Folding Nuclei of the scFv Fragment of an Antibody†

Christian Freund,‡ Annemarie Honegger,‡ Peter Hunziker,‡ Tad A. Holak,§ and Andreas Plückthun*,‡

Biochemisches Institut, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland, and Max-Planck-Institut für Biochemie, D-82152 Martinsried, FRG

Received November 21, 1995; Revised Manuscript Received April 5, 1996⊗

ABSTRACT: The folding kinetics of the variable domains of the phosphorylcholine-binding antibody McPC603, combined into a scFv fragment [V_H-($Gly_4Ser)_3$-V_L], were investigated by the use of fluorescence spectroscopy, nuclear magnetic resonance (NMR), and mass spectrometry (MS). All three methods gave evidence for the occurrence of a major kinetic intermediate during the refolding of the denatured, oxidized scFv fragment. This intermediate is formed within the first 30 s of folding and comprises exchange-protected amide protons of hydrophobic and aromatic amino acids, most of which are localized within the inner β-sheet of the V_L domain. In the subsequent slow step, most of the amide protons become protected with rate constants that are very similar for residues of both domains. These data are in agreement with the MS results, which indicate a cooperative folding event from the intermediate to the native state of the scFv fragment.

While antibody variable domains have become one of the structurally best characterized groups of proteins because of their central role in molecular recognition in the immune system (Alzari et al., 1988; Davies & Padlan, 1990), little is known about the structure formation during their folding *in vitro*. The interaction of the domains is known to be important for the stability and functionality of the antibody fragments of which they form part (Hochmann et al., 1976; Searle et al., 1995). The two variable domains may be expressed as heterodimeric, noncovalently linked Fv fragments (Skerra & Plückthun, 1988; Riechman et al., 1988) or as single-chain Fv[1] fragments (scFv's; Bird et al., 1988; Huston et al., 1988), where the two domains are connected by a flexible peptide linker. A simple two-state model can be used to fit the thermodynamic data from reversible solvent denaturation studies of Fv and scFv fragments in many cases (Pantoliano et al., 1991; Knappik & Plückthun, 1995), suggesting a coupling of the two variable domains into a cooperative folding unit.

Kinetic data from different antibody fragments reveal a more complex behavior. The rate-limiting step for the folding of a C_L domain (Goto & Hamaguchi, 1982), an immunoglobulin light chain (Lang & Schmid, 1988), or a Fab fragment (Lilie et al., 1993) of an antibody has been shown to be proline isomerization. Isolated V_L domains were shown to refold rapidly (Goto et al., 1979; Tsunenaga et al., 1987) and were suggested to undergo a retarded folding in the presence of the C_L domain. Complete folding of denatured, oxidized Fv (Knappik & Plückthun, 1995) and scFv fragments is known to be slow (Pantoliano et al., 1991), but the mode of interaction of the two domains during the folding reaction remains to be elucidated.

In order to investigate the nature of possible structured intermediates of antibody variable domains and to address the question of cooperativity at the level of individual amino acids, we made use of the recently obtained NMR assignments of most of the ^{15}N and ^{1}H resonances of the Fv and scFv fragments of the antibody McPC603 (Freund et al., 1994). The analysis of H/D exchange experiments by NMR methods allows the characterization of folding intermediates in terms of secondary structure formation (Udgaonkar & Baldwin, 1988; Roder et al., 1988), while MS analysis of the same experiments reveals the population of such intermediates (Miranker et al., 1993) and allows sequential and parallel pathways to be distinguished.

The intermediates of a number of α and α/β proteins have been characterized by pulsed hydrogen exchange techniques and NMR (Udgaonkar & Baldwin, 1988; Roder et al., 1988; Bycroft et al., 1990; Briggs & Roder, 1992; Lu & Dahlquist, 1992; Radford et al., 1992; Jennings & Wright, 1993; Mullins et al., 1993, Jacobs & Fox, 1994, Jones & Matthews, 1995) and have been shown in most cases to form rapidly as compared to the time taken for the completion of folding. Two all-β-sheet proteins have been characterized by H/D experiments in combination with NMR analysis (Valery et al., 1993; Koide et al., 1993), and intermediates with native-like β-sheet structure were identified. No member of the immunoglobulins has been studied by these techniques so far, however. Since antibody variable domains are one of the prototypes of the immunoglobulin fold (Bork et al., 1994), they might provide an interesting model system for studying the folding of this protein family. We therefore applied hydrogen exchange experiments in combination with

† This work was supported by the Schweizerische Nationalfonds (31-37717.93), the Bundesamt für Bildung und Wissenschaft (93.0198, EG BIO2-CT-0367), the ZENECA Strategic Research Fund, and the Deutsche Forschungsgemeinschaft (SFB).

* Author to whom correspondence should be addressed. Telephone: (+41-1) 257-5570. Fax: (+41-1) 257-5712. E-mail: plueckthun@biocfebs.unizh.ch).

‡ Universität Zürich.

§ Max-Planck-Institut.

⊗ Abstract published in *Advance ACS Abstracts,* June 1, 1996.

[1] Abbrevations: CDR, complementarity-determining region; C_L, constant domain of the light chain of an antibody; Fab, antigen binding antibody fragment; Fv, antibody fragment composed of the variable heavy (V_H) and variable light (V_L) chain domains; HSQC, heteronuclear single-quantum coherence; MS, mass spectrometry; NMR, nuclear magnetic resonance; PPI, peptidyl−prolyl *cis*−*trans* isomerase; scFv fragment, single-chain Fv fragment (a genetically linked Fv fragment of the topology V_H-linker-V_L); TPPI, time-proportional phase incrementation.

S0006-2960(95)02764-4 CCC: $12.00

NMR and MS analysis to the folding of an antibody scFv fragment.

EXPERIMENTAL PROCEDURES

Protein preparation. The scFv fragment was expressed and purified as described previously (Freund et al., 1993). ^{15}N-Labeled samples were obtained by growth of the bacteria in defined medium containing [^{15}N]ammonium chloride as the sole nitrogen source. The isolated variable domains were obtained by chain separation of the corresponding Fv fragment (Freund et al., 1994). The periplasmic peptidyl *cis*−*trans* isomerase from *Escherichia coli*, which is the product of the *rot*A gene, was prepared using the procedure of Liu and Walsh (1990).

