Meyer, J., Gagnon, J., Sieker, L. C., van Dorsselaer, A., & Moulis, J.-M. (1990) *Biochem. J. 271*, 839.
Mukund, S., & Adams, M. W. W. (1990) *J. Biol. Chem. 265*, 11508–11516.
Mukund, S., & Adams, M. W. W. (1991) *J. Biol. Chem. 266*, 14208.
Oh, B.-H., & Markley, J. L. (1990) *Biochemistry 29*, 3993.
Papavassiliou, P., & Hatchikian, E. C. (1985) *Biochim. Biophys. Acta 810*, 1.
Paulptit, R. A., Karlsson, R., Picot, D., Jenkins, J. A., Niklaus-Reimer, A., & Jansonius, J. N. (1988) *J. Mol. Biol. 199*, 525.
Pinantini, U., Sorensen, O. W., & Ernst, R. R. (1982) *J. Am. Chem. Soc. 104*, 6800.
Saeki, K., Yao, Y., Wakabayashi, S., Shen, G.-J., Zeikus, J. G., & Matsubara, H. (1989) *J. Bacteriol. 171*, 4736–4741.
Seki, Y., Seki, S., Satoh, M., Ikeda, A., & Ishimoto, M. (1989) *J. Biochem. (Tokyo) 106*, 336.
Shimizu, F., Ogata, M., Yagi, T., Wakabayashi, S., & Matsubara, H. (1989) *Biochimie 71*, 1171.
Sieker, L. C., Stenkamp, R. E., Jensen, L. H., Prickril, B., & LeGall, J. (1986) *FEBS Lett. 208*, 73.
South, T. L., Blake, P. R., Sowder, R. C., III, Arthur, L. O., Henderson, L. E., & Summers, M. F. (1990) *Biochemistry 29*, 7786.
South, T. L., Blake, P. R., Hare, D. R., & Summers, M. F. (1991) *Biochemistry 30*, 6342.
States, D. J., Haberkorn, R. H., & Ruben, D. J. (1982) *J. Magn. Reson. 48*, 286–292.
Stetter, K. O. (1986) in *The Thermophiles: General, Molecular and Applied Microbiology* (Brock, T. D., Ed.) pp 39–74, John Wiley, New York.
Stetter, K. O., Fiala, G., Huber, G., Huber, R., & Segerer, G. (1990) *FEMS Microbiol. Rev. 75*, 117–124.
Summers, M. F., South, T. L., Kim, B., & Hare, D. R. (1990) *Biochemistry 29*, 329.
Voordouw, G. (1988) *Gene 69*, 75.
Watenpaugh, K. D., Sieker, L. C., Herriott, J. R., & Jensen, L. H. (1973) *Acta Crystallogr. B29*, 943.
Watenpaugh, K. D., Sieker, L. C., & Jensen, L. H. (1979) *J. Mol. Biol. 131*, 509.
Weber, K., Pringle, J. R., & Osborn, M. (1972) *Methods Enzymol. 26*, 2–27.
Werth, M. T., Kurtz, D. M., Jr., Moura, I., & LeGall, J. (1987) *J. Am. Chem. Soc. 109*, 273.
Woolley, K. J., & Meyer, T. E. (1987) *Eur. J. Biochem. 163*, 161–166.
Wuthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley, New York.
Yasunobu, K. T., & Tanaka, M. (1973) in *Iron–Sulfur Proteins* (Lovenberg, W., Ed.) Vol. II, pp 27–130, Academic Press, New York.
Yu, L. P., LaMar, G. N., & Rajarathan, K. (1990) *J. Am. Chem. Soc. 112*, 9527.
Zuber, H. (1988) *Biophys. Chem. 29*, 171.
Zwickl, P., Fabry, S., Bogedain, C., Haas, A., & Hensel, R. (1990) *J. Bacteriol. 172*, 4329.

Construction of a Synthetic Gene for an R-Plasmid-Encoded Dihydrofolate Reductase and Studies on the Role of the N-Terminus in the Protein†

Lisa J. Reece,‡ Robert Nichols,‡ Richard C. Ogden,§ and Elizabeth E. Howell*,‡
Department of Biochemistry, University of Tennessee, Knoxville, Tennessee 37996-0840, and The Agouron Institute, 505 Coast Boulevard South, La Jolla, California 92037

Received May 24, 1991; Revised Manuscript Received August 1, 1991

ABSTRACT: R67 dihydrofolate reductase (DHFR) is a novel protein that provides clinical resistance to the antibacterial drug trimethoprim. The crystal structure of a dimeric form of R67 DHFR indicates the first 16 amino acids are disordered [Matthews et al. (1986) *Biochemistry 25*, 4194–4204]. To investigate whether these amino acids are necessary for protein function, the first 16 N-terminal residues have been cleaved off by chymotrypsin. The truncated protein is fully active with k_{cat} = 1.3 s^{-1}, $K_{m(NADPH)}$ = 3.0 μM, and $K_{m(dihydrofolate)}$ = 5.8 μM. This result suggests the functional core of the protein resides in the β-barrel structure defined by residues 27–78. To study this protein further, synthetic genes coding for full-length and truncated R67 DHFRs were constructed. Surprisingly, the gene coding for truncated R67 DHFR does not produce protein in vivo or confer trimethoprim resistance upon *Escherichia coli*. Therefore, the relative stabilities of native and truncated R67 DHFR were investigated by equilibrium unfolding studies. Unfolding of dimeric native R67 DHFR is protein concentration dependent and can be described by a two-state model involving native dimer and unfolded monomer. Using absorbance, fluorescence, and circular dichroism techniques, an average ΔG_{H_2O} of 13.9 kcal mol^{-1} is found for native R67 DHFR. In contrast, an average ΔG_{H_2O} of 11.3 kcal mol^{-1} is observed for truncated R67 DHFR. These results indicate native R67 DHFR is 2.6 kcal mol^{-1} more stable than truncated protein. This stability difference may be part of the reason why protein from the truncated gene is not found in vivo in *E. coli*.

Dihydrofolate reductase (DHFR;[1] EC 1.5.1.3) catalyzes the NADPH-dependent reduction of 7,8-dihydrofolate (DHF) to 5,6,7,8-tetrahydrofolate. Since tetrahydrofolate is required for the synthesis of thymidylate, purine nucleosides, methionine, and other metabolic intermediates (Kraut & Matthews,

† This research was supported by NIH Grant GM35308 (to E.E.H.) and NSF Grant DMB8318244 (to R.C.O.).
‡ University of Tennessee.
§ The Agouron Institute.

[1] Abbreviations: DHFR, dihydrofolate reductase; TMP, trimethoprim; DHF, dihydrofolate; GdnHCl, guanidine hydrochloride.

