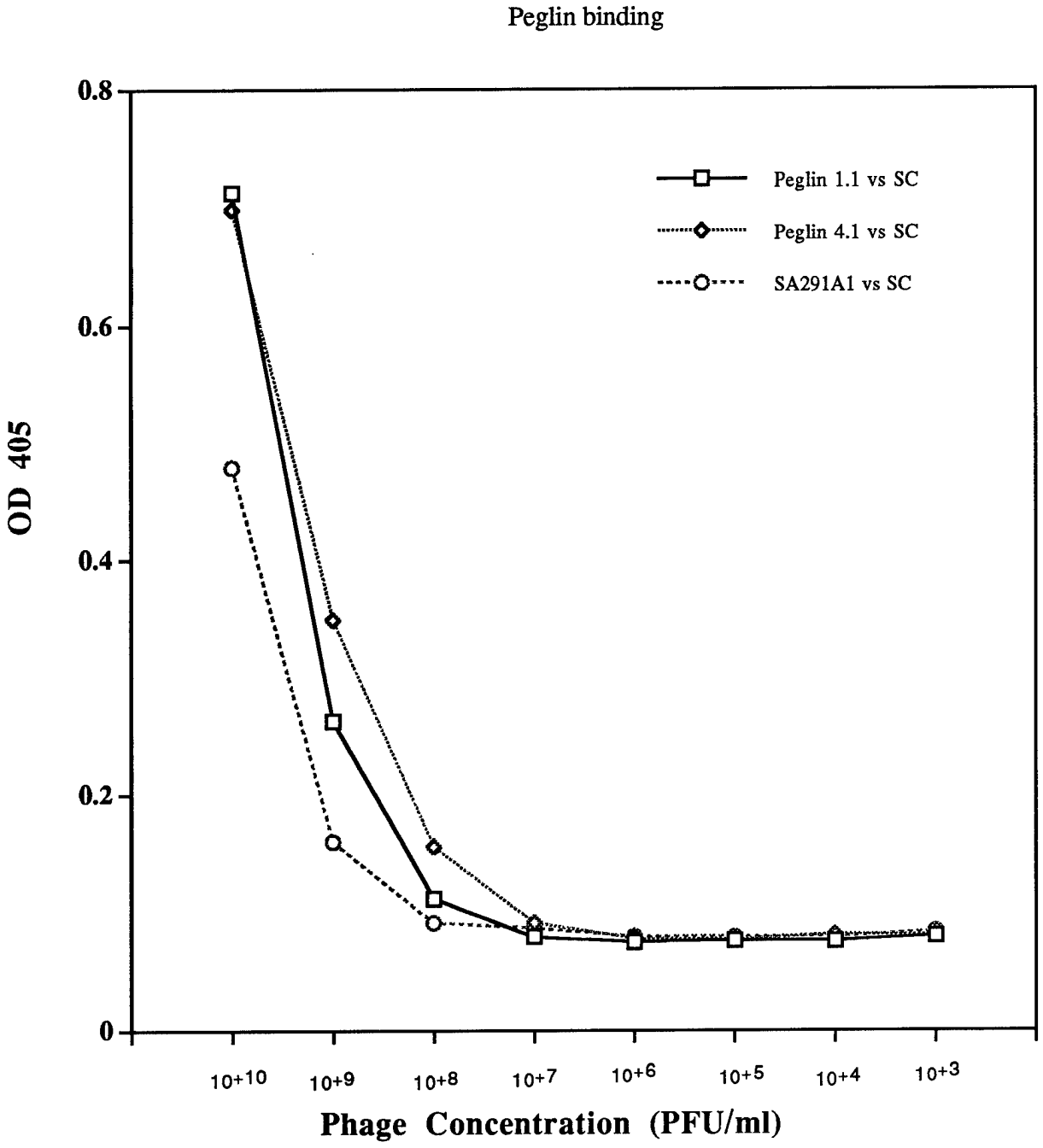


Figure 9. Binding to Wells Coated with Subtilisin. The two peglin constructs both bind to the native target for eglin c, subtilisin, more readily than does a phage displaying a peptide which binds to streptavidin. Since all proteins bind to some extent to this proteinase we wanted to determine relative binding efficencies. Binding at high phage concentrations is due to non-specific binding.

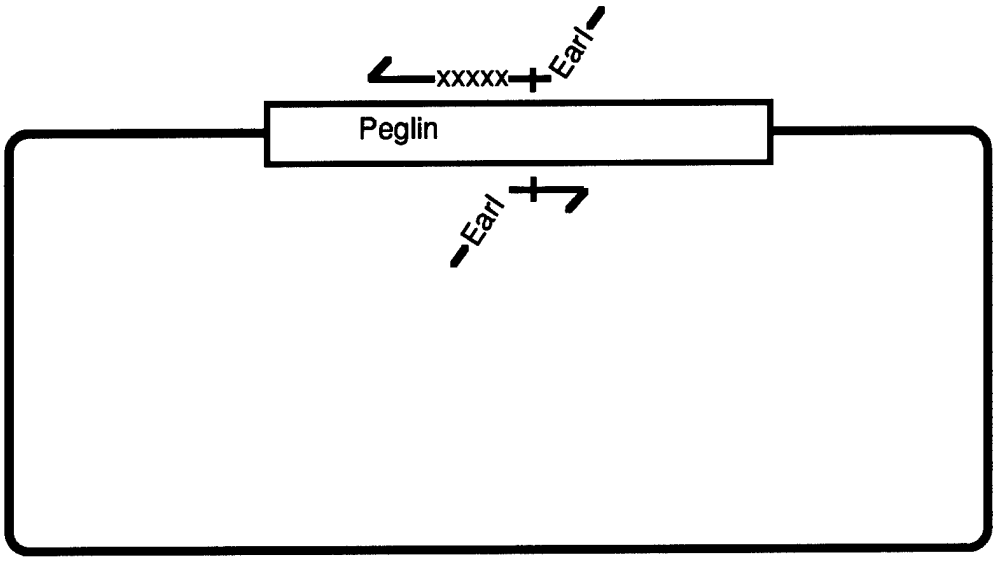


Figure 10. Construction Plan for Library COnstruction Using Ear I Sites. The Ear I restriction enzyme cuts outsie of it's recoginition site. Hence one can construct PCR primers which contain Ear I sites and sequences which are homologous to the target sequence downstream of the recognition site to allow cutting at any location within the peglin gene that is desired. The 'xxx' region in the upper primer indicates the region with randomized residues and hence this region would not hybridize to the peglin gene. The primer is placed by hybridization regions upstream and downstream of the region to be randomized. Long PCR technology allows one to make a product in which the entire plasmid can be reconstructed by cleaving the amplification product with Ear I and then annealing and ligating the product.

