Protein Engineering vol.8 no.1 pp.81–89, 1995

Engineered turns of a recombinant antibody improve its *in vivo* folding

Achim Knappik[1] and Andreas Plückthun[2]

Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland
[1]Present address: Morphosys GmbH, Frankfurter Ring 193a, D-80807 München, Germany

[2]To whom correspondence should be addressed

Using recombinant antibodies functionally expressed by secretion to the periplasm in *Escherichia coli* as a model system, we identified mutations located in turns of the protein which reduce the formation of aggregates during *in vivo* folding or which influence cell stability during expression. Unexpectedly, the two effects are based on different mutations and could be separated, but both mutations act synergistically *in vivo*. Neither mutation increases the thermodynamic stability *in vitro*. However, the *in vivo* folding mutation correlates with the yield of oxidative folding *in vitro*, which is limited by the side reaction of aggregation. The *in vivo* folding data also correlate with the rate and activation entropy of thermally induced aggregation. This analysis shows that it is possible to engineer improved frameworks for semi-synthetic antibody libraries which may be important in maintaining library diversity. Moreover, limitations in recombinant protein expression can be overcome by single amino acid substitutions.

Key words: E.coli/F(ab) and Fv fragments/protein engineering/protein expression/protein folding

Introduction

Folding *in vivo* can be the limiting process in the production of many recombinant proteins. It has been shown that aggregation during *in vivo* and *in vitro* folding is a specific off-pathway process involving folding intermediates (Jaenicke, 1987; Wetzel, 1994). For a few proteins it has been possible to isolate point mutations which induce or suppress aggregation (Tsai *et al.*, 1991; Jappelli *et al.*, 1992; Mitraki and King, 1992; Chrunyk and Wetzel, 1993; Chrunyk *et al.*, 1993; Dale *et al.*, 1994), and the solution of this problem may offer the possibility of improving production of industrially important proteins (Mitraki and King, 1992). We undertook an investigation of the molecular basis of this limitation using recombinant antibody fragments secreted into the periplasmic space of *Escherichia coli* as a model system. This class of structurally well characterized β-barrel proteins offers an almost unlimited database of proteins which have similar sequence and structure but have been found to fold in *E.coli* in rather different proportions.

The efficient expression of antibody fragments in bacteria is clearly of great technological importance. The routinely achievable amounts of recombinantly expressed antibody will be decisive for the viability of this technology for many medical and technological applications. In this paper we show that there are two principal factors which are limiting the expression yields. The first can be directly related to folding problems of the antibody fragments in the periplasmic space of *E.coli*, and the second is a consequence of the stress imposed by the antibody on the bacterial cell secreting it. In its mildest form it leads to periplasmic leakiness and then to lysis with the occasionally observed plasmid loss; in severe cases it leads to plasmid rearrangements and loss of clones from libraries. Since such problems are not universally observed for high-level secretion of proteins in *E.coli*, it follows that they are a consequence of the antibody sequence and structure. Here we describe experiments that have helped us to analyze and solve some of these problems.

While many expression experiments of antibody fragments in bacteria have now been reported (Plückthun, 1994), a comparison of functional yields is made essentially impossible by the use of many different vector systems, strains, experimental conditions and types of antibody fragment. Moreover, results from high cell density experiments are occasionally directly compared with shake-flask experiments; however, this is not useful for drawing conclusions about the components of the expression system. Here we analyze the expression of different antibody fragments under identical conditions and show that the single most important factor is the primary sequence of the antibody variable domains. We then show that the crucial features within the sequence can be elucidated and that it is possible to engineer an antibody of mediocre expression behavior into one of superior properties without interfering with its binding properties.

The basis for our study was the conclusion derived from previous work that it is the periplasmic folding process itself which limits the whole expression process (Skerra and Plückthun, 1991; Knappik *et al.*, 1993). This suggestion was based on the observations that (i) substantial amounts of periplasmic, correctly processed, albeit insoluble, protein were formed and (ii) increased overproduction of total protein was easily achieved but led to only very modest increases in soluble processed protein. Only small amounts of soluble precursor could be detected when a strong promoter was used. It is likely that the precursor would be stable enough to observe it if it were present in large amounts, since it has been seen in certain mutants (Glockshuber *et al.*, 1992). The fact that high-level production leads to the formation of insoluble protein is also observed with many other proteins (Shatzman, 1990), and the high achievable amounts of *total* protein show that the levels, production rates and degradation rates of plasmid, mRNA and protein are not the primary culprits of low yields of functional protein, although they may of course be contributing factors. Furthermore, the observation of processed, insoluble protein suggests that the protein does not have a problem in transport through the membrane, since it has apparently come into contact with the signal peptidase on the periplasmic face of the inner membrane. Taken together, it appears that a side reaction of folding, namely aggregation, limits the yield *in vivo*. In this paper we show that the *in vitro* folding yields under oxidative folding conditions of

the antibody mutants constructed and examined do correlate directly with the soluble expression yields *in vivo*.

In previous work we have also investigated the effect of external catalysis on the antibody fragments folding in the periplasm (Knappik *et al.*, 1993). While the antibody fragments absolutely require the *E.coli* disulfide-forming protein DsbA *in vivo*, its overexpression does not increase the yield for any fragment tested. This suggests that DsbA plays an active role in the folding process but is not limiting under the conditions examined. Similarly, the overexpression of proline-*cis-trans*-isomerase (PPIase) had only a marginal effect on the yield of one of the fragments. The aggregation processes appear, therefore, to occur before the action of these proteins, or at least to be independent of their action.

Periplasmic folding, and in particular the question of periplasmic chaperones, have been reviewed recently (Wülfing and Plückthun, 1994a). While indirect evidence for the existence of a periplasmic chaperone was obtained (Wülfing and Plückthun, 1994b), the efficiency of this system is still unclear. It appears that the most effective strategy to obtain high-yield folding of periplasmic proteins is to engineer the protein itself. Here we propose that the crucial determinants are loops in the structure, at least in β-barrel proteins.

Materials and methods

Materials and strains

For all cloning and expression experiments the *E.coli* K12 strain JM83 [*ara* (*lac-pro* AB) *rpsL* (*str*A) *thi* (φ80 *lacZ*ΔM15), Yanisch-Perron *et al.*, 1985] was used. The plasmid pAK19 (Carter *et al.*, 1992) was obtained from P.Carter (Genentech, South San Francisco, CA).