0006-2960/91/0430-10895$02.50/0

1987), inhibition of chromosomally encoded DHFR results in blockage of DNA synthesis and ultimately cell death. Therefore, the active-site inhibitor trimethoprim (TMP), which selectively inhibits bacterial DHFR, has been used as a broad-spectrum antibiotic (in combination with sulfonamides).

Clinical resistance to TMP has been observed and is typically due to production of novel DHFRs encoded by R-plasmids (Amyes & Smith, 1974; Skold & Widh, 1974). Numerous studies have delineated at least seven groups of R-plasmid DHFRs designated type I through type VII (Pattishall et al., 1977; Fling et al., 1988; Sundstrom et al., 1987, 1988; Wylie et al., 1988; Amyes et al., 1989). For recent reviews, see Amyes (1989) and Huovinen (1987).

Type II DHFR is particularly interesting as it is genetically unrelated to chromosomal DHFR. Three variants of type II DHFRs have been found and named R67, R388, and R751 DHFR (Flensburg & Steen, 1986; Brisson & Hohn, 1984; Swift et al., 1981). Most of the sequence differences between these type II DHFRs cluster within 21 residues at the N-terminus and the last 2 residues of the C-terminus. Type II DHFRs confer virtually complete TMP resistance to host bacteria (K_i = 0.15 mM; Amyes & Smith, 1976) even though the K_m for dihydrofolate is close to that for chromosomal *Escherichia coli* DHFR (5.8 vs 1.2 μM). The catalytic efficiency (k_{cat}/K_m) of R67 DHFR is decreased 100-fold when compared to *E. coli* chromosomal DHFR (Joyner et al., 1984; Smith & Burchall, 1983). Also, recent NMR studies of a R388 DHFR variant (Brito et al., 1990, 1991) suggest the NMNH moiety of NADPH binds to the active site in a syn conformation. This is in contrast to the anti conformations observed for NADPH binding in all known chromosomal DHFRs.

R67 DHFR has a molecular weight of approximately 34000 and is composed of four identical monomers (Smith et al., 1979). The monomer is 78 amino acids long, and it bears no sequence homology to chromosomal DHFR and shows no immunological cross-reactivity (Stone & Smith, 1979; Fling & Elwell, 1980).

X-ray crystal structures of dimeric and tetrameric forms of R67 DHFR have recently been obtained (Matthews et al., 1986; Matthews, personal communication). No homology between either the overall structure or the active-site area is observed when compared to chromosomal DHFR. In dimeric R67 DHFR, each monomer is essentially an up and down six-stranded β-barrel, except the fifth strand is missing. Because of interactions between β-strands, strands B, C, and D from one monomer associate with the equivalent strands in the second monomer and form a third β-barrel.

In the unpublished tetramer structure, an unusual pore passes through the center of the protein (Matthews et al., unpublished results). Difference density maps describing bound folate and/or NADPH indicate the center of the pore is the locus for both substrate and cofactor binding. Of especial interest is the implication that one active site exists per tetramer and that residues from each monomer contribute to binding in this single active site. While a shared active site between protomers is not surprising, the presence of only one active site per multimer is unusual. This unusual type of binding has also been seen in HIV protease, a dimer possessing a single active site (Wlodawer et al., 1989). Additionally, a single 2,3-bisphosphoglycerate molecule binds to the central pore in tetrameric deoxyhemoglobin (Arnone, 1972; Kilmartin, 1976), and (−)-carvone or 2-methoxy-4-isopropylpyrazine binds to the interface in dimeric olfactory binding protein (M. Amzel, personal communication).

Table I: A Comparison of Kinetic Parameters for N-Terminal Variants of R67 DHFR (pH 7.0)

enzyme	k_{cat} (s^{-1})	$K_{m(DHF)}$ (μM)	$K_{m(NADPH)}$ (μM)
native R67 DHFR	1.3 ± 0.073	5.8 ± 0.015	3.0 ± 0.060
chymotrypsin-truncated R67 DHFR	1.5 ± 0.12	7.2 ± 0.15	2.7 ± 0.030
synthetic Met-Ile R67 DHFR	1.1 ± 0.075	5.0 ± 0.004	2.6 ± 0.080
chromosomal *E. coli* DHFR	29 ± 1.0	1.1 ± 0.019	0.94 ± 0.36

When R67 DHFR is cleaved by chymotrypsin, the truncated protein (lacking 16 N-terminal residues) is fully active. To study this unusual enzyme, we have constructed synthetic genes coding for full-length and truncated R67 DHFRs. In this paper, we contrast the ability of the two synthetic R67 DHFR genes to confer trimethoprim resistance upon *E. coli* and describe folding studies designed to explore whether truncated R67 DHFR is less stable than native R67 DHFR.

Materials and Methods

DNA Manipulations. Oligonucleotides used to construct a synthetic R67 DHFR gene were synthesized on an Applied Biosystems Model 380A instrument. Standard cloning procedures were used to anneal the oligonucleotides, ligate the cassette into a vector, and transform the DNA into *E. coli* (Maniatis et al., 1982).

Single- and double-stranded DNA sequencing was performed using a Sequenase kit from United States Biochemical Corp. Primers used for dideoxy sequencing included a universal primer (Promega) as well as the synthetic oligonucleotides used to construct the gene. Two compressions were observed in the synthetic DNA sequence and were only readable using dITP reaction mixes from the Sequenase kit.

A gene coding for native R67 DHFR cloned in pUC4 (named P700) was the generous gift of Lynn Elwell (Wellcome Research Laboratories, Research Triangle Park, NC).

Protein Expression. Protein expression levels were evaluated by dividing the R67 DHFR concentration by the total soluble protein concentration. R67 DHFR concentrations were obtained by assaying the activity of sonicated cell lysates. A specific activity of 2.0 units/mg of protein was used (see Table I); 20 nM TMP was added to the assay to inhibit any chromosomal DHFR activity that might be present. The respective K_i's of R67 DHFR and *E. coli* chromosomal DHFR are 0.15 mM and 20 pM (Amyes & Smith, 1976; Stone & Morrison, 1986). Total protein concentrations were measured by a Bradford dye assay (Bio-Rad).

Antibodies were prepared by immunizing a New Zealand white rabbit with 70 μg of R67 DHFR (expressed as tetramer) according to Weaver et al. (1991). Western blots were performed by running SDS–PAGE gels, blotting the proteins onto a nitrocellulose filter, and probing for R67 DHFR expression with a 1:1000 dilution of sera. Experimental details were as described in Weaver et al. (1991). Preimmune sera was used as a control as it did not react with R67 DHFR. If R67 DHFR expression was not observed, plasmid DNA was transformed into an La protease deficient strain (CAG626, *lon*; C. A. Gross, personal communication) and expression reevaluated.

