Protein Science (1999), *8*:1571–1591. Cambridge University Press. Printed in the USA.

REVIEW

The hydrogen exchange core and protein folding

RENHAO LI AND CLARE WOODWARD

Department of Biochemistry, University of Minnesota, St. Paul, Minnesota 55108

(RECEIVED September 22, 1998; ACCEPTED May 10, 1999)

Abstract

A database of hydrogen-deuterium exchange results has been compiled for proteins for which there are published rates of out-exchange in the native state, protection against exchange during folding, and out-exchange in partially folded forms. The question of whether the slow exchange core is the folding core (Woodward C, 1993, *Trends Biochem Sci 18*:359–360) is reexamined in a detailed comparison of the specific amide protons (NHs) and the elements of secondary structure on which they are located. For each pulsed exchange or competition experiment, probe NHs are shown explicitly; the large number and broad distribution of probe NHs support the validity of comparing out-exchange with pulsed-exchange/competition experiments. There is a strong tendency for the same elements of secondary structure to carry NHs most protected in the native state, NHs first protected during folding, and NHs most protected in partially folded species. There is not a one-to-one correspondence of individual NHs. Proteins for which there are published data for native state out-exchange and ϕ values are also reviewed. The elements of secondary structure containing the slowest exchanging NHs in native proteins tend to contain side chains with high ϕ values or be connected to a turn/loop with high ϕ values. A definition for a protein core is proposed, and the implications for protein folding are discussed. Apparently, during folding *and* in the native state, nonlocal interactions between core sequences are favored more than other possible nonlocal interactions. Other studies of partially folded bovine pancreatic trypsin inhibitor (Barbar E, Barany G, Woodward C, 1995, *Biochemistry 34*:11423–11434; Barber E, Hare M, Daragan V, Barany G, Woodward C, 1998, *Biochemistry 37*:7822–7833), suggest that developing cores have site-specific energy barriers between microstates, one disordered, and the other(s) more ordered.

Keywords: hydrogen exchange; NMR; protein folding; slow exchange core

In native proteins, the group of backbone amide hydrogens slowest to out-exchange by a folded state mechanism tend to cluster in mutually packed elements of secondary structure; these define a submolecular domain we call the slow exchange core. Slow exchange core elements usually contain NHs protected from exchange early during folding; apparently, the regions of the protein most resistant to exchange in the native state are also the regions most likely to favor organized structure early in folding. We proposed that "the slow exchange core is the folding core" of proteins (Kim et al., 1993; Woodward, 1993), and noted that if this is general, a number of significant implications follow. Our original suggestion was based on the few cases available at the time; since

Reprint requests to: Clare Woodward, Department of Biochemistry, University of Minnesota, 1479 Gortner Avenue, St. Paul, Minnesota 55108; e-mail: clare@biosci.cbs.umn.edu.

Abbreviations: k_{obs}, observed exchange rate constant; k_N, rate constant for exchange by the folded state mechanism; k_D, rate constant for exchange by the unfolding mechanism; k_{cx}, rate constant for exchange from a random conformation peptide (unprotected exchange); k_{calc}, rates computed by the method of Bai et al. (1993); k_u and k_f, rate constants for cooperative unfolding and folding; K_{eq}, equilibrium constant for cooperative global denaturation; β, probability that an NH is available for exchange; EX2/EX1, limiting cases for a model for hydrogen exchange; k_1, k_2, and k_{exg}, rate constants for the EX2/EX1 model; $\Delta G(HX)$, free energy difference between native and denatured states determined from the ration of k_{obs}/k_{calc}; $\Delta\Delta G(HX)$, difference between ΔG for wild-type vs. mutant determined from the ratio of $k_{obs,wt}/k_{obs,mutant}$; N, native state; PF, partially folded state; P/C, pulsed exchange or competition experiment; p/c pr., probe NHs in a pulsed exchange or competition experiment; P_f and P_d, more ordered and more disordered microstates of one NH detected as separate cross peaks in NMR experiments; ACBP, acyl coenzyme A binding protein; α-LA, α-lactalbumin; apoMb, apomyoglobin; barnase, *Bacillus amyloliquefaciens* ribonuclease; BPTI, bovine pancreatic trypsin inhibitor; CI2, chymotrypsin inhibitor 2; cyt *c*, cytochrome *c*; DHFR, dihydrofolate reductase from *Escherichia coli*; GB1, B1 immunoglobulin-binding domain of streptococcal protein G; GdmCl, guanidinium chloride; HEWL, hen egg-white lysozyme; IL-1β, interleukin-1β subunit; LB1, B1 immunoglobulin-binding domain of peptostreptococcal protein L; NH, amide proton; NMR, nuclear magnetic resonance; OMTKY3, turkey ovomucoid third domain; pdTp, thymidine 3′,5′-bisphosphate; PDB, Protein Data Bank; RNase A, ribonuclease A; RNase H, ribonuclease H; RNase T1, ribonuclease T1; scFv fragment, single-chain antibody fragment composed of V_H and V_L domains; SNase, staphylococcal nuclease; TCA, trichloroacetate; TFE, 2,2,2-trifluoroethanol; V_H domain, antibody variable heavy domain; V_L domain, antibody variable light domain.

then the relevant literature has increased considerably. We have compiled exchange data for all proteins with published rates for native state out-exchange and exchange protection during folding. Out-exchange rates in partially folded species are available for some proteins. Here, we analyze the database, discuss implications of the results, and re-evaluate earlier conclusions. The validity of comparing out-exchange rates to protection rates is discussed. Also reviewed are the proteins that permit a comparison of slow exchange core elements and high ϕ values. We begin with a background discussion of issues in analyses of hydrogen exchange.

Discussion

Hydrogen exchange in proteins

Out-exchange from native or partially folded states

Hydrogen isotope exchange provides a unique probe of protein dynamics. The native state ensemble fluctuates about an average approximated by the crystal structure, and on many time scales samples numerous additional conformers, many near the average in structure and energy, and some rare and far from the average in structure and/or energy. Hydrogen isotope exchange reports a subset of protein internal fluctuations. Intramolecularly H-bonded amides with no solvent contact in the crystal structure, nevertheless, undergo solvent isotope exchange with rate constants that range from those approaching k_{cx}, the rate constant of an analogous NH in a random conformation peptide, to those that are smaller than k_{cx} by many orders of magnitude. Exchange rates are sometimes expressed as protection factors, equal to the ratio k_{calc}/k_{obs}, where k_{obs} is the observed rate constant and k_{calc} is the exchange rate constant for an NH in a small peptide of equivalent sequence computed from empirical, nearest neighbor rules (Bai et al., 1993). The most rapidly exchanging amides are on the surface; however, some surface protons exchange with rate constants that are an order of magnitude less than k_{calc} (Tüchsen & Woodward, 1985), demonstrating that NHs may be "protected" from free exchange even when accessible to solvent and not intramolecularly H-bonded in the crystal structure.

Exchange occurs by two processes, a folded state mechanism and a global unfolding mechanism, as diagramed in Figure 1A. Each NH may exchange by the first mechanism under one set of conditions, and by the second under another. Likewise, in one protein under a single set of conditions, some NHs may exchange by one mechanism while others exchange by the second, or a mixture of both. The two mechanisms are distinguished by their temperature dependence. Failure to take two-process exchange into account has led to incorrect attribution of exchange characteristics of one regime to motions of the other regime (e.g., exchange by the unfolding mechanism incorrectly ascribed to folded state motions). Tritium exchange experiments led to the first proposal that some NHs exchange from the folded state while others exchange by an unfolding mechanism (Rosenberg & Chakravarti, 1968; Woodward & Rosenberg, 1971b). NMR experiments showed that one NH may exchange by either of the two mechanisms depending on solvent conditions (Woodward & Hilton, 1980; Woodward et al., 1982).

The unfolding mechanism is better understood than the folded state mechanism, since global cooperative denaturation is extensively characterized. The folded state mechanism is a subject of literature debate, and alternative models have been offered. Exchange of buried amides, however slow, implies that native state fluctuations populate conformations in which the exchanging NH reacts with H_2O and catalyst ions (H^+ or OH^-). In molecular dynamics simulations of native proteins, intramolecular H-bonds in secondary structure commonly break and reform noncooperatively on the psec time scale, and their breakage/reformation is not necessarily rate limiting once solvent gains access to buried regions. Three models for the folded state mechanism are commonly discussed. In the "penetration" model, multiple, small, noncooperative, internal fluctuations create ensembles of interconverting conformations of varying protection, and provide transient access of solvent to buried NHs (references in Woodward et al., 1982). In the "local unfolding" model, a helix undergoes cooperative breakage of H-bonds and exchangeable species have locally unfolded

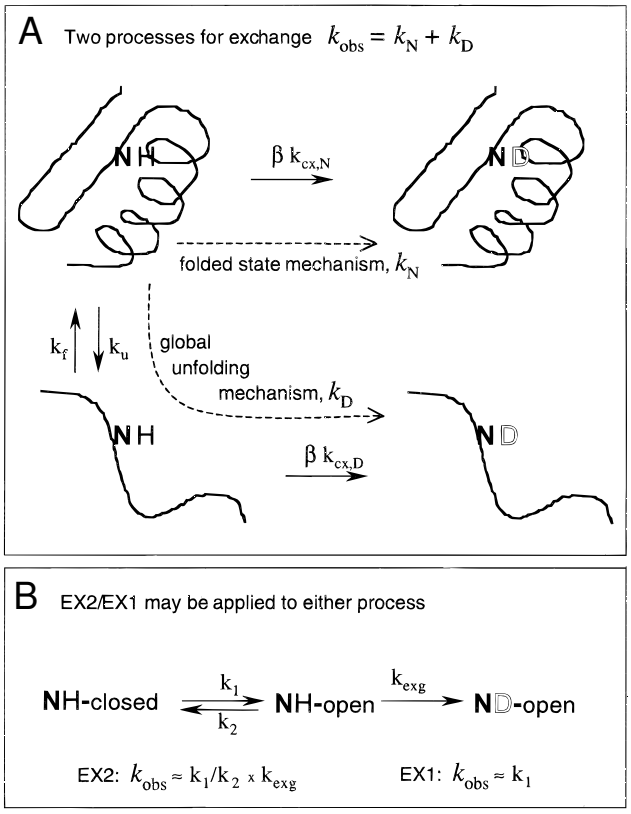

Fig. 1. Out-exchange of one NH from the native state. **A:** The two process model (Woodward et al., 1982). Under native conditions, for each NH, the observed rate constant k_{obs} for exchange is the sum of exchange rate constants for the folded state mechanism k_N and the global unfolding mechanism k_D. For the folded state mechanism, the rate constant for the chemical step is $k_{cx,N}$, and for the global unfolding mechanism, the rate constant for the chemical step is $k_{cx,D}$. β expresses the probability of contact of the exchanging NH with water and catalyst (the conformational part). For the unfolding mechanism $\beta = 1$ when the unfolded state is sufficiently disordered to make no conformational contribution to the exchange rate; $\beta < 1$ when the unfolded state has "residual" structure that slows exchange. "Protection" from exchange can arise from conformational effects embodied in β, or from chemical effects on $k_{cx,N}$ or $k_{cx,D}$ (e.g., local electrostatic field, strained geometry, or other). Global unfolding and refolding rate constants are k_u and k_f. The same two-process model may be used to analyze exchange from a partially folded protein. **B:** The EX2/EX1 formalism (Hvidt & Nielsen, 1966). A closed form, N-closed, is in equilibrium with an open form, N-open, with interconversion rate constants k_1 and k_2; k_{exg} is the rate constant for exchange from the open form.

secondary structure (Englander & Kallenbach, 1984). In a third model, exchange occurs from first-excited states that can have a slightly higher free energy, yet very different conformations than the native state, and could be produced by the types of protein motions invoked in both penetration and local unfolding mechanisms (Miller & Dill, 1995). No experiments thus far allow us to unambiguously eliminate penetration or local unfolding as an explanation of folded state exchange. Experiments that might distinguish between penetration vs. local unfolding have been suggested, e.g., periodicity of exchange rates in amphiphilic secondary structure, exchange rates of the same NH in solution and in crystals, the pH dependence of histidine C-2 exchange, and the temperature dependence of adjacent NHs (Hilton et al., 1981; Woodward et al., 1982). Regarding the first, a penetration model in which exchangeable species are on average approximated by the crystal structure predicts that NHs in an amphiphilic helix on the interior side will exchange slower than on the exterior, and exchange rates will show an $i, i + 3$ or 4 periodicity (similarly β-sheet can show $i, i + 1$ periodicity). Since then, exchange rates for a number of helices have been reported. Some show clear exchange periodicity, and some do not; in general, amphiphilic helices with slower exchanging NHs show the most prominent periodicity. It is possible that NHs in less mobile interior regions exchange by a penetration mechanism, while NHs in more flexible regions exchange via larger structural deformations.

The EX2/EX1 analysis developed by Linderstrøm-Lang, Hvidt, and associates (Hvidt & Nielsen, 1966) posits a pre-equilibrium between "closed" and "open" conformations of the same NH, as shown in Figure 1B. The equilibrium favors the closed form, and exchange takes place only from the open form. In the EX2 limit where $k_2 \gg k_1 + k_{exg}$, the observed exchange rate constant k_{obs} is $\approx (k_1/k_2) \cdot k_{exg}$; in the EX1 limit where $k_2 \ll k_{exg}$, $k_{obs} \approx k_1$. In the EX2/EX1 model, N-open is fully exposed to solvent and k_{exg} is approximated by the exchange rate constant of a model peptide (taken at the time to be polyalanine), and the open/close equilibrium is pH-independent. An EX2 mechanism accounts for the observation that most native state exchange rate constants have the same pH dependence as model peptides, since the observed rate constant is proportional to k_{exg}, the term in which the pH dependence resides. In present day applications of the EX2 formalism, the value of k_{exg} is usually taken as equal to k_{calc}, the exchange rate constant computed for a small peptide with the same neighboring amino acids (Bai et al., 1993).