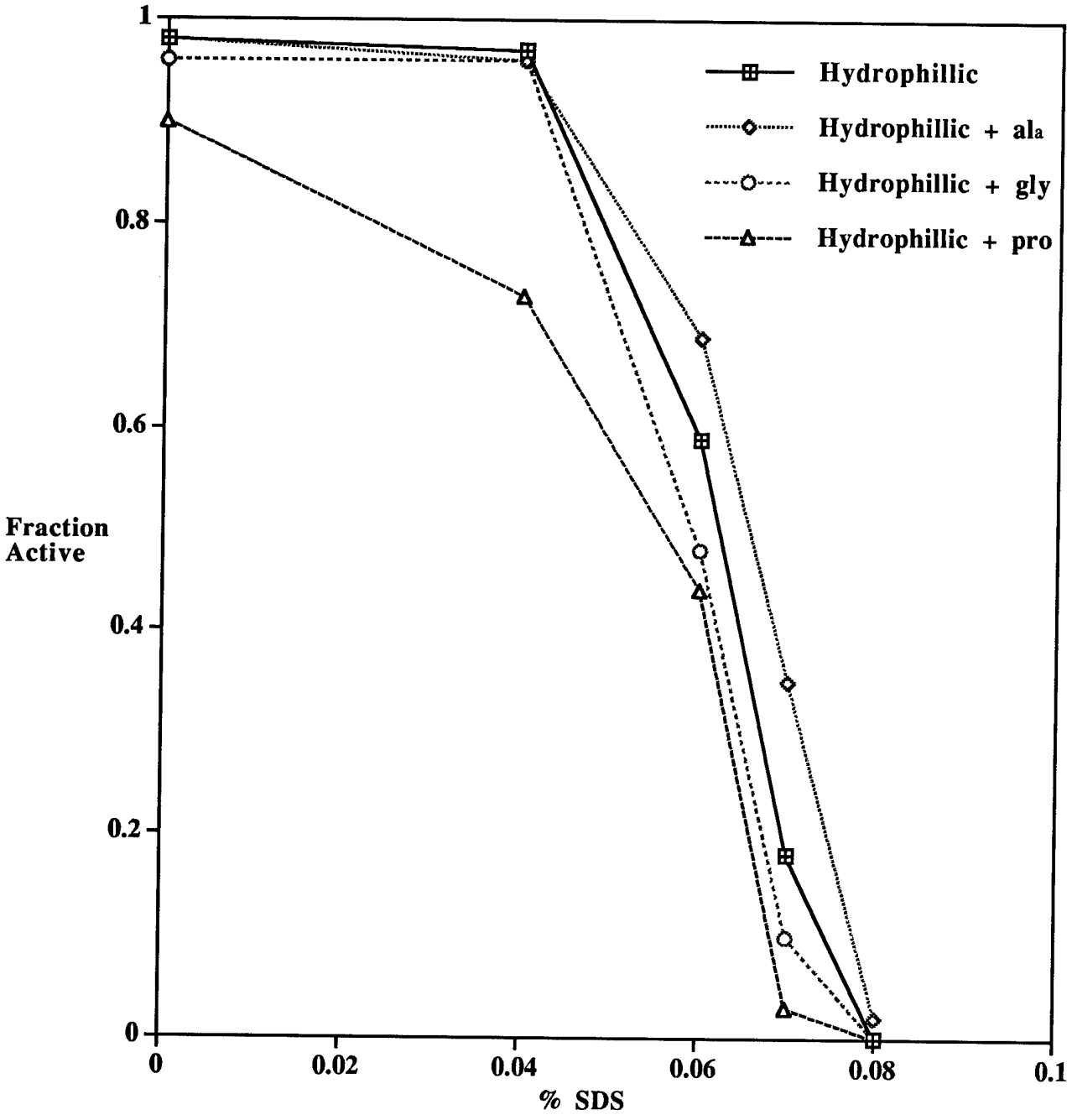


Figure 11. Library Activity in Increasing Denaturant Concentrations. Individual isolates from each liibrary is tested for activity in various concentrations of SDS. A clone is scored positive when its Vmax is greater or equal to 1/2 Vmax of wild type eglin at zero denaturant.

APPENDIX

Draft Manuscript

Thermal Denaturation of Eglin c as a Function of Protein Concentration

by

Jennifer Waldner, Marshall Hall Edgell, and Gary J. Pielak

Thermal Denaturation of Eglin c as a Function of Protein Concentration and pH[†]

[†] This work was supported by US Army Medical Research and Material Command under DAMD17-94-J-4270. J.C.W. was partially supported by a G.A.A.N.N. Fellowship from the US Department of Education.

Jennifer C. Waldner[‡], Marshall Hall Edgell[§], & Gary J. Pielak[‡*]

[‡]Department of Chemistry, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3290,

[§]Department of Microbiology and Immunology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599-3290

[*]Corresponding author: voice (919)966-3671; FAX 966-3675; email gary_pielak@unc.edu

Running Title: Denaturation, Sedimentation, Non-ideality

Abbreviations

AUC, analytical ultracentrifugation; CD, circular dichroism; CI2, chymotrypsin inhibitor II; DSC, differential scanning calorimetry; LB, Luria broth; MW_{app}, apparent molecular weight, rms, root mean square deviation; β, $\Delta H_{vH}/\Delta H_{cal}$; ΔC_p, change in heat capacity upon denaturation; ΔG_D, Gibbs free energy of denaturation; ΔH_D, enthalpy of denaturation; ΔH_{cal}, ΔH_D from calorimetry; ΔH_{vH}, ΔH_D from van't Hoff analysis; $\Delta J_{D,cal}$, ΔG_D and ΔS_D determined by using ΔH_{cal}; $\Delta J_{D,vH}$, ΔG_D and ΔS_D determined by using ΔH_{vH};; ΔS_D, entropy of denaturation

Abstract

We studied the eglin c denaturation by using circular dichroism spectropolarimetry and differential scanning calorimetry. At low protein concentration, thermal denaturation is consistent with a two-state model. At concentrations greater than several hundred micromolar, however, the thermodynamic parameters depend on protein concentration in a way that suggests native-state aggregation. Using analytical ultracentrifugation, we demonstrate that thermodynamic non-idealities explain the high protein-concentrations data. We conclude that the observed increase in T_m with increasing protein concentration is not a result of native state association but, instead, a result of thermodynamic non-ideality.

Fundamental to the rational design of proteins is that all information necessary for a protein to fold is contained in its amino acid sequence (Anfinsen, 1973). To this end, understanding how amino acid sequences effect protein stability is essential. Biophysical characterization measures protein stability by determining the free energy of denaturation, ΔG_D.

Several characteristics make eglin c, a serine protease inhibitor from the leach *Hirudo medicinalis* (Seemüler et al., 1977), a good model for protein stability studies. Eglin c is a monomeric single domain protein containing no metal ion binding sites and no disulfides. It can be expressed in *Escherecia coli* in hundred-milligram quantities and its three-dimensional structure is known to high resolution (Bode et al., 1986; McPhalen and James, 1988; Hyberts and Wagner, 1990; Heinz et al., 1991; Hipler et al., 1992; Hyberts et al., 1992). Eglin c belongs to the potato inhibitor I family of proteins. Another member of this family is chymotrypsin inhibitor II (CI2), a biophysically well characterized protein (Jackson and Fersht, 1991; Jackson et al., 1993; elMasry and Fersht, 1994). Because the structures of eglin c and CI2 are similar [the root mean square (rms) difference between their x-ray crystal structures is 1.68 Å (McPhalen and James, 1988)] the thermal denaturation of

eglin c and CI2 are expected to be similar. It has been shown that CI2 denatures by way of a two-state mechanism. Bae and Sturtevant studied the thermal denaturation of eglin c using differential scanning calorimetry (DSC)(1994). They report that eglin c denaturation is reversible and independent of protein concentration.

DSC is a powerful tool for studying protein stability because it gives a direct measure of ΔH_D, the enthalphy of denaturation (Privalov, 1979). Additionally, DSC data can be analyzed using van't Hoff analysis if a model, most commonly a two-state model, is assumed. By comparing ΔH_D from van't Hoff analysis (ΔH_{vH}) to ΔH_D measured calorimetrically (ΔH_{cal}), information is obtained about the unfolding reaction. If ΔH_{vH} equals ΔH_{cal}, then unfolding is consistent with the proposed model. Additionally, DSC provides a global measure of protein denaturation, in contrast to optical methods that may record isolated changes. DSC does have disadvantages compared to optical methods such as circular dichroism spectropolarimetry (CD). For instance, DSC's low sensitivity demands the use of higher protein concentrations. Unfortunately, highly concentrated protein solutions often exhibit thermodynamic non-ideality which can complicate data analysis.

To confirm the validity of the two-state model, the effect of protein concentration on thermodynamic parameters must be considered. If the temperature at which the thermal transition is half over, T_m, increases with protein concentration, native state oligomerization is implied. Alternately, a decrease in T_m with protein concentration implies denatured state oligomerization (Sturtevant, 1987). Protein oligomerization can be quantitated using analytical ultracentrifugation, AUC. By monitoring the distribution of protein at sedimentation equilibrium using AUC, the molecular mass, interaction stoichiometry and non-ideality can be determined.

We studied the thermal denaturation of eglin c by using CD and DSC. AUC was also used to explore protein oligomerization and thermodynamic non-ideality. We report a detailed thermodynamic analysis of eglin c thermal denaturation. Our results differ from those previously reported by Bae and Sturtevant,(1994).

Materials and Methods

Protein Purification and Expression

The eglin c gene was synthesized and inserted into the pET17b plasmid (Novagen) utilizing the *Eco*RI and *Nde*I restriction enzyme sites to produce the vector, ME007. For protein expression, ME007 was transformed into the *Escherecia coli* strain BL21(DE3)pLysS (F^-*ompT* $r_B^-m_B^-$; Grodberg & Dunn, 1988). A single colony was used to inoculate 50 mL of Luria broth (LB) containing 100 μg ml^{-1} carbenicillin and 34 μg ml^{-1} chloramphenicol and grown to an $OD_{600\ nm}$ of 0.6 with shaking at 37 °C. The cells were then pelleted at 5000 x g for 15 min and used to inoculate 1 L of fresh LB containing 100 μg ml^{-1} carbenicillin and 34 μg ml^{-1} chloramphenicol. The culture was incubated at 37 °C with shaking to an $OD_{600\ nm}$ of 0.8. The cells were pelleted as described above, resuspended in 1 L of LB containing fresh antibiotics, and induced with 0.4 mM isopropyl β-D-thiogalactopranoside. Three hours after induction, the cells were harvested by centrifugation. The cell pellet was stored at -70 °C overnight.

Cells were lysed by the freeze-thaw method (Moffatt & Studier 1987; Studier 1991) and suspended in 100 mL 50 mM Tris-HCl, pH 8.5 and 5 mM EDTA. The cell lysate was treated with 10

μg μL^{-1} deoxyribonuclease for 1 h at 37 °C followed by centrifugation at 10,000 x g for 15 min. The supernatant was removed and dialyzed overnight against 50 mM glycine-HCl, pH 3.0, followed by centrifugation at 10,000 x g for 15 min. The eglin c-containing supernatant was purified using a Sephadex G-75 column equilibrated in 50 mM glycine-HCl, pH 3.0. Fractions analyzed by SDS-PAGE showed eglin c as a single band when stained with Coomasie blue.

Calorimetry

Experiments were performed in 50 mM glycine-HCl buffer at pH 1.5, 2.0, 2.5, 3.0 and 3.3. Under these conditions the protein is soluble and no aggregation is observed on heating. Samples were prepared by dialysis overnight against the desired buffer. The samples were then sterile filtered and degassed immediately prior to loading the sample cell.