Protein Purification, Cleavage, and Kinetic Analysis. E. coli [strains SK383 and/or JM107 (Zieg et al., 1978; Yanish-Perron et al., 1985)] containing various cloned R67 DHFR genes were grown in a modified version of TB media (Tartof & Hobbs, 1987) containing 0.017 M KH_2PO_4, 0.07 M K_2HPO_4, 12 g/L Bactotryptone, 24 g/L Bacto-yeast extract, 4

mL/L glycerol, and 2 g/L NaCl. Cells were grown to late stationary phase in the presence of 200 μg of ampicillin/mL plus 50 μg of TMP/mL and were lysed by addition of 1 M NaOH until pH 12 was reached (Vermersch et al., 1986). Cell debris was removed by centrifugation, and the lysis solution was titrated to pH 2.0. Precipitated material was again removed and the solution returned to pH 7.4. After dialysis, the protein solution was applied to a 1.0 × 29 cm PBE chromatofocusing column (Pharmacia) equilibrated in 0.025 M imidazole, pH 7.4, and R67 DHFR was eluted using a 1:10 dilution of Polybuffer 74, pH 5.0 (Pharmacia). In a final purification step, the protein solution was chromatographed on a 2.5 × 88 cm G-75 Sephadex column. Protein was stored as lyophilized powder.

Contamination of R67 DHFR by chromosomally encoded DHFR from host *E. coli* is not a problem as chromosomally encoded DHFR does not survive the NaOH and HCl addition steps. Even if chromosomal DHFR did survive these steps, it does not coelute with R67 DHFR during PBE and G-75 column chromatography.

N-Terminal peptide sequencing of native R67 DHFR was performed by the Molecular Biology Research Facility at the University of Tennessee, and sequencing of the Met-Ile N-terminal variant was kindly done by Charles Murphy, University of Tennessee Medical Center, Knoxville.

A truncated version of R67 DHFR was prepared by mixing 10 mg of R67 DHFR with 5 mg of NADPH in 50 mM KH_2PO_4 + 1 mM EDTA, pH 8.0, buffer followed by addition of 10 mg of immobilized chymotrypsin (Sigma). This solution was mixed overnight at 4 °C and then mixed at room temperature for 6 h the next day. The extent of the reaction was monitored by SDS gel electrophoresis. Immobilized chymotrypsin was removed from the digestion mix by filtration through a 0.2-μm Acrodisc (Gelman), and the solution was dialyzed to remove NADPH plus short peptide fragments. Truncated R67 DHFR was readily separated from any uncut R67 DHFR using a PBE chromatofocusing column. The N-terminal sequence of the larger peptide fragment was provided courtesy of Brian Tack, Scripps Clinic, La Jolla, CA.

Cleavage of R67 DHFR by CNBr treatment in 70% formic acid was performed as previously described (Stone et al., 1979). The extent of the reaction was monitored by SDS–PAGE.

Unfolding Studies. Unfolding of native and truncated R67 DHFR by guanidine hydrochloride (GdnHCl) was measured using difference spectroscopy as previously described (Villafranca et al., 1987). Briefly, split cell cuvettes were used to monitor the tryptophan environment by changes in the absorbance at 290 nm. Unmixed spectra were obtained and subtracted from mixed spectra. The samples were allowed to equilibrate at 30 °C for ≥10 min prior to recording the mixed spectra. The buffer used was 50 mM acetic acid + 50 mM MES + 100 mM Tris buffer, pH 5.0 (Ellis & Morrison, 1982). GdnHCl concentrations were calculated using an equation relating refractive index to concentration (Pace et al., 1990) and refractive indexes were measured using a Zeiss refractometer. Unfolding was fully reversible as monitored by recovery of enzyme activity as well as UV absorbance.

Protein unfolding was also monitored by the loss of secondary structure at 217 nm using a Jasco J-40A spectropolarimeter with a cell path length of 1 cm (Ropson et al., 1990; Gittleman & Matthews, 1990). Additionally, protein unfolding was monitored by changes in the tryptophan fluorescence using a Perkin Elmer LS-5B fluorometer. Excitation of tryptophans at 290 nm (slit 3) allowed monitoring of the emission at 340 nm (slit 5) using a cell path length of 1 cm. These studies were performed at room temperature.

The experimental data were fit to a two-state model of unfolding:

$$D \rightleftharpoons 2U$$

where D is dimeric R67 DHFR species and U is unfolded monomer. The equilibrium constant describing this reaction is $K = [U]^2/[D]$. Gittleman and Matthews (1990), in a study of dimeric *trp* aporepressor unfolding, used this two-state model and derived an equation describing F_{app}, the fraction of unfolded form:

$$F_{app} = [(K^2 + 8K[P_{tot}])^{1/2} - K]/4[P_{tot}] \quad (1)$$

where $[P_{tot}]$ describes total protein concentration and equals 2[D] + [U]. According to Pace et al. (1990), two additional equations, $K = \exp(-\Delta G/RT)$ and $\Delta G = \Delta G_{H_2O} + M_G \times [GdnHCl]$, can be substituted into the above equation to calculate M_G and ΔG_{H_2O}. For these equations, R is the gas constant, T is the temperature, ΔG is the apparent free energy difference between folded and unfolded forms at a specific GdnHCl concentration, ΔG_{H_2O} is the free energy in the absence of GdnHCl, and M_G is the slope describing the dependence of ΔG_{H_2O} on GdnHCl concentration.

The observed changes in fluorescence, absorbance, or CD signals can be related to F_{app} by an equation from Pace et al. (1990):

$$F_{app} = \frac{Y_{obs} - Y_N}{Y_U - Y_N} \quad (2)$$

where Y_{obs} is the observed optical value at a particular GdnHCl concentration and Y_N and Y_U are the calculated values for the native and unfolded forms, respectively, at the same denaturant concentration. A linear dependence of Y_{obs} on GdnHCl concentration was observed in the pre- and posttransition base-line regions. Linear extrapolations from these base lines yielded concentration-independent values for Y_N and Y_U (i.e., Y_N^0 and Y_U^0). Concentration-dependent values for Y_N and Y_U can be calculated from the equations:

$$Y_N = Y_N^0 + M_N[GdnHCl] \quad (3a)$$

$$Y_U = Y_U^0 + M_U[GdnHCl] \quad (3b)$$

where M_N and M_U are the slopes of the pre- or posttransition phases (native and unfolded forms, respectively) (Santoro & Bolen, 1988).

