Core-Directed Protein Design. I. An Experimental Method for Selecting Stable Proteins from Combinatorial Libraries†

Michael D. Finucane, Mihriban Tuna, Jennifer H. Lees, and Derek N. Woolfson*

Centre for Biomolecular Design and Drug Development, School of Biological Sciences, University of Sussex, Falmer, Brighton BN1 9QG, U.K.

Received April 1, 1999; Revised Manuscript Received June 21, 1999

ABSTRACT: The design of proteins represents a significant challenge to modern-day structural biology. A major obstacle here is the specification of well-packed hydrophobic cores to drive the folding and stabilization of the target. Computational approaches have been used to alleviate this by testing alternate sequences prior to the production and characterization of a few proteins. Here we present the experimental counterpart of this approach. We selected stable variants from a library of ubiquitin hydrophobic-core mutants as follows. Hexahistidine-tagged proteins were displayed on the surface of phage. These protein-phage were immobilized onto Ni-coated surfaces. The bound fusion-phage were treated with protease to remove unstable or poorly folded proteins. Stable phage fusions were eluted and infected into *Escherichia coli*, which allowed amplification for further selection, sequencing, or protein expression. Two Ni-derivatized supports were tested: Ni−NTA chips for surface plasmon resonance (SPR) and Ni−NTA agarose beads. SPR had an advantage in that the selection process could be monitored directly. This allowed individual clones and experimental conditions to be tested rapidly prior to preparative panning of the library, which was carried out using Ni−NTA agarose beads. We demonstrate the method by selecting stable core mutants of ubiquitin, the characterization of which is described in the following paper [Finucane, M. D., and Woolfson, D. N. (1999) *Biochemistry 38*, XXXXX−XXXXX]. As our method selects only on the basis of structure and stability, it will be of use in improving the stabilities and structural specificities of proteins of de novo design, and in establishing rules that link sequence and structure.

In de novo protein design attempts are made to construct amino acid sequences that adopt prescribed folds. There are two principal reasons for conducting such studies. First, by designing proteins we test hypotheses of how sequence relates to structure. As sequence data are accumulating much faster than structural information, it is becoming even more critical to determine rules that aid protein structure prediction and modeling. Second, successful protein designs would present routes to new proteins with ranges of structures and functions that extend beyond those found in nature.

In recent years, several successful designs have been reported for a number of simple protein architectures (*1*). More usually however, designed proteins do not adopt a single, well-defined structure, but exhibit characteristics of molten-globule intermediates (*2*−*4*) such as the binding of hydrophobic dyes, broad thermal unfolding curves rather than two-state transitions, and poor protection of amide protons with respect to solvent exchange. These studies have led to the concept of negative design (*5*, *6*), which is the idea that while rules should be used in designing a target structure, features must also be added to guard against alternative, energetically similar structures. The production of molten-globule states is the extreme of this in which the design produces an ensemble of related structures with similar folding free energies. A probable cause of this particular type of structural heterogeneity is poor specificity in the designed hydrophobic core (*7*−*9*). Thus, a potential solution to this aspect of the protein-design problem is to engineer or to select proteins with complementary core packing.

Recently, computational methods have proven to be effective in designing proteins with enhanced specificity and stability (*10*, *11*). These methods use algorithms to search combinations of side chains and identify those that pack together most efficiently within the context of a given backbone conformation. However, natural proteins can accommodate alternative packing arrangements by slight rearrangements of the backbone (*12*). Thus, unless the protein backbone can be allowed to readjust with the side chain alterations, equally attractive folding alternatives may be missed. Indeed, it is probable that alternative combinations are potentially as stable as the optimal packs based on the targeted backbone configuration. For example, Lim and Sauer recover mutants of λ repressor with enhanced thermostability (e.g., L36 L40 L47), but decreased DNA binding activity in vivo, which is attributed to the likely adoption of a slightly altered structure (*13*, *14*). Therefore, if design algorithms do not allow the backbone configurations to alter, they may not yield the most stable proteins (*15*, *16*).

These points have implications for another method for engineering proteins, namely, the selection of mutants from libraries based on the activity or binding ability of the wild-

† This work was supported by the BBSRC of the U.K. (Grant B04756). M.D.F. thanks the Royal Society for travel funds to present part of this work at the IUBMB in San Francisco in 1997.

* Corresponding author. Telephone: +44 1273 678214. Fax: +44 1273 678433. E-mail: dek@biols.sussex.ac.uk.

10.1021/bi990765n CCC: $18.00
Published on Web 08/17/1999

type protein (*13*, *17*−*20*). If altered packing arrangements cause minor structural rearrangements, as in λ repressor, this may decrease or abolish activity even though the mutations might result in an increase in stability. As a result, stabilizing mutations may not be recovered under functional selection. Another possibility is that unstable, poorly folded proteins may well exhibit activity by on-site-assembly processes (*21*). Thus, methods that uncouple function and stability would have considerable value in protein design and in examining the links between sequence and structure. Here we present such an experimental protocol and apply it to the core-directed redesign of ubiquitin.

The method is the counterpart of the aforementioned computational methods as stable variants are selected from combinatorial libraries of core mutants. Although sequence space cannot be covered to the same extent as with the computational methods, many more protein variants are tested experimentally by our method, and each is allowed to find its (side chain and backbone) low-energy conformation, which, at the moment, is not possible in silico. Our approach differs from foregoing experimental selection methods in that it is structure-stability based, and is not limited by the requirement of preserving some wild-type activity or binding property. The method utilizes phage display and is based on the concept that stably folded proteins resist proteolysis more than unfolded or partly folded structures (*22*). The procedure involves (1) the generation of multiple core mutants of the target protein, (2) display of this library on the surface of filamentous phage, and (3) selection for the stably folded mutants by challenges with protease (*23*).