The molar extinction coefficient of eglin c was determined by quantitative amino acid analysis and checked by the difference spectrum method (Edelhoch, 1967; Gill and von Hipple, 1989). These two values agree within experimental uncertainty yielding an ϵ_{280} of 12,430 $M^{-1}cm^{-1}$. The protein concentration was determined by measuring the absorbance, in a 1.00 mm path-length cuvette, so

that no sample dilution was necessary. Except for concentration-dependence experiments, protein concentrations of 100 - 130 μM were used.

Calorimetry was performed with a MCS-Differential Scanning Calorimeter from MicroCal Inc., Northampton MA. The dialysis buffer was placed in both the sample and reference compartments. Several thermograms were acquired to establish an instrumental baseline. The sample buffer was then replaced with the eglin c solution. Samples were scanned from 10 to 90 °C. Except for scan rate-dependence studies, a scan rate of 60 °C h^{-1} was used.

Analysis of Calorimetric Data

The appropriate buffer baseline was subtracted from the thermogram and the data were converted to molar excess heat capacity using the protein concentration and a cell volume of 1.21 mL. A baseline through the pre- and post- transition was subtracted to eliminate any changes in heat capacity between the folded and denatured states. Thermograms were analyzed as a non-two-state transition using MicroCal Origin software. Protein-concentration-dependent data were also analyzed with a model that includes reversible dimerization in the folded and denatured states (Liggins et al., 1994).

Circular Dichroism Spectropolarimetry

Data were acquired with an Aviv model 62DS spectropolarimeter equipped with a five-position sample chamber. Thermal denaturations were carried out in 50 mM glycine-HCl at 1 °C intervals, with ≈6.0 min between temperatures. The protein concentration was 60 - 70 μM. The ellipticity at 227 nm was followed from 5 to 70 °C for pH 1.5 and 2.0, and from 30 to 90 °C for pH 2.5, 3.0 and 3.3. Reversibility was checked by returning the heated samples to the initial temperature and repeating the denaturation experiment. Data were fit to a two-state model with linear baselines and thermodynamic parameters were obtained as described by Cohen and Pielak (1994).

Analytical Ultracentrifugation

Sedimentation equilibrium experiments were performed with a Beckman XLA analytical ultracentrifuge at the Macromolecular Interactions Facility at UNC-Chapel Hill. Protein concentrations were 700, 350, 175, 54, 27, and 13 μM in 50 mM glycine-HCl buffer, pH 3.0. The samples were centrifuged at 22,000, 27,000, 40,000, and 45,000 rpm. The overlay of successive absorbance scans monitored at 280 nm was used to define equilibrium. The

concentration distribution of protein at equilibrium, c(r), is defined by:

$$c(r) = c_o e^{\sigma\xi} \qquad (1)$$

where

$$\sigma = \frac{M(1-J\rho)\omega^2}{RT} \qquad (2)$$

c_o is the initial protein concentration in moles L^{-1}, MW_{app} is the apparent molecular weight in g mol^{-1} , $\bar{v}$ is the partial specific volume in ml g^{-1} calculated from the amino acid composition, ρ is the buffer density in g ml^{-1}, ω is the angular velocity in radians s-1, and ξ is a function of the radial distance. Our data were fit to a modified form of equation 1 which includes the second virial coefficient, B,

$$c(r) = c_o e^{\sigma\xi - 2B[c(r)]} \qquad (3)$$

using Nonlin (Johnson et al., 1981; Laue, 1995) to find MW_{app}.

Results

Calorimetry

Reversibility was determined by comparing thermograms from successive scans of the same sample (Figure 1). Denaturation is reversible because ΔH_{cal} from successive scans retains ≈95% of the transition enthalpy of the previous scan. If the protein sample is rapidly cooled immediately after the start of the denatured baseline, ΔH_{cal} is >95% of the value from the previous scan.

The system is at equilibrium when denaturation and folding are much faster than the scan rate. Under these conditions, T_m and $\Delta H_{vH}/\Delta H_{cal}$ will be independent of scan rate (Lepock et al., 1992). To ensure our data were obtained at equilibrium, T_m and β were acquired as a function of scan rate. The values of T_m and β do not change with scan rate (Table 1) showing that our data represent equilibrium measurements.

The thermal denaturation of eglin c was monitored between pH 1.5 and 3.3. Typical data are shown in Figure 2 along with the fit to a non-two-state transition. Thermograms are symmetric at all protein concentrations (Marky and Breslaur, 1987) and show a positive value for ΔC_p, the change in heat capacity upon denaturation. The average values of T_m , ΔH_{vH} and ΔH_{cal} at five

different pH values are given in Table 2.

The ratio $\Delta H_{vH}/\Delta H_{cal}$, is a measure of how closely denaturation follows the model used to obtain ΔH_{vH} (Sturtevant, 1987). In our case, if the ratio is unity, then the data are well described by a two-state denaturation model. Values for $\Delta H_{vH}/\Delta H_{cal}$ are reported in Table 2. The uncertainty was estimated by applying propagation of error analysis using the measured uncertainties in ΔH_{vH} and ΔH_{cal}. At low protein concentration, denaturation is consistent with a two-state model because $\Delta H_{vH}/\Delta H_{cal}$ is unity.

To utilize a two-state model, denaturation must be independent of protein concentration. Therefore, eglin c denaturation was monitored between 80 and 700 μM. Over this concentration range, T_m increases 10 °C and ΔH_{cal} increases 30 kcal mol^{-1}. Because ΔH_{vH} is concentration independent but ΔH_{cal} increases with protein concentration, $\Delta H_{vH}/\Delta H_{cal}$ (Figure 3). These results do not agree with the previous report of Bae and Sturtevant (1995). They report no differences in T_m, or ΔH_{cal} when protein concentration is increased from 150 to 500 μM.

CD-detected thermal denaturation

Reversibility was checked by returning the heated samples to

the initial temperature and repeating the denaturation experiment (data not shown). Denaturation is structurally reversible because >95% of the native signal returns upon cooling to the initial temperature. Denaturation is thermodynamically reversible because the parameters obtained from the first and second scan are identical within the uncertainty of the measurements.

Thermal denaturation was monitored between pH 1.5 and 3.3. Figure 4 shows plots of fraction denatured versus temperature along with the fit to a two-state model, producing T_m and ΔH_{vH}. The average values of T_m and ΔH_{vH} at five different pH values are given in Table 3.

Determination of ΔC_p

In principle, ΔC_p can be determined from one DSC thermogram. In practice, however, this method is unreliable (Connelly et al., 1991; Liggins et al., 1994; Liu and Sturtevant, 1996; Betz et al., 1997). A more accurate ΔC_p is obtained by measuring the temperature dependence of ΔH_D over a range of pH values. Because calorimetric data was analyzed both directly and by van't Hoff analysis, the temperature dependence of both ΔH_{vH} and ΔH_{cal} were used(Figure 5) to determine ΔC_p. Values of ΔC_p using ΔH_{vH} and ΔH_{cal}

from DSC are 0.84 ± 0.07 kcal mol^{-1} K^{-1} and 0.82 ± 0.19 kcal mol^{-1} K^{-1}, respectively. Plots of ΔH_{vH} versus T_m from CD-monitored thermal denaturation yield a ΔC_p of 0.84 ± 0.07 kcal mol^{-1} K^{-1} (Figure 5). All these ΔC_p values are identical within the measured uncertainty and agree with the value of 0.73 kcal mol^{-1} K^{-1} reported by Bae and Sturtevant. Our ΔC_p values also agree with values for CI2, 0.789 kcal mol^{-1} K^{-1}, (Jackson and Fersht, 1991.

pH effects

The data in Tables 2 and 3 show that T_m and ΔH_D are affected by pH. To characterize the pH dependence of thermal denaturation, the ionization of the native and denatured states must be considered. The number of protons bound to the denatured state minus the number bound to the native state, $\Delta\nu$, can be estimated using equation 4 (Privalov and Ptitsyn, 1969; Ptitsyn and Birshtein, 1969):

$$\Delta\nu = \frac{\Delta H_D}{2.3RT_m^2} \bullet \frac{\delta T_m}{\delta pH} \qquad (4)$$

Values of $\Delta\nu$ were calculated at each pH by using ΔH_D and T_m from

Table 3. The $\delta T_m/\delta pH$ term was calculated by fitting a second-order polynomial to plots of T_m versus pH (Figure 6A) and evaluating the first derivative at each T_m. Figure 6B shows Δv is greatest at pH 3.3 and decreases with decreasing pH.

Values for ΔH_D at 300K for each pH were calculated using equation 5:

$$\Delta H_{D,T} = \Delta H_m + \Delta C_p(T - T_m) \quad (5)$$

where ΔH_m is ΔH_D at T_m. Uncertainties in ΔH_D were calculated by applying propagation of error analysis to equation 5. Because the ionization enthalpies of glycine and carboxylates are small, no corrections were made to the observed enthalpies reflecting these ionization heats.