All of the above equations were combined into one equation, and unfolding curves were fit using a nonlinear least-squares fit program (NLIN) in the Statistical Analysis Systems program package (Sas Institute Inc., Cary, NC). According to Santoro and Bolen (1988), this type of nonlinear fit provides a better estimate of the error for ΔG_{H_2O} as errors associated with deriving pre- and posttransition slopes will be incorporated into the ΔG_{H_2O} estimate. Equation 4 was used to fit the experimental data. If the standard error associated with M_N or M_U was larger than two standard deviations (i.e., 95% confidence interval), the value was removed from the equation and the fit redone. Goodness of fit was assessed as per Motulsky and

$$Y_{obs} = [\{[(Y_U + M_U[GdnHCl]) - (Y_N + M_N[GdnHCl])] \times [\{\exp[-(\Delta G_{H_2O} + M_G[GdnHCl])/RT]\}^2 + 8\{\exp[-(\Delta G_{H_2O} + M_G[GdnHCl])/RT]\}[P_{tot}]]^{1/2} - \{\exp[-(\Delta G_{H_2O} + M_G[GdnHCl])/RT]\}\}/4[P_{tot}]] + Y_N + M_N[GdnHCl] \quad (4)$$

Ransnas (1987). The midpoint of the denaturation curve was obtained by solving for [GdnHCl] when F_{app} was set equal to 0.5 using the computer program MAPLE, from the Symbolic Computation Group, Waterloo, Ontario.

Kinetic Analysis. Steady-state kinetic data were obtained as previously described (Howell et al., 1990). In brief, assays were performed at 30 °C in a polybuffer containing 50 mM acetic acid + 50 mM MES + 100 mM Tris + 10 mM β-mercaptoethanol, pH 7.0 (Ellis & Morrison, 1982). Both NADPH and dihydrofolate concentrations were varied at subsaturating levels to obtain k_{cat}, $K_{m(NADPH)}$, and $K_{m(DHF)}$ values. Addition of enzyme started the assay. Ranges of DHF and NADPH concentrations used were 4.3–23 and 4.2–31 μM, respectively.

Concentrations were determined spectrophotometrically using molar extinction coefficients of 28 000 M^{-1} cm^{-1} at 282 nm for DHF (Blakley, 1960) and 6220 M^{-1} cm^{-1} at 340 nm for NADPH (Penner & Frieden, 1985). The molar extinction coefficient for DHF utilization by DHFR was 12 300 M^{-1} cm^{-1} at 340 nm (Baccanari et al., 1975). Enzyme concentration was initially determined using the biuret procedure (Gornall et al., 1949). Native and truncated R67 DHFR concentrations were then routinely measured using extinction coefficients of 1.82 and 2.09 mg^{-1} cm^{-1}, respectively.

RESULTS

What Is the Minimal Functional Unit of R67 DHFR? In an initial set of experiments designed to define the minimum polypeptide length capable of maintaining catalytic activity, both enzymic and chemical cleavage procedures were used to produce truncated versions of R67 DHFR. Treatment of R67 DHFR with chymotrypsin yields a single cut in the primary sequence when the active site of the enzyme is protected with NADPH. Sequence analysis of the purified, large peptide fragment gave the unique N-terminal sequence Val-Phe-Pro-Ser-Asn-Ala indicating chymotrypsin cuts R67 DHFR uniquely after Phe16. While other potential chymotrypsin cleavage sites exist in R67 DHFR, for example, Phe18, they are apparently not accessible when NADPH is bound.

Chymotrypsin-truncated R67 DHFR retains full enzymic activity as shown in the kinetics section below. Truncated enzyme also maintains its tetrameric structure as monitored by molecular sieving chromatography (data not shown). Finally, X-ray crystallography studies using truncated R67 DHFR indicate it crystallizes as a tetrameric species (Matthews et al., unpublished results).

In contrast to chymotrypsin cleavage after Phe16, cleavage at the unique methionine residue (M26) in R67 DHFR by cyanogen bromide treatment in 70% formic acid results in an inactive peptide. Whether the inactivity of this peptide is due to an inability to refold or whether the refolded peptide was inactive was not investigated. Native R67 DHFR does partially survive formic acid treatment as 50% activity remains after 24-h exposure to 70% formic acid.

The above results indicate a short section of the N-terminus is not necessary for activity in R67 DHFR and allow us to bracket a cleavage point between Phe-16 and Met-26 as defining the smallest *functional* unit in R67 DHFR. Since the R67 DHFR monomer is only 78 amino acids long, removal of 16 amino acids from the N-terminus constitutes loss of 20% of the protein sequence.

Kinetic Characterization of R67 DHFR Variants. Kinetic analysis of native R67 DHFR at pH 7.0 gives a k_{cat} value of 1.3 ± 0.073 s^{-1}, a $K_{m(dihydrofolate)}$ of 5.8 ± 0.015 μM, and a $K_{m(NADPH)}$ of 3.0 ± 0.060 μM as shown in Table I. These values were obtained using subsaturating concentrations of both NADPH and DHF, and they confirm literature k_{cat} and K_m values (Pattishall et al., 1977; Smith & Burchall, 1983; Morrison & Sneddon, 1990). R67 DHFR is approximately 100-fold less efficient than chromosomal *E. coli* DHFR.

In Table I, the kinetic parameters for native R67 DHFR are compared to two N-terminal R67 DHFR variants as well as to chromosomal *E. coli* DHFR values. Both the Met-Ile R67 DHFR derived from the full-length synthetic gene and truncated R67 DHFR have almost identical values for k_{cat} and K_m, indicating the N-terminus does not participate in the catalytic reaction.

To determine whether the kinetic values were equivalent due to removal of several N-terminal residues, the native (i.e., Met-Glu) and synthetic (Met-Ile) R67 DHFRs were sequenced. For the Met-Ile variant, 72% of the protein population possessed Met at the N-terminus and 28% possessed Ile at the N-terminus. For native R67 DHFR, 97% of the protein population possessed Met at the N-terminus and 3% possessed Glu.

Synthetic Gene Construction. Two variants of a synthetic R67 DHFR gene have been assembled. The first variant codes for a shortened version of R67 DHFR where the point of truncation was determined by chymotrypsin treatment of the native protein. The second variant of the R67 DHFR gene codes for a full-length protein. Figure 1 summarizes the cloning steps involved in constructing these two genes.