Fluorescence Measurements. All measurements were carried out using a Shimadzu RF-5000 spectrofluoriphotometer with an excitation wavelength of 295 nm for the scFv fragment and 280 nm for the V_L domain. In the case of the scFv fragment, the emission was monitored at 331.4 nm, the wavelength of the fluorescence maximum of the native scFv fragment. For the V_L domain the emission was recorded at 338 nm, the fluorescence maximum of the spectrum of the V_L domain recorded directly after dilution of the denatured protein into refolding buffer. Proteins were denatured for at least 4 h at room temperature in 4 M guanidinium chloride, and refolding was initiated by a 1:100 dilution into 35 mM HEPES buffer, pH 8.0, at 10 °C. In the case of the scFv fragment, 5 mM phosphorylcholine was added to the buffer solution. In one experiment, 1 μM *E. coli* peptidyl−prolyl *cis*−*trans* isomerase was added to the refolding buffer, and the renaturation of the isolated V_L domain was monitored in the presence of the enzyme. Protein concentrations were 0.18 μM in the case of the scFv fragment, 1 μM for the V_L domain, and 0.08 μM for the V_H domain. The experimental curves were fitted by a sum of exponential functions (Kaleidagraph Software).

H/D Exchange Experiments. Five to six milligrams of the purified scFv fragment (^{15}N-labeled for the NMR experiments) was denatured overnight at 4 °C in 1 mL of D_2O buffer, containing 4 M guanidinium chloride and 20 mM borate, pH 8.0. The pH of D_2O solutions was not isotope corrected. For the initiation of the folding reaction, the protein was diluted 1:100 into 0.4 M arginine and 0.1 mM phosphorylcholine, pH 8.0 in D_2O. After various times, a further 1:5 dilution into a H_2O buffer containing 0.12 M potassium phosphate and 0.1 mM phosphorylcholine, pH 4.0, was performed. The final pH was 5.2. In the case of the reference sample, the first dilution was done into H_2O buffer at pH 8.0 to allow all deuterated amide protons to exchange. The reaction was then allowed to proceed to completion (4 h at 10 °C). After concentration of the solution using an Amicon A8200 cell, the protein was dialyzed against 5 mM KH_2PO_4 buffer for the NMR analysis and against 10 mM ammonium acetate buffer, pH 5.2, for the MS analysis. In the case of MS analysis, the protein was further purified by gel filtration on a PD-10 column. The samples were then lyophilized and dissolved in D_2O (12 h at 4 °C) for the NMR experiments and in H_2O for the MS experiments.

MS Analysis. Molecular masses were determined using ESI-MS by flow injection of the scFv fragment into the ion source of a Sciex API III instrument. The ion spray voltage was approximately 5000 V, and the nebulizer gas pressure was 40 psi. Just before injection, the lyophylized samples were dissolved in 10 mM ammonium acetate buffer, and acetic acid was added to 1%. For the analysis of the data, it was taken into account that the final refolding buffer for the reference sample (see Figure 2a) contained no residual D_2O while the other samples (Figure 2b−d) had a residual D_2O content of 20%. This means that the slowly exchanging amide protons, which become protected in the slow folding reaction, are 20% deuterated. The equation $y + (46 - y)0.2 = 18$ gives the number (y) of amide protons which become protected in the intermediate due to the formation of stable secondary structure. This results in a calculated number of 11 amide protons. The weak additional peaks in all spectra with molecular masses increased by 60 mass units indicate possible acetic acid adducts of the respective proteins (Smith et al., 1990).

NMR Spectra. All spectra were recorded on a Bruker DRX 600 spectrometer equipped with a *Z*-gradient unit. The temperature was 27 °C, and the samples were dissolved in D_2O. Relative protein concentrations were determined by comparing the methyl resonances at −1.0 ppm in 1D spectra recorded with 64 scans. ^{15}N−^{1}H correlation spectra were recorded using a gradient enhanced version of a HSQC experiment [FHSQC (Mori et al., 1995)]. The carrier was positioned on the water resonance. Residual water suppression was achieved by the application of a WATERGATE 3−9−19 refocusing pulse (Sklenár et al., 1993) with a pulse interval of 200 μs in the final BACK INEPT step, allowing optimal inversion of the amide resonances. The gradient strengths were 25%, 10%, and 80% of a maximum gradient power of 30 G/cm for the gradients G_1, G_2, and G_3, respectively. The delay for transferring proton magnetization to nitrogen in the INEPT step was set to 2.25 ms. Decoupling during aquisition was achieved using the GARP pulse sequence (Shaka et al., 1985). To obtain phase sensitive spectra, the TPPI method was used (Marion & Wüthrich, 1993). A data matrix of 4K × 128 points was aquired. Data were processed to a final size of 2K × 256 points using the in-house written software CC-NMR (Cieslar et al., 1993).

For each peak analyzed in the HSQC spectra, the corresponding peak in the reference spectrum was chosen to represent 100% proton occupancy. The intensities of the peaks were determined by measuring the peak heights since the line widths of individual peaks did not change in the spectra of different samples. Rate constants for the individual peaks were determined by applying single-exponential fits to the plots of signal intensities (measured as peak heights) versus time of folding (until the dilution step into H_2O, see above). Protection within the intermediate (in percent) was determined by using the equation $(I_0 - I_t)/(I_0 - I_\infty) \times 100$, where I_0 denotes the peak height in the H_2O reference sample, I_t the peak height at the respective time point, and I_∞ the calculated peak height at time ∞.