EX2/EX1 analyses are often made of both exchange by the folded state mechanism and exchange by the unfolded mechanism; only in the latter case is it clearly warranted. Application of EX2/EX1 formalism to the global unfolding mechanism is straightforward since there are independent determinations of the unfolding and folding rate constants k_u and k_f, as well as the equilibrium constant for unfolding/folding $K_{eq} = k_u/k_f$. In this case, "NH-closed" is the native state, and "NH-open" is the globally denatured state pictured in Figure 1A, and $k_1 = k_u$, $k_2 = k_f$. If in the denatured state the protein in the vicinity of the exchanging NH is sufficiently disordered, then $\beta = 1$ and $\beta k_{cx,D} \approx k_{calc}$, $k_{obs} \approx K_{eq} \cdot k_{calc}$ and $\Delta G(HX) \approx -RT \ln(k_{obs}/k_{calc})$. The global unfolding mechanism, with or without residual structure in the denatured state, is characterized for a number of proteins; most often the mechanism is EX2, but EX1 behavior is also reported (e.g., Woodward & Hilton, 1980; Clarke & Fersht, 1996; Arrington & Robertson, 1997). For NHs exchanging by the unfolding mechanism, the temperature and denaturant dependence of K_{eq} explain the comparable dependences of k_{obs} (e.g., Hilton et al., 1981; Kim & Woodward, 1993; Orban et al., 1995; Swint-Kruse & Robertson, 1996). For mutant proteins altered in stability, the ratio of k_{obs} (unfolding mechanism, wild-type) to k_{obs}(unfolding mechanism, mutant) yields $\Delta\Delta G(HX)$, the difference between mutant and wild-type values for ΔG for global denaturation; $\Delta\Delta G(HX)$ is in agreement with $\Delta\Delta G$ determined by other methods (Kim et al., 1993 and references therein).

$\Delta G(HX)$ computed from rate constants for the unfolding mechanism is often, but not always, the same as ΔG for global denaturation determined by methods such as calorimetry or chemical denaturation. Assuming that the latter are in fact correct, the simplest explanation of why, for the unfolding mechanism, $\Delta\Delta G(HX)$ but not $\Delta G(HX)$ may agree with comparable global denaturation values from other methods is that $\beta k_{cx,D} \neq k_{calc}$. If, for example, the denatured state retains nonrandom structure around the exchanging NH or has large hydrophobic side chains near the exchanging NH, then $\beta < 1$, and $\Delta G(HX)$ computed as $-RT \ln(k_{obs}/k_{calc})$ does not give the free energy change for global unfolding. However, even when $\beta k_{cx,D} \neq k_{calc}$, the correct value of $\Delta\Delta G$ for wild-type vs. mutant will be calculated from the ratio of observed rate constants for the unfolding mechanism, because no value for k_{calc} is used; in the relationship $\Delta\Delta G(HX) = -RT(k_{obs,WT}/k_{obs,mutant})$ the $k_{cx,D}$ terms cancel out (Kim et al., 1993 and references therein).

For folded state exchange, there is no independent measure of the putative pre-equilibrium transition(s), whether numerous small internal motions as in the penetration model, or cooperative excursions as in the local unfolding model. The pH dependence expected for EX2 is observed for most, perhaps all, native state exchange, and often an EX2 formalism is used to extract from k_{obs}(folded state mechanism) a value of $\Delta G(HX)$ for a presumed open/close transition of the *folded* state, using again the relationship $\Delta G(HX) = -RT(k_{calc}/k_{obs})$. However, for folded state exchange this carries an embedded assumption, that $k_{cx,N} = k_{calc}$, which we think is not warranted (Woodward et al., 1982). The implicit assumption is that local environmental effects on the NH in an "exchangeable" conformation are the same as for an NH freely exposed to solvent in a small peptide. In the penetration model this is highly unlikely, as the NH is in an interconverting ensemble of conformations of varying "exchangeability" and approximated on average by the crystal structure. In any model for folded state exchange, including local unfolding, it cannot be tacitly assumed that for an NH in exchangeable conformations $k_{cx,N} = k_{calc}$, especially in view of the fact that for many surface protons $k_{obs} \ll k_{calc}$ (Tüchsen & Woodward, 1985).

Reporting hydrogen exchange data as a series of $\Delta G(HX)$ values for all amides in the protein is not advisable because it has led to confused interpretations since, while the unfolding mechanism *may* reasonably be expressed this way, the folded state mechanism is not. In publications, observed exchange rate constants should be stated explicitly because the actual exchange rate constant is lost if $\Delta G(HX)$ is reported but k_{calc} is unspecified (wobble in the k_{calc} value is due primarily to temperature and pH corrections). In our view of folded state exchange, $k_{cx,N} \neq k_{calc}$, and at present we do not have experimental or calculated $k_{cx,N}$ values. The most useful parameters for expressing hydrogen exchange data are k_{obs} for folded state exchange, and k_{obs}, $\Delta G(HX)$, and $\Delta\Delta G(HX)$ for the unfolding mechanism.

The approximate preservation among NHs of rank order of rates at varying temperature and pH is characteristic of hydrogen exchange in proteins (Woodward & Rosenberg, 1971a; Gregory

et al., 1986). This means that even when the temperature is raised or lowered, or when pH is changed, the slowest exchanging NHs in a native protein remain about the same. Approximate preservation of rank order was shown by tritium-hydrogen methods in which the average number of NHs exchanged per protein molecule is measured as a function of time after transfer to tritium solvent. In NMR-detected, deuterium-hydrogen exchange experiments, if for folded state exchange, rank order is approximately the same at all temperatures then Arrhenious plots of $\ln k_{obs}$ vs. $1/T$ for individual NHs will, for the most part, tend not to cross one another (at temperatures below onset of the unfolding regime).

Identification of the slow exchange core

The slow exchange core is identified by NHs last to exchange *by the folded state mechanism*. Slow exchange core sequences are stretches of residues that bracket the slowest NHs. Slow exchange core elements are the secondary structural elements containing slow exchange core sequences. When identifying core elements from hydrogen exchange, it is most useful to consider all the types of data in Table 1. Confidence that a slow exchange core element is correctly identified is increased when the element has several NHs in the "slowest exchanging" category.

Rate constants for the folded state mechanism k_N are usually well dispersed over many orders of magnitude, while rate constants for the unfolding mechanism k_D are similar since global unfolding is a common step. As discussed below with reference to Figures 3C and 3D, folded state exchange rates are often not sufficiently resolved to permit identification of the slow exchange core. This usually occurs in two ways, when for a large number of NHs experimental time $\ll$ exchange time, or alternatively when for a large number of NHs the *un*folding mechanism is predominant ($k_D > k_N$). Either shortcoming can usually be overcome by sliding the observation window toward slower folded state rates, that is by manipulating solution conditions and bearing in mind the two process model (Fig. 1A) and the approximate preservation of exchange rank-order when pH or temperature is varied. When a large number of NHs are "too slow to measure," increasing the pH often moves k_N values of additional NHs into the observation window without a switch in mechanism (i.e., without a switch from $k_N > k_D$ to $k_N < k_D$). Raising pH above 4, but keeping below a pH that significantly destabilizes the protein, often works in this regard because k_N is increased by an order of magnitude for each pH unit but k_D is not significantly affected (because k_u/k_f is not significantly affected). When so many NHs exchange by the unfolding mechanism that the last few cannot be identified, then one may increase the number of resolved k_N values by working with a stabilized mutant, so that for some slower NHs the mechanism changes from $k_N < k_D$ to $k_N > k_D$. Also, lowering temperature can switch some NHs from the $k_N < k_D$ regime to the $k_N > k_D$ regime because the activation enthalpy is greater for k_D than for k_N (Woodward et al., 1982). In summary, in less stable proteins (e.g., ubiquitin at pH 3.5) many NHs exchange by the unfolding mechanism, and lowering temperature while raising pH may increase the number of measurable k_N values. In very stable proteins (e.g., BPTI), many k_N values are too small to measure; raising pH and/or temperature often brings k_N values into the observable range, provided global stability is not sufficiently decreased to cause a switch from $k_N > k_D$ to $k_N < k_D$ for the target NHs. Note that the slow exchange core is not necessarily identified by *all* the NHs exchanging by global unfolding; however, if most NHs have exchanged by the folded state mechanism, the last few protons may in fact exchange by the unfolding mechanism under these experimental conditions.

Do slow exchange core sequences encompass the full length of an element of secondary structure? Slow exchange core sequences are usually less than a full strand or helix, and they comprise segments that interact nonlocally with other elements. For small to medium length secondary structural elements, most of the element is included in core sequences. In this paper, we discuss the data in terms of core elements (entire secondary structural elements), since the currently available data do not support a more precise breakdown into subelement sequences.

Are NHs in turns and loops among the very slowest to exchange? Usually no; while they may be slow in the native state, they are generally more flexible and more rapidly exchanging than the slowest exchanging NHs in packed secondary structure. Occasionally, an NH in a turn or loop is in the very slowest group (e.g., residue 25 in barnase), and although these are invariably intramolecularly H-bonded, the reason for very slow exchange is not obvious from inspection of the crystal structure. In several proteins, there is evidence that a turn(s) initiates folding (e.g., Barbar et al., 1995; Gu et al., 1997; Grantcharova et al., 1998). If "the slow exchange core is the folding core," should we expect NHs in turns that are fold-initiating, or nucleating, to be in the slowest exchanging group? Some turns thought to be nucleation sites do have NHs that exchange slower than other turns/loops in the molecule, but not as slow as NHs in core sheet/helix. For example, native-like NOEs involving the reverse turn (25–28) are observed in unfolded BPTI (Pan et al., 1995), and this turn is apparently a nucleation site for formation of the core antiparallel β-sheet (Barbar et al., 1998). In native BPTI, NHs in the 25–28 turn exchange faster than in core elements (β1, β2, and first turn of the α-helix), but slower than in other turns/loops. In the sense that the term was coined (Woodward, 1993), the folding core is the product of nucleation and is composed of sequences that favor nonlocal interactions early in folding. Nucleating turns connect core sheet/helix elements.

Lumry et al. have identified protein "knots" from several properties of proteins, including slow exchange, and have suggested a number of functional roles for knots in protein folding and enzyme catalysis (Lumry, 1991, 1995; references in Kim et al., 1993). They have also suggested models for small motions that may be responsible for a penetration mechanism (Lumry & Rosenberg, 1975).

Pulsed exchange and folding competition experiments

Experiments diagramed in Figure 2 identify NHs that acquire slower exchange rates during folding (reviewed in Englander & Mayne, 1992; Woodward, 1994; Gladwin & Evans, 1997). Our group term for pulsed-labeling or competition experiments is P/C experiments. With a combination of NMR spectroscopy and rapid mixing methods, folding is followed in the time range of milliseconds and longer (see Table 1 in Supplementary material in Electronic Appendix). The fast limit is the dead time of the quenched- or stopped-flow apparatus, normally 4–15 ms. In competition methods (Fig. 2A), unfolded, deuterated protein is rapidly diluted into refolding buffer in H_2O. At each NH the competing processes are isotope exchange and protection due to folding. After refolding, the extent of hydrogen isotope exchange is determined by NMR; less exchange means greater protection during folding. As pH is raised, k_{cx} increases while the folding rate is often not significantly

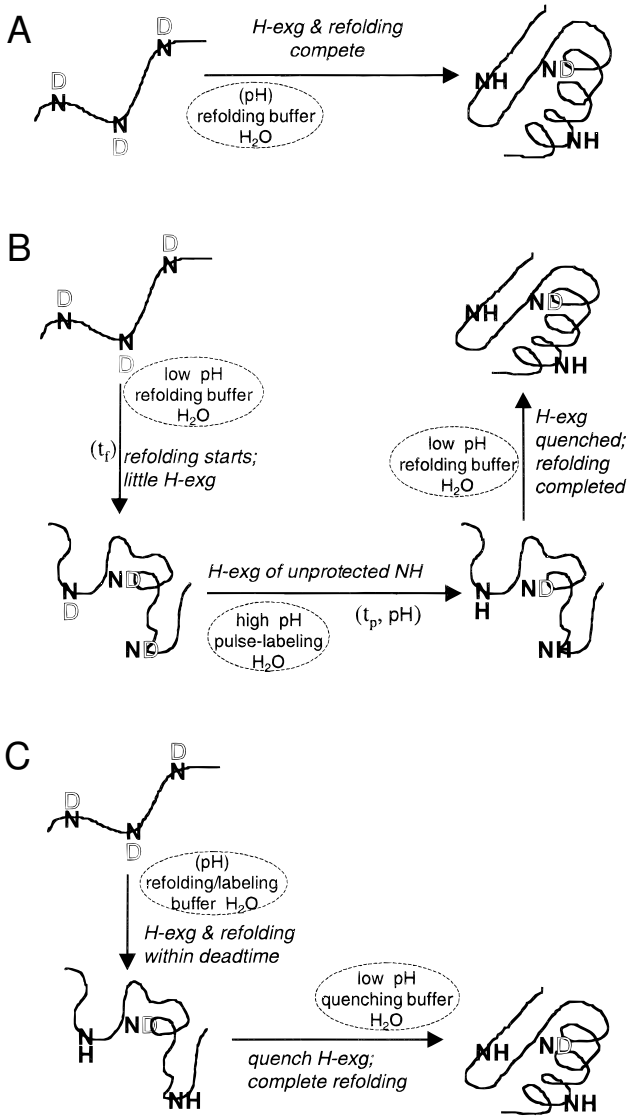

Fig. 2. Methods for measuring hydrogen exchange during folding. Exchangeable amides are represented at three locations along a polypeptide chain. **A:** Competition method. **B:** Pulsed-labeling method. **C:** Dead time labeling method. Variables are in parenthesis: t_f, the refolding time; t_p, the pulse labeling time, usually milliseconds; and pH. "Jump" buffers are enclosed by dotted lines.

changed, and the number of NHs protected during folding decreases. Amides with highest exchange protection at the highest folding pH are presumably in parts of the protein that fold first.

In the pulsed-labeling method (Fig. 2B), unfolded and deuterated protein is first diluted into a low pH refolding buffer, to start refolding while minimizing hydrogen exchange. After a refolding time period t_f, the sample is pulsed with a high pH H_2O buffer. The length of the pulse t_p and the pH are chosen so that unprotected amide deuterons are maximally exchanged during the pulse. After the pulse, pH is lowered to complete folding and quench further exchange. Typically, the quench step jumps pH to 3.5–4.5 where out-exchange is at a minimum. NHs exchanging slowly enough to be probes for the experiment are subsequently monitored by NMR after folding is complete. Exchange as a function of t_f is determined. Amides that show protection earliest are presumably in regions of the protein that fold first. A recently developed hybrid method monitors protection within the few milliseconds of refolding (Gladwin & Evans, 1997; Fig. 2C). When unfolded and deuterated protein is mixed with refolding/labeling buffer in H_2O, refolding and exchange compete. After the first few milliseconds, which normally corresponds to the dead time of mixing (giving rise to the name, dead time labeling), a low pH quenching buffer is added to stop exchange and complete refolding. The protection pattern is estimated by varying the pH of the refolding/labeling buffer.