A synthetic gene coding for a truncated version of R67 DHFR was constructed from 8 fully overlapping oligonucleotides ranging in length from 30 to 55 residues and possessing 9 base overlaps. Following purification by polyacrylamide gel electrophoresis, the 5′ ends of the oligonucleotides were phosphorylated (except for two oligos that were at the extreme 5′ ends of the cassette) and ligated. The gene cassette was then ligated into M13mp8 cut with *Eco*RI and *Bam*HI. Following transformation into JM103, clear plaques were screened with one of the internal oligonucleotides labeled at the 5′ terminus with ^{32}P. Of four candidates picked for sequencing, one had the desired sequence as determined by dideoxy sequencing. The resulting construct was named M13-sR67-20. The gene coding for truncated R67 DHFR was then cloned behind a constitutive, mutant (up) promoter for the *E. coli* chromosomal DHFR gene (Smith & Calvo, 1980, 1982). Specifically, a 193 bp *Sau*3A (partial digest)–*Eco*RI fragment from the truncated R67 DHFR gene was inserted into an approximately 3400 bp fragment obtained from *Bcl*I and *Eco*RI (partial) digestion of plasmid pAg200, where pAG200 is pUC8 containing a mutant (D27N) chromosomal DHFR gene (Villafranca et al., 1983). The correct sequence was confirmed by dideoxy sequencing and the construct named pCBPF1.

The synthetic gene coding for R67 DHFR was tailored to fit downstream from a mutant (up) promoter of the *E. coli* chromosomal DHFR gene using a *Bcl*I restriction site (Smith & Calvo, 1980, 1982). This mutant promoter was originally isolated in response to genetic pressure using trimethoprim (TMP), an active-site-directed inhibitor of chromosomal DHFR. The promoter produces large quantities of protein and has been used in expression of cytochrome *c* peroxidase (Fishel et al., 1987) and thymidylate synthetase (J. E. Villafranca, personal communication). Use of the *Bcl*I cloning site, however, requires the synthetic R67 DHFR gene code for an N-terminal sequence of Met-Ile. In the synthetic gene for truncated R67 DHFR, the N-terminal sequence is Met1-Ile17-Phe18-Pro19-Ser20 etc. This new N-terminal coding sequence is identical to the N-terminal sequence of R67

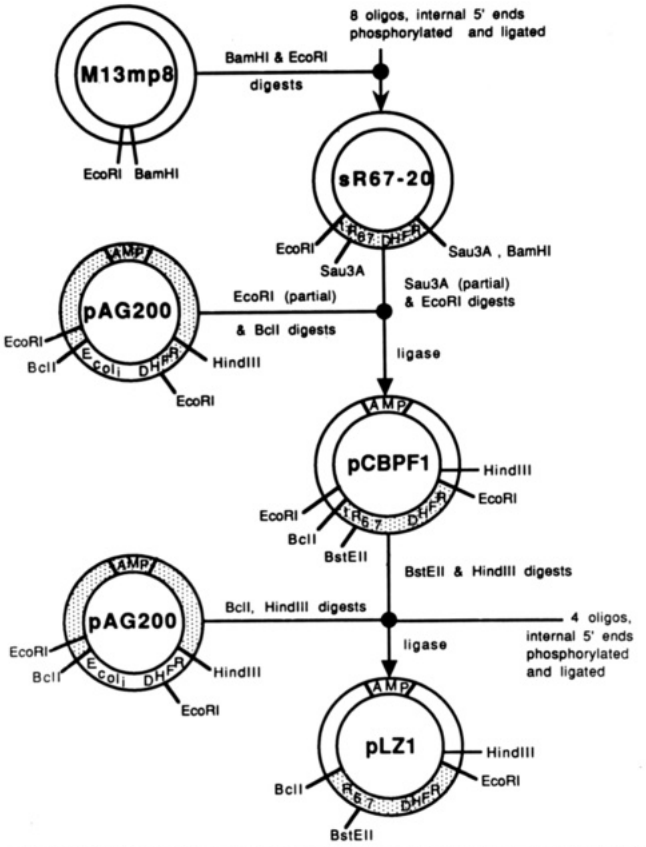

FIGURE 1: Construction of a plasmid (pLZ1) containing a synthetic R67 DHFR gene coding for a full-length R67 DHFR with an N-terminal sequence of Met-Ile. A truncated R67 DHFR gene was made first and is abbreviated tR67 DHFR in plasmids sR67-20 and pCBPF1. Using the *Bcl*I site at the 5′ end of the gene and either an *Eco*RI or a *Hin*dIII site at the 3′ end of the gene, both truncated R67 DHFR and Met-Ile R67 DHFR were cloned behind a mutant (up) promoter associated with *E. coli* chromosomal DHFR (Calvo & Smith, 1980, 1982). The parent plasmid for pAG200, and thus pCBPF1 and pLZ1, is pUC8. For more details, see Results.

DHFR truncated by chymotrypsin treatment except for the addition of methionine and the conservative substitution of isoleucine for Val17.

To restore the coding sequence for 16 N-terminal amino acids to the truncated R67 DHFR gene, 4 additional oligonucleotides were used. The oligonucleotides were phosphorylated (except for the two oligos at the extreme 5′ ends of the cassette) and ligated. This short cassette was mixed with a 317 bp *Bst*EII–*Hin*dIII fragment from the truncated synthetic R67 DHFR gene and ligated into a ~3200 bp fragment obtained by *Bcl*I–*Hin*dIII digestion of pAG200 containing the mutant (up) promoter of the *E. coli* chromosomal DHFR gene. The gene sequence was verified by dideoxy sequencing and the vector named pLZ1. The sequence of the full-length synthetic gene and unique restriction sites are shown in Figure 2.

The synthetic gene coding for truncated R67 DHFR did not confer TMP resistance upon *E. coli* and did not produce any observable protein as monitored by either Coomassie stains of denaturing gels, activity stains of nondenaturing gels, or Western blots of denaturing gels. For the latter, neither truncated protein nor smaller protein fragments were observed when blots were probed by an antibody specific for R67 DHFR. The synthetic gene coding for truncated R67 DHFR was also transformed into *E. coli* strain CAG626 to see if the La protease deficient strain (*lon*) would increase protein production. Again, no protein or protein fragments were observed using Western blot techniques.

```
        10                  30                  50
ATGATCAGGAGCTCTAATGAAGTTTCGAATCCAGTTGCCGGCAATTTTGTATTCCCTTCG
------                  ------
  BclI                  AsuII
M I R S S N E V S N P V A G N F V F P S

        70                  90                  110
AACGCCACGTTTGGTATGGGTGACCGCGTACGTAAGAAATCCGGCGCCGCCTGGCAAGGT
                    ------  ------  ------  ------
                    BstEII  SnaBI   NarI
N A T F G M G D R V R K K S G A A W Q G

        130                 150                 170
CAGATTGTCGGGTGGTACTGCACAAATTTGACCCCTGAGGGCTACGCCGTCGAGTCTGAG
                                    -------
                                    MstII
Q I V G W Y C T N L T P E G Y A V E S E

        190                 210                 230
GCTCACCCGGGCTCAGTACAGATCTATCCTGTTGCGGCGCTTGAACGCATCAACTAATAG
------  ------  ------
SmaI    BglII
A H P G S V Q I Y P V A A L E R I N EndEnd

AATTC
-----
EcoRI
```

FIGURE 2: Synthetic gene sequence for the full-length Met-Ile variant of R67 DHFR. Unique restriction enzyme sites are underlined, and every 10th base is marked. The peptide translation is shown below the DNA sequence. Except for our substitution of Ile2 for Glu2, the peptide sequence is identical to that for native R67 DHFR (Brisson & Hohn, 1984). The N-terminal gene sequence for truncated R67 DHFR is ATG,ATC,TTC,CCT..., and the corresponding peptide sequence would be Met1-Ile17-Phe18-Pro19-etc.