In our method, gene fusions are prepared to direct the display of hexahistidine-tagged proteins on the surface of filamentous phage (Figure 1). The tag allows protein-phage to be tethered to a variety of Ni-derivatized supports, and the tag-protein-phage arrangement facilitates protease selection as follows; protease treatment cleaves any unstable, poorly folded linker proteins, which severs phage carrying such mutants from the support allowing them to be removed from the system. An advantage of this strategy is that the Ni support can be a Ni−NTA chip used for surface plasmon resonance (SPR)[1] in a BIACORE instrument. This allows the processes of fusion-phage binding and protease selection to be monitored directly and rapidly. However, because binding capacity is limited by the available surface of the chip, we developed an alternative procedure for panning large protein-phage libraries. This utilizes Ni agarose beads where capacity is limited only by the bed volume used. Regardless of the support used, we found that the protocol was able to discriminate and select stable, folded proteins from mixtures of proteins and peptides. Once selected, stable variants can be expressed free from the phage fusion and purified quickly using the hexahistidine tag. We illustrate the utility of our method by selecting, preparing, and characterizing stable variants of ubiquitin that have multiply mutated hydrophobic cores (*39*).

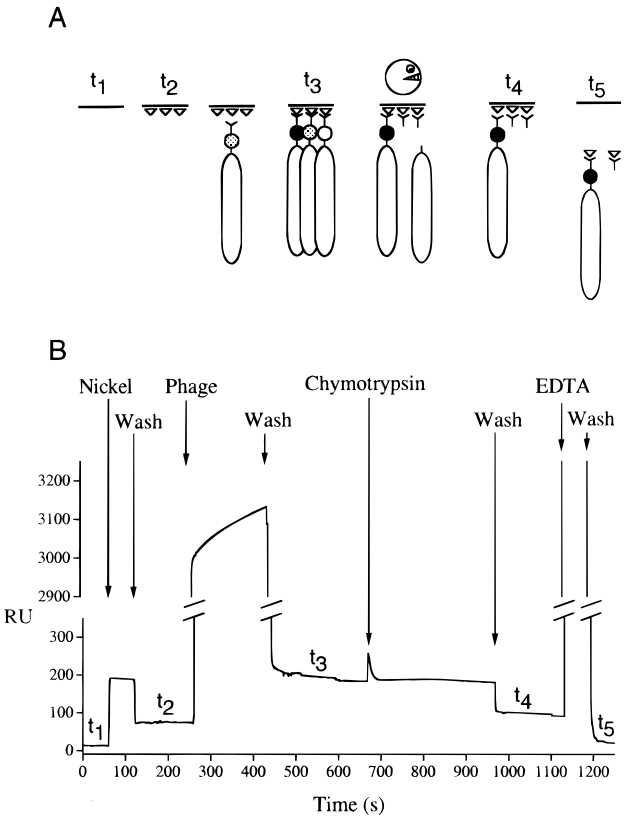

FIGURE 1: Principle of the protease selection procedure. (A) Schematic depiction of the method. Phage-displaying proteins with different stabilities (represented by circles with different degrees of shading) are tagged with hexahistidine tags (Y-shaped tags). These tags allow the phage to bind to a Ni-coated surface (▽). Subsequent treatment with protease cleaves the less stable proteins (shaded and white circles) and releases their phage partners from the surface. Remaining phage displaying more stable proteins (●) are then eluted and retrieved. Each of these steps is labeled with a time (t_1−t_5), which corresponds to an event in the surface plasmon resonance experiment whose results are shown in panel B. (B) Sensorgram trace of a typical SPR experiment carried out in a BIACORE instrument. At time t_1, the resonance intensity [measured in resonance units (RU)] is at a minimum, and the intensity can be normalized to zero. As buffer containing additional components such as Ni ions, phage, or enzyme is passed over the chip, the resonance intensity changes, because of associated changes in the refractive index. Ni is added first to derivatize the chip. When washing with normal running buffer (HPS) is resumed (t_2), the resonance intensity drops, but to a new equilibrium value that is higher than the starting value (t_1), indicating that material (Ni) has been bound. Phage are then added, which bind through their displayed hexahistidine tags. Resumption of the HPS wash sees the baseline level of resonance intensity increase a second time, indicating that phage have bound. Addition of the chymotrypsin-containing solution is followed by a decline in RU as phage are stripped from the surface by proteolysis. After resumption of the HPS wash, the new equilibrium value t_4 is lower than t_3, indicating a loss of phage from the surface. Finally, an EDTA solution is applied to strip all Ni from the surface and, thus, any remaining phage. This restores the resonance intensity (t_5) to approximately the initial value (t_1).

Ubiquitin has a well-defined and highly stable structure (*24*, *25*). It resists proteolysis and provided a positive control for our experiments. To contrast, we prepared two alternative polypeptides designed to be less stable than the wild-type protein (WT-UBQ). The first was a multiple-core mutant of ubiquitin (AL_7) in which core positions 3, 5, 13, 15, 17, 26, and 30 were replaced by leucine, and methionine-1 was replaced by alanine. This mutant was expressed only at very

[1] Abbreviations: AL_7, M1A/I3L/V5L/I13L/V17L/V26L/I30L multiple mutant of mammalian ubiquitin; FLEXI, peptide GSGSG-SYGSGSGG; NTA, nitrilotriacetic acid; SPR, surface plasmon resonance; SDS−PAGE, denaturing polyacrylamide gel electrophoresis; WT, wild-type; UBQ, ubiquitin.

low levels compared with WT-UBQ and was also totally lost in the protease selection studies (*39*). These properties provide good evidence that the AL_7 mutant was destabilized, although no direct stability measurement was possible. The second destabilized, negative control was a Gly-Ser-based peptide (FLEXI) that harbored a single chymotrypsin cleavage site. For a final, thorough test of core-directed design, and in particular to examine if we were able to select protease-resistant clones, we prepared a library of core mutants in which the eight positions highlighted in AL_7 were replaced by combinations of the five hydrophobic residues, Phe, Ile, Leu, Met, and Val.

During the preparation of this and the following paper, two independent groups presented alternative methods for selecting stable proteins (*26*, *27*), which also combine phage display with proteolysis. The methods are similar to each other, but distinct from our procedure, in that target proteins are placed between the infectious (N1) and envelope-anchoring (N2) domains of the minor coat protein (pIII). Thus, proteolysis of the insert severs the N1 domain from phage and renders them noninfective. The two groups demonstrate their method in different model systems. Kristensen and Winter show that they can select for one barnase mutant in the presence of another more destabilized variant (*27*), while Sieber et al. (*26*) apply their method to select stabilized variants of RNAse T1 that are mutated at three noninteracting sites at the surface of the protein.