RESULTS

Fluorescence Kinetics. The folding kinetics of the scFv fragment measured by tryptophane fluorescence is shown in Figure 1a, where the oxidized and denatured protein was diluted 1:100 into renaturation buffer. There is a fast phase ($k_1 = 0.324 \pm 0.11$ min^{-1}) with a quenching of fluorescence,

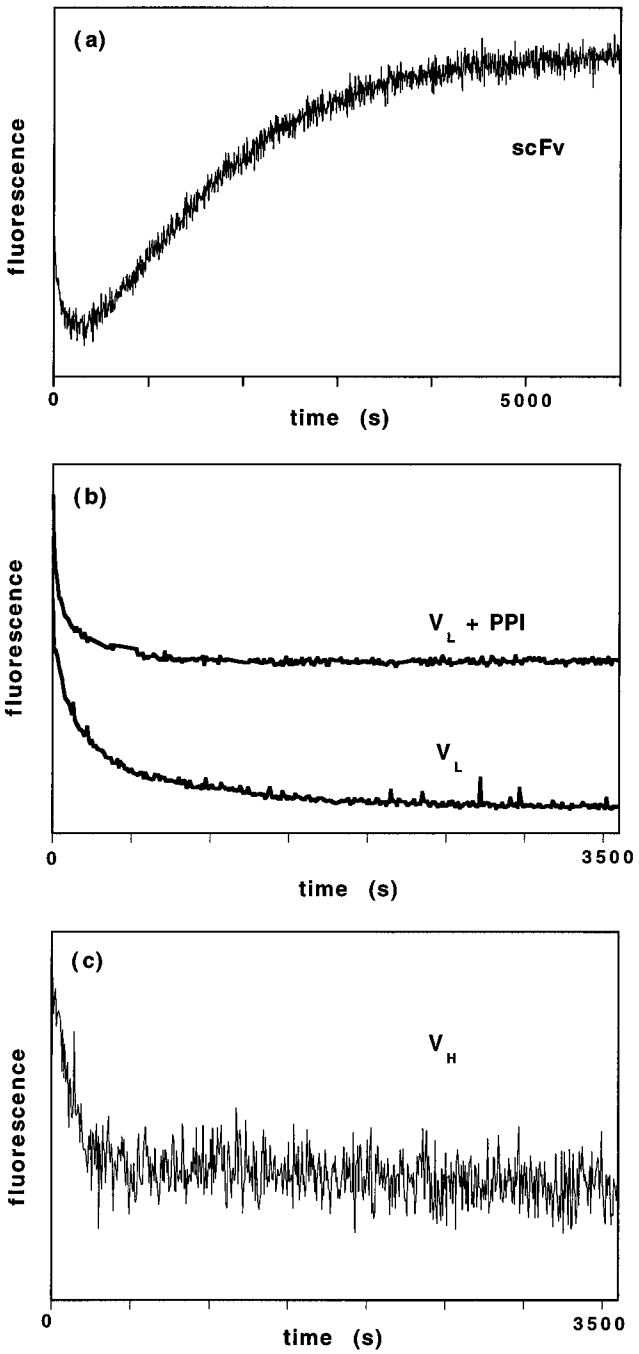

FIGURE 1: Fluorescence kinetics of the scFv fragment (a) as well as the isolated variable domains, V_L (b) and V_H (c), respectively. In panel b, the influence of *E. coli* peptidyl−prolyl *cis*−*trans* isomerase on the folding of the V_L domain is also shown (upper trace).

followed by a slow gain of fluorescence ($k_2 = 0.048 \pm 0.0022$ min^{-1}) at the wavelength of the emission maximum of the native scFv fragment (331.4 nm). These kinetics are very similar to the corresponding kinetics of the Fv fragment (Knappik & Plückthun, 1995), suggesting that the linker does not change the folding behavior of the variable domains (Freund et al., 1993). In contrast to the scFv fragment, the isolated V_L or V_H domains (Figure 1b,c) do not show such a slow phase with a fluorescence gain.

The overall folding of the isolated V_L domain can be fitted by a double exponential. Proline isomerization is the rate-limiting step (Figure 1b), since the addition of the enzyme peptidyl−prolyl *cis*−*trans* isomerase accelerates both kinetic phases (Figure 1b, upper trace).

For the isolated V_H domain, it is not clear whether the native state is reached after the initial loss of fluorescence (Figure 1c). Only a slight shift of the emission maximum from 348 nm in the fully denatured state to 345 nm is observed, while the emission maximum of the native scFv fragment is at 331.4 nm. At higher concentrations than the ones used in the fluorescence experiments (0.08 μM), the V_H domain tends to aggregate and all attempts to isolate the V_H domain in sufficient amounts to perform CD or NMR measurements under nondenaturing conditions were unsuccessful. It is therefore most likely that under renaturation conditions the isolated V_H domain collapses to a compact native-like state which needs the presence of the V_L domain and the interface provided by it to be stabilized and escape off-pathway aggregation. Since three of the tryptophans in V_H are localized at the interface between the domains, we suggest that the slow kinetic phase of the scFv fragment reflects the interaction of the domains and the formation of the native interface.

MS Analysis of H/D Exchange Experiments. In order to determine whether significant populations of intermediates occur along the slow folding pathway of the scFv fragment, we carried out an analysis of the H/D exchange experiments by mass spectrometry. Folding of the denatured, deuterated, and oxidized scFv fragment was initiated by dilution into D_2O buffer, pH 8.0, and allowed to proceed for various lengths of time. Then a further 5-fold dilution was performed into H_2O buffer (final pH 5.2) in order to allow refolding to be completed and H/D exchange of still unprotected amide protons to take place. As is shown in Figure 2a, the mass of the reference sample in H_2O is 27 089 ± 3 Da, which is in close agreement with the theoretical mass of 27 087 Da. Already at the first time point (30 s), an intermediate is formed with a mass of 27 107 ± 3 Da (Figure 2b). Taking into account a background of 20% deuteration of the amide protons protected late in the refolding reaction, this leaves 11 ± 3 amide protons becoming protected within the intermediate. During the slow phase, no further intermediate becomes significantly populated, as can be seen from Figure 2c. After 10 min of refolding, only the masses of the intermediate and the fully amide protected scFv fragment could be detected. The latter has a mass of 27 135 ± 3 Da (Figure 4d), meaning that 46 ± 3 protons become protected during the whole folding reaction.