What does protection early during folding mean? Typically in P/C experiments, protection rates are 4- to 100-fold slower than k_{calc}. This does not imply any particular structure; rather it means that, for any or all of a number of reasons, exchange of that NH is detectably slower than for the same NH in a small peptide of equivalent sequence. Possible structural reasons for early protection cover the gamut, and include favoring of secondary structure for a significant fraction of the time in the region of the exchanging NH, favoring of other local or nonlocal contacts, and so on. Exchange protection patterns in P/C experiments do not tell us whether exchange-slowing structure in the vicinity of the protected NH is native like, although it may be; we can surmise only that structure around a protected NH is nonrandom.

Is the comparison valid?

Do technical limitations bias the outcome of a comparison of native state out-exchange and pulsed-labeling/competition results? Is a correlation built into the experiment because the reporter groups in P/C experiments are NHs that exchange slowly in the native state around pH 3.5–4, while those that exchange rapidly in the native state are not monitored? We answer no, because the number of amide probes, though limited, is large (see Table 1), and the probes are distributed throughout the protein molecule. This point is clearly illustrated in Figure 3, where for each set of experimental P/C results, there is a visual comparison to the NHs actually monitored. For each protein in Figure 3, the lower left structure labeled "p/c pr." shows in light red the residues that are probes in each pulsed-labeling/competition experiment. In the lower right structure labeled "P/C," dark red residues are those protected early.

Another question that may be raised concerns differences in pH or temperature of out-exchange vs. pulsed exchange experiments. Hydrogen exchange rate constants are sensitive to the environment, especially pH and temperature (Woodward et al., 1982). However, it is not individual rate constants, but the slowest exchanging protons that matter in identifying the slow exchange core. As discussed, approximate rank order is maintained in hydrogen exchange (Woodward & Rosenberg, 1971a; Gregory et al., 1986). Even under conditions where exchange mechanisms apparently switch (e.g., Clarke & Fersht, 1996), the slowest exchanging NHs remain largely the same. Since the magnitude of the apparent protein folding/unfolding rate constants and phase amplitudes are sensitive to pH and temperature, one might ask if protection patterns observed in P/C experiments are sensitive to pH or temperature. P/C experiments at different pH values are reported for cyt *c*; although folding kinetics are different at pH 6.2 vs. pH 5, the N- and C-helices are the first to gain protection at either pH, except for histidine residues involved in non-native heme binding reactions (Elöve et al., 1994). Under different jump conditions of pH 2 to 5, cyt *c* folds very rapidly, with time constant ~15 ms, and all probes show uniform protection at 5–95 ms (Sosnick et al., 1994).

Table 1. *Summary of hydrogen exchange data for the comparison between the slow exchange core elements, P/C elements, and PF elements* [a]

Proteins						Out-exchange experiments[b]		Pulsed-labeling/competition experiments[b,c]			
Name	Structure class	Number of amino acids	S-S bonds	Subdomain	Bound cofactor	Experimental condition	NHs slowest to exchange out[d]	Folding pH, temperature	Number of probes	NHs first to gain protection[d]	References
ApoMb	All-α	147	None	One[e]	None	N, 5 °C pH 6.0	V10, L11, W14, V17 (αA); I30 (αB); I112, V114, L115 (αG)	5 °C pH 6.1	38	L9–L11, W14, V17 (αA); L29, I30, F33 (αB); K102, L104, I107, A110, I112–L115 (αG); K133, F138, I142, A143 (αH)	Hughson et al. (1990) Jennings and Wright (1993)
						PF, 5 °C pH 4.2	L9–L11, V17 (αA); I107, I112, V114 (αG); I142 (αH)				Hughson et al. (1990)
						PF, 4 °C pH 4.2 5 mM TCA	L11, W14, V17 (αA); I107, A110, I112, V114 (αG); R139 (αH)				Loh et al. (1995)
Cyt *c*	All-α	112	None	One	Heme	N, 20 °C pH 7.0	F10 (N-α); L94–K99 (C-α)	10 °C pH 6.2	35	K8–V11 (N-α); E92, L94–K99 (C-α)	Roder et al. (1988) Jeng et al. (1990)
								10 °C pH 5.0	34	I9, V11 (N-α); R91, A96, Y97 (C-α)	Elöve et al. (1994)
						PF[f], 20 °C pH 2.2 1.5 M NaCl	V11, A15 (N-α); M65 (60's α); L94–K99 (C-α)				Jeng et al. (1990)
Protein A B-domain	All-α	58	None	One	None	N, 20 °C pH 7.0	R27, L34 (α2); A48–K50 (α3)	5°C pH 5.0	20	Y14–L17 (α1); L34–K35 (α2); A48–K50 (α3)	Bai et al. (1997)
HEWL	$\alpha+\beta$	129	4	Two	None	N, 30 °C pH 7.5	M12 (αA); W28–A31 (αB); I58 (β3); A95, K96, I98 (αC)	pH 4.5–9.5	65 (34α, 31β)[g]	36 amides (33 are in α-domain, 3 are in β-domain[h])	Miranker et al. (1991) Radford et al. (1992a)
								20 °C pH 5.5	48	M12 (αA); V29 (αB); W63; I78; K96, V99 (αC); I124 (3_{10})	Radford et al. (1992b)
						PF, 27 °C pH 2.0 50% TFE	A11 (αA); A32 (αB); N93–K96 (αC); W111, K114 (αD)				Buck et al. (1993) Buck et al. (1994)
						D, −13 °C pH 2.0 3.7 kbar	A9, A11, M12 (αA); V29, A32 (αB); W63–N65, I78; C94 (αC); W108, W111, R114, C115 (αD); I124 (3_{10})				Nash and Jonas (1997)
T4 lysozyme	$\alpha+\beta$	164	None	Two	None	N, 20 °C pH 3.5–7.8	M6 (αA); F67, L79 (αC); Y88 (αD); A98–F104 (αE); I150, T152–R154 (αH)	20 °C pH 6.0[i]	84	I17, Y18 (β1); Y25, I29 (β2); I58 (β4); R95, A98–M102, F104 (αE)	Lu and Dahlquist (1992) Anderson et al. (1993)

SNase	$\alpha+\beta$	149	None	Two	Ca^{2+} pdTp	N, 37 °C pH 5.5[j]	T22, K24–M26 (β2); F34 (β3); L37	22 °C pH 4.0[k]	39	L25 (β2); Q30, F34 (β3)	Loh et al. (1993) Jacobs and Fox (1994)
RNase A	$\alpha+\beta$	124	4	One	None	N, 35 °C pH 6.5	F46, V47 (β1); Q55, C58 (α3); Y73, Q74 (β3); M79, I81 (β4); I106, V108, A109 (β5)	10 °C pH 4.0	27	V47-H48 (β1); V54 (α3); V63 (β2); C72, Y73 (β3); I81, C84 (β4); K98, I106, V108 (β5); V116, V118, H119 (β6)	Udgaonkar and Baldwin (1990) Wang et al. (1995)
						D, 10 °C pH 2.0 4.2 kbar	E49; S59 (α3); C72 (β3); M79, C84, R85 (β4); I106 (β5); E111				Zhang et al. (1995)
						D, −17 °C pH 2.0 3 kbar	E49; S59 (α3); C72 (β3); D83, C84 (β4); E111				Nash et al. (1996) Nash and Jonas (1997)
RNase T1	$\alpha+\beta$	104	2	One	None	N, 25 °C pH 5.6	Y57, W59, I61 (β2); R77–N81 (β3); Q85, A87 (β4)	25 °C pH 5.0	24	A19, L26 (α); R77 (β3); A87 (β4); H92; F100, V101 (β5)	Mullins et al. (1993) Mullins et al. (1997)
Barnase	$\alpha+\beta$	110	None	One	None	N, 33 °C pH 5.6	L14 (α1); I25 (loop); A74 (β2);L89 (β3); Y97 (β4)	25 °C pH 6.3	29	(See Electronic supplementary material)	Bycroft et al. (1990) Matouschek et al. (1992) Perrett et al. (1995)
RNase H	$\alpha+\beta$	155	None	One	Mg^{2+}	N, 27 °C pH 5.5	L49, M50–I53, A55, L56 (αA); L107 (αD)	25 °C, pH 5.5–7.5	59	A51, A52, V54–L56 (αA)	Yamasaki et al. (1995)
RNase H*[1]	$\alpha+\beta$	155	None	One	Mg^{2+}	N, 25°C pH 5.1	M47, A51, A52, V54–L56 (αA); W104, L107 (αD)	25 °C pH 5.5		A51, A52, V54–E57 (αA); I66, L67 (β4); W104, Q105, L107, A110, L111 (αD)	Chamberlain et al. (1996) Raschke and Marqusee (1997)
						PF, 4 °C pH 1.3	A51, A52, V54–E57 (αA)				Dabora et al. (1996)
Ubiquitin	$\beta+\alpha$	76	None	One	None	N, 22 °C pH 3.5	I3-V5, T7 (β1); V17 (β2); I23, V26, A28–I30 (α1); I44 (β3); L56, Y59 (3_{10})	25 °C pH 5.0	26	(Most probes are protected early.)	Briggs and Roder (1992) Pan and Briggs (1992)
						PF, 22 °C pH 2.7 60% methanol	I3–K6 (β1); I13, V17(β2); V26, I30 (α1); Q40; I44, F45 (β3); L56, Y59 (3_{10}); I61; L67, L69, V70 (β5)				Pan and Briggs (1992)
						PF, pH 2.0 room temperature 60% methanol	I3–T7 (β1); I13, L15, V17 (β2); V26, A28, E34 (α1); L67, L69, V70 (β5)				Harding et al. (1991)
CI2	$\beta+\alpha$	64	None	One	None	N, 33 °C pH 5.3	K11; I20, L21 (α); I30, L32 (β2); V47, L49, F50, V51 (β3)				Itzhaki et al. (1997) Neira et al. (1997)
GB1	$\beta+\alpha$	56	None	One	None	N, 25 °C pH 5.7	L5–I6 (β1); E27, F30 (α); T44 (β3); T51–V54 (β4)	5 °C pH 4.0	26	T25, A26 (α); K50, T51 (β4)	Kuszewski et al. (1994) Orban et al. (1995)

(*continued*)

Table 1. *Continued*

Proteins						Out-exchange experiments[b]		Pulsed-labeling/competition experiments[b,c]			
Name	Structure class	Number of amino acids	S-S bonds	Subdomain	Bound cofactor	Experimental condition	NHs slowest to exchange out[d]	Folding pH, temperature	Number of probes	NHs first to gain protection[d]	References
LB1	$\beta+\alpha$	62	None	One	None	N, 60 °C pH 11	I4, A6, L8, I9 (β1); F20 (β2); A31–D36 (α); L56, I58, K59 (β4)	pH 8.5–10	24	A6 (β1); S29, A33, Y34 (α)	Yi and Baker (1996) Yi et al. (1997)
BPTI	All-β	58	3	One	None	N, pH 3.5	R20–Y23 (β1); Q31, F33 (β2); F45 (β3)	pH 4.0–7.5	8	Y21–Y23 (β1); Q31, F33, Y35 (β2); F45 (β3)	Woodward and Hilton (1980) Roder and Wüthrich (1986)
						PF, 1 °C pH 4.6[m]	I19, Y21–Y23 (β1); Q31, F33, Y35 (β2)				Barbar et al. (1995)
CTX III	All-β	60	4	One	None	N, 25 °C pH 3.2	Y22–M24 (β3); I39 (β4); V52, C54 (β5)	5 °C pH 3.0	32	K23 (β3); I39 (β4); V49; V52, C53 (β5); D57, R58 (loop)	Sivaraman et al. (1998)
IL-1β	All-β	153	None	One	None	N, 36 °C pH 6.1	Y68, L69, V72 (β6); Q81, E83 (β7); V100, F101 (β8); E113, A115 (β9); F117; Y121 (β10)	4 °C pH 5	47	F42 (β4); Y68–S70, V72 (β6); T79, Q81, E83 (β7); F99–I104 (β8); E111, E113, S114 (β9); F117; W120, Y121, S123 (β10)	Driscoll et al. (1990) Varley et al. (1993)
scFv	All-β	252	2	Two	None	N, 27 °C pH 6.9	V13 (Lβa); C23 (Lβb); A40–Q43 (Lβc); S71 (Lβd); T78, I81, V84 (Lβe); A90, V91, Y93, C94 (Lβf); M34, Q39 (Hβc); I48, A49, S51 (Hβc′); S63 (Hβc″); F70, I71, D75 (Hβd); Y82, Q84, R89 (Hβe); D92, A94, I95 (Hβf); V118 (Hβg)	10 °C pH 8.0	49	L39–Y42 (Lβc); I54 (Lβc′); A90, V91, Y93 (Lβf); I95 (Hβf)	Freund et al. (1996) Freund et al. (1997)

α-LA (human)	$\alpha+\beta$	123	4	Two	Ca^{2+}	N, 15 °C pH 6.3 3 mM Ca^{2+}	T33 (αB); C61 (loop); C73, C77, F80 (3_{10}); C91–K94 (αC)	Schulman et al. (1995)
				n.d.	None	PF, 5 °C pH 2	L26, I27, M30, T33 (αB)	Schulman et al. (1995)
α-LA (guinea pig)	$\alpha+\beta$	123	4	n.d.	None	PF, 25 °C pH 2	W26, C28–F31 (αB); I95, L96 (αC)[n]	Chyan et al. (1993)
Lysozyme (equine)	$\alpha+\beta$	129	4	Two	Ca^{2+}	N, 25 °C pH 4.5 10 mM $CaCl_2$	A9, K11–K13 (αA); N27, W28, M31, A32 (αB); F38, N39 (β2); N61, C65 (β5); S93–A95 (αC)	Morozova et al. (1995) Morozova-Roche et al. (1997)
				n.d.	None	PF, 5 °C pH 2	A9, K13 (αA); W28, M31, A32 (αB); W111 (αD)	Morozova et al. (1995) Morozova-Roche et al. (1997)
				n.d.	None	PF, 25 °C pH 2	A9 (αA); W28, A32, E35 (αB); W111 (αD)	Morozova-Roche et al. (1997)
Tendamistat	All-β	74	2	One	None	N, 50 °C pH 3	A23 (β1(2)); C27; V33, V35–Y37 (β2(2)); Y46, V48 (β2(3)); A71 (β2(1))	Wang et al. (1987) Schönbrunner et al. (1996)
						PF, 25 °C pH 3 70% TFE	S21, Q22, D24 (β1(2)); C27; K34–V36 (β2(2)); D40; T41, E42, C45 (β2(3)); C73 (β2(1))	Schönbrunner et al. (1996)

[a] Abbreviations used in this table: α-LA, α-lactalbumin; apoMb, apomyoglobin; barnase, *Bacillus amyloliquefaciens* ribonuclease; BPTI, bovine pancreatic trypsin inhibitor; CI2, chymotrypsin inhibitor 2; CTX III, cardiotoxin analogue III; cyt *c*, cytochrome *c*; D, cold and/or pressure denatured state; GB1, B1 immunoglobulin-binding domain of streptococcal protein G; HEWL, hen egg-white lysozyme; IL-1β, interleukin-1β subunit; LB1, B1 immunoglobulin-binding domain of peptostreptococcal protein L; N, native state; pdTp, thymidine 3',5'-bisphosphate; PF, partially folded state; RNase A, ribonuclease A; RNase H, ribonuclease H; RNase T1, ribonuclease T1; scFv, single-chain antibody fragment composed of the V_H and V_L domains; SNase, staphylococcal nuclease; TCA, trichloroacetate; TFE, 2,2,2-trifluoroethanol.