In contrast, the full-length, synthetic gene (Met-Ile N-terminus) did confer TMP resistance upon host bacteria and did produce protein. While the P700 plasmid encoding native R67 DHFR typically produces 2.5–3% of the soluble protein as R67 DHFR, our full-length, synthetic gene produces 2% of the soluble protein as R67 DHFR. These expression levels are comparable.

These initial results suggest a prime function of the N-terminus in R67 DHFR may be to facilitate protein folding and/or to minimize protein degradation in the *E. coli* cytosol.

Protein Folding Monitored by Denaturation in Guanidine Hydrochloride. The stabilities of native and truncated R67 DHFR were compared at pH 5.0 where the normal tetrameric structure of R67 DHFR is unstable and dissociates to form dimers (Nichols et al., 1991, and unpublished results).[2] Initiating denaturation studies with the dimeric form of R67 DHFR allows us to eliminate any signal changes associated with tetramer dissociation.

[2] A pK_a of 6.2 (at 29 μM R67 DHFR expressed as tetramer) has been observed in a plot of K_{av} versus pH where K_{av} describes elution from a Superose 12 FPLC column (Nichols et al., 1991, and unpublished results). From a recently obtained crystal structure of the tetrameric form of R67 DHFR (Matthews et al., unpublished results), this pK_a likely describes titration of His62/162/262/362 residues at the dimer–dimer interface such that protonation of His results in dissociation of tetramer to dimers. We also note that Trp38/138/238/338 occur at the dimer–dimer interface and a red shift in tryptophan fluorescence is observed upon titrating the protein from pH 8 to 5. This is consistent with a change in environment (hydrophobic to hydrophilic) accompanying tetramer dissociation. We obtain a pK_a of 6.0 describing this fluorescence titration (at 8 μM R67 DHFR expressed as tetramer). Finally, we note the titration of chymotrypsin-truncated R67 DHFR is unaltered when compared to native R67 DHFR (as monitored by a fluorescence titration curve from pH 8 to 5). This indicates the N-terminus does not play a significant role in the dimer to tetramer equilibrium.

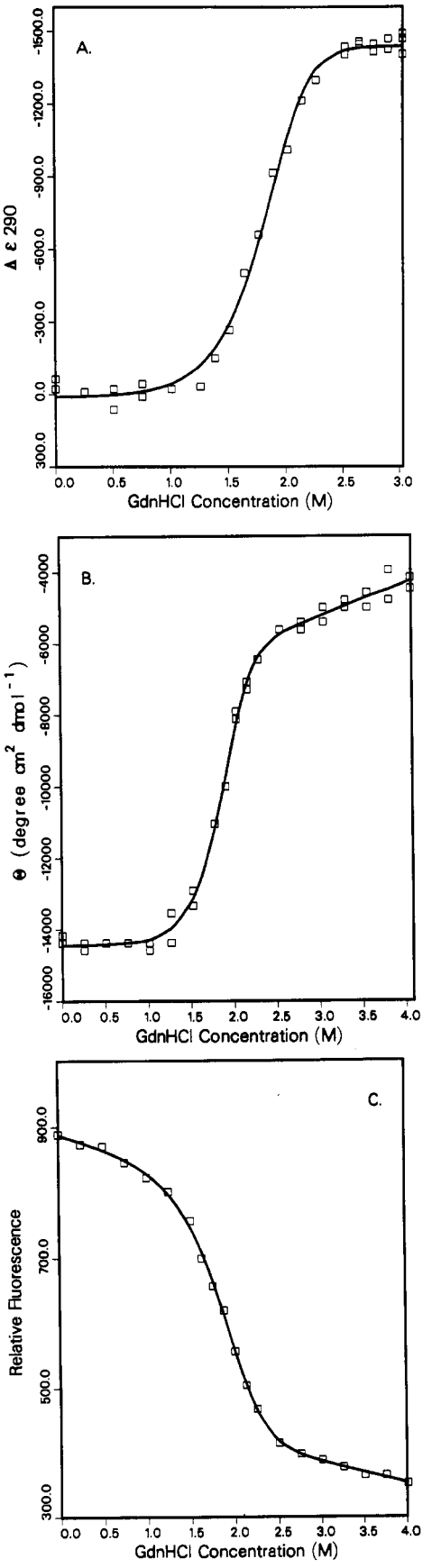

FIGURE 3: Equilibrium unfolding of native R67 DHFR monitored by (A) difference absorbance at 290 nm, (B) molar ellipticity at 217 nm, and (C) tryptophan fluorescence at 340 nm. Protein concentrations (expressed as monomer) were 94, 12, and 12 μM for panels A, B, and C, respectively. Theoretical curves were predicted by nonlinear fitting to eq 4 as described under Materials and Methods.

Figure 3 shows sharp unfolding curves for native R67 DHFR when changes in absorbance, circular dichroism, and fluorescence signals are monitored. The lines shown fit a two-state model for unfolding described by eq 4 under Materials and Methods. If denaturation curves were described by a three-state model:

$$D \rightleftharpoons 2M \rightleftharpoons 2U$$

where D is dimer, M is folded monomer, and U is unfolded monomer, one might expect the denaturation curves to be biphasic as seen for dimeric aspartate aminotransferase (Herold & Kirschner, 1990). This is not observed, suggesting folded monomer is not a significantly populated state during unfolding. Also within our error range, the unfolding curves performed at equivalent protein concentrations using different optical techniques are coincident, suggesting that complete, not partial unfolding is being monitored.

If dimeric R67 DHFR were to dissociate prior to the unfolding transition zone, unfolding would then monitor the M $\rightleftharpoons$ U transition. Since this reaction is unimolecular, all or part of the unfolding curves would be independent of protein concentration. In contrast, if dimeric R67 DHFR is significantly populated during the unfolding transition, the reaction would be bimolecular and the midpoint of unfolding dependent on protein concentration. The latter case can be seen in Figure 4 where a higher concentration of R67 DHFR is more stable to unfolding conditions.