NMR Results. In order to obtain structural information about the folding reaction, we performed H/D exchange experiments and subsequently analyzed these by NMR spectroscopy. The refolding reaction was performed in a manner similar to that used in the MS experiments, but uniformly ^{15}N-labeled scFv fragment was used for all experiments. Spectra were taken for samples in which the H/D exchange competition was started at 0 s, 30 s, 1 min, 2 min, 5 min, 10 min, 30 min, and 60 min after initiation of folding into D_2O buffer. The ^{15}N−^{1}H correlation spectra were taken in D_2O, so that only slowly exchanging protons were detected. The results for the 49 slowly exchanging protons analyzed are summarized in Table 1. As can be seen, a number of residues become amide protected to a significant degree after a refolding time of 30 s. There are a few residues with a large degree of amide protection (>40%) after 30 s of refolding (Leu L39, Ala L40, Trp L41, Tyr

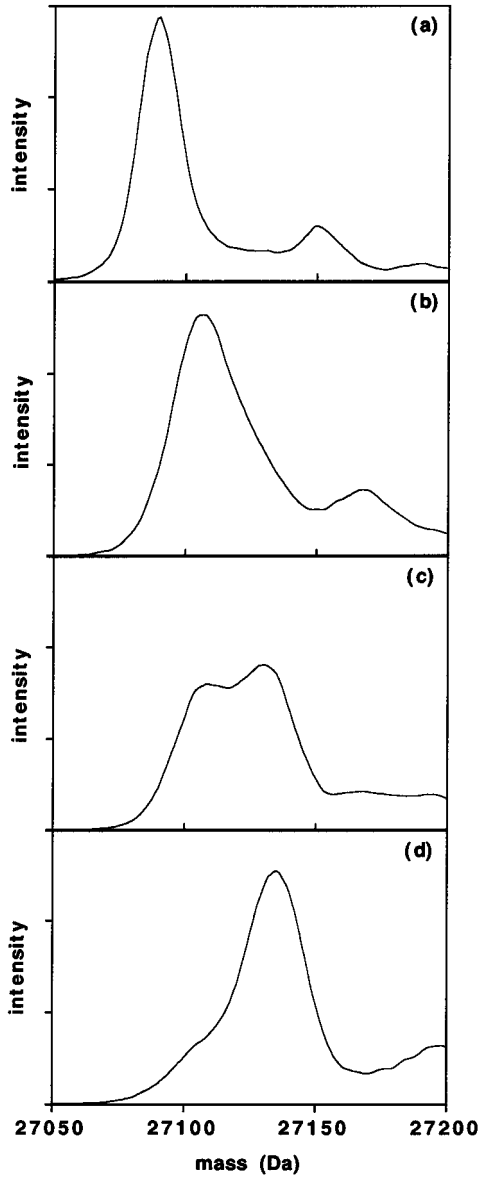

FIGURE 2: Results of the electron spray mass spectrometrical (ES/MS) analysis of the H/D experiments showing the reconstructed masses of the scFv fragment at different time points of refolding. In panel a the reference sample diluted directly into H_2O buffer is shown, in panel b the protein was allowed to refold for 10 s, in panel c for 10 min, and in panel d for 60 min in D_2O, before the dilution into H_2O buffer. The peak masses (in Da) were 27 089 ± 3 (a), 27 107 ± 3 (b), 27 108 ± 3, and 27 132 ± 3 (c), and 27 135 ± 3 (d).

L42, Ile L54, Ala L90, Val L91, Tyr L93, Ile H95) and a number of residues with still significant protection (25−35%: Gln L43, Ile L81, Gln H39, Gln H84, Tyr H82, Ala H94, Phe H109). Two more residues gave protections between 10 and 20% (Ile H71 and Val H118). All other residues analyzed showed no significant protection after 30 s of refolding. Representative curves of amide protection versus time are shown in Figure 3. The amide protection curves of residues with negligible protection after 30 s of refolding could be fitted quite well by a single exponential (Figure 3a). The corresponding curves of residues with significant protection after 30 s of refolding (>25%) could also be fitted by single exponentials in most cases if the first point (0 s) was left out from the analysis (Figure 3b−c). The significant protection of a number of amide protons suggests that these residues contribute to secondary structure stabilization in a major folding intermediate of the scFv fragment, which is in agreement with the results obtained by mass spectrometry.

Table 1: Protection of Individual Backbone Amide Protons after 30 s of Refolding and Rate Constants of the Slow Phase of Folding

residue[a]	location[b]	% protection within the intermediate	rate constant of slow phase (min^{-1})
Tyr L93[c]	inner β-sheet	95	
Ile H95[c]	inner β-sheet	76	
Val L91[d]	inner β-sheet	73	0.038 ± 0.024
Tyr L42[c]	inner β-sheet	67	
Ile L54[c]	inner β-sheet	66	
Ala L90[d]	inner β-sheet	56	0.055 ± 0.018
Trp L41[d]	inner β-sheet	48	0.06 ± 0.021
Leu L39[c]	inner β-sheet	46	
Ala L40[d]	inner β-sheet	43	0.085 ± 0.046
Tyr H82[d]	outer β-sheet	34	0.092 ± 0.044
Gln L43[d]	inner β-sheet	32	0.051 ± 0.018
Ile L81[d]	outer β-sheet	30	0.049 ± 0.026
Ala H94[d]	inner β-sheet	29	0.058 ± 0.019
Gln H39[d]	inner β-sheet	27	0.056 ± 0.010
Phe H109[c]	CDR3	27	
Gln H84[c]	outer β-sheet	25	
Ile H71[d]	outer β-sheet	16	0.10 ± 0.02
Val H118	last β-strand	13	0.069 ± 0.025
Ser L9	first β-strand	<10	0.06 ± 0.026
Leu L11	first β-strand	<10	0.05 ± 0.018
Val L13	first β-strand	<10	0.058 ± 0.012
Glu L17	outer β-sheet	<10	0.067 ± 0.011
Thr L20	outer β-sheet	<10	0.064 ± 0.012
Cys L23	outer β-sheet	<10	0.03 ± 0.011
Glu L61	CDR2	<10	0.066 ± 0.007
Phe L68	outer β-sheet	<10	0.056 ± 0.014
Thr L69	outer β-sheet	<10	0.059 ± 0.014
Thr L78	outer β-sheet	<10	0.044 ± 0.018
Val L84	outer β-sheet	<10	0.065 ± 0.009
Glu L85	loop	<10	0.053 ± 0.01
Ala L86	loop	<10	0.096 ± 0.022
Glu L87	loop	<10	0.053 ± 0.008
Lys L109	last β-strand	<10	0.044 ± 0.016
Leu L110	last β-strand	<10	0.066 ± 0.013
Val H5	first β-strand	<10	0.062 ± 0.011
Ser H21	outer β-sheet	<10	0.064 ± 0.011
Ile H48	inner β-sheet	<10	0.041 ± 0.022
Ala H49	inner β-sheet	<10	0.059 ± 0.019
Ser H51	inner β-sheet	<10	0.052 ± 0.016
Ser H63	CDR2	<10	0.052 ± 0.012
Ala H64	CDR2	<10	0.03 ± 0.018
Arg H69	CDR2	<10	0.055 ± 0.011
Phe H70	outer β-sheet	<10	0.057 ± 0.01
Ser H73	outer β-sheet	<10	0.049 ± 0.011
Asp H75	outer β-sheet	<10	0.058 ± 0.007
Arg H89	loop	<10	0.068 ± 0.012
Asp H92	loop	<10	0.041 ± 0.012
Thr H119	last β-strand	<10	0.066 ± 0.011
Val H120	last β-strand	<10	0.066 ± 0.01