From the cardinal parameters, (ΔH_D, T_m, and ΔC_p), ΔG_D was calculated using a modified form of the integrated Gibbs-Helmholtz equation (Elwell and Schellman, 1977):

$$\Delta G_{D,T} = \Delta H_m\left(1 - \frac{T}{T_m}\right) - \Delta C_p\left[\left(T_m - T\right) + T\ln\left(\frac{T}{T_m}\right)\right] \quad (6)$$

The uncertainty in ΔG_D was estimated by applying propagation of error analysis to equation 6 (Cohen and Pielak, 1994). Values of ΔG_D increase with increasing pH (Figure 7). Because ΔH_D is assumed to be pH independent, changes in ΔG_D are due to the pH

dependence of ΔS_D. Values of ΔS_D were calculated using equation 7:

$$\Delta S_D = \Delta S_m + \Delta C_p \ln\left(\frac{T}{T_m}\right) \quad (7)$$

where ΔS_m is the entropy of denatration at T_m. Plots of $T\Delta S_D$ versus pH are shown in Figure 7.

Analytical Ultracentrifugation

To characterize the native state oligomerization of eglin c, sedimentation equilibrium experiments were completed at varying protein concentrations. Global fitting of the 54, 27, and 13 μM data to equation 3 yielded a value MW_{app}, of 7500 ± 800 D, which is in agreement with the monomer molecular weight of 8200 D, and a value for B of 0.05 M^{-1}. The rms of the fit is 0.01, within the accepted noise value of the instrument (Laue, 1995). When the data were fit separately, MW_{app} decreased from 9300 D to 6700 D upon increasing the protein concentration from 13 to 54 μM.

Sedimentation equilibrium experiments were also attempted at 175, 350, and 700 μM but the samples did not reach equilibrium after 16 h at 27,000 rpm. This observation is inconsistent with the low concentration data. Only the data collected at 35,000 and 40,000 rpm were analyzed. As seen at low protein

concentration, there is a trend of decreasing MW_{app} with increasing protein concentration, but specific values for MW_{app} were not determined, because the data do not fit the model (see discussion).

Discussion

two-state behavior

The DSC and CD detected thermal denaturations show that eglin c undergoes a cooperative transition. At low protein concentration, denaturation appears to be a two-state process where only the native and denatured states are significantly populated. Comparison of ΔH_{vH} and ΔH_{cal} obtained from DSC thermograms provides useful information about the denaturation process. Here the mean ratio of ΔH_{vH} to ΔH_{cal} from Table 2, assuming a two-state process, is 1.01 ± 0.07. This suggests that eglin c undergoes cooperative two-state denaturation. This is in agreement with denaturation data collected for CI2 which shows the average ratio of $\Delta H_{vH}/\Delta H_{cal}$ is 0.98 ± 0.03 (Jackson and Fersht, 1991).

These results differ from those of Bae and Sturtevant (1995). They report a value for $\Delta H_{vH}/\Delta H_{cal}$ of 1.18 ± 0.02 for eglin c. Comparison of the data in Table 2 to the values reported by Bae and Sturtevant show that ΔH_{vH} values are in agreement but values for ΔH_{cal} are not. Therefore, this discrepancy may be a result of using different methods to quantitate protein concentration, as ΔH_{cal} is protein

concentration dependent.

Determination of ΔC_p

Inspection of Figure 1 shows that the heat capacity of the native state is less than that of the denatured state. This increase in heat capacity upon denauration is the hallmark of the hydrophobic interaction (Privalov and Gill, 1989). The heat capacity change is thought to arise from the hydration of non-polar residues which are exposed upon denaturation. The change in heat capacity upon denaturation therefore reflects the change in the degree of exposure of hydrophobic residues (Pfeil and Privalov, 1976; Privalov and Gill, 1988).

Values of ΔC_p for proteins lie in the range between 10.4 cal mol^{-1} K^{-1} $residue^{-1}$ (ribonuclease A) and 17.8 cal mol^{-1} K^{-1} $residue^{-1}$ (myogobin)(Privalov and Gill, 1988). The average value determined for eglin c, 11.7 cal mol^{-1} K^{-1} $residue^{-1}$ is at the lower end of the range. Small proteins, such as eglin c, which are unable to form a large hydrophobic core due to their size, exhibit small ΔC_p values.

Inspection of Figure 1, shows that the slopes of the pre- and post- transition baselines are different. This observation suggests that ΔC_p changes with temperature. This is in contrast

with ΔC_p values obtained from ΔH_D versus T_m plots (Figure 5). These plots are fit to a straight line with a high correlation coefficient suggesting that ΔC_p is temperature independent. The reason for this discrepancy is unknown but has been observed by others (Liggins et al., 1994; Kitamura and Sturtevant, 1989; Connelly et al., 1991).

pH effects

As indicated in Figure 7, ΔG_D is affected by pH over the range studied. To characterize the pH dependence of thermal denaturation, the ionization of the native and denatured states must be considered. As the pH is lowered, more groups on the protein become protonated, increasing the charge on the native and the denatured states. Proton uptake occurs during denaturation when carboxyl groups involved in salt links in the native protein take up a proton during unfolding. Generally, carboxyl groups are protonated at pH 3.5, however due to the net positive charge on the protein at pH 2-3, some carboxyl groups may have an abnormally low pK_a in the native state (Anderson et al., 1990). These low pKa values result in a decrease in Δv with decreasing pH.

Protein concentration effects; evidence for thermodynamic non-ideality

The increase in T_m with increasing protein concentration (Figure 3) suggests association of the native state. We tested this possibility by fitting our data to an extended two-state model which permits dimerization of the native state, the denatured state or both (see Materials and Methods). In all cases, the data were not well fit with dimerization constants. The rms deviation between the data and the model is typically 38% of the temperature maximum of the transitional heat capacity, compared to a rms of 1.5% for a two-state model. Clearly, the two-state model provides a better fit. Furthermore, fitting of the data with dimerization constants, gave negligible dimerization in the native state and extensive dimerization in the denatured state. This result contradicts the trend we see of increasing T_m with increasing protein concentration.

Analyzing the symmetry of DSC thermograms provides information about the denaturation process. Symmetric DSC thermograms are indicative of two-state denaturation while asymmetric curves are indicative of end-state aggregation (Marky and Breslaur, 1987). Eglin c thermograms are symmetric at all protein concentrations

suggesting two-state denaturation without native state association. In light of these results we can not explain the increase in T_m with increasing protein concentration as a native state aggregation effect.

To confirm that native eglin c is monomeric and not undergoing native state association, AUC experiments were completed at varying protein concentrations. An increase in the MW_{app} with increasing protein concentration demonstrates protein oligomerization whereas a decrease in the MW_{app} with increasing protein concentration indicates thermodynamic non-ideality (Laue, 1995). Thermodynamic non-ideality results from nonfavorable interactions between solute (protein) molecules resulting in an altered concentration distribution of protein (Yphantis and Roark, 1972). Under these conditions, data are fit using equation 3 which includes the second virial coefficient, B. Global fitting of the 13, 27 and 54 μM data to equation 3 show that eglin c is monomer. Additionally, when the data were fit separately MW_{app} decreased with increasing protein concentration indicating thermodynamic non-ideality.

Ultracentrifugation experiments were also completed at the higher protein concentrations corresponding to those used for calorimetry. Fitting these data to equation 3 produced

erroneously low values for the MW_{app} and an unacceptable rms deviation, indicating that the high concentration data do not fit the model. Yphantis and Roark show that equation 3 cannot be utilized for proteins exhibiting large non-idealities (1972).

Thermodynamic non-idealities result from either charge-charge repulsion (Donnan effect) or excluded volume effects. Thermodynamic non-idealities arising from charge-charge repulsion occur if the macromolecule (protein) carries a significant charge. During sedimentation of charged proteins, the buffer ions must redistribute to maintain electroneutrality, coupling the sedimentation of the protein molecules and the buffer ions. This coupling results in low values for the MW_{app} (Yphantis and Roark, 1972). Our experiments were performed at pH 3.0 where eglin c carries a +10 charge. Generally, charge-charge repulsions are eliminated at high buffer concentrations (0.1 to 1 M)(Teller, 1973). Our experiments were performed in 50 mM buffer, the same buffer concentration used for denaturation experiments. The low buffer concentration used and the high charge associated with eglin c makes our system a candidate for non-idealities resulting from charge-charge repulsion.

Non-idealities also result from excluded volume effects. Excluded volume effects describe the different degree of

solvation of a protein when dissolved with different cosolutes (Schachman and Lauffer, 1949; Jacobsen et al., 1996). This solvation affect is detected in the ultracentrifuge because proteins sediment at different rates depending on the size of the cosolute present. In our case, eglin c would be acting to excluded its own volume. To test this idea, we have calculated the effect crowding would have on ΔG_D using scaled particle theory (Berg, 1990). We found that excluded volume affects are insignificant, contributing < 0.2 kcal mol^{-1} to ΔG_D at the highest protein concentration.

Our AUC data show our system to be affected by thermodynamic non-ideality. The thermodynamic non-ideality results from charge-charge repulsions due to the low pH and low buffer concentration used. AUC detects non-idealities using colligative properties. Therefore any technique based on colligative properties, *e.g.* calorimetry, is subject to these same non-ideal effects. Because eglin c is monomeric at all protein concentrations, we believe that the increase in T_m associated with increased protein concentration is not a native state aggregation effect but is a result of thermodynamic non-ideality.