MATERIALS AND METHODS

Materials. Restriction enzymes, T7 DNA polymerase, T4 DNA ligase, and T4 polynucleotide kinase were obtained from New England Biolabs, Boehringer Mannheim, or Gibco-BRL. α-Chymotrypsin (EC 3.4.21.1, type II from bovine pancreas) and buffering agents were obtained from Sigma Chemical Co., and trypsin (from bovine pancreas) was obtained from Aldrich Chemical Co. Ni−NTA agarose beads were purchased from Quiagen, and used according to the manufacturer's directions. Oligonucleotides were synthesized on an ABI 381A machine and purified using COP cartridges (Cruachem) according to the manufacturer's directions.

Preparation of the Ubiquitin-Encoding pIII Fusion Construct. The DNA sequence of ubiquitin was amplified from a rat DNA library using the PCR and standard protocols. The primers (5′-GTACTGCAGGGAATGCAGATCTTCGTGAAGACC and 5′-AATCTCGAGACCGCCGCCCCTCAGGCGGAGGAC) included 5′ *Pst*I and 3′ *Xho*I restriction sites and terminal glycine codons. The PCR product was gel-purified and ligated into the pGEM-T vector (Promega) prior to subcloning into the in-house vector, pCANTABB, using the *Pst*I and *Xho*I sites. pCANTABB was derived from the phage-display vector pCANTAB5 by adding the coding sequence for a hexahisitidine tag in-frame with, and 5′ to, the gIII sequence. The final pCANTABB-WT-UBQ construct directed the expression of the protein fusion: gIII periplasmic leader sequence−hexahistidine tag−LQG−ubiquitin−GLDQQ−pIII. The plasmid had two suppressible amber (TAG) codons between the ubiquitin gene and pIII, which allowed expression of hexahistidine-tagged proteins without the pIII fusion partner in a nonsuppressor strain of *Escherichia coli* (HB2151). A potential shortcoming of such a system is that suppressor strains (XL1-blue) may not be 100% efficient in suppressing the amber codons and reading through to the phage coat protein. This may reduce the effective size of the protein library expressed on phage. Thus, it is important to construct and confirm that the starting DNA libraries have a high level of redundancy.

Construction of the AL_7, $HisT^-$, and FLEXI-Fusion Constructs. The pCANTABB-AL_7 construct was made by Kunkel mutagenesis (*28*) according to the manufacturer's protocols (Kunkel mutagenesis kit, Bio-Rad). Three different oligonucleotides were simultaneously annealed to uracil-containing single-stranded DNA derived from pCANTABB-WT-UBQ: 5′-G GGT CTT CAA GAA GAG CTG CGC TCC CTG CAG G-3′, 5′-T GGG CTC CAA CTC CAG GGT GAG GGT CTT GCC-3′, and 5′-T ATC CTG GAG CTT GGC CTT CAA GTT CTC GAT-3′. The $HisT^-$ construct was made from pCANTABB-WT-UBQ by removing the coding region for the hexahistidine tag using Kunkel mutagenesis and the primer 5′-CTG CGC TCC CTG CAG GGC CAT GGC CGG CTG-3′. pCANTABB-FLEXI was made from pCANTABB-WT-UBQ using sticky feet-directed mutagenesis (*29*) and the primer 5′-GAG ACC GCC GCC ACT ACC ACT ACC ATA ACT ACC ACT ACC ACT ACC CTG CAG GTG ATG-3′. In addition, the ampicillin resistance encoded by pCANTABB-WT was changed to chloramphenicol resistance in pCANTABB-FLEXI by inserting a gene for acetyltransferase into that for β-lactamase.

Characterization of the Plasmids and Fusion Proteins. After construction, all plasmids were confirmed by sequencing from the coding regions for the histidine tag through to gIII, either manually using Sequenase (Amersham) or on an ABI automated sequencer running the dye-terminator method. Phage display of histidine-tagged ubiquitin was confirmed by detection of a component that reacted with an anti-ubiquitin antibody (Sigma Chemical Co.), a monoclonal anti-polyhistidine antibody (Sigma Chemical Co.), and an anti-pIII antibody (Mo Bi Tech GmbH) using SDS−PAGE and Western blotting. In addition, DNA transformed into XL1-blue cells directed the expression of a protein in the periplasmic fraction that also reacted with all three antibodies.

Preparation and Counting of Fusion Phage. Phage were prepared by (1) transforming XL1-blue *E. coli* with the appropriate plasmid followed by (2) infecting a log phase culture of the transformed cells with VCSM13 helper phage (∼10-fold ratios of phage to *E. coli* were used). After 1 h, kanamycin was added to maintain the plasmids for overnight growth. The cells were spun down and the phage precipitated from the supernatant with 20% (w/v) polyethylene glycol/2.5 M NaCl before being resuspended in phosphate buffer. Before use, the suspension was centrifuged briefly to remove insoluble debris. Phage were counted by infecting an aliquot into log-phase XL1-blue *E. coli* and plating onto agar containing the appropriate species (chloramphenicol or ampicillin) to select bacteria infected with the appropriate phagemid.

Surface Plasmon Resonance (SPR) Methods. Experiments were carried out on BIACORE 2000 and 3000 instruments, using standard procedures and buffers, and NTA-derivatized chips. Typical SPR experiments had the following steps (Figure 1): (1) NTA chip surfaces were charged with nickel ions by passing 20 μL of 500 μM $NiSO_4$ in HPS running buffer [10 mM HEPES, 0.15 M NaCl, 0.005% Surfactant

P20 (Biacore AB), and 50 μM EDTA (pH 7.4)] over the surface for 60 s; (2) phage were bound via their hexahistidine tags to the nickel surface; (3) a further brief wash with HPS buffer was applied prior to washing with 10 μM chymotrypsin in HPS buffer which released some phage from the surface; (4) a short period of washing with HPS buffer was used to establish the new baseline; and (5) the surface was regenerated by washing with HPS buffer incorporating 0.35 M EDTA at pH 8.3, which removed the nickel from the NTA surface. The extent of proteolysis was determined by calculating the ratio of the drop in resonance intensity (RU) during the hydrolysis period (steps 3 and 4 above and the period from t_4 to t_3 in Figure 1B) to the total amount of phage bound immediately before cleavage (step 2 above and the period from t_3 to t_2 in Figure 1B) also measured in RU. All SPR experiments within the BIACORE instruments were carried out at 25 °C, while biopanning of the library performed on the BIACORE chips outside the machines was carried out at 37 °C.