The free energy of unfolding in the absence of GdnHCl was calculated to be 13.5 ± 0.525 kcal mol^{-1} by difference UV, 15.5 ± 0.576 kcal mol^{-1} by CD measurements, 13.2 ± 0.500 kcal mol^{-1} (12 μM DHFR) by fluorescence measurements, and 13.3 ± 0.635 kcal mol^{-1} (1.2 μM DHFR) by fluorescence measurements. The values for other fitting parameters are given in Table II. Our average value for ΔG_{H_2O} (13.9 kcal mol^{-1}) is in the range of published values for dimeric DNA binding proteins. A ΔG_{H_2O} of 11.1 kcal mol^{-1} for unfolding of dimeric arc repressor to unfolded monomer (Bowie & Sauer, 1989), 11.5 kcal mol^{-1} for unfolding of the dimerization domain of transcription factor LFB1 to unfolded monomer (De Francesco et al., 1991), and an average ΔG_{H_2O} of 23.2 kcal mol^{-1} for unfolding of dimeric *trp* aporepressor to unfolded monomer (Gittleman & Matthews, 1990) have been obtained.

When unfolding of truncated R67 DHFR was investigated, a decrease in protein stability was observed. Figure 5 compares the unfolding profiles for native and truncated DHFR monitoring absorbance and fluorescence changes. Again, plots of F_{app} versus [GdnHCl] are depicted for ease of comparison. ΔG_{H_2O} for truncated R67 DHFR is calculated to be 10.6 ± 0.204 kcal mol^{-1} for unfolding monitored by difference UV and 12.0 ± 0.578 kcal mol^{-1} for unfolding monitored by fluorescence. A difference ΔG ($=\Delta\Delta G$) of 2.6 kcal mol^{-1} is observed when the average stabilities of truncated and native R67 DHFR are compared with native enzyme being more stable.

The average ΔG_{H_2O} values of 13.9 and 11.3 kcal mol^{-1} for native and truncated R67 DHFR describe unfolding equilibrium constants of 5.04×10^{-11} and 4.25×10^{-9} M, respectively. These results suggest the N-terminal residues participate in some fashion in stabilizing the β-barrel structure of R67 DHFR.

Discussion

What Sequences Comprise the Essential Core of Type II R-Plasmid-Encoded DHFRs? Three variants of type II DHFR gene (R67, R388, R751) have been cloned and sequenced (Brisson & Hohn, 1984; Swift et al., 1981; Zolg &

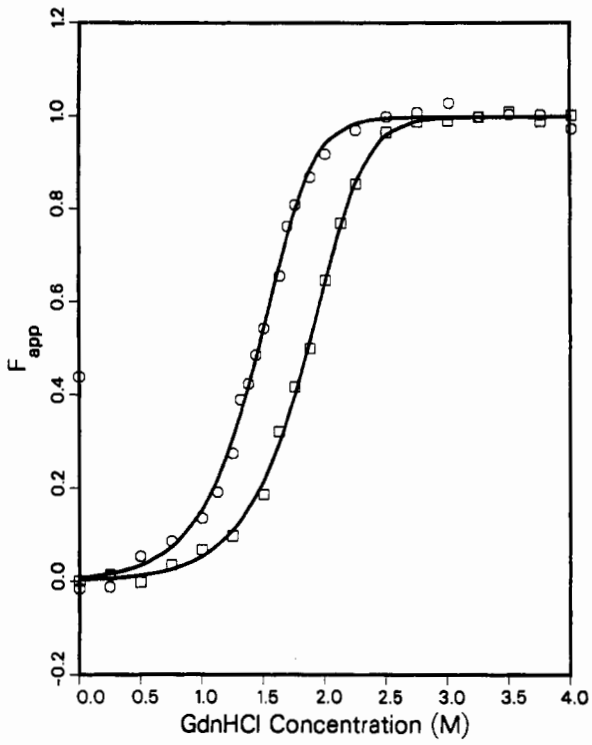

FIGURE 4: Concentration dependence of native R67 DHFR denaturation monitored by alterations in tryptophan fluorescence. Data for 12 μM R67 DHFR (expressed as monomer) are described by (□) points and 1.2 μM R67 DHFR by (○) points. The data were converted to F_{app} to facilitate comparison.

Hanggi, 1981; Flensburg & Steen, 1986). Most of the sequence differences between these type II DHFRs cluster within 21 residues at the amino terminus where 11 amino acid variations are observed. In contrast, there are only 6 differences in the next 65 residues, suggesting this region of the gene is conserved and encodes the essential core of the protein.

Another experiment supporting the idea of an essential core associated with type II DHFRs was performed by Vermersch and Bennett (1988). They truncated the first 18 codons of the R388 DHFR gene and observed a TMP-sensitive phenotype. Upon adding eight codons back to the R388 DHFR gene, they regained a TMP-resistant phenotype.

Additionally, the electron density map for a 2.8-Å X-ray structure of a dimeric form of R67 DHFR indicates 16 residues at the amino terminus of the protein are completely disordered (Matthews et al., 1986). Matthews et al. suggested that since the first 10 amino acids are hydrophilic (with the exception of Val8), residues 1–10 would probably be readily accessible to solvent.

However, the most definitive evidence for an essential, functional core in type II DHFRs is our finding that R67 DHFR truncated at the N-terminus by 16 residues retains full enzymic activity. This result in conjunction with CNBr cleavage at Met26 (which results in nonfunctional protein) suggests the essential functional core begins somewhere between Phe16 and Met26.

Construction of a Synthetic Gene. As a prelude to studying this novel enzyme using enzyme engineering techniques, we have synthesized the R67 DHFR gene as a fully overlapping DNA duplex sequence. The advantages of this approach are numerous as the degeneracy of the genetic code allows us to introduce several restriction sites unique to M13 or pUC8. By judicious choice of the restriction enzyme sites, the protein was divided into self-contained structural modules, i.e., β-strands, that can be manipulated individually. These features increase