Acknowledgements

We thank Chris Lombardo for assistance with ultracentrifugation experiments, Tom Laue for help interpreting the centrifugation data, David Cohen for assistance with the DSC, Dorothy Erie for helpful discussions, Anne Broadwater for inserting the eglin c gene into the pET17b plasmid, and the Pielak group for helpful discussions.

References

Anfinsen, C. B. (1973) Science 181, 223-230.

Bae, S. and Sturtevant, J. M. (1995) *Biophys. Chem. 55*, 247-252.

Berg, O. G. (1990) *Biopolymers 30*, 1027-1037.

Bode W., Papamokos, E., Musil, D., Seemueller, U., and Fritz, H. (1986) *EMBO J. 5*, 813-818.

Cohen, D. S., and Pielak, G. J. (1994) *Protein Sci. 3*, 1253-1260.

Connelly, P., Ghosaini, L., Hu, C., Kitamura, S., Tanaka, A., and Sturtevant, J. M. (1991) *Biochemistry 30*, 1887-1891.

Edelhoch, H. (1967) *Biochemistry 6*, 1948-1954.

elMasry, N. F. and Fersht, A. R. (1994) *Protein Eng.* 7, 777-782.

Elwell, M. L. and Schellman, J. A. (1977) *Biochim. Biophys. Acta. 494*, 367-383.

Gill, S.C., and von Hipple, P.H. (1989) *Anal. Biochem 182*, 319-326.

Heinz, D. W., Priestle, J. P., Rahuel, J., Wilson, K. S., and Grütter, M. G. (1991) *J. Mol. Biol. 217*, 353-371.

Hipler, K., Priestle, J. P., Rahuel, J. and Grütter, M. G. (1992) *FEBS Lett. 309*, 139-145.

Hyberts, S. G. and Wagner, G. (1990) *Biochemistry 29*, 1465-1474.

Hyberts, S. G., Goldberg, M. S., Havel, T. F., and Wagner, G. (1992) *Protein Sci. 1*, 736-751.

Jackson, S. E., and Fersht, A. R. (1991) *Biochemistry 30*, 10428-10435.

Jackson S. E., Moracci, M., elMasry, N., Johnson, C. M., and Fersht, A. R. (1993) *Biochemistry 32*, 11259-11269.

Jacobsen, M. P., Wills, P. R., and Winzor, D. J. (1996) *Biochemistry 35*, 13173-13179.

Johnson, M. L., Correla, J. J., Yphantis, D. A., and Halvorson, H. R. (1981) *Biophys J. 36*, 575-588.

Laue, T. M. (1995) *Methods Enzymol. 259*, 427-453.

Lepock, J. R., Ritchie, K. P., Kolios, M. C., Rodahl, M., Heinz, K. A., and Kruuv, J. (1992) *Biochemistry 31*, 12706-12712

Liggins, J. R., Sherman, F., Mathews, A. J., and Nall, B.T. (1994) *Biochemistry 33*, 9209-9219.

Liu, Y. And Sturtevant, J. M. (1996) *Biochemistry 35*, 3059-3062.
Kitamura, S., and Sturtevant, J. M. (1989) *Biochemistry 28*, 3788-3792.

Marky, L. A. And Breslauer, K. J. (1987) *Biopolymers 26*, 1601-1620.

McPhalen, C. A. and James, M. N. G. (1988) *Biochemistry 27*, 6582-6598.

Moffatt, B. A. and Studier, F. W. (1987) *Cell 49*, 221.

Pfeil, W. And Privalov, P. L. (1976) *Biophys. Chem. 4*, 33-40.

Privalov, P. L. And Ptitsyn O. B. (1969) *Biopolymers 8*, 559-571.

Privalov, P. L. (1979) *Adv. Protein Chem. 33*, 167-241.

Privalov, P. L. And Gill, S. J. (1988) *Adv. Protein Chem. 39*, 191-234.

Ptitsyn, O. B. And Birshtein, T. M. (1969) *Biopolymers 7*, 435-445.

Schachman, H. K. And Lauffer, M. A. (1949) *J. Am. Chem. Soc. 71*, 536-541.

Seemüller, U., Meier, M., Ohlsson, K., Müller, and Fritz, H. (1977). *Hoppe-Seyler's Z. Physiol. Chem. 358*, 1105-1117.

Studier, F. W. (1991) *J. Mol. Biol. 219*, 37-44.

Sturtevant, J. M. (1987) *Ann. Rev. Phys. Chem. 38*, 463-483.

Tanaka, A., Flanagan, J., Sturtevant, J. M. (1993) *Protein Sci 2*, 567-576.

Teller, D. (1973) *Methods Enzymol. 27*, 346-441.

Yphantis, D. A and Roark, D. E. (1972) Biochemistry 11, 2925-2934.

Table 1: Scan rate dependence of T_m and β[a].

Scan rate (°C h^{-1})	T_m (±0.1°C)[b]	β (±0.06)[b]
20	62.7	0.97
60	63.1	1.17
90	62.6	0.98

[a]50 mM glycine-HCl, pH 3.0, 130 μM eglin c. [b]uncertainties were determined as described in the text.

Table 2: Thermodynamic Parameters From DSC[a]

pH	$T_m \pm 0.1$[b] (°C)	$\Delta H_{vH} \pm 0.7$[b] (kcal mol^{-1})	$\Delta H_{cal} \pm 5.0$[b] (kcal mol^{-1})	$\beta \pm 0.07$[c]
1.5[d]	45.4	49.7	45.6	1.09
2.0[d]	48.1	54.8	58.3	0.94
2.5[d]	56.8	63.1	61.3	1.03
3.0[e]	66.5	69.6	71.0	0.98
3.3[d]	75.0	72.9	72.9	1.00

[a] Values of ΔH_{vH} and ΔH_{cal} are reported at T_m. [b]Uncertainties are the standard deviation of the mean from three trials at pH 3.0. Uncertainties at other pH values are assumed to be that of pH 3.0. [c]Uncertainties in β were determined as described in the text. [d]The average values from two or [e]three repetitions.

Table 3: Thermodynamic Parameters From CD[a]

pH	$T_m \pm 1.6$[b] (°C)	$\Delta H_{vH} \pm 6.3$[b] (kcal mol^{-1})
1.5 [c]	45.1	45.3
2.0 [c]	47.3	43.8
2.5 [d]	54.6	51.2
3.0 [e]	62.2	57.4
3.3 [d]	69.0	64.3

[a]ΔH_{vH} is reported at T_m. [b]Uncertainties are the standard deviation of the mean from five repetitions at pH 3.0. Uncertainties at other pH values are assumed to be that of pH 3.0. [c]The average values from two, [d]three, or [e]five repititions.

Figure Legends

Figure 1: Reversibility of eglin c thermal denaturation in 50 mM glycine-HCl, pH 3.0. From top to bottom, the first through fifth thermograms for the same sample. Each thermogram retains 95% of the calorimetric enthalpy of the preceding scan.

Figure 2: Typical thermogram obtained with 130 μM eglin c in 50 mM glycine-HCl, pH 3.0. The baseline through the pre- and post-transition was subtracted to eliminate any changes in heat capacity between the folded and denatured states. Also shown is the curve from non-linear least squares fitting. The rms deviation of the data from the theoretical curve is 1.5% of the temperature maximum of the transitional heat capacity.