Panning Using Ni Agarose as a Support Matrix. Ni−NTA agarose beads (Qiagen) were suspended in sonication buffer [50 mM Na_2HPO_4 and 300 mM NaCl (pH 7.8)], and all pre-elution washes were carried out using sonication buffer with an additional 20 mM imidazole to minimize nonspecific binding. Elution of the bound phage was accomplished using sonication buffer with 250 mM imidazole. See ref *39* for further details of these experiments. The proteolysis reactions in these biopanning experiments were performed at 37 °C.

RESULTS

Analysis of Specific, Monoclonal Phage-Display Constructs Using Surface Plasmon Resonance

In phage-display pools of protein-phage are enriched for certain mutants through selection by applying specific challenges. If conditions are not sufficiently stringent, selection will not be achieved. If, on the other hand, selection conditions are too stringent, too many clones will be lost during early rounds of panning (*30*). It is useful, therefore, to be able to monitor this process directly to optimize selection conditions. Surface plasmon resonance (SPR) is a technique that allows direct monitoring of the interaction of biomolecules with ligands captured on the surfaces of derivatized chips. As such, it is ideally suited to monitoring phage binding and release in real time (*31*). The capture method that we chose was the interaction of hexahistidine affinity tags with chelated nickel ions, which is compatible with SPR and can be transferred to a wide variety of support systems that are commercially available. Furthermore, once selected, proteins can be purified easily using the hexahistidine tags.

Specificity of Phage Binding. To ensure that the histidine tags were fully exposed on the surface of the phage-displayed protein, and to test that phage were binding specifically to the chip surface through the tags, phage displaying hexahistidine-tagged wild-type ubiquitin (WT-UBQ-phage) were passed across two consecutive flow cells of a Ni−NTA chip in the BIACORE instrument (see Figure 1 for an example of the method). The second cell was precharged with nickel ions, but the first was not. After washes with phage followed by standard running buffer, a significant increase [typically ≈200−300 resonance units (RU)] in the SPR signal above the starting value was observed for the second, Ni-charged cell, but not for the underivatized cell (data not shown). This indicated that phage were only binding to the sensor chip surface through direct interactions with Ni. Additional experiments demonstrated that WT-UBQ-phage without histidine tags did not bind to the surface (data not shown).

Specificity of Protease Cleavage. Filamentous phage generally resist proteases (*32*). However, the minor coat protein, pIII, to which the displayed proteins were fused, can be susceptible to certain proteases, including chymotrypsin (*33*). If proteolysis of pIII were significant in our experiments, the proposed method would either fail altogether or not yield stabilized proteins reliably. Similar problems would be met if target residues in unfolded variants of the displayed protein were not sufficiently exposed to allow attack by protease. To exclude these possibilities, we determined (1) whether infection-competent phage were released from the surface on exposure to protease and (2) if this release depended on the sequence of the target protein linking the histidine tag and pIII.

As illustrated in Figure 1B, phage were released from the surface of a Ni−NTA chip considerably more rapidly with protease present than with washes of running buffer alone. The released phage were able to infect *E. coli*, indicating that at least some of the pIII proteins survived proteolysis. Similarly, phage that survived proteolysis and remained bound to the surface of the Ni−NTA chip could be eluted, and these too were found to re-infect *E. coli*. To demonstrate directly that the release was caused by proteolysis, and also to calibrate how much protease was required for efficient proteolysis, we determined how phage release depended on the chymotrypsin concentration. WT-UBQ-phage was bound to four lanes of a Ni−NTA chip in the BIACORE instrument, and each lane was washed with a different chymotrypsin concentration. As judged by the drop in resonance intensity over a 100 s incubation, the amount of phage released from the surface depended linearly on the logarithm of the chymotrypsin concentration between 0.1 and 10 μM (Figure 2A).

To verify that the protease cleavage was linker-specific, we compared the proteolysis of two phage constructs using SPR in a BIACORE instrument (Figure 2B). The protein-phage used were the aforementioned WT-UBQ-phage and a construct in which ubiquitin was replaced completely by a flexible, Gly-Ser-based linker (FLEXI) that contained a single tyrosine cleavage site. These were bound to two adjacent lanes of a Ni−NTA chip in the BIACORE instrument and washed with chymotrypsin. A dramatic difference in the rates of hydrolysis of the two linkers was clear upon addition of 10 μM chymotrypsin (Figure 2B); only a fraction (≤30%) of the WT-UBQ-phage was lost from the surface of the chip over a 300 s incubation, whereas virtually all of the FLEXI linker was hydrolyzed within 50 s. We repeated this experiment and included, on a separate lane on the chip surface, another protein-phage variant that had a destabilized ubiquitin mutant (AL_7-UBQ) as the linker between the hexahistidine tag and pIII. In this case, the 50 s pulse of chymotrypsin removed 20% of the WT-UBQ-phage, 80% of the AL_7-UBQ, and 90% of the phage with the FLEXI linker. Thus, protease treatment followed by SPR in the BIACORE instrument was reproducible and quantifiable and

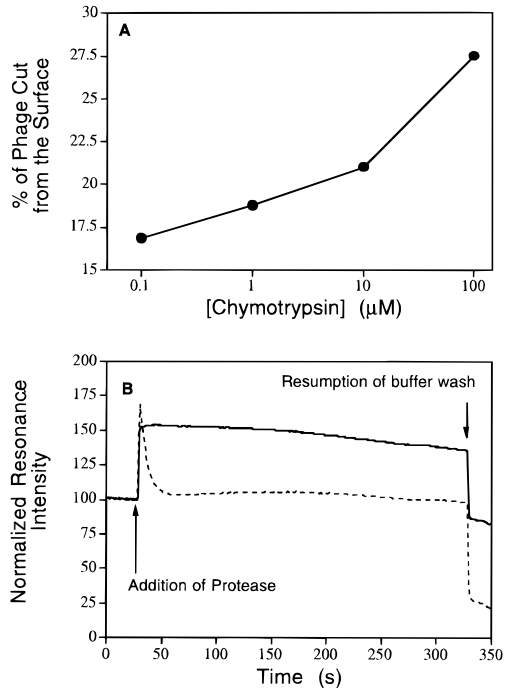

FIGURE 2: Specificity of protease cleavage. (A) The amount of phage cleaved from the Ni surface in SPR as a function of protease concentration of protease. The % cleaved was calculated as $(t_3 - t_4)/(t_3 - t_2) \times 100$ from experiments described in the legend of Figure 1. (B) Sequence dependence of the cleavage profile as judged by the change in resonance intensity. The solid trace shows the cleavage of phage displaying WT-UBQ, whereas the dashed line is for FLEXI-phage. The data have been normalized such that 100 on the *Y*-axis is equivalent to 100% of the phage bound immediately before proteolysis. After exposure to chymotrypsin for 300 s, 83% of the WT-UBQ-phage remained bound to the BIACORE chip surface, and in contrast, only 23% of the FLEXI-phage remained.