Construction of plasmids

All *in vivo* experiments were carried out in the pHJ series of vectors (Knappik *et al.*, 1993; Knappik and Plückthun, 1994). They were constructed from our previous vectors (Skerra and Plückthun, 1991; Knappik *et al.*, 1993) and contain unique restriction sites at the 5′ and 3′ ends of the antibody genes, facilitating subcloning of antibody genes and domain shuffling between the F(ab)–, Fv– and single-chain Fv– vectors (Knappik and Plückthun, 1994; A.Knappik, unpublished data). Additionally, both genes for the heavy and the light chains carry short versions of the FLAG™ tag (Prickett *et al.*, 1989; Knappik and Plückthun, 1994) attached to the N-terminus of the mature protein to allow detection of the chains in immunoblots, regardless of the particular antibody fragment expressed. The genes of the huMab4D5 antibody (Carter *et al.*, 1992) were inserted in the pHJ vectors by PCR amplifying the coding regions of the variable and the constant domains independently using the vector pAK19 as a template.

The mutant antibody fragments H1–H10 were constructed by site-directed mutagenesis (Kunkel *et al.*, 1987) using the vector pLisc_SE (Knappik *et al.*, 1993) as a single-stranded template and up to six oligonucleotides per reaction. After screening and sequencing, the mutated part of the gene was gel-purified and ligated into the respective vector. Mutant H11 was constructed by combining the mutations H1 and H3; mutant H12 was constructed by removing the P40A mutation in mutant H10. Complementarity-determining region (CDR) grafting of the CDRs of McPC603 onto the huMab4D5 framework was carried out by the recursive PCR gene synthesis protocol described by Prodromou and Pearl (1992). All constructs were verified by DNA sequencing.

Monitoring cell stability during expression

Growth curves and leakiness data were obtained as follows. 20 ml of LB medium containing 100 μg/ml ampicillin and 25 μg/ml streptomycin were inoculated with an *E.coli* JM83 overnight culture harboring the plasmid encoding the respective antibody fragment and incubated at room temperature (24°C) until an OD_{550} of 0.5 was reached. IPTG was added to a final concentration of 1 mM and the incubation was continued for several hours. Every 30–60 min after induction the OD_{550} was measured. At the same time, 100 μl of the culture were removed and the cells pelleted by centrifugation (4 min at 6000 r.p.m.). In the supernatant, the β-lactamase activity was measured (O'Callaghan *et al.*, 1972) to quantify the degree of cell leakiness. After 3 h of induction, an aliquot of the culture was removed and the cells were lysed (see below). The β-lactamase activity in the cell extract was measured to verify that differences in the β-lactamase activity in the medium were not due to differences in the total amount of β-lactamase. The total amount of β-lactamase was found to deviate by ~20% between the clones in repeated experiments.

Cell lysis and immunoblotting

Cell cultures harboring the plasmids encoding the respective antibody Fv fragments were grown and induced (see above). After 3 h of induction, 2–5 ml of the cultures, depending on the OD_{550}, were harvested by centrifugation and lysed using a modified protocol given by Wetzel *et al.* (1991). The cell pellets (normalized to the same OD_{550}) were frozen for at least 1 h at −20°C and lysed by the addition of 500 μl lysis buffer (10 mM glycylglycine, 1 mM EDTA, 2 M urea, pH 7.5) to the frozen cells, and incubated on ice until thawed (20 min). 5 μl hen-egg lysozyme (250 μg) were added and the suspension was incubated for 10 min at room temperature. After adding 20 μl of 100 mM $MgCl_2$ and 10 μl of DNase I (1 mg/ml) and a further incubation of 10 min at room temperature, the lysates were centrifuged (5 min at 14 000 r.p.m.) and the supernatants containing the soluble fractions were withdrawn. The pellets were shaken at 4°C in 535 μl lysis buffer until resuspended to give the insoluble fractions. This procedure was chosen because it gave the same distribution of soluble and insoluble antibody fragments as lysing the cells with a French press, which was used for preparative antibody expression. The β-lactamase activity was measured in all fractions. The fractions were assayed by reducing SDS–PAGE with the samples normalized to the same β-lactamase activity to account for differences in plasmid copy number as well as cell leakiness. After blotting and immunostaining the gel using the anti-FLAG antibody M1 (Prickett *et al.*, 1989) as the first antibody, an Fc-specific anti-mouse antiserum conjugated to horseradish peroxidase as second antibody and a chemoluminescent detection assay as described elsewhere (Ge *et al.*, 1994), the lanes were scanned and quantitated by laser densitometry. All experiments were carried out at least in duplicate.

Thermodynamic stability

Fv fragments were purified using phosphorylcholine–Sepharose affinity chromatography (Skerra and Plückthun, 1988). The concentration was estimated photometrically using an extinction coefficient calculated according to Gill and von Hippel (1989). Equilibrium denaturation curves were obtained by denaturing 0.1 μM protein in borate buffered saline/phosphorylcholine (BBS/PC) buffer (200 mM sodium tetraborate, pH 8.0, 160 mM NaCl, containing 5 mM phosphorylcholine) and increasing amounts of urea (0.0–6.8 M

urea in 36 steps) in a total volume of 1.8 ml. After incubating the samples for 15 h at 10°C, the fluorescence spectra were recorded at 10°C from 300 to 400 nm with an excitation wavelength of 280 nm. The emission wavelength of the fluorescence peak shifted from 331.4 to 348.3 nm during denaturation (maximal deviation in repeated measurements 0.3 nm) and was used as the dependent variable. The denaturation was found to be reversible (data not shown). The data were fitted according to Pace (1990), assuming a two-state equilibrium system, since the fluorescence of the Fv fragment of McPC603 is dominated largely by V_H (>90%), which contains four Trp and nine Tyr, compared with one Trp and five Tyr in V_L (Glockshuber *et al.*, 1992). The 'fraction unfolded' was calculated from the fit values using the equation:

$$F_U = [y_f + (m_f \cdot [D]) - y]/[(y_f + m_f \cdot [D]) - (y_u + m_u \cdot [D])],$$

where y_f is the y intercept for the native conformation, y_u is the y intercept for the unfolded conformation, m_f is the slope of the pre-transition baseline, m_u is the slope of the post-transition baseline and [D] is the denaturant concentration. The parameters were obtained by fitting the data according to Pace (1990).

Renaturation kinetics

Purified Fv fragments were denatured for at least 3 h at 10°C in either BBS/PC (200 mM sodium tetraborate, 160 mM NaCl, 5 mM phosphorylcholine, pH 8.0) containing 7 M urea (denaturation under oxidized conditions) or BBS/PC containing 7 M urea and 100 mM dithioerythreitol (denaturation under reducing conditions). The denatured protein was diluted 90-fold in a cuvette containing refolding buffer (BBS/PC for refolding of the oxidized protein or BBS/PC containing 5 mM oxidized glutathione for refolding of the reduced protein), leading to a final protein concentration of 0.11 µM. During refolding, the solution was maintained at 10°C and mixed with a magnetic stirrer. Refolding was monitored for 3 h by scanning the fluorescence every 90 s from 300 to 400 nm (excitation wavelength: 280 nm). The data were analyzed by two methods: (i) the peak-shift from 348.3 to 331.4 nm during refolding was calculated by the estimation of the maximum emission wavelength of every scan. These data indicate the time-course of folding into a 'native-like' structure, where 'native-like' is defined by the property of having an emission maximum of ~331.4 nm; and (ii) the fluorescence intensities at 331.4 nm were estimated from the scans. These values represent the yield of the folding process, assuming that a decrease in fluorescence intensity at 331.4 nm during refolding monitors aggregation or loss by adsorption to the cuvette while an increase in fluorescence at that wavelength monitors the conversion into the folded protein.