[b] In both out-exchange and pulsed labeling/competition experiments, there are cases of the same species studied under different conditions. For example, there are two studies of partially folded apoMb at pH 4.2, with and without TCA.

[c] When pH is given as a range, the data referenced are from a competition experiment, rather than a pulsed-labeling experiment; the difference between pulsed exchange and competition experiments is explained in the text.

[d] The secondary structural elements containing the residues listed are given in parentheses. For instance, αA means helix A; C-α means the C-terminal helix.

[e] One means that the protein is a single domain protein; n.d. means "not determined."

[f] This species is referred to by authors as acid-denatured state.

[g] Of the 65 probes monitored in competition experiment, 34 are located in the α-domain of HEWL and 31 are in β-domain.

[h] The three highly protected amides in the β-domain are: W63, C64, I78 (Miranker et al., 1991).

[i] The data are reported for C54T/C97A of T4 lysozyme.

[j] The data are reported for H124L of SNase, in both unligated and ligated forms.

[k] The data are reported for P117G refolding without Ca^{2+} and pdTp, but in the presence of 400 mM sodium sulfate.

[l] RNase H* is a mutant (C13A/C63A/C133A) of *E. coli* RNase H.

[m] The data are reported for $[14–38]_{Abu}$ of BPTI.

[n] The indole hydrogen of W26 has the highest protection factor of >920 (Chyan et al., 1993).

Review of the exchange literature

Comparable native state out-exchange and pulsed-labeling/competition results are published for 16 proteins; for about half of these, out-exchange rates for partially folded species are also reported. Data from these systems are summarized in Table 1, where proteins are sorted by structural class. Residue numbers of the slowest exchanging NHs, secondary structural elements in which

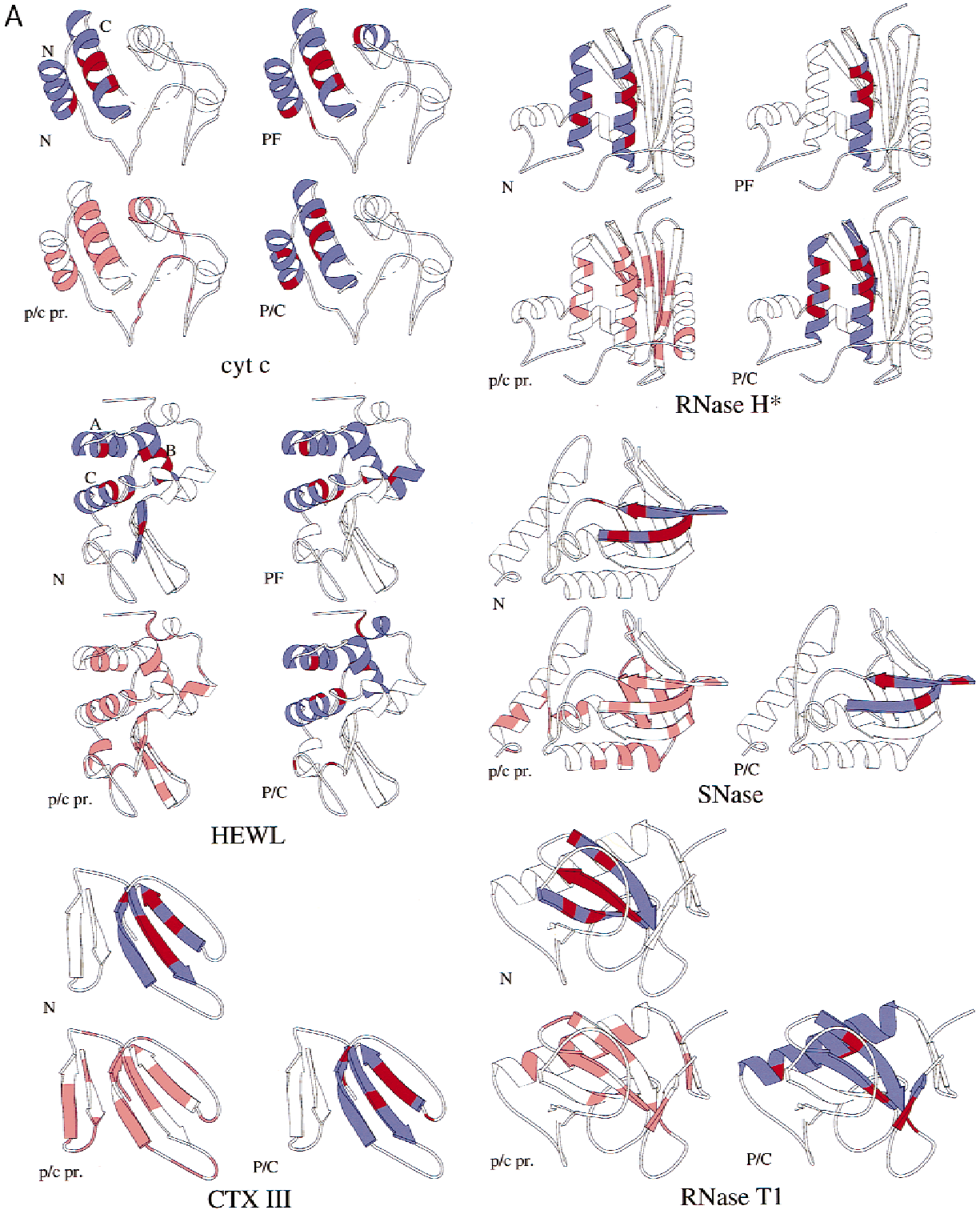

Fig. 3. *See caption on next page.*

the slowest exchanging NHs are located, and experimental conditions are listed in columns under "out-exchange experiments," with separate entries for N, the native state, and PF, a partially folded state. Residue numbers of NHs protected first during folding, secondary structural elements in which they are located, and the total number of NHs monitored in each pulsed-labeling/competition experiment, are listed in columns under "pulsed-labeling/competition experiments." Cut-off values distinguishing the "slowest exchanging" NHs, and the "first protected" NHs are, when possible, those in the literature cited; if cut-offs are not specified in a paper, they were chosen in line with the conclusions of the authors. Each protein is covered in detail in Supplementary material in Electronic Appendix.

The data in Table 1 are summarized graphically in Figure 3; results for N-state out-exchange, partially folded state out-exchange, and pulsed-labeling/competition experiments are mapped onto ribbon structures labeled, respectively, N, PF, and P/C. NHs listed in Table 1 are red, while the rest of the secondary structure element in which they are located is dark blue. Figure 3 also shows the NHs monitored in each pulsed-labeling/competition experiment; these are light red in the structure labeled "p/c pr" (for P/C probes). The elements of secondary structure carrying NHs most protected in the native state, first protected during folding, and most protected in partially folded species, we term, respectively, slow exchange core elements, P/C elements, and PF elements.

Hydrogen exchange studies of denatured states are not systematically covered here, although this is a growing and interesting area (Jeng et al., 1990; Robertson & Baldwin, 1991; Lu & Dahlquist, 1992; Radford et al., 1992a; Buck et al., 1994; Arcus et al., 1995; Mori et al., 1997). Particularly notable are two cold and pressure denatured states that have hydrogen exchange patterns similar to P/C protection patterns (Nash et al., 1996; Nash & Jonas, 1997). These are included as "D" entries under lysozyme and RNase A in Table 1, and for RNase A in Figure 3 as D, rather than PF, in the upper right structure.

According to the nature of the published data, the proteins are divided into four categories, corresponding to Figures 3A–D. For proteins in Figure 3A, the data permit the most unambiguous comparison of slow exchange core to P/C protection rates; folded state

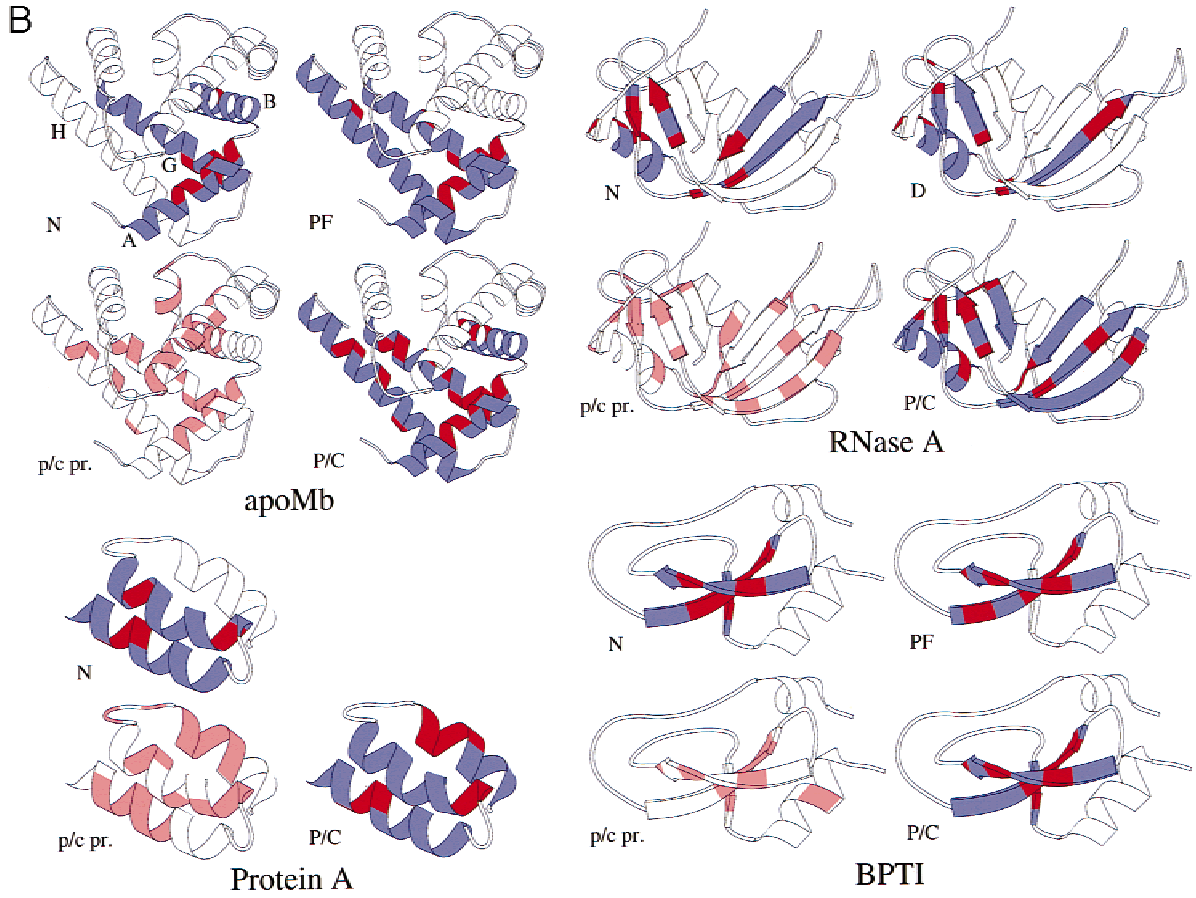

Fig. 3. Comparisons of slow exchange core elements, P/C elements, and PF elements. For each protein, residues listed in Table 1 are red; the secondary structural elements in which they are located are blue. "N" shows results for native state out-exchange experiments; "PF" for partially folded out-exchange experiments; and "P/C" for pulsed-labeling/competition experiments. The structure labeled "p/c pr." shows probes monitored in the P/C experiment (light red). When there are two or more similar studies, results of only one experiment are displayed. Proteins are divided into four categories (**A–D**) as discussed in the text. Ribbon diagrams are drawn with MOLSCRIPT (Kraulis, 1991) for PDB coordinates of apoMb, 1mbo; cyt *c*, 1hrc; B-domain of protein A, 1bdd; HEWL, 1hel; RNase A, 1rbx; RNase H, 2rn2; ubiquitin, 1ubi; BPTI, 5pti; SNase, 1stn; T4 lysozyme, 2lzm; RNase T1, 9rnt; GB1, 1pga; LB1, 2ptl; CTX III, 2crt; IL-1β, 1i1b; scFv, 2mcp. (*Figure continues on next page.*)

out-exchange rates are sufficiently resolved to permit identification of "the slowest," and there are resolved P/C protection rates, rather than cutoffs. In SNase, CTX III, and cyt *c*, the slow exchange core elements are the same as P/C elements. In HEWL, the four P/C elements include three of the four slow exchange core elements; however, loop NHs 63, 64, and 78 are protected early in P/C experiments but are not among the very slowest to out-exchange. In RNase H, the P/C elements are the slow exchange core elements plus an additional one. In RNase T1, to quote the authors, "All of the slowest exchanging amide residues are located in strands 2–4