clearly discriminated folded proteins (WT-UBQ) from destabilized proteins (AL_7) and fully unfolded peptides (FLEXI).

Measurement and Comparison of Rates of Proteolysis. In SPR, large changes in resonance intensity can accompany relatively small changes in the composition of the wash buffer. For instance, in our studies, adding 10 μM chymotrypsin usually resulted in a rise of about 100 RU, which was comparable to the intensity changes caused by the binding and loss of phage. Potentially, this hampers following the kinetics of the cleavage step. To test this, we added a chymotrypsin solution in brief pulses followed by washes with HPS running buffer. In this way, the resonance intensity before and after the chymotrypsin pulse could be compared to measure the amount of phage that was cleaved. With pulses kept sufficiently short to allow some phage to remain bound to the chip surface, this process could be repeated and pseudokinetics determined (Figure 3). A slight disadvantage was that a small amount of material appeared to be lost from the surface during the washing steps which was probably due to a low steady-state dissociation of the histidine tag from the Ni, and of Ni from the NTA chip surface; it is possible that this could be eliminated by using a longer histidine tag, although we have not tested this possibility.

After a total of 93 s of contact with chymotrypsin, 87% of the WT-UBQ-phage and 25% of the FLEXI-phage remained bound to the surface (Figure 3). This compared

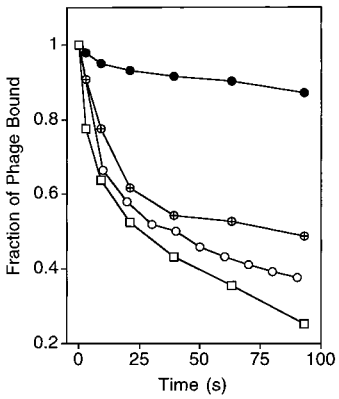

FIGURE 3: Pulse proteolysis of protein-phage bound to a BIACORE sensor chip. In this experiment, changes in SPR signal were measured after pulses with 10 μM chymotrypsin. After each pulse, washing was resumed to re-establish equilibrium and determine the amount of phage remaining. In the figure, cumulative loss of phage from the surface is plotted against the cumulative exposure time to protease: (●) WT-UBQ-phage, (⊕) LIB-UBQ-phage, (○) AL_7-UBQ-phage, and (□) FLEXI-phage.

with 83% (WT-UBQ-phage) and 23% (FLEXI-phage) after exposure for 300 s to the same concentration of chymotrypsin (10 μM) in a continuous wash experiment (Figure 2). Thus, the relative extent of WT-UBQ-phage and FLEXI-phage proteolysis was similar in both experiments, and confirmed that simple time courses with one pulse of protease gave an accurate representation of the relative stability of different clones.

Nonetheless, to be sure about this point, we compared WT-UBQ, FLEXI, and AL_7-phage with a fourth sample, LIB-UBQ-phage, which displayed a library of hydrophobic core mutants of ubiquitin. In the pulsed experiment, none of the protein-phage preparations exhibited simple, single-exponential kinetics. This was probably due to a combination of things, including the aforementioned steady-state dissociation of the histidine tag from the Ni−NTA surface, residual proteolysis at additional sites in the polypeptide, and the fact that LIB-UBQ-phage contained many clones, presumably with different stabilities and susceptibilities to protease. However, the traces (Figure 3) clearly showed that proteolysis of the WT-UBQ and LIB-UBQ linkers tended toward similar rates, which were slower than that for FLEXI-phage. This encouraging result suggested that variants whose average stability was similar to that of WT-UBQ were being selected from the library.

Loading Capacity of Ni−NTA Chips Compared with Library Size. In phage display, it is imperative to ensure adequate representation of each clone, or else underrepresented but otherwise fit clones may be lost. Although this has no bearing on the use of SPR in comparing the stability of individual clones as described above, it may affect panning experiments with large libraries. To determine the size of the library that could be panned using this technique, we washed WT-UBQ-phage over the surface of one lane of an Ni−NTA SPR chip until no further increase in resonance intensity was observed. The resonance intensity that was achieved was typically about 300 RU, although it did vary in the range of 250−350 RU over several experiments. As 1000 RU corresponds to ≈1 ng/mm^2 of bound material (BIACORE AB, personal communication) and each lane has a surface area of 1 mm^2, 300 RU represents ≈300 pg of

phage. The M_r of phage is $\approx 1.5 \times 10^7$, and therefore, each lane binds $\approx 12 \times 10^6$ individual clones. In addition, we determined the coverage of a single lane experimentally to be 1.5×10^6 CFU by recovering and counting several samples of phage. The calculation and this experiment placed a clear limit on the size of the library that could be panned using SPR in BIACORE, which is significantly lower than library sizes normally accessible through phage display ($\approx 10^9$ clones). In our case, the library size was quite modest; eight sites were mutated to combinations of five amino acids, giving a potential library size of 390 625 mutants, and we would thus have expected to have approximately a 10-fold redundancy in panning this library by SPR.

In Vitro Selection from a Library of Ubiquitin Mutants Displayed on Phage (LIB-UBQ-Phage)

As the area used in each lane of the chip (1 mm^2) is a fraction of the total surface area (49 mm^2), once we had determined the optimal cleavage conditions, we removed the chip from the instrument and panned the ubiquitin library on the entire chip surface. This removed the advantage of being able to monitor the binding, cleavage, and elution steps, but enhanced the size of the library that could be panned by $\approx$50-fold, which increased our confidence of binding several copies of each mutant. Because the support surface is unaffected by removal from the system, conditions that had been optimized within the machine could be transferred seamlessly.