Thermal denaturation

For measuring the thermal denaturation, 20 µl of purified Fv fragments were diluted into a cuvette containing 1.78 ml of BBS/PC buffer (200 mM sodium tetraborate, 160 mM NaCl, 5 mM phosphorylcholine, pH 8.0), prewarmed to the respective temperature and stirred with a magnetic stirrer. The final protein concentration was carefully adjusted to 1.0 µM in all experiments and for all mutants. The time-course of formation of aggregates was monitored with a fluorescence spectrometer (Shimadzu, RF-5000) at an excitation/emission wavelength of 400 nm. The kinetics of aggregation were found to be concentration-dependent and followed a pseudo-first-order reaction at the chosen protein concentration of 1.0 µM. The apparent rate constant k_{obs} was calculated by fitting the data using the equation:

$$F = F_{[\infty]} \cdot [1 - e^{-k(t - t_0)}],$$

where t is the time, t_0 is the apparent starting time-point of aggregation, and F and $F_{[\infty]}$ are the emission values at the time-points t and $t_{[\infty]}$, respectively.

Results

Development of the analysis system

We chose a model system consisting of the murine anti-phosphorylcholine antibody McPC603 (Satow *et al.*, 1986; Skerra and Plückthun, 1988) and the humanized version of the anti-HER2 antibody 4D5 (Carter *et al.*, 1992; Eigenbrod *et al.*, 1993). Despite sharing a high degree of sequence homology, they seemed to differ to a large extent in expression titers.

An alignment of the sequences of hu4D5 and McPC603 Fv, based on a structural definition for the CDRs, showed that 67 of 89 framework residues are identical in V_L (75%) and 72 of 92 framework residues are identical in V_H (78%). Nevertheless, the yield of recombinantly expressed and purified McPC603 F(ab) was ~0.2 mg per liter of bacterial culture and OD (Skerra and Plückthun, 1991), while the reported yield of the hu4D5 F(ab) fragment was ~50-fold higher (Carter *et al.*, 1992).

Analysis systems were developed to investigate the fate of the antibody proteins *in vitro* and *in vivo*. A shortened improved FLAG epitope (Prickett *et al.*, 1989; Knappik and Plückthun, 1994) on the N-terminus of the heavy and/or light chain, combined with a quantitative luminescence-based Western blot (Ge *et al.*, 1994), were used to quantify antibody protein, in both the soluble and insoluble forms, produced and transported to the periplasm of the bacteria. Correct assembly of the protein was determined independently by antigen affinity chromatography which does not allow either chain to bind separately. The effect on the *E.coli* growth physiology was measured by determining growth rates and the leakage of periplasmic β-lactamase into the growth medium after induction.

Comparison of hu4D5 and McPC603 antibody expression

The extraordinary expression levels of the hu4D5 F(ab) fragment were reported for high cell density fermentation (Carter *et al.*, 1992), while the McPC603 fragments were mostly studied in shake flasks (Skerra and Plückthun, 1991; Knappik *et al.*, 1993). Additionally, the growth media, the expression vectors and the *E.coli* strains were different in both systems. Therefore it was not possible to compare the available data directly or to give any statements about the reasons for the differences in expression titers.

For these reasons it was necessary to first investigate the antibody expression in identical *E.coli* host/vector systems. We PCR-cloned the coding regions of the variable and the constant domains of the hu4D5 F(ab) gene independently, using the vector pAK19 as a template, into the pHJ vector series (leading to Fv–, F(ab)– and scFv– expression vectors) and could then compare the expression behavior of the McPC603– and hu4D5 antibody fragments directly.

We found that in identical *E.coli* host/vector systems an enormous increase in the amount of folded periplasmic protein of hu4D5 compared with McPC603 was found, and this was observed independent of whether Fv, scFv or F(ab) fragments were studied. However, we also observed a marked physiologi-

cal difference: cells expressing McPC603 became leaky and eventually lysed, and major parts of the expressed molecules were found as insoluble aggregates (see below). These unfavorable effects on folding and growth physiology were absent when the cells expressed the corresponding hu4D5 fragments. We concluded therefore that the primary sequence alone must be responsible for the differences in both expression yield and cell stability during expression.

Point mutations in the McPC603 antibody framework improve its expression yield

We then wished to find out whether these effects could be ascribed to individual amino acid substitutions between the two antibodies. After domain switching experiments had shown that both the growth and leakiness phenotype and the folding problems were associated with V_H (A.Knappik and A.Plückthun, unpublished results), a number of single and multiple mutations (Table I) were constructed and analyzed, first in the form of the Fv fragment. We limited this analysis to framework regions using a structural definition of the CDRs, with a view to defining a generally useful antibody framework. The selection of the mutations was based on a comparison of the framework sequences of the McPC603 and hu4D5 V_H sequences, and any changes introduced during the humanization of 4D5, which itself appeared to improve folding (Carter *et al.*, 1992).

Indeed we found from this analysis that certain point mutations improved the expression behavior of the McPC603 antibody. Surprisingly, the growth/leakiness phenotype and the folding properties could be separated. The substitution P40A (H1 mutant; Table I) led to a slight decrease in leakiness and improved cell growth (with no influence on the amount of insoluble protein), while double mutant H3 (S63A/A64D) eliminated most of the insoluble protein yet had no effect on cell physiology (Figure 1). The H10 mutant, containing all eight mutations tested (Table I), showed a further improved phenotype: the ratio of soluble to insoluble protein was identical to H3, but the leakiness of cells during induction was much more reduced than with the H1 mutant. As a control, the Fv fragment carrying all except the P40A mutation (H12; Table I) was tested and found to lead to a leaky phenotype (data not shown). Therefore we concluded that the improved expression behavior of mutant H10 is due to the combination of mutations H1 and H3. We tested this hypothesis by constructing the mutant Fv fragment H11 which carried both the H1 and the H3 mutations (Table I). Indeed we found that this combination

Table I. Mutations introduced in the McPC603 V_H framework

Mutant	P40A 40	I48V 48	S63A 60	A64D 61	I71T 68	V72I 69	Q78K 75	A87S 82B
H1	●							
H2		●						
H3			●	●				
H4							●	
H5					●	●	●	
H6								●
H7					●	●	●	●
H8		●			●	●	●	
H9	●	●	●	●			●	●
H10	●	●	●	●	●	●	●	●
H11	●		●	●				
H12		●	●	●	●	●	●	●

Each line represents a different protein carrying the mutations indicated. The residues are numbered according to the McPC603 V_H sequence (Satow *et al.*, 1986). The second line gives the residue numbers according to Kabat *et al.* (1991) and is provided to simplify comparison of the positions with other antibodies.