[a] Residue numbers are given according to the consecutive numbering of the PDB file 2MCP. Residues of the V_L domain are marked by an uppercase L, residues of the V_H domain by an uppercase H. [b] Locations within the secondary structure elements are given: the inner β-sheet comprises strands c, c′, c″, f, and g, the outer β-sheets strands a, b, d, and e of each domain (see also Figure 4). The first and last b-strands (strands a and g) are listed separately, since they "connect" the two sheets of each domain by backbone hydrogen bonds. [c] Residues with significant protection after 30 s of folding which could not be fitted properly, since signals were getting to weak in all spectra except the 0 s spectrum. [d] Residues for which the first point was left out for the exponential fitting.

Structure of the Intermediate. The localization of the highly protected and moderately protected amide protons in the intermediate is shown in Figure 4. As can be seen from the chain topology, most of the amide protons which become

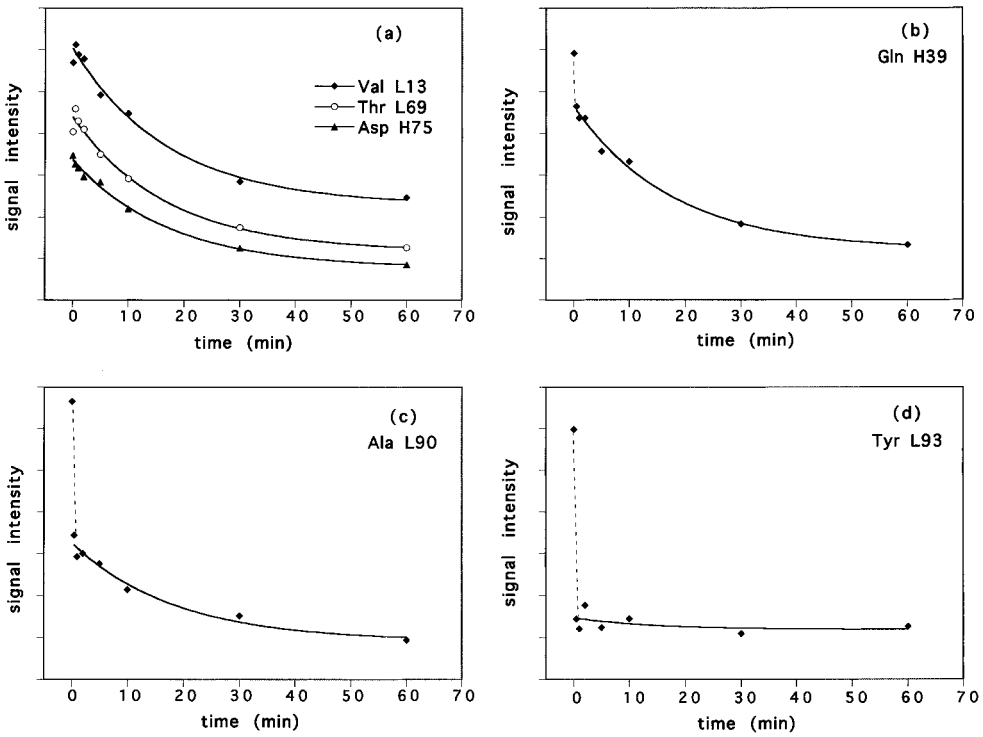

FIGURE 3: Results of the H/D exchange experiments analyzed by NMR. The signal intensities of the respective NMR peaks are plotted versus the time the protein was allowed to refold in D_2O buffer before the second dilution into H_2O buffer was performed. In panel a, three representative residues with no significant protection in the intermediate are shown. In panel b a residue with moderate protection (Gln H39), in panel c a residue with significant protection, and in panel d a residue with high protection in the intermediate is displayed. Single-exponential fits are drawn as solid lines in the respective graphs.

highly protected in the intermediate belong to hydrophobic and aromatic amino acids within the β-strands c, c′, and f of the V_L domain. The side chains of these residues are either part of the buried core of the V_L domain (Leu L39, Trp L41, Ile L54, Ala L90) or point into the V_L-V_H interface (Ala L40, Tyr L42, Val L91, Tyr L93). One additional residue (Gln L43), adjacent to a stretch of highly protected amides, is protected somewhat less (32%). Most of the highly protected amides belong to residues within the V_L domain which have mutual hydrogen bonds in the native protein (Figure 4), suggesting the formation of native-like β-sheet structure in the intermediate. In the V_H domain, only Ile H95 becomes amide protected to a large extent in the intermediate. This residue is topologically at the same position as Val L91, which also becomes highly protected in the intermediate. Figure 5 shows the localization of the residues in the native structure which become highly protected within the intermediate.