We bound the library of ubiquitin core mutants to the entire NTA chip. This was then treated with protease, and the cleaved phage were washed off and put to one side prior to elution of bound phage. Phage released in these two ways were counted by infection into *E. coli* and plating out. We enriched separately for both stable and unstable proteins as follows. After the first round of panning, phage were divided into two fractions, those cleaved from the chip (which we labeled "unstable") and those that resisted proteolysis ("stable"). Both pools were re-infected into *E. coli* and amplified. In subsequent rounds, both pools were panned individually, and again the fraction that remained uncleaved from the stable pool was retained, as was the fraction cleaved in the unstable pool. In this way, the protease resistance of the stable pool was continually increased, and that of the unstable pool continually decreased. Consistent with this, over three rounds of panning, the ratio of cut to bound phage for the two pools diverged (Figure 4). Further rounds of selection did not increase this divergence significantly. This demonstrated that the method could select both stable and unstable populations of clones from libraries based on protease resistance.

Several of the phage were isolated from the final round of panning for stable "bound" clones. These were amplified individually, confirmed as encoding full-length ubiquitin mutants using colony PCR, and tested for protease resistance using SPR back in the BIACORE instrument. Although marginally less resistant than WT-UBQ-phage ($25 \pm 8\%$ of which were cleaved from the chip over the time course of the experiment), the selected clones were much more stable ($27 \pm 10\%$ cleavage) than the parent AL_7-UBQ mutant ($70 \pm 5\%$ cleavage), FLEXI ($73 \pm 9\%$ cleavage), and the initial library when taken as a whole ($53 \pm 16\%$ cleavage). Thus, stable ubiquitin mutants could be selected from a pool of multiply mutated variants on the basis of protease resistance.

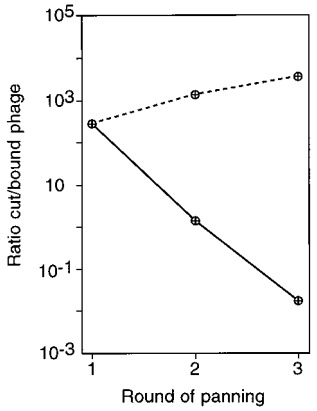

FIGURE 4: Panning of a library of ubiquitin-phage on the surface of a BIACORE sensor chip, but outside the BIACORE instrument. The ratio of phage cut from the chip to that which remained bound is plotted over three rounds of panning. The selection was carried out to enrich both for stabilized proteins (solid line) and for destabilized proteins (dashed line).

Panning Large Phage-Displayed Libraries Using Ni Agarose Beads. We estimated the coverage over an entire Ni−NTA chip to be $\approx 10^8$ phage. This is not adequate for selection from the larger libraries possible in phage display. To combat this, we transferred the technology to a system with a larger binding capacity. As an example, we tested Ni agarose beads as a support matrix. In this case, only the amount of beads used limits the capacity in panning libraries of protein-phage.

Specificity of Hexahistidine-Tagged Protein-Phage Binding to Ni Agarose Beads. Again we tested for specificity of phage binding to the support to be sure that it depended on the Ni−hexahistidine interaction. This was especially important with agarose beads because of the possibility of nonspecific binding, either via surface interactions or through the phage becoming partially trapped within the agarose. To eliminate these possibilities, we tested for discrimination between two WT-UBQ-phage, one of which carried an N-terminal hexahistidine tag and another that did not (termed here WT-UBQ and $HisT^-$-UBQ, respectively). These phage also carried different antibiotic resistances to allow them to be distinguished in the following experiment. Equal amounts of the two phage were mixed and bound to Ni agarose beads. These were washed repeatedly with noneluting buffer and the supernatants recovered to determine the amounts of phage that were not binding to the beads. Consistent with unbound phage being removed, successive washes released fewer and fewer of both types of phage until a slower steady-state dissociation of phage from the matrix was achieved. This behavior is shown for WT-UBQ-phage in Figure 5A. When the buffer was changed to an elution buffer that contained 250 mM imidazole, there was an immediate 20-fold rise in the number of WT-UBQ-phage found in the supernatant (data not shown). By contrast, there was only a 3-fold increase in the concentration of the $HisT^-$-UBQ-phage. This was consistent with phage binding predominantly through the hexahistidine tag as required.

Protease Selection of Bead-Bound Protein-Phage. As with the Ni−NTA chip, it was essential to establish that (1) protease released phage from Ni agarose beads and (2)

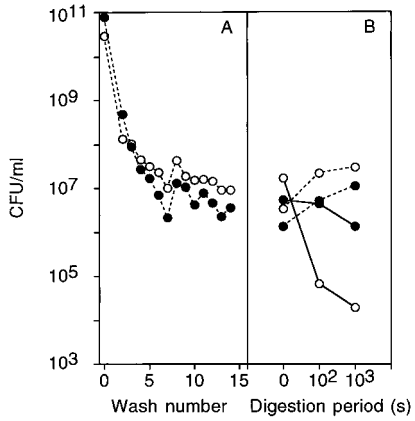

FIGURE 5: Protease cleavage of phage bound to Ni agarose beads. (A) Fall off of nonspecific and weak binding phage during washes of beads after initial binding of phage. The number of phage released into the supernatant [measured as colony-forming units per milliliter (CFU/mL)] was determined by infection into *E. coli*, followed by plating and counting. (B) Proteolysis of bead-bound phage. After reaching the steady-state release (A), beads were resuspended in buffer containing 10 μM chymotrypsin. Prior to this, some phage were counted to determine a zero time point. Aliquots were removed from the protease/bead mixture at intervals logged on the *X*-axis, and phage in the supernatant and remaining bound to the beads were counted. For WT-UBQ-phage and FLEXI-phage, the number of phage in the supernatant increased with time of exposure to chymotrypsin. Similarly, the amount of bead-bound phage declined with exposure time. The release of phage from the beads into the supernatant was more rapid for FLEXI-phage: (●) WT-UBQ-phage, (○) FLEXI-UBQ, (—) bead-bound phage, and (- - -) phage measured in the supernatant.