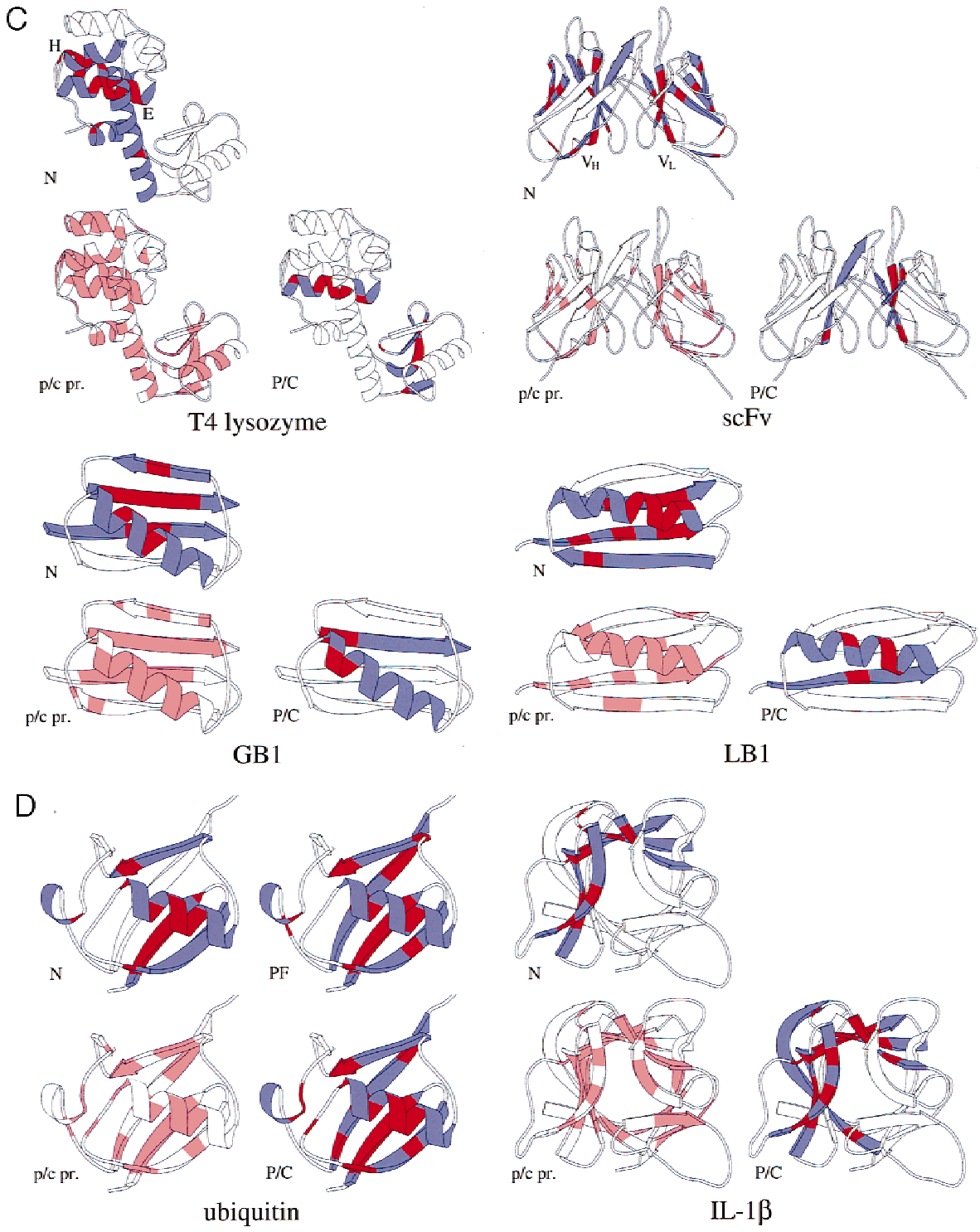

Fig. 3. *Continued.*

of the central β-sheet and these residues are protected first in the early stages of folding. The residues that have somewhat lower rate constants for protection in early folding are predominantly found in the α-helix and the first strand of the small β-sheet" (Mullins et al., 1997). In RNase T1, β3 has five consecutive NHs in the slowest exchanging group (R77–N81), and one of these, R77, has a protection rate qualitatively more rapid than all others. This suggests β3 as the central core element in RNase T1. In summary, for the proteins in Figure 3A, there is a strong tendency for the same elements of secondary structure to be slow exchange core elements, P/C elements *and* PF elements. The correlation is not perfect, but a clear trend is apparent. As expected, there is not a one-to-one correspondence of individual NHs most protected in native and partially folded states, and first protected during folding.

For proteins in Figure 3B, folded state out-exchange rates are fairly well resolved but P/C protection rates are reported as cutoffs or are monitored for only a few NHs. For apoMb, all P/C NHs in Table 1 are protected in $<$5 ms (other categories are $\sim$1 and $\sim$2.5 s). ApoMb P/C elements are the slow exchange core elements plus an additional helix (αH). The high stability of RNase A makes it difficult to pin point core NHs. For RNase A, 11 NHs are in the "slowest" category (Table 1) and all exchange by the unfolding mechanism, but we cannot distinguish the slowest among these 11. The P/C cutoff for RNase A is any protection factor $>$1,000 after 0.4 s of folding; NHs that fit this category are large in number and widely distributed in the protein, and the N vs. P/C comparison is not as informative as the N vs. D comparison (below). In B-domain of protein A, although the native state out-exchange rates identify helices 2 and 3 as the slow exchange core elements, the protein folds so rapidly that all three helices in the protein are well protected within the dead time of folding experiments. In BPTI, out-exchange rates are well characterized but P/C results are from early pioneering experiments. Within the limitations of the P/C data, the slow exchange core elements of BPTI are the same as the P/C elements.

Figure 3C shows proteins for which the P/C rates are spread well enough to identify NHs that are first protected, but the slowest out-exchanging group is very large, under the conditions of the experiment. In scFv, out-exchange was followed for three days, after which numerous NHs are too slowly exchanging to measure. P/C experiments with scFv show earliest protection in three strands of V_L and one strand of V_H. In LB1 and GB1 numerous NHs out-exchange very slowly under the experimental conditions, and the slowest among these cannot be identified unambiguously. In both proteins, the helix and one strand of sheet contain NHs that are first protected. In LB1, β1 and α have the largest number of slowest exchanging NHs, and in a folding competition experiment (Yi et al., 1997), β1 and α also have the first protection, followed closely by an NH in β2 (not labeled in Fig. 3C). In T4 lysozyme, NHs too slow to be measured under the experimental conditions are distributed in five helices. Helix αE in the C-terminal domain stands out with seven consecutive NHs "too slow to be measured" and the earliest protection rates; αE is apparently the primary core element of T4 lysozyme. Also in the C-terminal domain, αH has three consecutive NHs too slow to measure under the experimental conditions but αH is not protected early; also strands in the N-terminal domain are protected in P/C experiments but their out-exchange rates are not in the slowest category.

Figure 3D shows proteins for which neither out-exchange rates nor P/C protection rates are well resolved. In IL-1β, the P/C elements are the slow exchange core elements plus one additional element. In ubiquitin, exchange rates are not well resolved under the experimental conditions, and neither a "slowest exchanging" group or a "first protected" group is distinguishable. However, the consecutive runs of slowest exchanging and first protected in β1 and α1 suggest that these elements are the central core of ubiquitin.

Partially folded species of about half the proteins in Figure 3 have been monitored for out-exchange; data are mapped onto upper right structures. PF elements tend to be either the same as slow exchange core elements or different by plus-or-minus one element. Of the seven proteins with both N and PF data (Fig. 3 and tendamistat in Supplementary material in Electronic Appendix), all are consistent with the generalization except ubiquitin, for which exchange rates are not well dispersed. In this context, the cold and pressure denatured state of RNase A (Nash et al., 1996; Nash & Jonas, 1997) is interesting. This D-state has considerable residual structure, and the elements carrying the slowest exchanging NHs in N are the same as the elements carrying the slowest exchange NHs in D, except for one β-strand (Fig. 3B).

Nonnative structure in PF states induced by alcohol or TFE has been reported for ubiquitin (Stockman et al., 1993), monellin (Fan et al., 1993), and β-lactoglobulin (Hamada & Goto, 1997). As the number of such cases grows, a detailed comparison of hydrogen exchange patterns should be done. The possibility of nonnative like secondary structure in PF and/or early species detected by P/C experiments is not contradictory to the hypothesis that the slow exchange core is the folding core. Our proposal is that nonrandom, nonlocal structure forms first in core sequences that also include NHs that are most refractory to exchange after folding is completed. Core sequences have the highest probability of sampling organized, nonlocal structure early during folding. The early structure may in most cases be native like but early nonnative-like structure is also possible. It has been suggested that nonnative PF conformations can have hydrogen exchange patterns similar to native ones (Alonso & Daggett, 1995).

Hydrogen exchange results are reported for human holo and apo α-LA, guinea pig apo α-LA, and the related proteins, equine holo and apo lysozyme (Table 1, Fig. 1 in Supplementary material in Electronic Appendix). Apo forms lacking specifically bound calcium were studied in a low pH compact denatured state also referred to as a molten globule (Kuwajima, 1989, 1996). In human α-LA, differences are noted between the slowest exchanging NHs in native holo protein and in partially folded apo protein (Schulman et al., 1995). However, holo vs. apo is not the appropriate comparison for this discussion; we would compare native vs. partially folded holo protein or native vs. partially folded apo protein, not native holo vs. partially folded apo. Removal of calcium in α-LA results in loss of structure within the β-domain and destabilization of a neighboring helix (Schulman et al., 1995), and these are the areas where exchange rate differences are observed (Fig. 1 in Supplementary material in Electronic Appendix).

In summary, the data in Table 1 and Figure 3 show a correlation between slow exchange core elements, P/C elements, and PF elements. The slowest exchanging NHs in native and in partially folded states, as well as NHs protected early during folding, tend to reside in the same secondary structure elements. There is *not* a one-to-one correspondence of individual NHs most protected in native and partially folded states and first protected during folding. Since the proteins surveyed vary in structural class, size, number of disulfide bonds, and bound ligand, it is reasonable to suggest that the generalization may hold for single domain soluble, glob-

ular proteins. Detailed discussion of each protein in Table 1 is in Supplementary material in Electronic Appendix.

ϕ values in slow exchange core elements

Experimental ϕ values express the combined effects of side-chain replacement on folding kinetics and energetics, relative to wild-type protein (Fersht et al., 1992). A ϕ value is the ratio of $\Delta\Delta G^{\ddagger}$ to $\Delta\Delta G$ (the difference between mutant and wild-type values of the folding activation free energy change $\Delta G^{\ddagger}$ and of the native/denatured Gibbs free energy change ΔG). A large ϕ value for a residue indicates that its replacement slows the rate limiting folding step more than another residue whose replacement has a similar effect on global stability, but lower ϕ value. High experimental ϕ values imply involvement of a side chain in interactions favored in rate limiting, early steps. More specific interpretations are made that ϕ values near unity indicate that interactions of the replaced residue are native-like in the folding transition state, while ϕ values near zero indicate that the replaced residue is disordered in the folding transition state (Fersht, 1993). Different amino acid substitutions at the same site often give very different ϕ values for the same residue. Also, ϕ values much less than zero are not uncommon and sometimes are interpreted as having structural implications similar to high positive ϕ values (Jackson et al., 1993; Daggett et al., 1996). Since interpretations of multiple ϕ values for the same side chain, and of ϕ values $\ll$ zero, are discretionary and include a number of assumptions, we consider a high experimental ϕ value as an indicator of probable involvement of the side chains in rate limiting steps of folding and leave aside for the moment more involved structural interpretations. Theoretical ϕ values or similar indicators are calculated for several proteins from statistical lattice-based models (Shoemaker et al., 1997; Portman et al., 1998) and molecular dynamics simulations (Daggett et al., 1996; Bond et al., 1997; Lazaridis & Karplus, 1997; Tirado-Rives et al., 1997).

Objections to our suggestion that the slow exchange core is the folding core from Clarke et al. and Fersht are based on their reading of the data for barnase and CI2 (Clarke et al., 1997; Fersht, 1998), and include interpretations of ϕ values and of hydrogen exchange. Our response to their hydrogen exchange issues is in Supplementary material in Electronic Appendix in sections on barnase and CI2. Here we focus on comparison of slow exchange with ϕ values. We have not suggested a one-to-one correlation between residues with the slowest NHs and highest ϕ values, and lack of one (Fersht, 1998) is not inconsistent with our conclusion. The question of interest to us is whether there are stretches of residues that contain NHs with slowest folded state exchange rates and side chains with highest ϕ values. The presently available data (Fig. 4) give an answer of yes. Proteins in Figure 4 are those for which experimental or calculated ϕ values and native state out-exchange rates are published. Secondary structure is depicted by aligned boxes at the top of each frame; the slowest exchanging protons are indicated as black stripes and slow exchange core elements are shaded gray. Experimental ϕ values are closed symbols; calculated ϕ values are open symbols.

In barnase, NHs of residues L14, I25, A74, L89, and Y97 have the slowest exchanging protons at pH 6.5 ($k_{obs} < 1 \times 10^{-5}$) and 7.6 ($k_{obs} < 2 \times 10^{-4}$) at 33 °C (Perrett et al., 1995). Slow exchange core elements are α1, β2, β3, and β4 (gray in Fig. 4). The loop between α1 and α2 contains one of the slowest exchanging protons, I25, but turns/loops are not presently included as slow exchange core elements (discussed above). Residues with ϕ of 0.8–1.0 are in α1 (D12, Y13, T16, Y17, H18), loop (N58, K62), β3 (I88, L89, S91, S92), and β4 (I96) (Serrano et al., 1992); residues I76 and I109 have high ϕ values for one mutant, but low values for another. We did not locate published ϕ values for a residue in β2 (W71–D75). We conclude from the barnase data in Figure 4 that the slow exchange core elements have the highest ϕ, with the caveat that we have no ϕ values for β2, and turns/loops are not included. The details of the loops are: (1) Residue I25 NH (black line between α1 and α2) has very slow exchange and the most negative ϕ value. As mentioned, some negative values are interpreted as having the same structural implications as $\phi \geq 1$ (Jackson et al., 1993; Daggett et al., 1996). (2) Loop residues N58 and K62 have $\phi \sim 1$, but hydrogen exchange rates are not listed for residues in loop S57–T70 (Perrett et al., 1995), presumably because they exchange very rapidly in the native state.

For CI2 (Fig. 4), the slowest exchanging amides are K11, I20, L21, I30, L32, V47, L49, F50, and V51 (Itzhaki et al., 1997; Neira et al., 1997). Slow exchange core elements are the helix, β2 and β3; K11 is just before the helix, which consists of residues S12–K24. Experimental ϕ values (closed symbols in Fig. 4) are highest for helix residues A16 ($\phi \sim 1$), K18 ($\phi = 0.7$), and E14 (multiple ϕ values of 1.2, 0.75, 0.2) (Itzhaki et al., 1995). For K11, the ϕ value is the most negative in CI2 ($\phi = -0.5$); other large negative values are observed for helix residues V19 ($\phi = -0.26$), D23 ($\phi = -0.25$), and K24 ($\phi = -0.35$ and 0.1). The negative ϕ value of V19 is explicitly interpreted as a reflection of native-like structure in the transition state (Jackson et al., 1993; Daggett et al., 1996). In summarizing CI2 ϕ values along with other considerations, Fersht et al. conclude that in addition to A16, residues that should be considered ordered in the transition state are S12, E14, E15 (helix), L49 (β3), and I57 (loop) (Itzhaki et al., 1995). Calculated values (Daggett et al., 1996; Shoemaker et al., 1997) are open symbols in Figure 4. They are highest for the helix, and in addition, residues in β2 and β3 also have ϕ values notably higher than the baseline; consistent results are reported in other calculations of CI2 folding (Lazaridis & Karplus, 1997). We did not find calculated ϕ values for K11 in the papers cited. In summary, of the three slow exchange core elements of CI2, the helix has residues with highest experimental and calculated ϕ values and most negative experimental ϕ values; the other two, β2 and β3, have residues for which calculated ϕ values are clearly higher than baseline.