Only one residue outside the inner β-sheet of the V_L domain, Ile L81, showed a moderate level of protection in the intermediate. In the case of the V_H domain, two residues with moderate protection in the intermediate were observed in the outer β-sheet (strand e), indicating a potential weak nucleus in the outer β-sheets of the two variable domains which is locally restricted and less stabilized than the inner β-sheet V_L nucleus. Two of the three other residues of the V_H domain which become moderately, but significantly protected in the intermediate, Gln H39 and Ala H94, are part of the inner β-sheet of the V_H domain and are localized near Ile H95, the only residue of the V_H domain with large protection in the intermediate. They are located at equivalent positions to those of Gln L43 and Val L91 of the inner β-sheet of V_L. Phe H109, a CDR 3 residue at the interface of the two domains, is moderately protected in the intermediate.

Slow Folding Reaction. The slow conversion of the intermediate to the native state is encompassed by the protection of most of the residues in the outer β-sheets of both variable domains (Figure 4 and Table 1). In V_L, the only residues of the inner β-sheet which become exclusively protected during this slow phase (Lys L109, Leu L110) are located within the last β-strand (Figure 4 and Table 1). This strand makes backbone hydrogen bonds to the first strand of the outer β-sheet and thereby closes the β-barrel like structure of immunoglobulin domains. In V_H, most residues of both β-sheets become protected only during this slow phase of folding, indicating the formation of stable, native secondary structure within the sheets as well as the adjustment of the sheets (protection of residues within strands a and g, Figure 4) to take place.

DISCUSSION

Cooperativity of Folding. The slow fluorescence phase of the scFv fragment and the amide protection curves of most slowly exchanging amino acids of both variable domains can be fitted by a single exponential with similar rate constants. This is in agreement with the results of the MS analysis of the H/D experiments and clearly demonstrates the cooperativity of the folding of the two domains from the structured intermediate to the native state. Whether this cooperativity is due to a slow rearrangement involving the formation of the native interface or reflects an intrinsically fast reaction in the scFv fragment, which is, however, dependent on the slow proline *cis*−*trans* isomerization within the V_L domain, remains to be shown and is currently under investigation (M. Jäger, unpublished experiments).

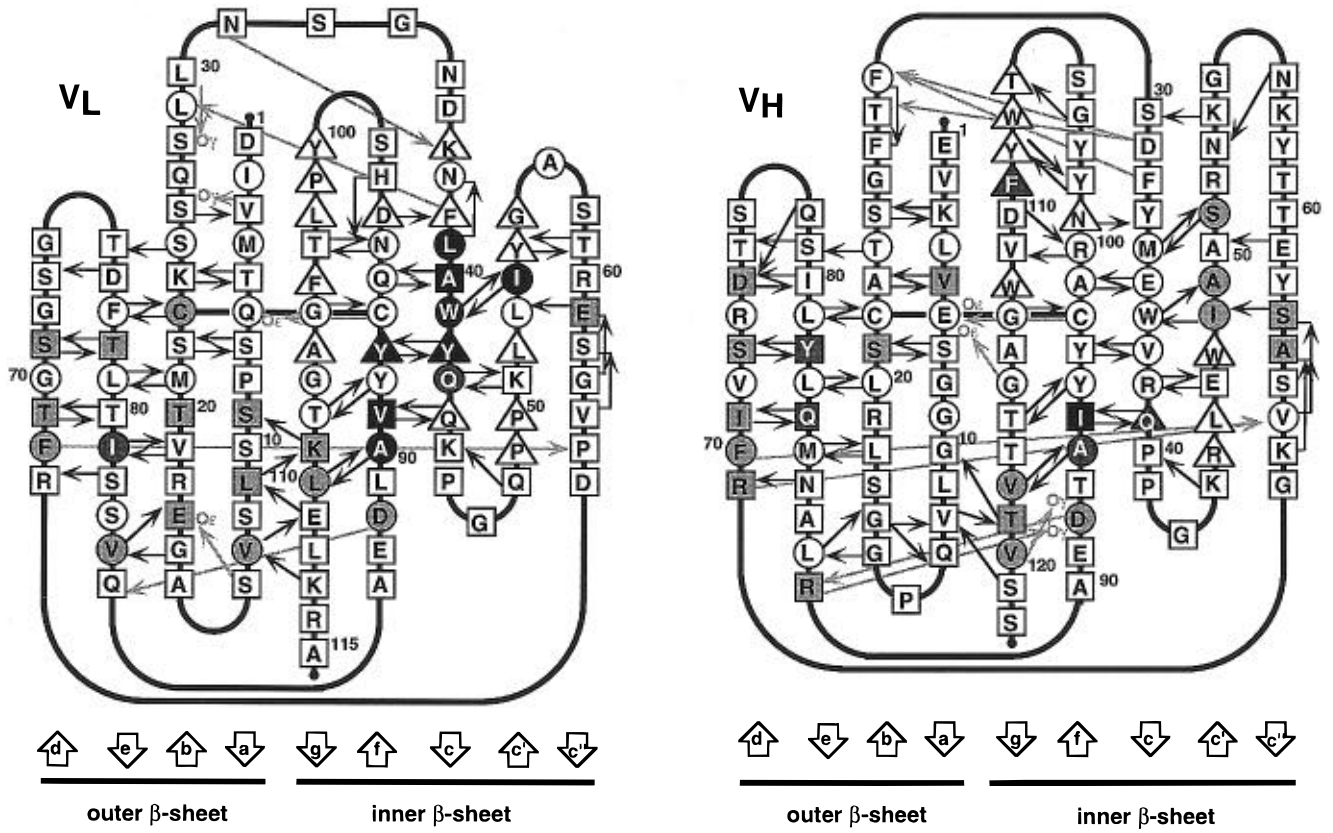

FIGURE 4: Chain topology of the V_L and the V_H domain. Residues are divided into three classes by geometric forms: Rectangles define residues with surface-exposed side chains, triangles represent residues with side chains becoming buried upon V_L-V_H association, and circles define residues with side chains with less than 10% accessible surface in the isolated V_L or the isolated V_H domain. The residues with amide protons which become highly protected within the intermediate (>40%) are shown in black and the amide protons with still significant protection (<25−35%) are colored dark gray. Residues with no significant protection in the intermediate are shown in light gray, and residues which exchanged too fast to be used for an analysis are colored white. Hydrogen bonds are displayed as arrows (NH → OC), with side chain oxygen acceptors labeled. β-strands are labeled consecutively from the N- to the C-terminus, and residues are numbered according to the consecutive numbering of the PDB file 2MCP.