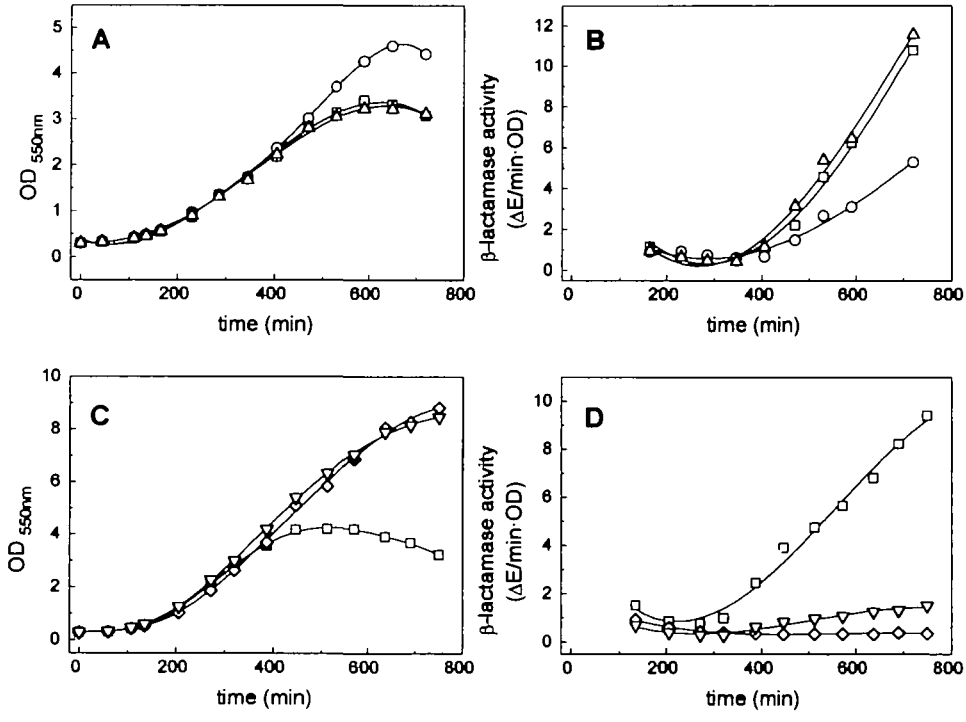

Fig. 1. Comparison of growth and leakiness phenomena of cells expressing various antibody Fv fragments. (**A**) Growth curves of cells expressing McPC603 wild-type (□), P40A (○) and S63A/A64D (△) (**B**) Corresponding leakiness data during induction, given as β-lactamase activity per OD in the culture supernatant. The symbols correspond to those in (A). (**C**) Comparison of growth curves of cells expressing the Fv fragments of McPC603 wild-type (□), the triple mutation P40A/S63A/A64D (▽) and the hu4D5 fragment (◇). (**D**) Corresponding β-lactamase activities in the medium. The symbols correspond to those in (C).

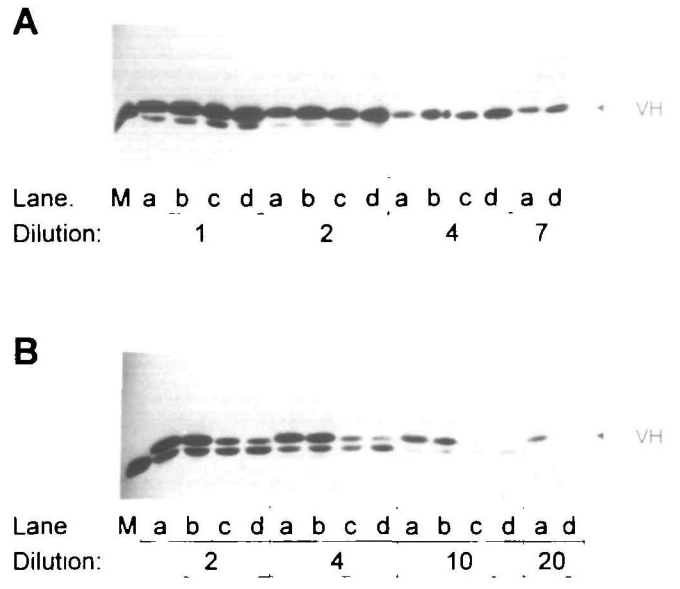

Fig. 2. Western blots showing the soluble (**A**) and insoluble (**B**) fractions of cell extracts, prepared as described in Materials and methods, expressing the Fv fragments of the McPC603 wild-type (a), the P40A mutation (b), the S63A/A64D mutation (c) and the triple mutation P40A/S63A/A64D (d). The amounts loaded on the gel were normalized to the same β-lactamase activity and diluted by the factor indicated in the panels. A dilution of 1 corresponds to ~100 µl of cell culture with an OD_{550} of 1. The first lane (M) shows the size marker. Since with the constructs used here the detection of the heavy chain in Western blots is 10-fold more sensitive than the detection of the light chain due to differences in the FLAG epitope (Knappik and Plückthun, 1994), only the heavy chain is visible in this type of blot (indicated by an arrow). The band visible below the V_H chain corresponds to cross-reacting hen-egg lysozyme used to lyse the cells.

Table II. Summary of *in vivo* properties of the mutated Fv fragments compared with the McPC603 wild-type Fv fragment

	P40A	S63A/ A64D	P40A/S63A/ A64D
Soluble (%)[a]	300	200	300
Insoluble (%)[a]	100	10	5
Increase in soluble/insoluble ratio	3-fold	20-fold	60-fold
Functional yield (%)[b]	300	200	300
Cell stability[c]	+	±0	+++
Volume yield (%)[d]	400	200	1000

[a]Relative amounts of soluble and insoluble mature V_H chain from the Fv fragment after 3 h induction at 24°C as estimated from densitometric scanning of Western blots (see Figure 2). The corresponding values of the wild-type are set to 100%.
[b]Functional yield (normalized to equal cell densities) given as the relative amount of Fv fragment purified from crude extracts by antigen affinity chromatography after 3 h induction at 24°C, compared with the McPC603 wild-type Fv fragment set at 100%.
[c]Stability during expression at 24°C of cells expressing mutated Fv fragments compared with cells expressing the McPC603 wild-type Fv fragment. See Figure 1 for details.
[d]Yield of mutated Fv fragments per liter cell culture, purified from crude extracts by antigen affinity chromatography after 'optimal' induction periods at 24°C is compared with the functional yield obtained with the McPC603 wild-type Fv fragment after 3 h of induction at 24°C. The 'optimal' induction period depends on the cell stability during induction and amounts to 5 h for the fragment containing the P40A mutation, 3 h for the fragment containing the S63A/A64D mutation and 12 h for the fragment containing the P40A/S63A/A64D mutation

had a synergistic effect, almost completely eliminating any insoluble protein (Figure 2 and Table II) and approaching the favorable growth behavior of cells expressing hu4D5 (Figure 1).