release depended on the targeted linker region between the tag and pIII. To facilitate these tests, the FLEXI-phage construct had a different antibiotic resistance, which allowed it to be panned with and distinguished from WT-UBQ-phage in competition experiments. Phage from the two constructs were combined and bound to Ni agarose beads. These were washed to reach the background release rate; a sample of beads was removed at time zero, and chymotrypsin was added. Aliquots of beads were withdrawn at 100 and 1000 s and treated with the broad-specificity protease inhibitor PMSF. Phage in the supernatants were counted to determine the proportion of each linker cleaved after these times. This revealed that phage release was enhanced enormously upon addition of protease, and that this change was more dramatic for FLEXI-phage than for WT-UBQ-phage. Consistent with this, more protease-resistant WT-UBQ-phage remained bound to the beads during and after the proteolysis reaction (Figures 5B and 6). After treatment with PMSF, the uncleaved phage were eluted from the beads using imidazole. The released phage were used to infect *E. coli* and plated out onto agar plates with different antibiotics. This allowed FLEXI-phage and WT-UBQ-phage to be distinguished and indicated a clear dominant selection of the WT-UBQ-phage throughout the time course of the reaction. This is seen particularly clearly in Figure 6.

Additional experiments with LIB-UBQ-phage showed that either chymotrypsin or trypsin was able to release phage from the beads. However, in the absence of protease, phage remained bound to the beads with no significant loss even after an overnight incubation. In the following paper, we describe how this technology was used to recover stable proteins from the LIB-UBQ-phage pool (*39*).

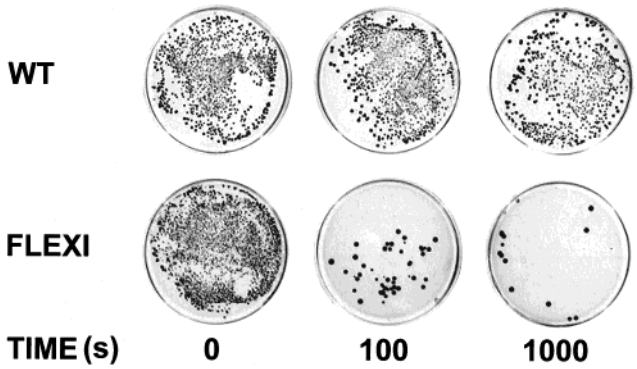

FIGURE 6: Visual comparison of the selection of WT-UBQ-phage vs FLEXI-phage. Aliquots of beads were removed from the proteolysis reaction mixture described in the legend of Figure 5 after 0, 100, and 1000 s, treated with PMSF, washed, and phage eluted with imidazole. The amounts of these phage were gauged by counting colonies after plating out infected *E. coli* onto ampicillin and chloramphenicol-containing agarose. *E. coli* that grew on the ampicillin plate contained the WT-UBQ construct, whereas those which grew on the chloramphenicol plate contained the FLEXI construct.

DISCUSSION

A key area for research in protein folding is the design and generation of new and useful proteins. A major difficulty here is not only in selecting proteins with the prescribed fold but also in choosing those that form stable structures. A lack of stability in engineered proteins has consequences not only for protein design but also for the production of proteins in high yield via heterologous expression; it has been shown that the expression rates and final yields of proteins are directly related to their thermal stabilities (*34*, *35*), and even in the periplasm, where the level of degradation is lower than in the cytoplasm, the extent of proteolysis can be reduced by increasing protein stability (*36*).

We present a new method for engineering stability into natural and designer proteins alike. The method combines the notion that correct packing of the hydrophobic core is crucial for protein stability (*37*) with a practical method for selecting stable proteins. The first step in our method involves choosing residues that form the hydrophobic core, and replacing them with combinations of other hydrophobic residues with different sizes and shapes. The second step involves selecting from this library of alternatives, those that are stable and folded. In this paper, we have described the development and testing of such a method.

Our selection method operates on a principle similar to substrate-phage (*32*). In substrate-phage, protease specificity is determined by expressing peptide variants between an affinity tag and pIII. Treatment of the phage-display libraries with protease results in the cleavage of peptides specific for the protease and loss of the affinity tag. Here we employ a similar approach, but select for full-length proteins that resist proteolysis and, therefore, are expected to be stable and folded (*23*). An alternative approach has recently been demonstrated by Sieber et al. (*26*) and by Kristensen and Winter (*27*), which does not employ an affinity tag in the selection process. Rather, the target sequence is inserted between the infectivity and phage-anchoring domain of the phage pIII, and thus, cleavage of the target leads to a loss in

infectivity of phage. This method requires the loss of all copies of the infectivity domain, so monovalent phage display cannot be used. In addition, all displayed copies of the target on the phage must be cleaved during the selection. Our selection protocol, on the other hand, allows the use of either monovalent or polyvalent phage display, and may be extended to using other coat proteins (e.g., pVIII) and different types of phage.

It is probable that our method offers further advantages over alternative procedures as the target protein is expressed N-terminally to pIII, instead of as an insert within pIII as in selectively infective phage (*26*, *27*). First, the proteins in our system are potentially more exposed to proteolysis as they only have one fusion partner instead of two, and hence the selection may be more efficient; second, after selection, proteins can be expressed free from pIII and ready for characterization or large-scale production simply by changing the bacterial strain that is used (*39*).

The use of an affinity tag is central to our method. It allows us to monitor both binding and release of phage by direct visualization using surface plasmon resonance, and to pan large libraries by switching to the high-capacity Ni agarose bead support. An additional advantage is that the N-terminal histidine tag can be used for purification of proteins once selected. Ni-coated ELISA plates are also commercially available and would give a further option for monitoring enrichment if desired. For our system, we demonstrated that, regardless of the specific Ni-surface employed, (1) binding of the fusion phage occurred directly through the N-terminal hexahistidine tag on the target protein, (2) cleavage of the phage from the supports depended on protease, and (3) protease resistance increased with the stability of the sequence inserted between the tag and phage coat protein.