FIGURE 5: Location of the residues which become highly protected within the intermediate in the native structure. The V_L domain ribbon is displayed in white, the V_H domain ribbon in light gray. Protected residues in the V_L domain are shown in dark gray, the protected Ile 95 in the V_H domain in black. In panel a the molecule is viewed from the antigen binding side downward, showing the interface as a tunnel in the middle of the two domains. In panel b the molecule is rotated 90° around the *x* axis, so the molecule is viewed from the side. As one can see, the side chains of the highly protected residues either point toward the core of the V_L domain or point into the domain interface.

Structure of the Intermediate and Folding Pathway. The structure formed in the intermediate has one contiguous region of amide protected residues which are localized in the inner β-sheet of the V_L domain (Figure 5). Many of the hydrophobic and aromatic residues of this region belong to the structurally highly conserved residues of immunoglobulin

variable domains (A. Honegger, unpublished results), indicating that the formation and presence of the observed folding intermediate might be of a more general nature. The V_H domain, which does not show such an extended nucleus in the intermediate, fails to fold into a stable, independent folding unit. However, there is no obvious structural reason for the observed differences of stabilized secondary structure in the intermediate as well as the native states of the two variable domains. The absence of some of the stable hydrogen bonds in the inner β-sheet of the V_H domain within the scFv fragment (residues 35−38) as compared to the V_L domain may either be due to some local flexibility of this region or to local destabilization resulting in low protection factors of these amide protons. This instability may already be reflected within the intermediate, and thus no contiguous amide-protected hydrophobic core within the V_H domain of the scFv fragment forms early in the folding reaction.

The isolated V_L domain does fold into a stable domain, which might at least partially be due to its potential to form homodimers. In agreement with this hypothesis, the V_L domain of the antibody McPC603 has been shown to crystallize as a dimer (Steipe et al., 1992), and NMR line widths of the isolated domain also suggest dimer formation (Freund et al., 1994). The formation of an intermediate with stabilized secondary structure, which seems to have a structure similar to the V_L domain within the scFv fragment, has been observed for the isolated V_L domain as well (Freund et al., unpublished results). This is consistent with our view that formation of an intermediate with stabilized secondary structure is a necessary step in the folding of isolated antibody variable domains.

Whether interface formation has taken place in the folding intermediate of the scFv fragment could not directly be observed in our experiments. Nevertheless, the large amide protection factors of all interface residues analyzed within the intermediate (Table 1 and Figure 4) indicates the formation of native-like structures at the interface and therefore suggests some mode of interaction of the two domains within the intermediate.

The slow folding kinetics of the McPC603 scFv fragment has been shown to proceed via an intermediate in a sequential manner, in accordance with the classical framework model (Kim & Baldwin, 1982). This intermediate is already formed within the first milliseconds of folding (C. Freund, unpublished observations) and therefore probably indicates intrinsic elements of early secondary structure formation within the individual domains. The protection patterns map four residues with high and moderate amide protection at equivalent postions of the inner β-sheet of the two variable domains (Val L91 and Ile H95, Ala L90 and Ala H94, see Table 1 and Figure 4), which might indicate the potential to form folding nuclei in these regions of the inner β-sheet to be a conserved feature of antibody variable domains.

It has been suggested that the folding of all-β-sheet proteins may be fundamentally different from that of other structural motifs (Dyson et al., 1992a,b), and a theoretical study by Finkelstein (1991) suggested that β-sheet structure formation is slow unless it is significantly stabilized. Two other β-sheet proteins have been studied to date by NMR analysis of H/D exchange experiments. For apoplastocyanin (Koide et al., 1993), proline isomerization was found to be rate determining for the folding reaction. In the case of interleukin 1β (Valery et al., 1993), however, the slow protection of a number of amide protons was attributed to the stabilization of native secondary structure and the final tight packing of side chains. This is in agreement with our results for the scFv fragment, where a hydrophobic core forms quite rapidly but the final stabilization of the structure is a slow event. Proline isomerization was found to be rate-determining in the case of the isolated V_L domain, and an exchange protected intermediate could be trapped (Freund et al., unpublished results). The isolated V_L domain thus behaves in a manner similar to the apoplastocyanin molecule, which also is a member of the "simple" greek key proteins (Richardson, 1981). In contrast, the V_H domain needs to form the native interface in order to become sufficiently stabilized. The V_L domain presumably prevents aggregation of exposed hydrophobic patches within an early V_H intermediate (Knappik & Plückthun, 1995).

It remains to be determined whether the elucidated folding mechanism is also valid *in vivo*. In the case of antibody expression in *E. coli*, it is well known that V_H domains often tend to form insoluble aggregates in the periplasm (Plückthun et al., 1987; Power et al., 1992). Since no general periplasmic chaperone has been characterized to date (Wülfing & Plückthun, 1994), it is probable that the V_L domain also assists in the folding of the V_H domain in *E. coli*. In the eukaryotic cell, this function may be taken over by chaperones such as the heavy chain binding protein BIP and the hsp90 protein Grp94 (Melnick & Argon, 1994), the existence of which may have reduced the need for V_H domains to be selected on the basis of their folding properties as single domains.

ACKNOWLEDGMENT

We thank Antonio Baici and Marcus Jäger for helpful discussions.

REFERENCES

Alzari, P. M., Lascombe, M. B., & Poljak, R. J. (1986) *Annu. Rev. Immunol. 6*, 555−580.

Bird, R. E., Hardman, F. D., Jacobson, J. W., Johnson, S., Kaufman, B. M., Lee, S. M., Lee, T., Pope, S. H., Riordan, G. S., & Whitlow, M. (1988) *Science 242*, 423−426.