The relative effect of the mutations on the *in vivo* folding yields (Table II) was found to be similar in the F(ab) and scFv fragments as in the Fv fragments, suggesting that the diversion from the folding pathway appeared to be governed by the folding of an individual domain. However, the effect of the P40A mutation on the lysis phenotype was much reduced in F(ab) and scFv fragments compared with Fv. We conclude that the lysis phenotype is not proportional to the amount of

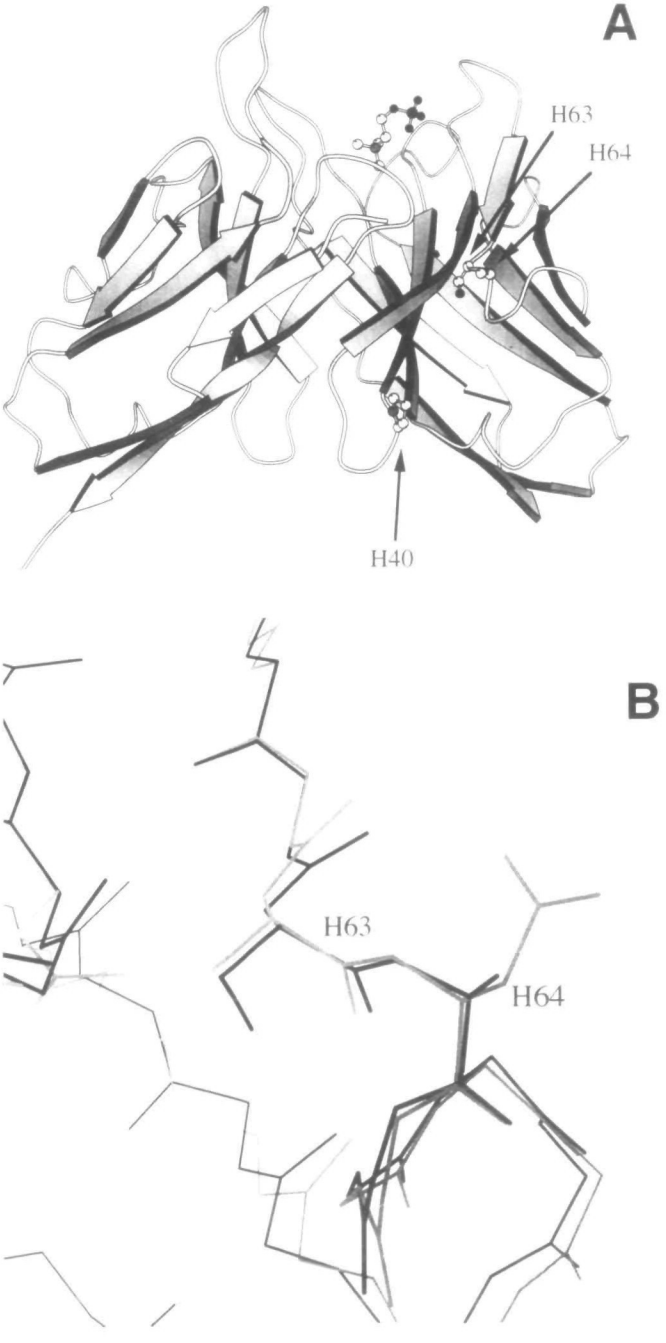

Fig. 3. (**A**) Location of the mutations in the structure of the phosphorylcholine binding antibody McPC603 (Satow *et al.*, 1986). (**B**) Superposition of the heavy chains of McPC603 (black) and hu4D5 (grey) (Eigenbrod *et al.*, 1993).

insoluble protein, and must be caused by the differential interaction of the various proteins with one or more cellular components during transport or folding.

Improved cell stability during induction of the expression of H11 mutant was significant at lower growth temperatures and was lost at 37°C, while the ratio of soluble to insoluble protein was independent of the growth temperature (data not shown). In contrast, cells expressing the hu4D5 Fv fragment remained stable at all temperatures tested. Therefore, additional crucial residues, either in the framework or in the CDRs, must be responsible for the differences in cell stability when the expression was carried out at 37°C. To distinguish between these two possibilities, we also grafted the CDRs of the McPC603 Fv antibody on the hu4D5 framework. This chimeric molecule consisted of the hu4D5 framework residues and the McPC603 CDRs. However, we found no additional improvements in yield and growth physiology with our assay systems (A.Knappik, H.Bothmann, K.Bauer and A.Plückthun, manuscript in preparation).

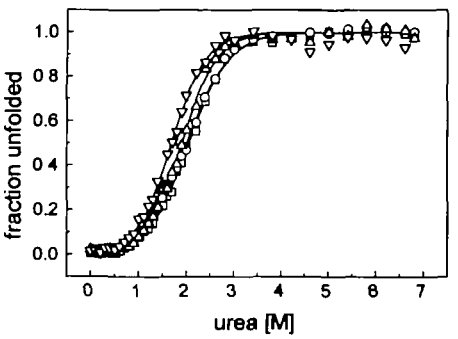

Fig. 4. Equilibrium denaturation curves of Fv fragments [McPC603 (□), P40A (○), S63A/A64D (Δ) and P40A/S63A/A64D (▽)]. The fraction unfolded was calculated from the wavelength of the fluorescence maximum (λ_{ex} = 280 nm), measured as a function of the denaturant concentration. The calculated values for *m* (the slope of the curve in the transition region) and the differences in $\Delta G(H_2O)$ are given in Table III.

We conclude therefore that the observed difference in stability between cells grown at 37°C and expressing the hu4D5 or the McPC603 Fv fragment containing the H11 mutations must be due to the different CDR sequences of the two antibodies. The mutations we found (Figure 3), however, are sufficient to eliminate the differences between the hu4D5 and the McPC603 *framework* in yield and cell physiology.