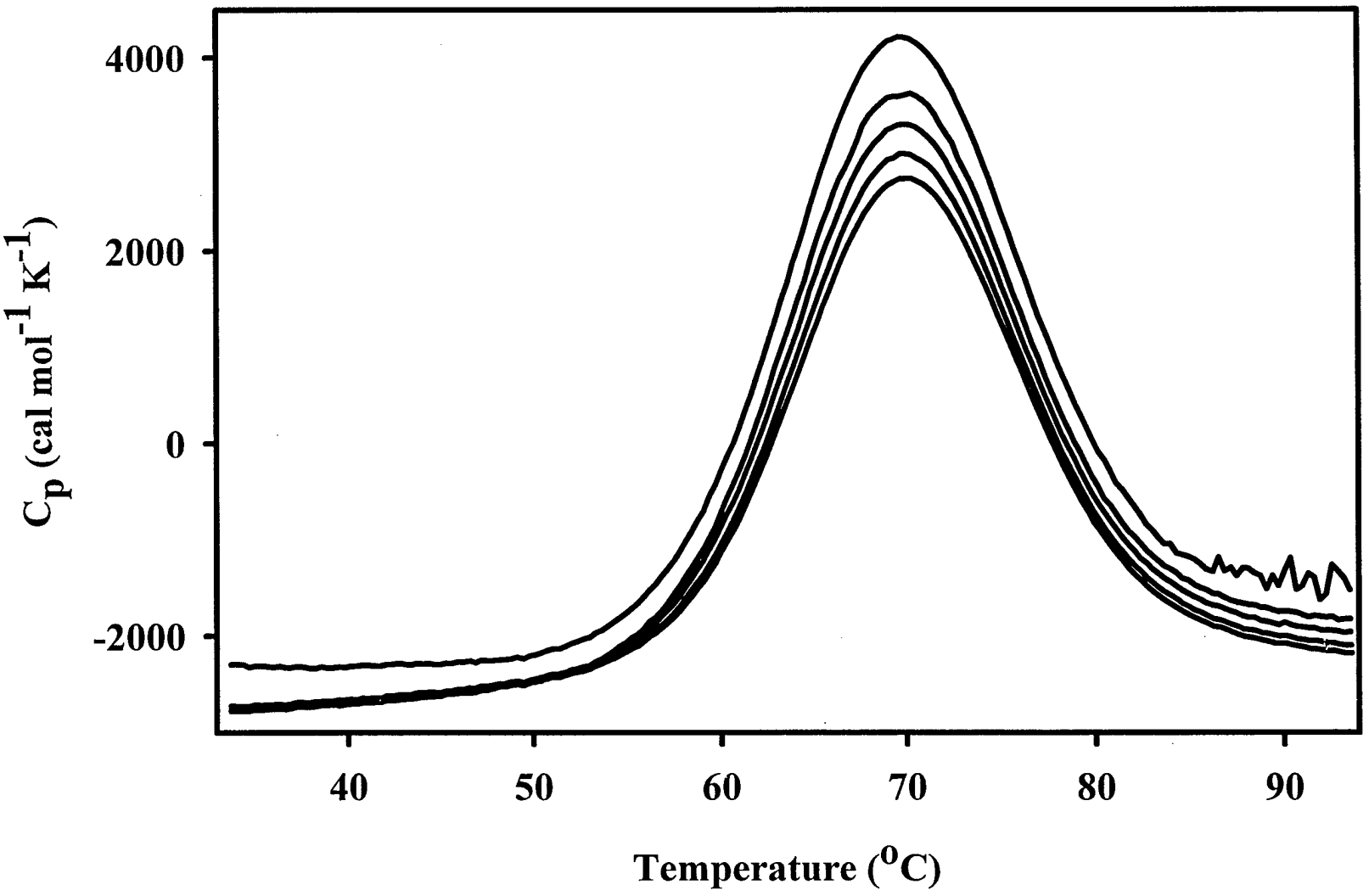
Figure 3: The concentration dependence of T_m and β. The uncertainty in β was determined as described in the text. The uncertainties in T_m are smaller than the symbols.

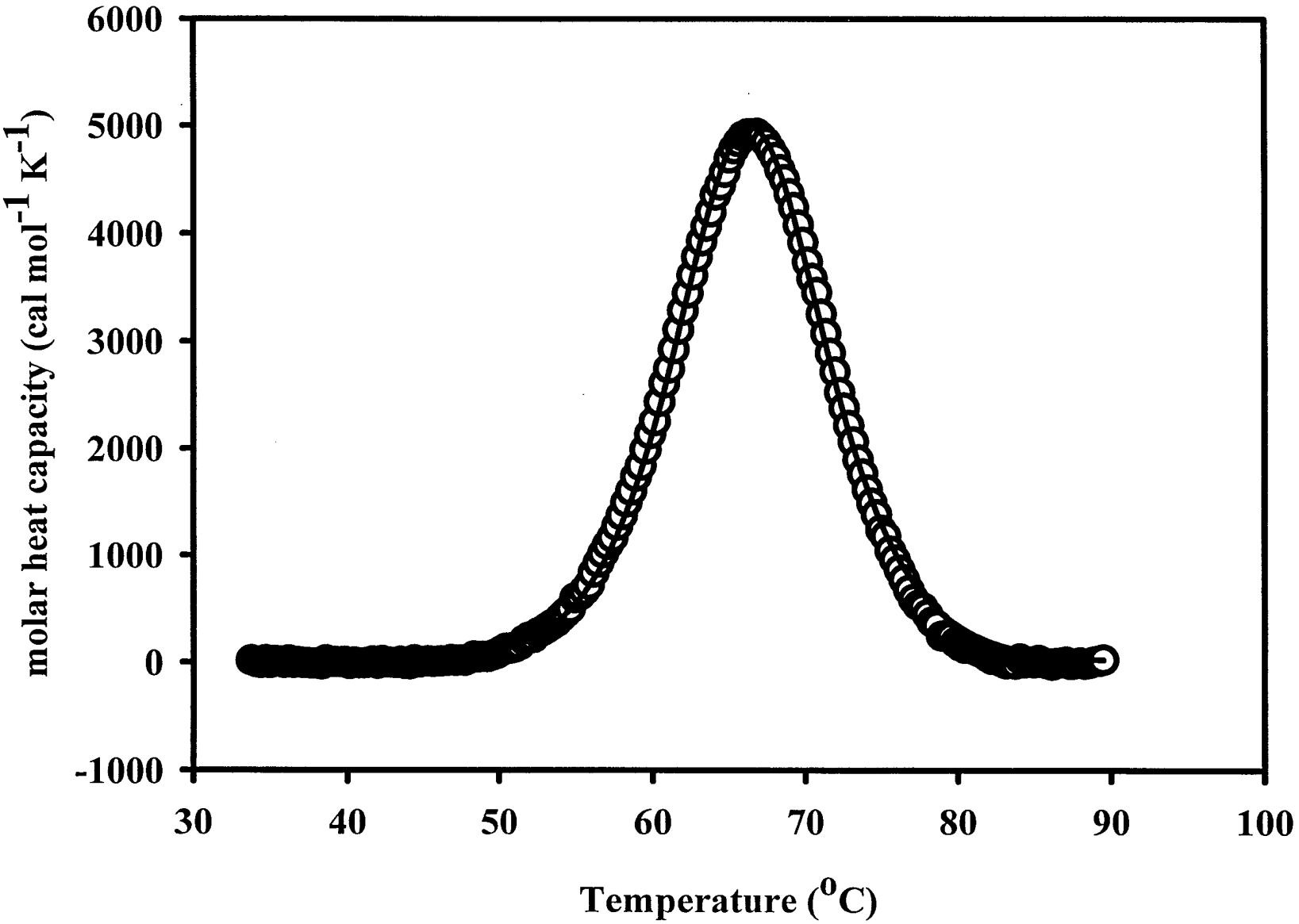
Figure 4: Fraction denatured-versus-temperature for CD-detected thermal denaturation at pH (∇)1.5, (○) 2.0, (□)3.0, and (Δ)3.3. Curves from the nonlinear least-squares fitting are also shown.

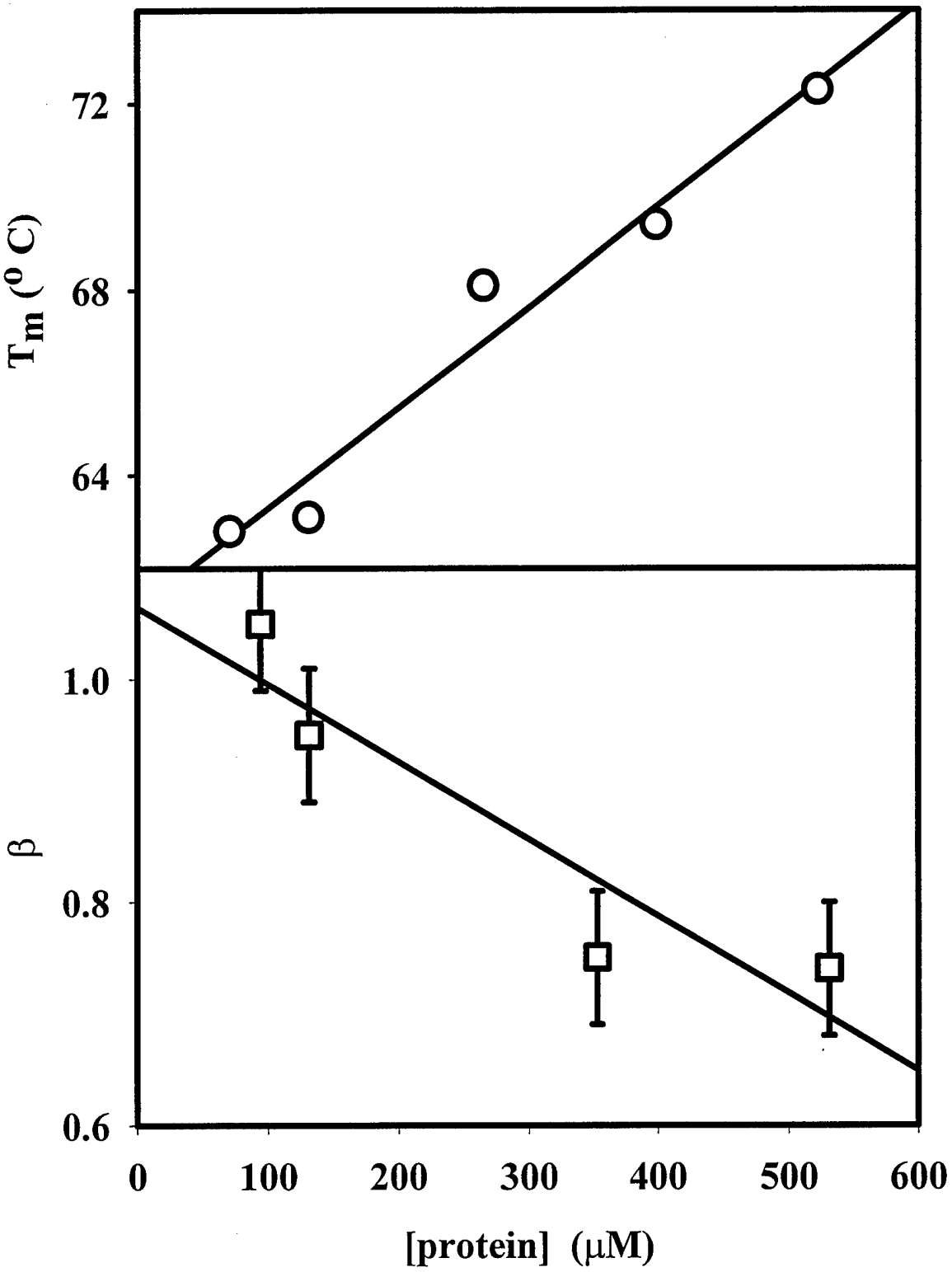
Figure 5: Plots of (O) ΔH_{cal}, (□) ΔH_{vH}(DSC), and (Δ) ΔH_{vH}(CD) versus T_m. The average values for each pH are shown (see Table 2). The error bars are the standard deviation of the mean from repitition of the same experiment.