For src SH3 domain, "the slowest exchanging NHs are located in two regions, at the base of the RT-loop (the second half of strand 1 and the diverging turn between strands 2 and 3) and in strands 4 and 5 which form a hairpin" (Grantcharova & Baker, 1997); their data are shown as black stripes in the SH3 frame of Figure 4. The three slowest exchanging NHs reported in an earlier study (Yu et al., 1993), Y14, A45, and I56, are shown by asterisks at the top of the frame. The highest ϕ value is for I34 in β3, and the next highest values are for residues in the turn between β2 and β3, and the loop between β4 and β5. In their ϕ value analysis (Grantcharova et al., 1998), the authors focus on the importance of the turn and the loop in nucleating SH3 folding; in addition they state, "segments of the src SH3 domain found to be most structured in the transition state also include the residues most protected from H-D exchange with the solvent." The nucleating loop and turn are connected on one or both sides to core strands.

CheY is relatively unstable, and NHs with the slowest folded state exchange are not well resolved; under the conditions of the experiment, many NHs exchange by the unfolding mechanism;

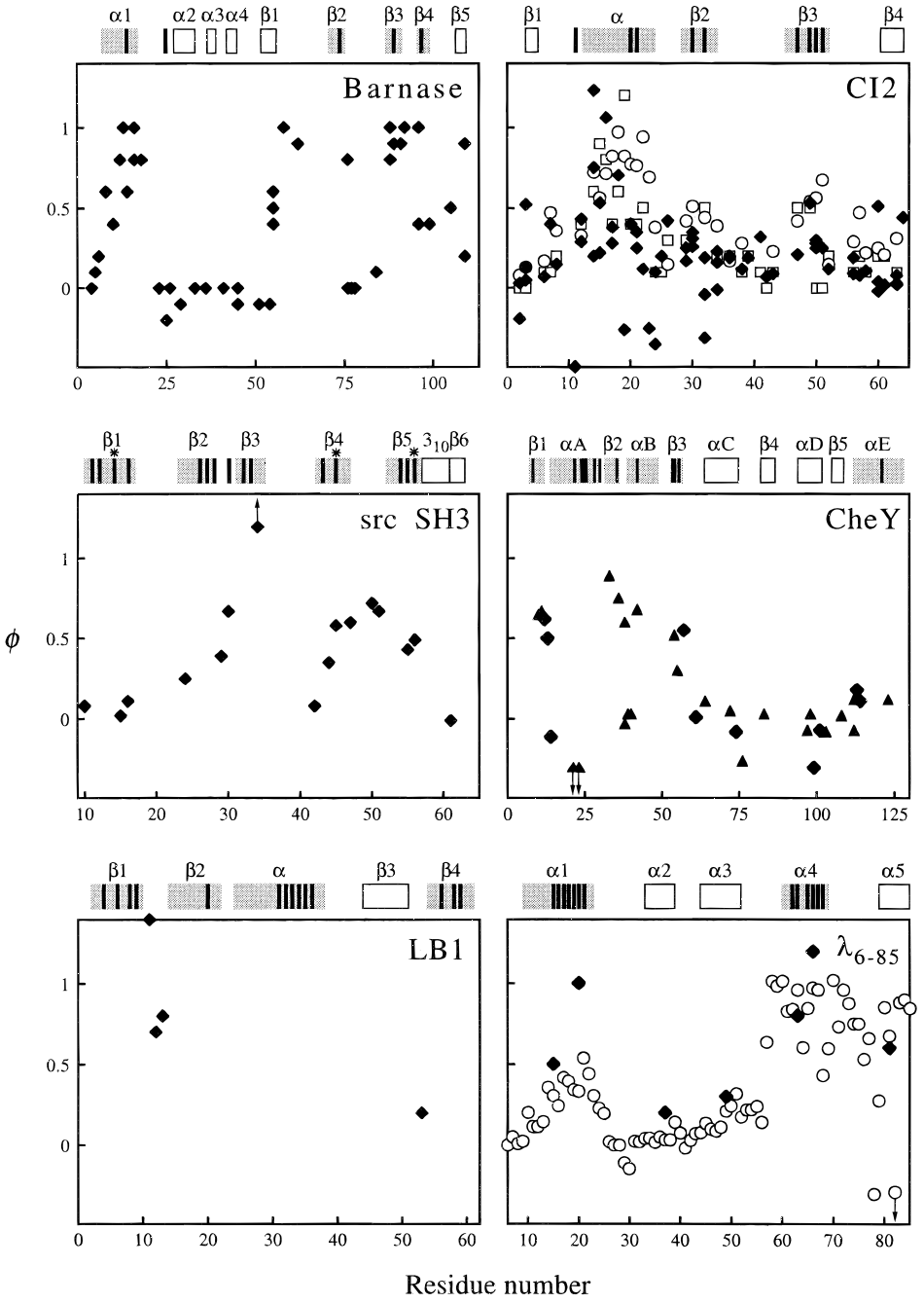

Fig. 4. Comparison of slow exchange core elements and ϕ values. For each protein, residue number is plotted against experimental ϕ values (closed symbols); for CI2 and λ_{6-85}, calculated ϕ values are also given (open symbols). Secondary structural elements are given by boxes at the top of each frame. The slowest exchanging protons are indicated as black stripes and slow exchange core elements are shaded gray (the slowest exchanging NHs for barnase, CI2, and LB1 are listed in Table 1). In frames for src SH3, CheY, and λ_{6-85}, arrows indicate off-scale ϕ values. For CI2, open circles are calculated $S_{3^{\circ}}$ values from Shoemaker et al. (1997), and open squares are calculated S values from Daggett et al. (1996). For src SH3 domain, $\phi_{F,eq}$ are plotted (Grantcharova et al., 1998). For CheY (López-Hernández & Serrano, 1996), experimental ϕ values for WT are closed diamonds and those for mutant F14N are closed triangles. Citations for other ϕ values, and for slowest exchanging NHs, are given in the text. Lists of calculated values for CI2 and λ_{6-85} were kindly provided by the authors (Shoemaker et al., 1997; Portman et al., 1998). Note: CI2 residue numbering begins after the first 19 N-terminal amino acids as in Itzhaki et al. (1995). CI2 secondary structural assignments are from the PDB coordinate file 2CI2, and are different from those in Itzhaki et al. (1995).

those with highest protection factors are shown by black stripes in Figure 4. However, helix A is clearly the slowest exchanging of all, as it has more NHs in the "slowest exchanging" category, and it has the four very slowest NHs in the molecule (Lacroix et al., 1997). The ϕ values plotted for CheY are for wild-type and a stabilized mutant. In assessing the correlation between ϕ values and slow exchange, the authors state that our suggestion of an empirical rule that slow exchange cores could be refolding cores is

"globally respected," although αE does not fit (López-Hernández & Serrano, 1996; Lacroix et al., 1997).

For LB1, the exchange data are shown in Figure 3A, and the 14 slowest out-exchanging NHs are shown again as black bars in Figure 4 (Yi & Baker, 1996; Yi et al., 1997). Four ϕ values are published, and these are particularly interesting in showing that the first turn (between β1 and β2) has high ϕ values, but the second turn (between β3 and β4) does not (Gu et al., 1997). Neither turn has the very slowest exchanging NHs in the native protein. The authors note that NHs in the first β-hairpin and helix exchange more slowly than NHs in the second hairpin and point out the correlation between hydrogen exchange and ϕ value results. The nucleating turns appear to connect core strands.

In λ_{6-85}, the first and fourth helices have the slowest exchanging amides (Hilser et al., 1998). Although the available number of experimental ϕ values are limited (Burton et al., 1997), both experimental and calculated ϕ values (Portman et al., 1998) generally fit well with hydrogen exchange data. That is, α1 and α4 have higher ϕ values than α2 and α3. However, α5, which is not a slow exchange core element, does have high ϕ values.

In summary, for a few proteins, published data permit comparison of slow exchange core elements and ϕ values. For the limited number of cases available, we conclude that slow exchange core elements tend to have residues with highest ϕ values, or be connected by turns/loops whose side chains have highest ϕ values. Agreement is not absolute, but a trend is clear. Another suggestive tendency is discernable in Figure 4: when core elements are helices, high ϕ values are within the element; when core elements are β-strands, high ϕ values are in turns/loops connected to the element.

What does a correlation between slow exchange core elements and high ϕ elements mean? Slowest exchange in native and partially folded states, highest P/C protection rates, and highest ϕ values are all qualitative pointers to core sequences. Out-exchange and P/C protection on one hand, and ϕ values on the other, provide complementary information about folding; the former from backbone groups that fold into secondary structure, and the latter from side chains that fold into loops/turns as well as secondary structure. We view high ϕ values as indicators of side-chain participation in interactions that are important in rate limiting steps of folding and make no further assumptions vis-à-vis the structures of the transition states. The folding core, defined as the out-exchange and P/C core (Woodward, 1993), has not been suggested to specify the nucleation site(s). If nucleation is the folding mechanism, then out-exchange and P/C cores are proximal products of nucleation.

Reevaluation of previous conclusions

In earlier papers, we introduced the phrase "folding core," suggested a correspondence between the slow exchange core and the folding core, and discussed the implications. The latter are reexamined in light of more extensive data in the following paragraphs; sentences from Woodward (1993) are in quotation marks.

1. "Folding is imprinted on the native structure. That is, a property of the folded state, internal mobility, indicates important events during folding." This is still the most intriguing result. There is no a priori expectation that measurements made on the final folded state identify features of the process by which that state is acquired. While intuitively comfortable, it is not predicted that the most stable region of the native state consists of sequences that most favor organized structure in early steps of folding. To point out that this apparently does occur is not a confusion of kinetics and thermodynamics, which in any case are related in complex processes of high dimensionality, such as protein folding. Rather, we propose that it reveals a new aspect of proteins, and its explanation may be part of a solution to the folding problem.

2. "Since the flexible loops are expected to fold last, the folding pathway would correspond approximately to the reverse order of out-exchange rates of NHs in the folded protein." An *approximate* correspondence is also apparent in the newer database: core elements tend to be the first to fold and they contain NHs with the lowest exchange rates when folding is completed; flexible loops tend to fold last and to contain NHs with highest exchange rates after folding is completed; other parts of the protein are in between. Specifics of the order of folding of secondary structural elements, and the order of out-exchange rates, are detailed for each protein in Supplementary material in Electronic Appendix. Briefly, (1) considering *only* core elements, generally there is good, but not perfect, correspondence between the order of exchange among slow exchange core elements, exchange among PF elements, and protection among P/C elements; (2) in cases where one slow exchange core element has slower exchanging NHs than all others, and/or a larger number of protons in the slowest exchanging category, that element is always the slowest exchanging PF element and the earliest protected P/C element; and (3) folding of *non*core secondary structural elements is not correlated with out-exchange rates. In some cases, NHs in secondary structural elements outside the slow exchange core, yet still rather slowly exchanging, are protected late during folding. For example, in β-strands of HEWL some amides have slower exchange rates than in helix D, but they are protected after helix D during folding.

3. "Protein segments in the slow exchange core determine the basic fold of the protein. Therefore, fragments or peptides corresponding to the slow exchange core should show native like structure." We still expect that the conformational ensemble of such peptides will favor native like structure, providing they do not have extensive self-aggregation.

4. "Methods for computer-assisted modeling of protein folding might usefully incorporate knowledge of the slow exchange core as an indicator of long-range interactions important in initial stages of protein folding . . . Identification of slow exchange core NHs in protein data bases could aid modeling of protein folding." There is a strong likelihood that knowledge of folding cores can be useful in computer-based predictions of tertiary structure from new sequences. The compilation of a useful database of slow exchange core elements is being aided by the increasing availability of hydrogen exchange rates for larger and more complex proteins from mass spectroscopy (Woodward, 1993; Chung et al., 1997; Smith et al., 1997; Deng & Smith, 1998; Matagne et al., 1998; Arrington et al., 1999).

5. "Partially folded proteins, and compact denatured states, are expected to be collapsed in regions corresponding to the slow exchange core." This is the strong trend for PF species in Table 1 and Figure 3. Regarding partially folded proteins, we do not consider the question of whether they are a side product vs. a true folding intermediate to be an critical distinction. For

the "new view" of folding on a funnel shaped rugged energy landscape (Dill & Chan, 1997), dichotomizing on- vs. off-pathway intermediates, like the "Levinthal paradox," is not a productive way to circumscribe the issue.

6. "Since out-exchange experiments are far less complex than refolding experiments, they would provide a simple way to identify the folding core." In our view, slow exchange core elements, P/C elements, PF elements, and ϕ values are all qualitative pointers to the folding core.

7. Our suggestion that the slow exchange core is the folding core is *not* correctly described as a "last out, first in hypothesis" as suggested elsewhere (Englander, 1998). The slowest exchanging protons identify a core that first favors nonlocal structure during folding; the order of folding vs. exchange rates of noncore elements of secondary structure is not addressed (cf. 2 above).

Implications of hydrogen exchange for protein folding

Several aspects of native structure suggest the existence of an essential core that endows a protein with its characteristic fold. With few exceptions, there is one fold type for each protein; alternatives are not observed. Amino acid substitutions often alter stability and/or structure around the substitution, but not the basic fold. Comparisons of evolutionarily related proteins demonstrate that the genetic message for a particular fold is not evenly encoded along the amino acid sequence, but rather is conserved in stretches of amino acids often far apart in the primary structure. Other parts of the protein may be highly modified or deleted altogether without significantly altering the basic native fold. Although the notion of a protein core as a three-dimensional, submolecular domain is a widely held concept, there is not a generally accepted, explicit definition. A common description of a hydrophobic core is a nonlocal cluster of buried hydrophobic side chains in the crystal structure. Allowed substitutions in such a core are inferred from the tolerance of protein folding to amino acid replacements (Bowie et al., 1990; Gassner et al., 1996). A somewhat different, structurally conserved core of protein families is identified in the Dali database (Holm & Sander, 1998).