Using our phage-display method, we were able to distinguish fully folded wild-type ubiquitin from a destabilized mutant of ubiquitin with a homogeneous (perleucine) hydrophobic core, and also from a flexible peptide linker with a single protease cleavage site. We also selected protease-resistant clones from a library of mutants. These particular experiments were conducted at moderate protease concentrations (10−100 μM) and at modest temperatures (37 °C). However, because phage resist a variety of more harsh conditions (*38*) more severe selection conditions are possible. Nevertheless, even when selection was carried out under the aforementioned conditions, proteins were recovered from the library that, like WT-UBQ, had high thermal stability with melting temperatures of >90 °C (*39*), which is presumably much greater than the stability of the starting AL_7 mutant.

In summary, we have developed a novel protocol for designing proteins with increased stability. This has immediate applications in the production of proteins of de novo design, and for improving the expression yields and thermal stabilities of existing proteins. The design of different phage-display constructs facilitated all stages in the process, from the visualization of the panning process to large-scale panning and protein recovery. In the following paper (*39*), we describe an application of the method that demonstrates that we can successfully recover stable proteins starting from a heavily compromised variant of ubiquitin with a homogenized core.

ACKNOWLEDGMENT

We thank Dave Chiswell (CAT) for the gift of the original pCANTAB5 phagemids used to prepare our constructs and Ian Badcoe, Tim Clackson, Jonathan Cox, and Tony Clarke for advice. We are grateful to BIACORE AB for the loan of SPR instruments.

REFERENCES

1. Schneider, J. P., Lombardi, A., and DeGrado, W. F. (1998) *Folding Des. 3*, R29−R40.
2. Betz, S. F., Raleigh, D. P., and Degrado, W. F. (1993) *Curr. Opin. Struct. Biol. 3*, 601−610.
3. Betz, S. F., Bryson, J. W., and Degrado, W. F. (1995) *Curr. Opin. Struct. Biol. 5*, 457−463.
4. Roy, S., Helmer, K. J., and Hecht, M. H. (1997) *Folding Des. 2*, 89−92.
5. Nautiyal, S., Woolfson, D. N., King, D. S., and Alber, T. (1995) *Biochemistry 34*, 11645−11651.
6. Betz, S. F., Liebman, P. A., and DeGrado, W. F. (1997) *Biochemistry 36*, 2450−2458.
7. Degrado, W. F. (1993) *Curr. Opin. Struct. Biol. 3*, 601−610.
8. Munson, M., Balasubramanian, S., Fleming, K. G., Nagi, A. D., O'Brien, R., Sturtevant, J. M., and Regan, L. (1996) *Protein Sci. 5*, 1584−1593.
9. Wu, L. C., and Kim, P. S. (1997) *Proc. Natl. Acad. Sci. U.S.A. 94*, 14314−14319.
10. Kono, H., Nishiyama, M., Tanokura, M., and Doi, J. (1998) *Protein Eng. 11*, 47−52.
11. Malakauskas, S. M., and Mayo, S. L. (1998) *Nat. Struct. Biol. 5*, 470−475.
12. Baldwin, E. P., Hajiseyedjavadi, O., Baase, W. A., and Matthews, B. W. (1993) *Science 262*, 1715−1718.
13. Lim, W. A., and Sauer, R. T. (1989) *Nature 339*, 31−36.
14. Lim, W. A., and Sauer, R. T. (1991) *J. Mol. Biol. 219*, 359−376.
15. Harbury, P. B., Tidor, B., and Kim, P. S. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 8408−8412.
16. Su, A., and Mayo, S. L. (1997) *Protein Sci. 6*, 1701−1707.
17. Milla, M. E., and Sauer, R. T. (1995) *Biochemistry 34*, 3344−3351.
18. Axe, D. D., Foster, N. W., and Fersht, A. R. (1996) *Proc. Natl. Acad. Sci. U.S.A. 93*, 5590−5594.
19. Gu, H. D., Yi, Q. A., Bray, S. T., Riddle, D. S., Shiau, A. K., and Baker, D. (1995) *Protein Sci. 4*, 1108−1117.
20. Gu, H. D., Kim, D., and Baker, D. (1997) *J. Mol. Biol. 274*, 588−596.
21. Pan, B., Deng, Z. W., Liu, D. J., Ghosh, S., and Mullen, G. P. (1997) *Protein Sci. 6*, 1237−1247.
22. Fontana, A., deLaureto, P. P., DeFilippis, V., Scaramella, E., and Zambonin, M. (1997) *Folding Des. 2*, R17−R26.
23. Finucane, M. D., Lees, J. H., and Woolfson, D. N. (1997) *FASEB J. 11*, 90.
24. Vijay-Kumar, S., Bugg, C. E., and Cook, W. J. (1987) *J. Mol. Biol. 194*, 531−544.
25. Woolfson, D. N., Cooper, A., Harding, M. M., Williams, D. H., and Evans, P. A. (1993) *J. Mol. Biol. 229*, 502−511.
26. Sieber, V., Plückthun, A., and Schmid, F. X. (1998) *Nat. Biotechnol. 16*, 955−960.
27. Kristensen, P., and Winter, G. (1998) *Folding Des. 3*, 321−328.
28. Kunkel, T. A. (1985) *Proc. Natl. Acad. Sci. U.S.A. 82*, 488−492.
29. Clackson, T., and Winter, G. (1989) *Nucleic Acids Res. 17*, 10163−10170.
30. Levitan, B. (1998) *J. Mol. Biol. 277*, 893−916.
31. Lasonder, E., Schellekens, G. A., and Welling, G. W. (1994) *Nucleic Acids Res. 22*, 545−546.
32. Matthews, D. J., and Wells, J. A. (1993) *Science 260*, 1113−1117.

33. Grant, R. A., Lin, T. C., Konigsberg, W., and Webster, R. E. (1981) *J. Biol. Chem. 256*, 539−546.
34. Parsell, D. A., and Sauer, R. T. (1989) *J. Biol. Chem. 264*, 7590−7595.
35. Inoue, I., and Rechsteiner, M. (1994) *J. Biol. Chem. 269*, 29241−29246.
36. Kwon, W. S., Da Silva, N. A., and Kellis, J. T., Jr. (1996) *Protein Eng. 9*, 1197−1202.
37. Richards, F. M., and Lim, W. A. (1993) *Q. Rev. Biophys. 26*, 423−498.
38. Smith, G. P., and Scott, J. K. (1993) *Methods Enzymol. 217*, 228−257.
39. Finucane, M. D., and Woolfson, D. N. (1999) *Biochemistry 38*, 11613−11623.

BI990765N