Figure 6: **A.** Dependence of T_m on pH from CD data. **B.** Δν-versus-pH plot calculated using equation 3 and values from Table 3. The curves for T_m and Δν are of no theoretical significance

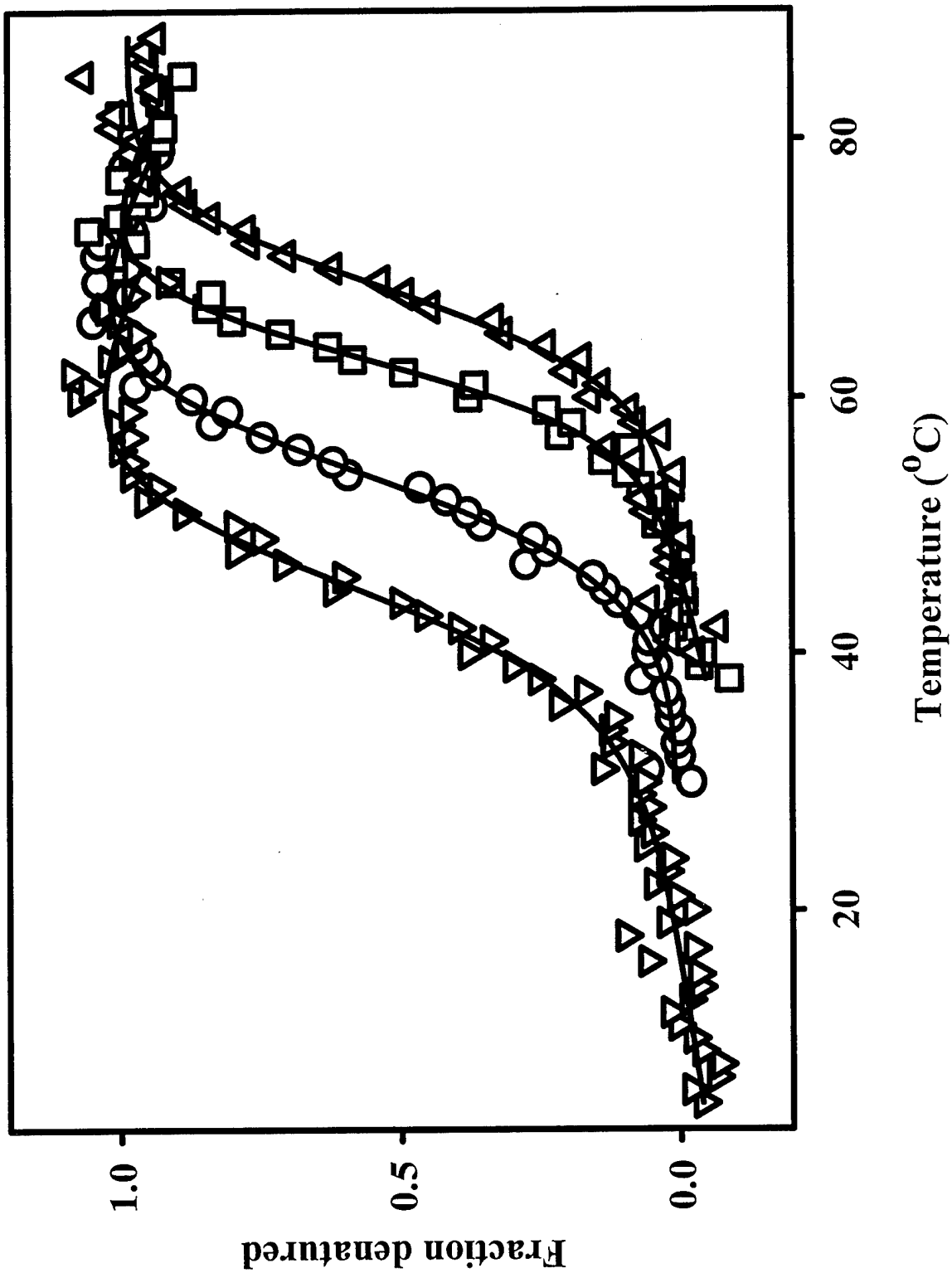
Figure 7: Plots of ΔG_D, ΔH_D, and $T\Delta S_D$ at 300 K versus pH. Calculations are as described in the text. Data are from Table 3. Uncertainties were calculated from propagation of error analysis. The curves for ΔG_D and $T\Delta S_D$ are of no theoretical significance.

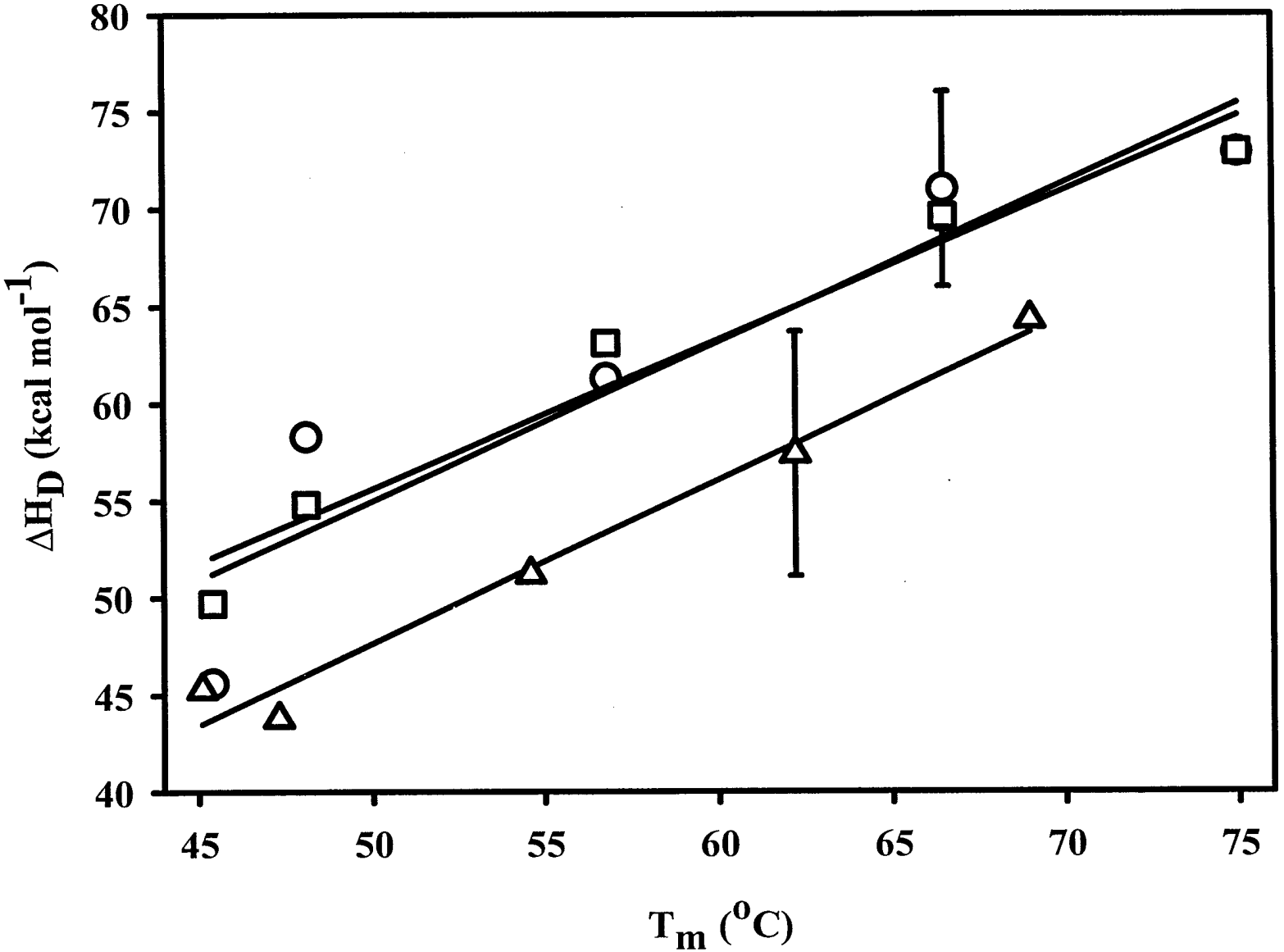

4000
2000
0
-2000
C_p (cal mol^{-1} K^{-1})
40
50
60
70
80
90
Temperature (^{o}C)