We propose a way to identify the essential protein core: sequences of about 3–10 residues that tend to (1) be part or all of a secondary structural element in the native state; (2) cluster nonlocally in the native state; (3) have NHs that are last to exchange by the folded state mechanism; (4) have NHs that are highly protected from exchange early during folding; (5) have the most ordered structure in partially folded proteins; and (6) contain side chains with high ϕ values or be flanked by a turn/loop with high ϕ values. The fundamental conclusion is that there *are* core sequences, nonadjacent stretches of residues for which ordered structure and mutual nonlocal interactions are most stable in the native state *and* during folding.

The core defined in this way is also expected to be a hydrophobic core; however, the well-packed interiors of proteins usually have several regions with extensive nonlocal, apolar contacts, but not all of these are in the core (e.g., two hydrophobic clusters in BPTI are discussed in the literature, but only one is in the core). The folding core, as coined in Woodward (1993) and used here, is not a nucleation site. If folding is nucleated, the folding core is the ensemble of proximal products of nucleation. Turns that may be nucleation sites are not presently identified in core elements, because the secondary structure we consider thus far is limited to sheet and helix, and because in native proteins turns/loops are usually more flexible and faster exchanging. Nucleating turns/loops connect core elements; they are expected to have high ϕ values, to contain NHs not in the very slowest exchanging or very first protected groups, and to be flanked on one or both sides by core elements.

Core sequences for which ordered structure and mutual nonlocal interactions are most stable in the native state and during folding are consistent with theoretical models for folding on a rugged energy landscape (Thirumalai & Woodson, 1996; Dill & Chan, 1997; Karplus, 1997; Onuchic et al., 1997), and with other models. Ideas from landscape models, sometimes called funnel models, have shifted mechanistic conceptualizations of folding away from specific obligate intermediates, to ensembles of interconverting conformations and multiple parallel pathways. During folding, ordered structure is more appropriately described as favored (in some fraction of the interconverting conformers), rather than formed (in all/most conformers). In the language of ensembles, *if the slow exchange core is the folding core, then over much of the energy landscape, as well as in the native state, interactions between core sequences are favored more than other possible nonlocal interactions*. Further, in one model of a developing core, partially folded BPTI, there are site specific energy barriers between microstates, one disordered, and the other(s) more ordered.

In the folding core scenario, the early core develops as a large family of interconverting conformers in which core sequences are more organized and noncore sequences are more disordered. Interactions between core sequences are more probable (stable) than other nonlocal interactions, but individual conformers vary in types and extent of interactions between core sequences. In the developing core, the more stable interactions between core sequences are expected to be native-like much of the time, but the idea does not exclude the possibility of nonnative structure at very early stages. In one model for an early developing core, partially folded BPTI produced by retaining a native disulfide between two loops, the more ordered regions are *not* near the disulfide cross-bridge, but rather are in core sequences. Our explanation is that the interloop disulfide eliminates more extended conformations from the unfolded distribution and thereby shifts the ensemble to more collapsed species, among which core interactions are favored. Further, this system reveals a new facet of developing cores. At *each* NH, two (and in one case three) microstates are detected as two (or three) NMR cross peaks for the same NH. One microstate, P_d, has characteristics of a "random coil," while the other, P_f, is more ordered. P_d and P_f are each composed of multiple, rapidly interconverting conformers. Slow P_d–P_f interconversion (millisecond or longer) means that *there is an energy barrier between two (and in one case three) microstates*. For NHs outside the core, P_f is not native-like; for NHs in the core, P_f is native-like. The relative population of P_d and P_f and their interconversion rates vary for different NHs throughout the molecule, indicating that P_d–P_f fluctuations are local and segmental (Barbar et al., 1995, 1998).

Signs of life that distinguish natural proteins from other polypeptides include folding into a native state, formation of a biological active site, and very slow exchange of some interior NHs. The latter properties are derivatives of the first, and a slow exchange core is diagnostic of a true native state. The wide range of folded state exchange rates reflects the variation in stability of intramolecular interactions within a native protein; the very slowest exchang-

ing NHs point to the regions of most favorable tertiary interactions. Protein-sized polypeptides of natural amino acids in random or designed sequences are readily produced by chemical or biological methods; typically they do not acquire native states comparable to natural proteins although recent advances in this area have been made (Walsh et al., 1999). Even when intramolecular collapse out competes intermolecular aggregation, large polypeptides with de novo designed sequences are often multiple, equilibrating conformations of similar energy but dissimilar structure. They may tend to favor target structure to some extent, but do not form a native state of significantly lower energy than all other compact states accessible to that peptide chain. At the practical level, the slowest exchanging amide protons show us where to begin identifying and designing core sequences, as a step toward production of new polypeptides that acquire a true native state. Other extensions of the idea are that core elements are the most likely to remain native-like in partially folded disease-related proteins, and the least likely to undergo specific rearrangements like domain swapping. As exchange data become available for more proteins, the identification of sequences with a high probability of being in the core could aid protein structure prediction in the emerging fields of structural genomics and proteomics (Montelione & Anderson, 1999; Woodward, 1999).

Supplementary material in Electronic Appendix

Contents: one text section, two tables, and one figure. The text contains individual summaries for proteins in Table 1. One table (filename "foldtime.txt") compiles folding times for pulsed-labeling experiments. The second table (filename "otherpro.txt") summarizes additional proteins for which partial hydrogen exchange data are available. A color figure summarizes the data from Table 1 for human and guinea pig α-lactalbumin, holo N, and apo PF equine lysozyme and tendamistat.

Acknowledgments

We thank D.W. Bolen for RNase A data; Leisha Mullins, C.N. Pace, and Frank Raushel for RNase T1 data; John Markley for revised out-exchange rates of SNase; Yuji Goto for β-lactoglobulin data; Kazuhiko Yamasaki for out-exchange rates of wild-type RNase H; Susan Marqusee for RNase H* data; Christian Freund and Andreas Plückthun for scFv fragment data; Qian Yi and David Baker for protein LB1 data; Chin Yu for CTX III data; Carlo van Meirlo for flavodoxin data; Ben Shoemaker, Shoji Takada, and Peter Wolynes for calculated ϕ values. This work was supported by NIH (GM26242 and GM51628).

References

Alonso DO, Daggett V. 1995. Molecular dynamics simulations of protein unfolding and limited refolding: Characterization of partially unfolded states of ubiquitin in 60% methanol and in water. *J Mol Biol 247*:501–520.

Anderson DE, Lu J, McIntosh L, Dahlquist FW. 1993. T4 lysozyme. In: Clore GM, Gronenborn AM, eds. *NMR of proteins*. Boca Raton: CRC Press. pp 258–304.

Arcus VL, Vuilleumier S, Freund SM, Bycroft M, Fersht AR. 1995. A comparison of the pH, urea, and temperature-denatured states of barnase by heteronuclear NMR: Implications for the initiation of protein folding. *J Mol Biol 254*:305–321.

Arrington C, Teesch L, Robertson A. 1999. Defining protein ensembles with native-state NH exchange: Kinetics of interconversion and cooperative units from combined NMR and MS analysis. *J Mol Biol 285*:1265–1275.

Arrington CB, Robertson AD. 1997. Microsecond protein folding kinetics from native-state hydrogen exchange. *Biochemistry 36*:8686–8691.

Bai Y, Karimi A, Dyson HJ, Wright PE. 1997. Absence of a stable intermediate on the folding pathway of protein A. *Protein Sci 6*:1449–1457.

Bai Y, Milne JS, Mayne L, Englander SW. 1993. Primary structure effects on peptide group hydrogen exchange. *Proteins Struct Funct Genet 17*:75–86.

Barbar E, Barany G, Woodward C. 1995. Dynamic structure of a highly ordered β-sheet molten globule: Multiple conformations with a stable core. *Biochemistry 34*:11423–11434.

Barbar E, Hare M, Daragan V, Barany G, Woodward C. 1998. Dynamics of the conformational ensemble of partially folded bovine pancreatic trypsin inhibitor. *Biochemistry 37*:7822–7833.

Bond CJ, Wong K-B, Clarke J, Fersht AR, Daggett V. 1997. Characterization of residual structure in the thermally denatured state of barnase by simulation and experiment: Description of the folding pathway. *Proc Natl Acad Sci USA 94*:13409–13413.

Bowie JU, Reidhaar-Olson JF, Lim WA, Sauer RT. 1990. Deciphering the message in protein sequences: Tolerance to amino acid substitutions. *Science 247*:1306–1310.

Briggs MS, Roder H. 1992. Early hydrogen-bonding events in the folding reaction of ubiquitin. *Proc Natl Acad Sci USA 89*:2017–2021.

Buck M, Radford SE, Dobson CM. 1993. A partially folded state of hen egg white lysozyme in trifluoroethanol: Structural characterization and implications for protein folding. *Biochemistry 32*:669–678.

Buck M, Radford SE, Dobson CM. 1994. Amide hydrogen exchange in a highly denatured state. Hen egg-white lysozyme in urea. *J Mol Biol 237*:247–254.

Burton RE, Huang GS, Daugherty MA, Calderone TL, Oas TG. 1997. The energy landscape of a fast-folding protein mapped by Ala → Gly substitutions. *Nat Struct Biol 4*:305–310.

Bycroft M, Matouschek A, Kellis JT Jr, Serrano L, Fersht AR. 1990. Detection and characterization of a folding intermediate in barnase by NMR. *Nature 346*:488–490.

Chamberlain AK, Handel TM, Marqusee S. 1996. Detection of rare partially folded molecules in equilibrium with the native conformation of RNase H. *Nat Struct Biol 3*:782–787.

Chung EW, Nettleton EJ, Morgan CJ, Gross M, Miranker A, Radford SE, Dobson CM, Robinson CV. 1997. Hydrogen exchange properties of proteins in native and denatured states monitored by mass spectrometry and NMR. *Protein Sci 6*:1316–1324.

Chyan CL, Wormald C, Dobson CM, Evans PA, Baum J. 1993. Structure and stability of the molten globule state of guinea-pig α-lactalbumin: A hydrogen exchange study. *Biochemistry 32*:5861–5891.

Clarke J, Fersht AR. 1996. An evaluation of the use of hydrogen exchange at equilibrium to probe intermediates on the protein folding pathway. *Folding Design 1*:243–254.

Clarke J, Itzhaki LS, Fersht AR. 1997. Hydrogen exchange at equilibrium: A short cut for analysing protein-folding pathways? *Trends Biochem Sci 22*:284–287.

Dabora JM, Pelton JG, Marqusee S. 1996. Structure of the acid state of *Escherichia coli* ribonuclease HI. *Biochemistry 35*:11951–11958.

Daggett V, Li A, Itzhaki LS, Otzen DE, Fersht AR. 1996. Structure of the transition state for folding of a protein derived from experiment and simulation. *J Mol Biol 257*:430–440.

Deng Y, Smith DL. 1998. Identification of unfolding domains in large proteins by their unfolding rates. *Biochemistry 37*:6256–6262.

Dill KA, Chan HS. 1997. From Levinthal to pathways to funnels. *Nat Struct Biol 4*:10–19.

Driscoll PC, Gronenborn AM, Wingfield PT, Clore GM. 1990. Determination of the secondary structure and molecular topology of interleukin-1β by use of two- and three-dimensional heteronuclear ^{15}N-^{1}H NMR spectroscopy. *Biochemistry 29*:4668–4682.

Elöve GA, Bhuyan AK, Roder H. 1994. Kinetic mechanism of cytochrome *c* folding: Involvement of the heme and its ligands. *Biochemistry 33*:6925–6935.

Englander SW. 1998. Native-state HX. *Trends Biochem Sci 23*:378.

Englander SW, Kallenbach NR. 1984. Hydrogen exchange and structural dynamics of proteins and nucleic acids. *Q Rev Biophys 16*:521–655.

Englander SW, Mayne L. 1992. Protein folding studied using hydrogen-exchange labeling and two-dimensional NMR. *Ann Rev Biophys Biomol Struct 21*:243–265.

Fan P, Bracken C, Baum J. 1993. Structural characterization of monellin in the alcohol-denatured state by NMR: Evidence for β-sheet to α-helix conversion. *Biochemistry 32*:1573–1582.

Fersht AR. 1998. A reply to Englander and Woodward. *Trends Biochem Sci 23*:379–381.

Fersht AR. 1993. Protein folding and stability: The pathway of folding of barnase. *FEBS Lett 325*:5–16.

Fersht AR, Matouschek A, Serrano L. 1992. The folding of an enzyme I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol 224*:771–782.