In vitro analysis of the mutant Fv fragments

We next sought to establish if the effects observed *in vivo* could be related to *in vitro* properties of the proteins. An *in vitro* characterization of the wild-type and mutant Fv fragments showed that these mutations did not increase the thermodynamic stability of the proteins, as measured by urea-induced fluorescence changes (Figure 4). When measuring the folding rate of the oxidized Fv fragments (containing the intradomain S–S bonds), a kinetic intermediate was detected by fluorescence changes, which was formed faster than precisely measurable by manual mixing and which slowly interconverted

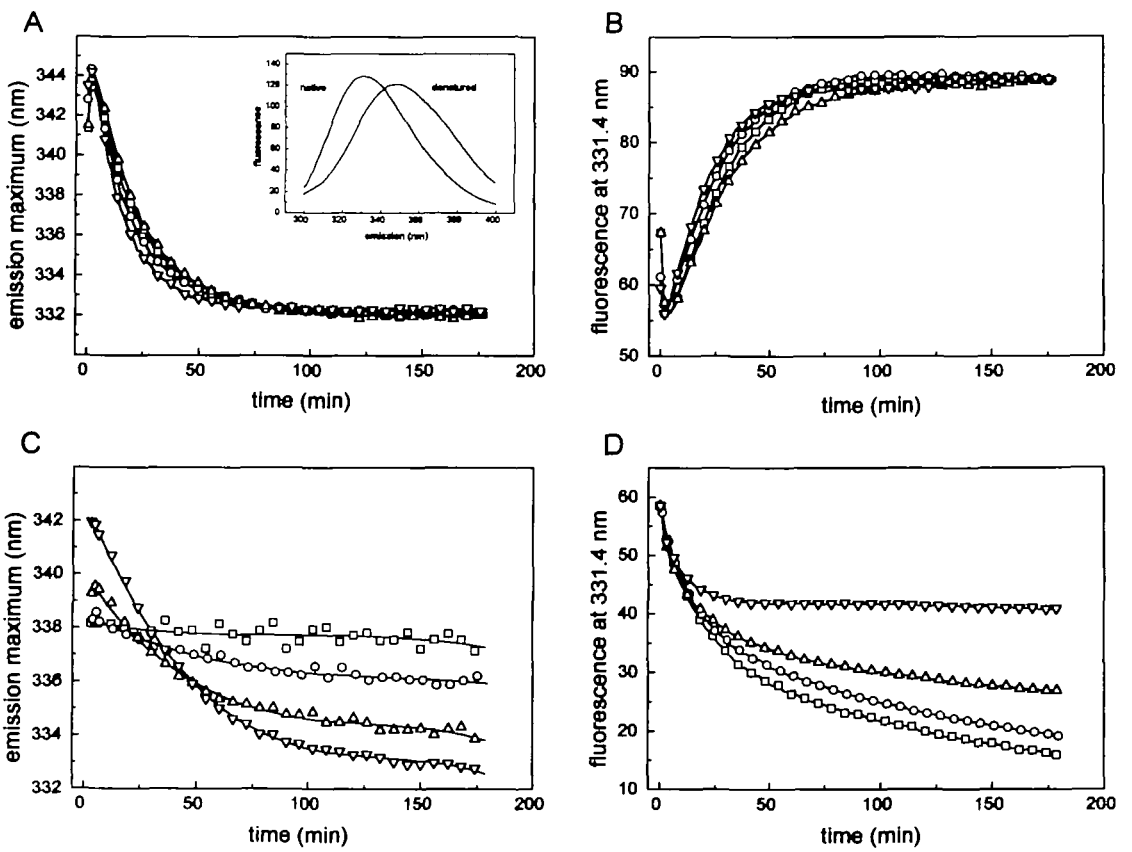

Fig. 5. Time-course of refolding of urea-denatured Fv fragments [McPC603 (□), P40A (○), S63A/A64D (Δ) and P40A/S63A/A64D (▽)] of either oxidized (**A** and **B**) or reduced (**C** and **D**) protein, monitored by fluorescence spectroscopy. (**A** and **C**) The shift of the emission peak during refolding is plotted against the time. The fully denatured protein shows an emission maximum at 348.3 nm, while the folded protein has an emission maximum at 331.4 nm (see small insert in A). (**B** and **D**) The amount of protein reaching a native-like state (as given in A and C) during refolding is plotted using the fluorescence intensity at 331.4 nm. See Materials and methods for details.

to the native state. The yields and rates of the slow folding step were again identical for all proteins examined (Figure 5A and B). To mimic the *in vivo* situation more closely, the experiment was repeated for the reduced proteins (Figure 5C and D). This time, the kinetics of the slow step differed dramatically between the proteins: while the McPC603 wild-type and (less dramatically) H1 mutant fragment did not fold during the time of measurement, but rather aggregated very rapidly, the H3 and particularly H11 mutant fragments interconverted slowly into a native-like structure (Figure 5C). Moreover, the yield also followed the *in vivo* results (Figure 5D). Interestingly, the synergistic effect observed *in vivo* with the triple mutant fragment could also be observed *in vitro*. It thus appears that a folding intermediate differs between the mutant proteins in its tendency to aggregate.

A similar ranking of the mutant fragments could be obtained by measuring by light scattering the rate of thermally induced aggregation, whose time-courses were consistent with a uni-bimolecular process when examined at different concentrations (data not shown) but followed first-order kinetics at 1.0 μM. Interestingly, since the slopes of the temperature dependence of the aggregation rate were identical for all mutants (Figure 6), the enthalpy of activation is similar for all proteins, while the entropy of activation (the intercept in Figure 6) depends on the nature of the amino acids at position H63 and/or H64, but not on that at H40.

Discussion

What is the mechanistic basis of the observed differences in expression yields? The side chain of AspH64 is exposed in the hu4D5 Fv fragment (Eigenbrod *et al.*, 1993) and other antibodies which carry this amino acid, and the main chain of McPC603 and hu4D5 is almost perfectly superimposable in this region (Figure 3). A plausible explanation for the different aggregation behavior of the mutants may thus be a greater ordering of water around AlaH64 in the ground state, compared with AspH64. Thus, less entropy of activation is required to overcome the barrier for forming the crucial transition state with increased hydrophobic surface, which leads to aggregation, since AlaH64 may start at a higher ground state level, characterized by higher ordering of solvent. The folding process in the bacterial periplasm mirrors the *in vitro* behavior surprisingly well, suggesting that under the experimental conditions there are no molecular chaperones present in sufficient amounts to efficiently prevent the aggregation reaction. Actually, this type of mutation seems to free the protein from the need of a chaperone.