Bork, P., Holm, L., & Sander, C. (1994) *J. Mol. Biol. 242*, 309−320.

Briggs, M. S., & Roder, H. (1992) *Proc. Natl. Acad. Sci. U.S.A. 89*, 2017−2021.

Bycroft, M., Matouschek, A., Kellis, J. T., Serrano, L., & Fersht, A. R. (1990) *Nature 346*, 488−490.

Cieslar, C., Ross, A., Zink, T., & Holak, T. A. (1993) *J. Magn. Reson. B101*, 97−101.

Davies, D. R., & Padlan, E. A. (1990) *Annu. Rev. Biochem. 59*, 439−473.

Dyson, H. J., Merutka, G., Waltho, J. P., Lerner, R. A., & Wright, P. E. (1992a) *J. Mol. Biol. 226*, 795−817.

Dyson, H. J., Sayre, J. R., Merutka, G., Shin, H.-C., Lerner, R. A., & Wright, P. E. (1992b) *J. Mol. Biol. 226*, 819−835.

Finkelstein, A. V. (1991) *Proteins: Struct., Funct., Genet. 9*,23−27.

Freund, C., Ross, A., Guth, B., Plückthun, A., & Holak, T. A. (1993) *FEBS Lett. 320*, 97−100.

Freund, C., Ross, A., Plückthun, A., & Holak, T. A. (1994) *Biochemistry 33*, 3296−3303.

Goto, Y., & Hamaguchi, K. (1982) *J. Mol Biol. 156*, 891−910.

Goto, Y., Azuma, T., & Hamaguchi, K. (1979) *J. Biochem. (Tokyo) 85*, 1427−1438.

Hochman, J., Gavish, M., Inbar, D., & Givol, D. (1976) *Biochemistry 15*, 2706−2710.

Huston, J. S., Levinson, D., Mudgett-Hunter, M., Tai, M.-S., Novotny, J., Margolies, M. N., Ridge, R. J., Bruccoleri, R. E., Haber, E., Crea, R., & Oppermann, H. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 5879−5883.
Jacobs, M. D., & Fox, R. O. (1994) *Proc. Natl. Acad. Sci. U.S.A. 91*, 449−453.
Jennings, P. A., & Wright, P. E. (1993) *Science 262*, 892−896.
Jones B. E., & Matthews, C. R. (1995) *Protein Sci. 4*, 167−177.
Kim, P. S., & Baldwin, R. L. (1982) *Annu. Rev. Biochem. 51*, 459−489.
Knappik, A., & Plückthun, A. (1995) *Protein Eng. 8*, 81−89.
Koide, S., Dyson, H. J., & Wright, P. E. (1993) *Biochemistry 32*, 12299−12310.
Lang, K., & Schmid, F. X. (1988) *Nature 331*, 453−455.
Lilie, H., Lang, K., Rudolph, R., & Buchner, J. (1993) *Protein Sci. 2*, 1490−1496.
Liu, J., & Walsh, T. (1990) *Proc. Natl. Acad. Sci. U.S.A. 87*, 4028−4032.
Lu, J., & Dahlquist, F. W. (1992) *Biochemistry 31*, 4749−4756.
Marion, D., & Wüthrich, K. (1983) *Biochem. Biophys. Res. Commun. 113*, 967−974.
Melnick, J., & Argon, Y. (1994) *Nature 370*, 373−375.
Miranker, A., Robinson, C. V., Radford, S. E., Aplin, R. T., & Dobson, C. M. (1993) *Science 262*, 896−899 .
Mori, S., Abeygunawardana, C., Johnson, M. O., & van Zijl, P. C. (1995) *J. Magn. Reson. B 108*, 94−98.
Mullins, S. L., Pace, C. N., & Raushel, F. M. (1993) *Biochemistry 32*, 6152−6156 (1993).
Pantoliano, M. W., Bird, R. E., Johnson, S., Asel, E. D., Dodd, S. W., Wood, J. F., & Hardman, K. D. (1991) *Biochemistry 30*, 10117−10125.
Plückthun, A., Glockshuber, R., Pfitzinger, I., Skerra, A., & Stadlmüller, J. (1987) *Cold Spring Harbor Symp. Quant. Biol. 52*, 105−111.
Power, B. E., Ivancic, N., Harley, V. R., Webster, R. G., Kortt, A. A., Irving, R. A., & Hudson, P. J. (1992) *Gene 113*, 95−99.
Radford, S. E., Dobson, C. M., & Evans, P. A. (1992) *Nature 358*, 302−307.
Richardson, J. S. (1981) *Adv. Protein Chem. 34*, 176−339.
Riechmann, L., Foote, J., & Winter, G. (1988) *J. Mol. Biol. 203*, 825−828.
Roder, H., Elöve, G. A., & Englander, S. W. (1988) *Nature 335*, 700−704.
Searle, S. J., Pedersen, J. T., Henry, A. H., Webster, D. M., & Rees, A. R. (1995) in *Antibody Engineering* (Borrebaeck, C. A. K., Ed.) pp 3−51, Oxford University Press, New York, Oxford.
Shaka, A. J., Barker, P. B., & Freeman, R. (1985) *J. Magn. Reson. 64*, 547−552.
Skerra, A., & Plückthun, A. (1988) *Science 240*, 1038−1041.
Sklenar, V., Peterson, R. D., Rejante, M. R., & Feigon, J. (1993) *J. Biomol. NMR 3*, 721−727.
Smith, R. D., Loo, J. A., Edmonds, C. G., Barinoga, C. J., & Udseth, H. R. (1990) *Anal. Chem. 62*, 882−889.
Tsunenaga, M., Goto, Y., Kawata, Y., & Hamaguchi, K. (1987) *Biochemistry 26*, 6044−6051.
Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694−699.
Valery, P., Gronenborn, A. M., Christensen, H., Wingfield, P. T., Pain, R., & Clore, G. M. (1993) *Science 260*, 1110−1113.
Wülfing, C., & Plückthun, A. (1994) *Mol. Microbiol. 12*, 685−692.

BI952764A