Freund C, Gehrig P, Holak TA, Plückthun A. 1997. Comparison of the amide

proton exchange behavior of the rapidly formed folding intermediate and the native state of an antibody scFv fragment. *FEBS Lett 407*:42–46.
Freund C, Honegger A, Hunziker P, Holak TA, Plückthun A. 1996. Folding nuclei of the scFv fragment of an antibody. *Biochemistry 35*:8457–8464.
Gassner NC, Baase WA, Matthews BW. 1996. A test of the jigsaw puzzle model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. *Proc Natl Acad Sci USA 93*:12155–12158.
Gladwin ST, Evans PA. 1997. Structure of very early protein folding intermediates: New insights through a variant of hydrogen exchange labelling. *Folding Design 1*:407–417.
Grantcharova VP, Baker D. 1997. Folding dynamics of the src SH3 domain. *Biochemistry 36*:15685–15692.
Grantcharova VP, Riddles DS, Santiago JV, Baker D. 1998. Important role of hydrogen bonds in the structurally polarized transition state for folding of the src SH3 domain. *Nat Struct Biol 5*:714–720.
Gregory RB, Dinh A, Rosenberg A. 1986. The effect of Tri-N-acetylglucosamine on hydrogen exchange in hen egg white lysozyme. *J Biol Chem 261*:13963–13968.
Gu H, Kim D, Baker D. 1997. Contrasting roles for symmetrically disposed β-turns in the folding of a small protein. *J Mol Biol 274*:588–596.
Hamada D, Goto Y. 1997. The equilibrium intermediate of β-lactoglobulin with non-native α-helical structure. *J Mol Biol 269*:479–487.
Harding MM, Williams DH, Woolfson DN. 1991. Characterization of a partially denatured state of a protein by two-dimensional NMR: Reduction of the hydrophobic interactions in ubiquitin. *Biochemistry 30*:3120–3128.
Hilser VJ, Dowdy D, Oas TG, Freire E. 1998. The structural distribution of cooperative interactions in proteins: Analysis of the native state ensemble. *Proc Natl Acad Sci USA 95*:9903–9908.
Hilton BD, Trudeau K, Woodward CK. 1981. Hydrogen exchange rates in pancreatic trypsin inhibitor are not correlated to thermal stability in urea. *Biochemistry 20*:4697–4703.
Holm L, Sander C. 1998. Touring protein fold space with Dali/FSSP. *Nucleic Acids Res 26*:316–319.
Hughson FM, Wright PE, Baldwin RL. 1990. Structure characterization of a partly folded apomyoglobin intermediate. *Science 249*:1544–1548.
Hvidt A, Nielsen SO. 1966. Hydrogen exchange in proteins. *Adv Protein Chem 21*:288–386.
Itzhaki LS, Neira JL, Fersht AR. 1997. Hydrogen exchange in chymotrypsin inhibitor 2 probed by denaturants and temperature. *J Mol Biol 270*:89–98.
Itzhaki LS, Otzen DE, Fersht AR. 1995. The structure of the transition state for folding of chymotrypsin inhibitor 2 analysed by protein engineering methods: Evidence for a nucleation-condensation mechanism for protein folding. *J Mol Biol 254*:260–288.
Jackson SE, elMasry N, Fersht AR. 1993. Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor 2: A critical test of the protein engineering method of analysis. *Biochemistry 32*:11270–11278.
Jacobs MD, Fox RO. 1994. Staphylococcal nuclease folding intermediate characterized by hydrogen exchange and NMR spectroscopy. *Proc Natl Acad Sci USA 91*:449–453.
Jeng M, Englander SW, Elöve GA, Wand AJ, Roder H. 1990. Structural description of acid-denatured cytochrome *c* by hydrogen exchange and 2D NMR. *Biochemistry 29*:10433–10437.
Jennings PA, Wright PE. 1993. Formation of a molten globule intermediate early in the kinetic folding pathway of apomyoglobin. *Science 262*:892–896.
Karplus M. 1997. The Levinthal paradox: Yesterday and today. *Folding Design 2*:S69–S75.
Kim KS, Fuchs JA, Woodward CK. 1993. Hydrogen exchange identifies native-state motional domains important in protein folding. *Biochemistry 32*:9600–9608.
Kim KS, Woodward C. 1993. Protein internal flexibility and global stability: Effect of urea on hydrogen exchange rates of bovine pancreatic trypsin inhibitor. *Biochemistry 32*:9609–9613.
Kraulis PJ. 1991. MOLSCRIPT: A program to produce both detailed and schematic plots of protein structures. *J Appl Crystallogr 24*:946–950.
Kuszewski J, Clore GM, Gronenborn AM. 1994. Fast folding of a prototypic polypeptide: The immunoglobulin binding domain of streptococcal protein G. *Protein Sci 3*:1945–1952.
Kuwajima K. 1989. The molten globule state as a clue for understanding the folding and cooperativity of globular-protein structure. *Proteins Struct Funct Genet 6*:87–103.
Kuwajima K. 1996. The molten globule state of α-lactalbumin. *FASEB J 10*:102–109.
Lacroix E, Bruix M, López-Hernández E, Serrano L, Rico M. 1997. Amide hydrogen exchange and internal dynamics in the chemotactic protein CheY from *Escherichia coli*. *J Mol Biol 271*:472–487.
Lazaridis T, Karplus M. 1997. "New view" of protein folding reconciled with the old through multiple unfolding simulations. *Science 278*:1928–1931.
Loh SN, Kay MS, Baldwin RL. 1995. Structure and stability of a second molten globule intermediate in the apomyoglobin folding pathway. *Proc Natl Acad Sci USA 92*:5446–5450.
Loh SN, Prehoda KE, Wang J, Markley JL. 1993. Hydrogen exchange in unligated and ligated staphylococcal nuclease. *Biochemistry 32*:11022–11028.
López-Hernández E, Serrano L. 1996. Structure of the transition state for folding of the 129 aa protein CheY resembles that of a smaller protein, CI-2. *Folding Design 1*:43–55.
Lu J, Dahlquist FW. 1992. Detection and characterization of an early folding intermediate of T4 lysozyme using pulsed hydrogen exchange and two-dimensional NMR. *Biochemistry 31*:4749–4756.
Lumry R. 1991. A study of enzymes, Volume 2. In: Kuby SA, ed. *Mechanism of enzyme action*. Boca Raton: CRC Press. pp 3–81.
Lumry R. 1995. On the interpretation of data from isothermal processes. *Methods Enzymol 259*:628–720.
Lumry R, Rosenberg A. 1975. The mobile defect hypothesis of protein function. *Col Int CNRS L'Eau Syst Biol 246*:55–63.
Matagne A, Chung EW, Ball LJ, Radford SE, Robinson CV, Dobson CM. 1998. The origin of the α-domain intermediate in the folding of hen lysozyme. *J Mol Biol 277*:997–1005.
Matouschek A, Serrano L, Meiering EM, Bycroft M, Fersht AR. 1992. The folding of an enzyme. V. $H/^2H$ exchange-nuclear magnetic resonance studies on the folding pathway of barnase: Complementarity to and agreement with protein engineering studies. *J Mol Biol 224*:837–845.
Miller DW, Dill KA. 1995. A statistical mechanical model for hydrogen exchange in globular proteins. *Protein Sci 4*:1860–1873.
Miranker A, Radford SE, Karplus M, Dobson CM. 1991. Demonstration by NMR of folding domains in lysozyme. *Nature 349*:633–636.
Montelione GT, Anderson S. 1999. Structural genomics: Keystone for a human proteome project. *Nature Struct Biol 6*:11–12.
Mori S, van Zijl PC, Shortle D. 1997. Measurement of water-amide proton exchange rates in the denatured state of staphylococcal nuclease by a magnetization transfer technique. *Proteins Struct Funct Genet 28*:325–332.
Morozova LA, Haynie DT, Arico-Muendel C, Van Dael H, Dobson CM. 1995. Structural basis of the stability of a lysozyme molten globule. *Nature Struct Biol 2*:871–875.
Morozova-Roche LA, Arico-Muendel CC, Haynie DT, Emelyanenko VI, Van Dael H, Dobson CM. 1997. Structural characterisation and comparison of the native and A-states of equine lysozyme. *J Mol Biol 268*:903–921.
Mullins LS, Pace CN, Raushel FM. 1993. Investigation of ribonuclease T1 folding intermediates by hydrogen-deuterium amide exchange—Two-dimensional NMR spectroscopy. *Biochemistry 32*:6152–6156.
Mullins LS, Pace CN, Raushel FM. 1997. Conformational stability of ribonuclease T1 determined by hydrogen-deuterium exchange. *Protein Sci 6*:1387–1395.
Nash D, Lee BS, Jonas J. 1996. Hydrogen-exchange kinetics in the cold denatured state of ribonuclease A. *Biochim Biophys Acta 1297*:40–48.
Nash DP, Jonas J. 1997. Structure of pressure-assisted cold denatured lysozyme and comparison with lysozyme folding intermediates. *Biochemistry 36*:14375–14383.
Neira JL, Itzhaki LS, Otzen DE, Davis B, Fersht AR. 1997. Hydrogen exchange in chymotrypsin inhibitor 2 probed by mutagenesis. *J Mol Biol 270*:99–110.
Onuchic JN, Luthey-Schulten Z, Wolynes PG. 1997. Theory of protein folding: The energy landscape perspective. *Ann Rev Phys Chem 48*:545–600.
Orban J, Alexander P, Bryan P, Khare D. 1995. Assessment of stability differences in the protein G B1 and B2 domains from hydrogen-deuterium exchange: Comparison with calorimetric data. *Biochemistry 34*:15291–15300.
Pan H, Barbar E, Barany G, Woodward C. 1995. Extensive non-random structure in reduced and unfolded bovine pancreatic trypsin inhibitor. *Biochemistry 34*:13974–13981.
Pan Y, Briggs MS. 1992. Hydrogen exchange in native and alcohol forms of ubiquitin. *Biochemistry 31*:11405–11412.
Perrett S, Clarke J, Hounslow AM, Fersht AR. 1995. Relationship between equilibrium amide proton exchange behavior and the folding pathway of barnase. *Biochemistry 34*:9288–9298.
Portman JJ, Takada S, Wolynes PG. 1998. Variational theory for site resolved protein folding free energy surfaces. *Phys Rev Lett 81*:5237–5240.
Radford SE, Buck M, Topping KD, Dobson CM, Evans PA. 1992a. Hydrogen exchange in native and denatured states of hen egg-white lysozyme. *Proteins Struct Funct Genet 14*:237–248.
Radford SE, Dobson CM, Evans PA. 1992b. The folding of hen lysozyme involves partially structured intermediates and multiple pathways. *Nature 358*:302–307.
Raschke TM, Marqusee S. 1997. The kinetic folding intermediate of ribonuclease H resembles the acid molten globule and partially unfolded molecules detected under native conditions. *Nature Struct Biol 4*:298–304.
Robertson AD, Baldwin RL. 1991. Hydrogen exchange in thermally denatured ribonuclease A. *Biochemistry 30*:9907–9914.

Roder H, Elöve GA, Englander SW. 1988. Structural characterization of folding intermediates in cytochrome *c* by H-exchange labelling and proton NMR. *Nature 335*:700–704.

Roder H, Wüthrich K. 1986. Protein folding kinetics by combined use of rapid mixing techniques and NMR observation of individual amide protons. *Proteins Struct Funct Genet 1*:34–42.

Rosenberg A, Chakravarti K. 1968. Studies of hydrogen exchange in proteins. The exchange kinetics of bovine carbonic anhydrase. *J Biol Chem 243*:5193–5201.

Schönbrunner N, Wey J, Engels J, Georg H, Kiefhaber T. 1996. Native-like β-structure in a trifluoroethanol-induced partially folded state of the all-β-sheet protein tendamistat. *J Mol Biol 260*:432–445.

Schulman BA, Redfield C, Peng Z-Y, Dobson CM, Kim PS. 1995. Different subdomains are most protected from hydrogen exchange in the molten globule and native states of human α-lactalbumin. *J Mol Biol 253*:651–657.

Serrano L, Matouschek A, Fersht AR. 1992. The folding of an enzyme III. Structure of the transition state for unfolding of barnase analyzed by a protein engineering procedure. *J Mol Biol 224*:805–818.

Shoemaker BA, Wang J, Wolynes PG. 1997. Structural correlations in protein folding funnels. *Proc Natl Acad Sci USA 94*:777–782.

Sivaraman T, Kumar TKS, Chang DK, Lin WY, Yu C. 1998. Events in the kinetic folding pathway of a small, all β-sheet protein. *J Biol Chem 273*:10181–10189.

Smith D, Deng Y, Zhang Z. 1997. Probing the non-covalent structure of proteins by amide hydrogen exchange and mass spectrometry. *J Mass Spectrom 32*:135–146.

Sosnick TR, Mayne L, Hiller R, Englander SW. 1994. The barriers in protein folding. *Nature Struct Biol 1*:149–156.

Stockman BJ, Euvrard A, Scahill TA. 1993. Heteronuclear three-dimensional NMR spectroscopy of a partially denatured protein: The A-state of human ubiquitin. *J Biomol NMR 3*:285–296.

Swint-Kruse L, Robertson AD. 1996. Temperature and pH dependences of hydrogen exchange and global stability for ovomucoid third domain. *Biochemistry 35*:171–180.

Thirumalai D, Woodson SA. 1996. Kinetics of folding of proteins and RNA. *Acc Chem Res 29*:433–439.

Tirado-Rives J, Orozco M, Jorgensen WL. 1997. Molecular dynamics simulations of the unfolding of barnase in water and 8 M aqueous urea. *Biochemistry 36*:7313–7329.

Tüchsen E, Woodward C. 1985. Hydrogen exchange kinetics of peptide amide protons at the bovine pancreatic trypsin inhibitor protein-solvent interface. *J Mol Biol 185*:405–419.

Udgaonkar JB, Baldwin RL. 1990. Early folding intermediate of ribonuclease A. *Proc Natl Acad Sci USA 87*:8197–8201.

Varley P, Gronenborn AM, Christensen H, Wingfield PT, Pain RH, Clore GM. 1993. Kinetics of folding of the all-β sheet protein interleukin-1β. *Science 260*:1110–1113.

Walsh ST, Cheng H, Bryson JW, Roder H, DeGrado WF. 1999. *Proc Natl Acad Sci USA 11*:5486–5491.

Wang A, Robertson AD, Bolen DW. 1995. Effects of a naturally occurring compatible osmolyte on the internal dynamics of ribonuclease A. *Biochemistry 34*:15096–15104.

Wang QW, Kline AD, Wüthrich K. 1987. Amide proton exchange in the alpha-amylase polypeptide inhibitor Tendamistat studied by two-dimensional 1H nuclear magnetic resonance. *Biochemistry 26*:6488–6493.

Woodward C. 1993. Is the slow exchange core the protein folding core? *Trends Biochem Sci 18*:359–360.

Woodward C. 1999. Advances in hydrogen exchange from mass spectrometry. *J Am Soc Mass Spectrom 10*:672–674.

Woodward C, Simon I, Tüchsen E. 1982. Hydrogen exchange and the dynamic structure of proteins. *Mol Cell Biochem 48*:135–160.

Woodward CK. 1994. Hydrogen exchange rates and protein folding. *Curr Opin Struct Biol 4*:112–116.

Woodward CK, Hilton BD. 1980. Hydrogen isotope exchange kinetics of single protons in bovine pancreatic trypsin inhibitor. *Biophys J 32*:561–575.

Woodward CK, Rosenberg A. 1971a. Studies of hydrogen exchange in proteins: The correlation of ribonuclease exchange kinetics with the temperature-induced transition. *J Biol Chem 246*:4105–4113.

Woodward CK, Rosenberg A. 1971b. Studies of hydrogen exchange in proteins: Urea effects on ribonuclease exchange kinetics leading to a general model for hydrogen exchange from folded proteins. *J Biol Chem 246*:4114–4121.

Yamasaki K, Ogasahara K, Yutani K, Oobatake M, Kanaya S. 1995. Folding pathway of *Escherichia coli* ribonuclease HI: A circular dichroism, fluorescence, and NMR study. *Biochemistry 34*:16552–16562.

Yi Q, Baker D. 1996. Direct evidence for a two-state protein unfolding transition from hydrogen-deuterium exchange, mass spectrometry, and NMR. *Protein Sci 5*:1060–1066.

Yi Q, Scalley ML, Simons KT, Gladwin ST, Baker D. 1997. Characterization of the free energy spectrum of peptostreptococcal protein L. *Folding Design 2*:271–280.

Yu H, Rosen MK, Schreiber SL. 1993. ^{1}H and ^{15}N assignments and secondary structure of the Src SH3 domain. *FEBS Lett 324*:87–92.

Zhang J, Peng X, Jonas A, Jonas J. 1995. NMR study of the cold, heat and pressure unfolding of ribonuclease A. *Biochemistry 34*:8631–8641.