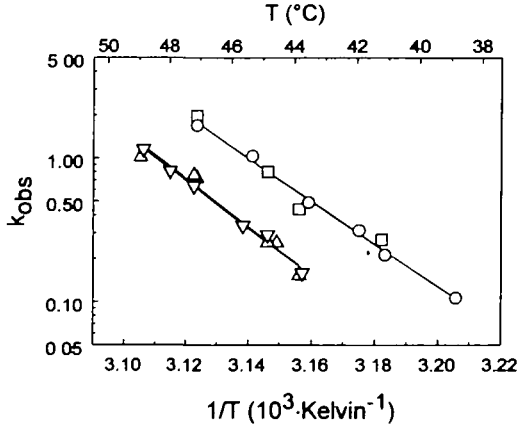

Fig. 6. Thermal denaturation of wild-type and mutant antibody Fv fragments [McPC603 (□), P40A (○), S63A/A64D (Δ) and P40A/S63A/A64D (▽)] measured with light scattering at 400 nm as a function of temperature.

While the mechanistic basis of the cell-stabilizing effect of the P40A mutation (H1) is not yet clear, it acts synergistically with the double mutation (H3) *in vivo* (Table II) and *in vitro* (Figure 5C and D), particularly if the proteins containing either the double (H3) or triple mutation (H11) are compared; thus both mutations may act on crucial steps of the same pathway. A speculative model consistent with the experimental data would state that there are two crucial intermediates during the *in vivo* folding pathway. The first is either itself toxic for the cells or leads to a toxic by-product, possibly by interacting with the transport apparatus of the cells. This effect of the lysis mediating intermediate can be reduced by introducing the P40A mutation in the McPC603 Fv fragment or, much more dramatically, by introducing this mutation in the Fv fragment which already contains the H3 mutations (Figure 1). The second intermediate leads to aggregation of the antibody fragments during *in vivo* and *in vitro* folding and can be prevented by the introduction of the H3 mutations.

The folding efficiency of proteins can thus depend on single amino acids which determine the kinetic partitioning between folding and aggregation. In other proteins, where similar mutations have been found (Jappelli *et al.*, 1992; Mitraki and King, 1992; Chrunyk *et al.*, 1993; Deng *et al.*, 1994), they were also found to lie in turns of the protein. We predict that the folding efficiency does not simply depend on a sufficient number of equidistant surface charges in the native structure, but that there are crucial exposed microdomains consisting of turns which, in a folding or unfolding intermediate, become partially detached from the protein and can, if sufficiently hydrophobic, provide an aggregation interface. *In vitro* studies with, for example, Trp synthase (London *et al.*, 1974), carbonic anhydrase (Cleland and Wang, 1990) and bovine growth hormone (Brems, 1988) are all consistent with the idea that it is an intermediate state, and neither the unfolded nor the native state, which leads to aggregation. Brems (1988) found that by adding helix-forming peptides from the growth hormone sequence to the refolding mixture of this protein, the formation of aggregates could be inhibited, suggesting that they may bind protectively at crucial sites for the aggregation reaction. These data are in excellent agreement with the results observed here and support the hypothesis that specific microdomains of a folding intermediate are susceptible to aggregation.

We suggest that a similar intermediate may be populated in thermal denaturation and in oxidative folding, because the mutations have a similar effect in both directions of folding. Since the NMR signals of the backbone amide proton have now been assigned for the Fv and single-chain Fv fragment

Table III. Summary of stability data of Fv fragments (obtained from Figure 4)

	McPC603 wild-type	P40A	S63A/A64D	P40A/S63A/A64D
m (kcal/mol/M)	1.34 ± 0.08	1.25 ± 0.07	1.52 ± 0.10	1.56 ± 0.10
$\Delta\Delta G(H_2O)$ (kcal/mol)	reference	0.23 ± 0.2	−0.14 ± 0.2	0.13 ± 0.2

The differences of the free energy of folding $\Delta\Delta G(H_2O)$ compared with the wild-type Fv fragment and the slope *m* in the transition region are shown. In this measurement, the protein concentration was 0.1 μM in all cases.

of McPC603 (Freund *et al.*, 1994), folding studies using NMR may give a very detailed picture of the folding and assembly pathway. There may thus be two solutions to the folding problem. One, as described here, is to inhibit aggregation by designing a charge in a crucial region likely to detach early from the protein or to attach late in the folding. Another may be to stabilize the interaction of this region with the rest of the protein. This strategy might have been operative in the disulfide mutants which stabilize T4 lysozyme against thermally induced aggregation, but which do not significantly change the overall ΔG of the molecule (Wetzel, 1987).

The mutations described here are of a different nature from those described by Hurle *et al.* (1994). These authors identified destabilizing mutations in V_L domains which appear to be frequent in those light chains found in amyloidosis but very infrequent in normal light chains. The free energy of folding was found to correlate roughly with the tendency to form amyloid-like aggregates. In contrast, the mutations we identified do not change ΔG within experimental error. Furthermore, the enhanced folding yields observed are not correlated at all with the distribution of these amino acids. The amino acids we replaced were not rare ones. It thus appears that there might be a selection against unstable molecules in the animal, but not necessarily for efficient folding. It rather seems that the eukaryotic organism has other means to overcome these problems.

The antibody folds after transport into the endoplasmic reticulum and it appears to be guided by at least two chaperones, Hsp70 (BiP) and Hsp90 (GRP94) (Melnick *et al.*, 1994). In addition, protein disulfide isomerase (PDI) has been shown to be involved in this process *in vivo* (Roth and Pierce, 1987). It is not yet known whether calnexin, a membrane-bound protein in the endoplasmic reticulum presumed to be involved in glycoprotein folding (Bergeron *et al.*, 1994), also participates in the folding process of the heavy chain, at least for some subclasses. In contrast, the chaperone content of the periplasmic space is probably far more modest, and there is no evidence for ATP in this compartment (Wülfing and Plückthun, 1994a,b). Rather than trying to recreate an identical environment as in the endoplasmic reticulum, which may be impossible, we attempted therefore to engineer the protein such that it no longer would be dependent on chaperones.

It is possible that, and it remains to be investigated whether, the susceptibility of the improved variants to disulfide isomerase and proline-*cis-trans*-isomerase overexpression has changed. Depending on the folding conditions, PDI may promote either folding or aggregation *in vitro* (Lilie *et al.*, 1994; Puig and Gilbert, 1994). While in the eukaryotic system Hsp70 and Hsp90 are probably the main factors in preventing premature aggregation, in the *E.coli* system it is the protein structure itself which is of primary importance. It can thus not be excluded that even better yields may now become possible, after the first bottleneck of aggregation has been identified and removed, by again increasing other catalysts in the system.

Recent impressive progress in constructing and screening libraries containing large antibody repertoires (Winter *et al.*, 1994) has revealed the problem that limitations in the maintenance and stability of the diversity, due to very different effects of distinct antibody sequences on the bacterial cell physiology, might lead to the elimination of clones. This study shows that in the future, semi-synthetic libraries based on engineered frameworks, but carrying the CDR variability, may overcome this problem, facilitating equal representations of clones and amplification and long-term maintenance of such libraries. Since the aspects of stability, folding and physiology can apparently be separated, such a design may now be possible. Particularly attractive might be the combination of the strategies described here with stabilizing disulfide mutations (Glockshuber *et al.*, 1990).