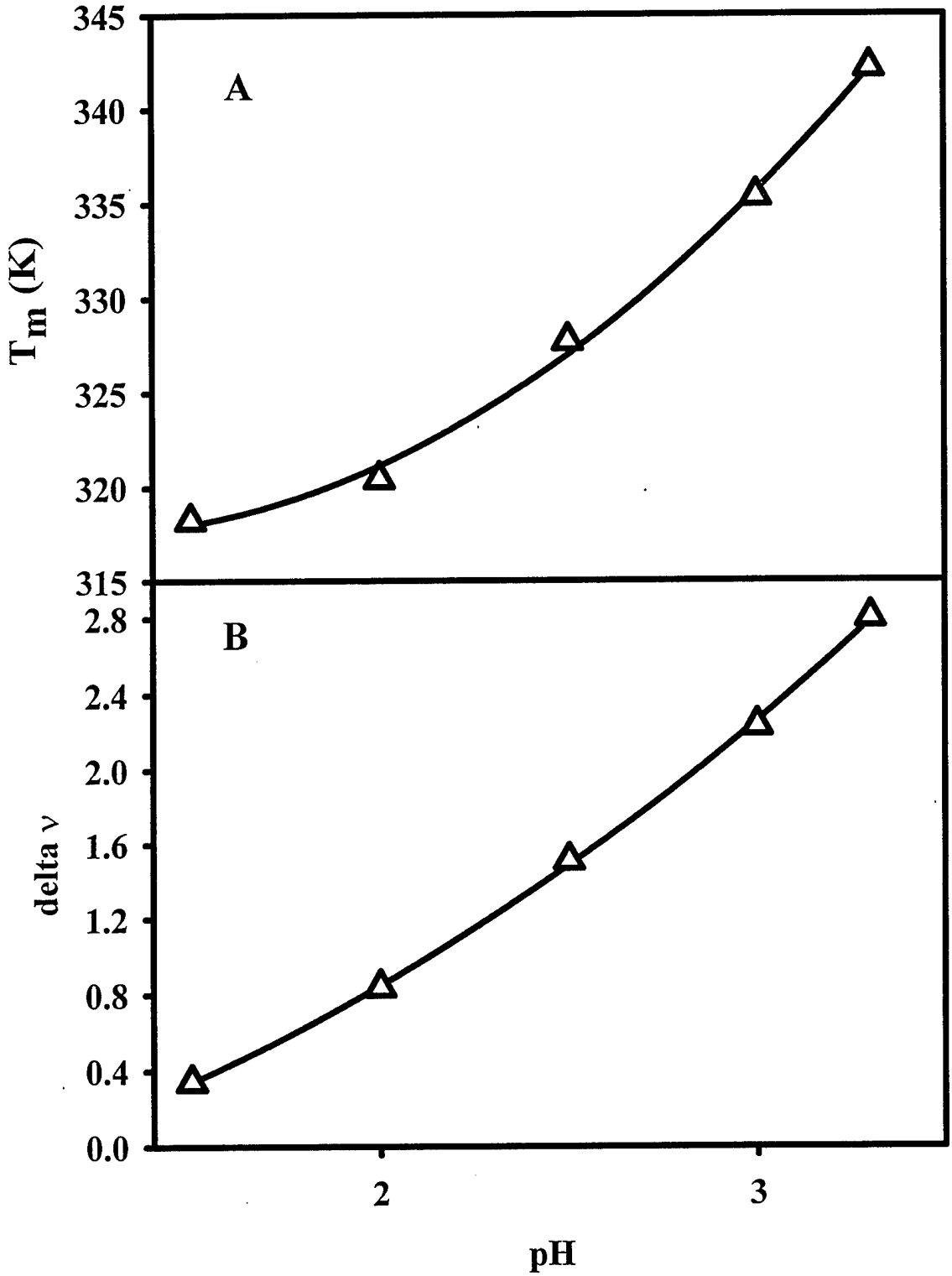

6000
5000
4000
3000
2000
1000
0
-1000
molar heat capacity (cal mol-1 K-1)
30
40
50
60
70
80
90
100
Temperature (oC)

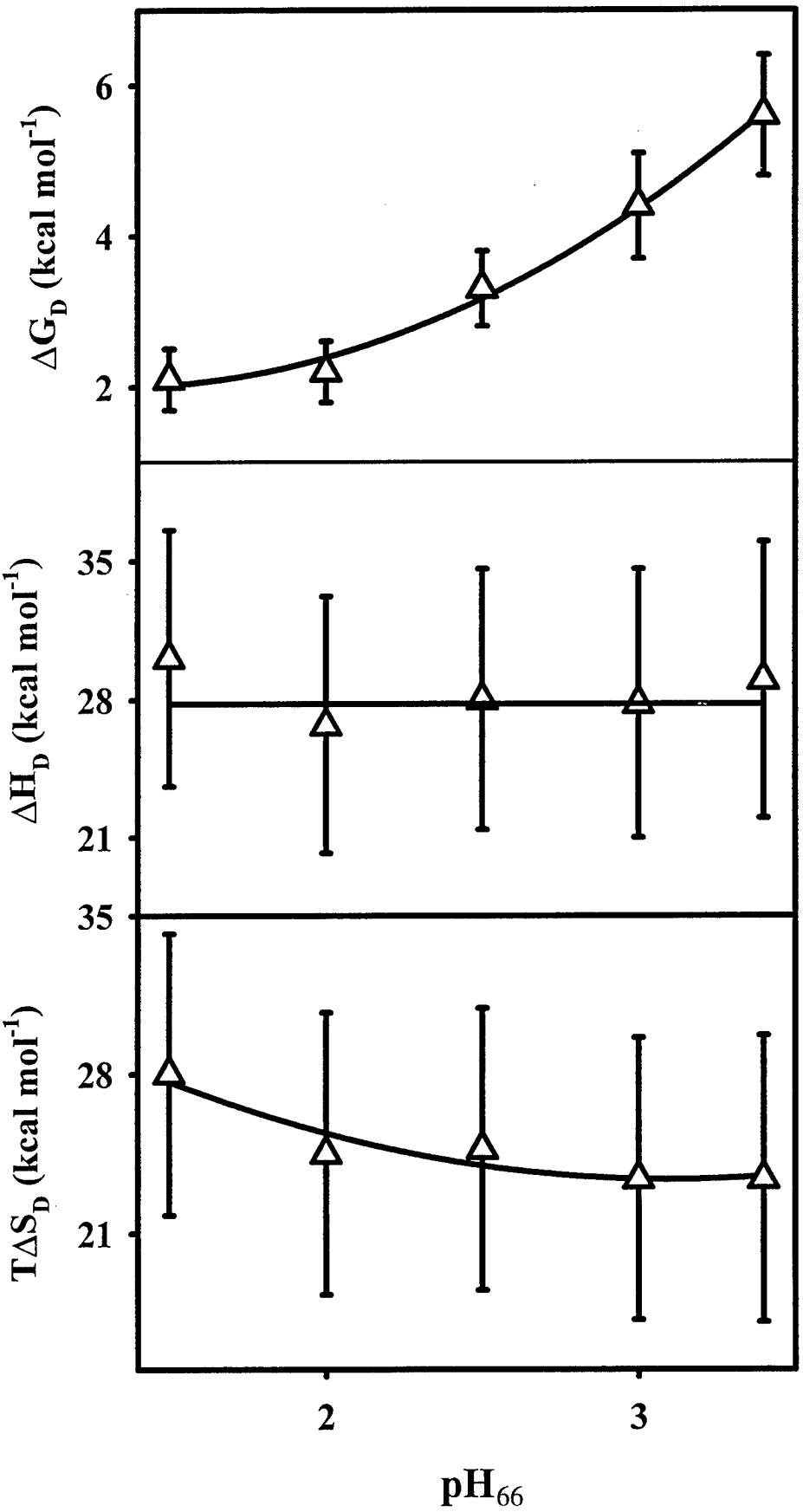

72
68
64
T_m (°C)
1.0
0.8
0.6
β
0
100
200
300
400
500
600
[protein] (μM)

Waldner et al., Figure 4




80
75
70
65
60
55
50
45
40
ΔH_D (kcal mol^{-1})
45
50
55
60
65
70
75
T_m (^{o}C)


A
B
T_m (K)
345
340
335
330
325
320
315
delta ν
2.8
2.4
2.0
1.6
1.2
0.8
0.4
0.0
2
3
pH


6
4
2
ΔG_D (kcal mol^{-1})
35
28
21
ΔH_D (kcal mol^{-1})
35
28
21
$T\Delta S_D$ (kcal mol^{-1})
2
3
pH_{66}