Moreover, it becomes apparent that mutations influencing the aggregation of proteins during folding are most often located in turns, at least in β-barrel proteins. An engineering approach as described here, as well as random mutagenesis focused on turn regions, may overcome potential folding problems of other recombinant proteins and, by evaluating different types of folding mutations as described here, may lead to a deeper insight into the mechanisms of protein folding.

Acknowledgements

We thank Konrad Bauer for help in the construction of the loop-grafted antibody genes, Dr Paul Carter (Genentech) for his kind gift of pAK19, and Claus Krebber and Dr Simon Moroney for useful discussions. This work was supported by the Schweizerische Nationalfonds grant 31-37717.93.

References

Bergeron,J.M., Brenner,M.B., Thomas,D.Y. and Williams,D.B. (1994) *Trends Biochem. Sci.*, **19**, 124–128.
Brems,D.N. (1988) *Biochemistry*, **27**, 4541–4546.
Carter,P. *et al.* (1992) *Biotechnology*, **10**, 163–167.
Chrunyk,B.A. and Wetzel,R. (1993) *Protein Engng*, **6**, 733–738.
Chrunyk,B.A., Evans,J., Lillquist,J., Young,P. and Wetzel,R. (1993) *J. Biol. Chem.*, **268**, 18053–18061.
Clelend,J.L. and Wang,D.I.C. (1990) *Biochemistry*, **29**, 11072–11078.
Dale,G.E., Broger,C., Langen,H., d'Arcy,A. and Stüber,D. (1994) *Protein Engng*, **7**, 933–939.
Deng,S., MacKenzie,C.R., Sadowska,J., Michniewicz,J., Young,N.M., Bundle,D.R. and Narang,S.A. (1994) *J. Biol. Chem.*, **269**, 9533–9538.
Eigenbrod,C., Randal,M., Presta,L., Carter,P. and Kossiakoff,A.A. (1993) *J. Mol. Biol.*, **229**, 969–995.
Freund,C., Ross,A., Plückthun,A. and Holak,T. (1994) *Biochemistry*, **33**, 3296–3303.
Ge,L., Knappik,A., Pack,P., Freund,C. and Plückthun,A. (1994) In Borrebaeck,C. (ed.), *Antibody Engineering. A Practical Approach*. Oxford University Press, in press.
Gill,S.C. and von Hippel,P.H. (1989) *Anal. Biochem.*, **182**, 319–326.
Glockshuber,R., Malia,M., Pfitzinger,I. and Plückthun,A. (1990) *Biochemistry*, **29**, 1362–1367.
Glockshuber,R., Schmidt,T. and Plückthun,A. (1992) *Biochemistry*, **31**, 1270–1279.
Hurle,M.R., Helms,L.R., Li,L., Chan,W. and Wetzel,R. (1994) *Proc. Natl Acad. Sci. USA*, **91**, 5446–5450.
Jaenicke,R. (1987) *Prog. Biophys. Biol.*, **49**, 117–237
Jappelli,R., Luzzago,A., Tataseo,P., Pernice,I. and Cesareni,G. (1992) *J. Mol. Biol.*, **227**, 532–543.
Kabat,E.A., Wu,T.T., Perry,H.M., Gottesman,K.S. and Foeller,C. (1991) *Sequences of Proteins of Immunological Interest*. 5th edition. National Institutes of Health, Bethesda, MD.
Knappik,A. and Plückthun,A. (1994) *BioTechniques*, **17**, 754–761.
Knappik,A., Krebber,C. and Plückthun,A. (1993) *Biotechnology*, **11**, 77–83.
Kunkel,T.A., Roberts,J.D. and Zakour,R.A. (1987) *Methods Enzymol.*, **154**, 367–382.
Lilie,H., McLaughlin,S., Freedman,R. and Buchner,J. (1994) *J. Biol. Chem.*, **269**, 14290–14296.
London,J., Skrzynia,C. and Goldberg,M.E. (1974) *Eur. J. Biochem.*, **47**, 409–415.
Melnick,J., Dul,J.L. and Argon,Y. (1994) *Nature*, **370**, 373–375.
Mitraki,A. and King,J. (1992) *FEBS Lett.*, **307**, 20–25.
O'Callaghan,C.H., Morris,A., Kirby,S.M. and Shingler,A.H. (1972) *Antimicrob. Agents Chemother.*, **1**, 283–288.
Pace,C.N. (1990) *Trends Biotechnol.*, **8**, 93–98.
Plückthun,A. (1994) In Rosenberg,M. and Moore,G.P. (eds), *The Pharmacology of Monoclonal Antibodies*. Springer Verlag, Berlin, Germany, Vol 113, pp. 269–315.
Prickett,K.S., Amberg,D.C. and Hopp,T.P. (1989) *BioTechniques*, **7**, 580–589
Prodromou,C and Pearl,L.H. (1992) *Protein Engng*, **5**, 827–829.

Puig,A. and Gilbert,H. (1994) *J. Biol. Chem.*, **269**, 7764–7771.
Roth,R.A. and Pierce,S.B. (1987) *Biochemistry*, **26**, 4179–4182.
Satow,Y., Cohen,G.H., Padlan,E.A. and Davies,D.R. (1986) *J Mol. Biol.*, **190**, 593–604.
Shatzman,A.R. (1990) *Curr. Opin. Biotechnol.*, **1**, 5–11.
Skerra,A. and Plückthun,A. (1988) *Science*, **240**, 1038–1041.
Skerra,A. and Plückthun,A. (1991) *Protein Engng*, **4**, 971–979.
Tsai,A.Y.M., Itoh,M., Streuli,M., Thai,T. and Saito,H. (1991) *J. Biol. Chem.*, **266**, 10534–10543.
Wetzel,R. (1987) *Trends Biochem. Sci.*, **12**, 478–482.
Wetzel,R. (1994) *Trends Biotechnol.*, **12**, 193–198.
Wetzel,R., Perry,L.J. and Veilleux,C. (1991) *Biotechnology*, **9**, 731–737.
Winter,G., Griffiths,A.D., Hawkins,R.E. and Hogenboom,H.R. (1994) *Annu. Rev. Immunol.*, **12**, 433–455.
Wülfing,C. and Plückthun,A. (1994a) *Mol. Microbiol.*, **12**, 685–692.
Wülfing,C. and Plückthun,A. (1994b) *J. Mol. Biol.*, **242**, 655–669.
Yanisch-Perron,C., Vieira,J. and Messing,J. (1985) *Gene*, **33**, 103–119.

Received August 23, 1994; accepted October 18, 1994