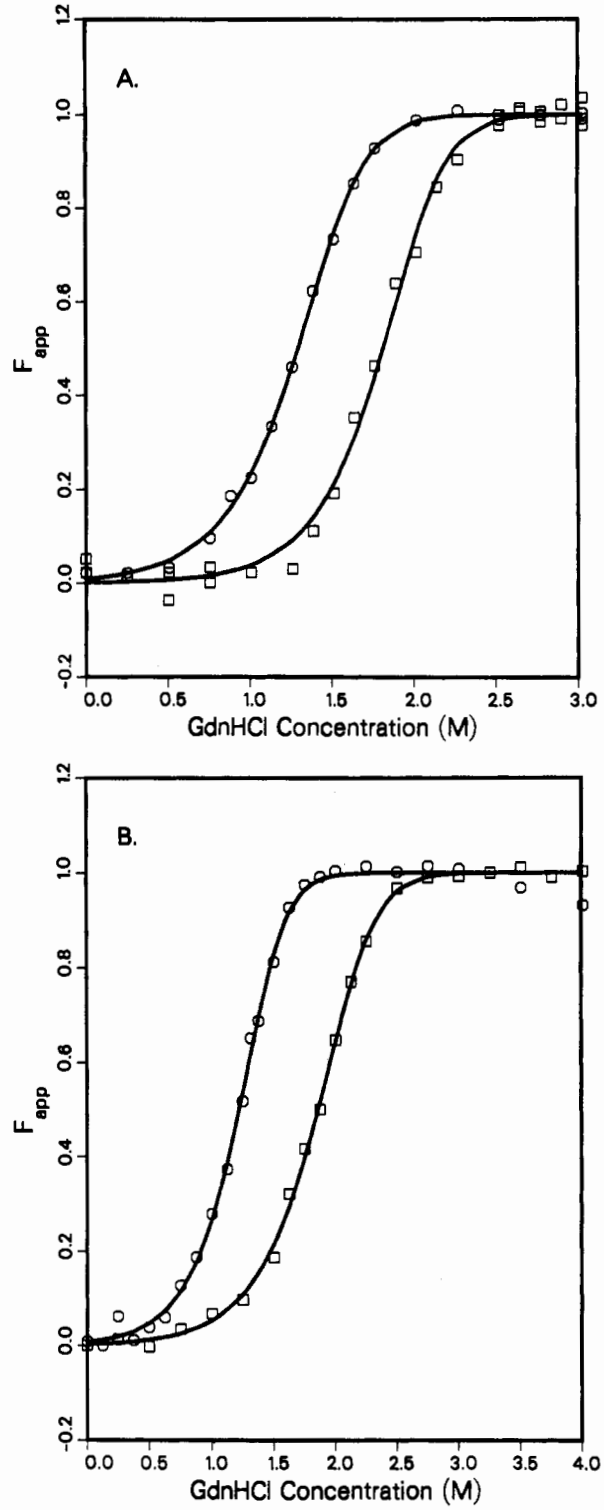

FIGURE 5: Comparison of equilibrium unfolding for native and truncated R67 DHFRs monitored by (A) difference absorbance and (B) tryptophan fluorescence. Native R67 DHFR data are described by (□) points and truncated R67 DHFR by (○) points. For absorbance measurements, native and truncated protein concentrations were 140 and 94 μM (expressed as monomer), respectively. For fluorescence measurements, the concentration for both native and truncated protein was 12 μM. The data were converted to F_{app} to facilitate comparison.

the versatility available in investigating structural features in R67 DHFR.

Since removing 16 N-terminal residues did not alter the catalytic activity of R67 DHFR, it seemed likely that a less drastic substitution at the N-terminus would also be functional. Therefore, we synthesized an N-terminal variant gene for R67

Table II: Fitting Parameters for Unfolding of R67 DHFR

condition[a]	Y_N^0	M_N (M^{-1})	Y_U^0	M_U (M^{-1})	ΔG_{H_2O} (kcal mol^{-1})	M_G (kcal mol^{-1} M^{-1})	midpoint[b] (M)
94 μM R67 DHFR, ϵ_{290}	11.7 ± 17.0	0	−1440 ± 13.6	0	13.5 ± 0.525	−4.34 ± 0.300	1.82
12 μM R67 DHFR, θ_{217}	−14400 ± 80.4	0	−8100 ± 388	947 ± 116	15.5 ± 0.534	−4.78 ± 0.285	1.85
12 μM R67 DHFR, F_{340}	889 ± 4.52	−38.6 ± 10.2	483 ± 18.9	−32.5 ± 5.54	13.2 ± 0.500	−3.51 ± 0.240	1.86
1.2 μM R67 DHFR, F_{340}	920 ± 5.61	−76.6 ± 21.4	454 ± 19.3	−40.7 ± 5.99	13.3 ± 0.634	−3.63 ± 0.360	1.46
140 μM truncated R67 DHFR, ϵ_{290}	37.8 ± 13.0	0	−1420 ± 55.8	104 ± 21.2	10.6 ± 0.204	−4.06 ± 0.159	1.30
12 μM truncated R67 DHFR, F_{340}	878 ± 5.29	−86.8 ± 24.9	350 ± 3.01	0	12.0 ± 0.578	−4.35 ± 0.374	1.22

[a] ϵ_{290}, θ_{217}, and F_{340} represent changes in tryptophan absorbance at 290 nm, molar ellipticity at 217 nm, and tryptophan fluorescence (excitation 290 nm, emission 340 nm) as described above. Y_N^0 and Y_U^0 are values for the optical parameters for folded and unfolded R67 DHFR at 0 M [GdnHCl]. M_N and M_U are the slopes associated with pre- and posttransition phases. ΔG_{H_2O} is the free energy difference between folded dimer and unfolded monomer. M_G is the slope describing the dependence of ΔG_{H_2O} on [GdnHCl]. These values were all obtained by nonlinear fitting as described above. [b] Midpoint is the [GdnHCl] when $F_{app} = 0.5$. This was calculated using the computer program MAPLE.

DHFR that differed from the native sequence by the substitution of Met-Ile for Met-Glu. This substitution was convenient in that it created a *BclI* site allowing the gene to be cloned behind a strong promoter. Natural variations in the N-terminal sequence of R-plasmid-encoded DHFRs are observed and include the N-terminal sequence Met-Asp (R751; Flensburg & Steen, 1986) and Met-Gly (R388; Brisson & Hohn, 1984). In addition, the N-terminus of a synthetic R388 DHFR gene (pV9134–124; Vermersch & Bennett, 1988) was truncated by seven residues such that a Met1 codon, immediately followed by an Ala8 codon, resulted in functional protein.

Our synthetic gene coding for a full-length Met-Ile variant of R67 DHFR yields a TMP-resistant phenotype and produces fully functional protein, as monitored by k_{cat} and K_m values. Surprisingly, however, the gene coding for a truncated R67 DHFR (analogous to the truncated functional protein created in vitro) does not yield a TMP-resistant phenotype. This result suggests either that the in vivo degradation rate has increased due to the altered N-terminal sequence and/or that the truncated protein is less stable. Even the use of a La protease deficient strain of *E. coli* did not facilitate protein production. Therefore, the in vitro stabilities of native and truncated DHFR were investigated.

Folding Changes Associated with Native and Truncated R67 DHFR. The GdnHCl denaturation curves presented above support a two-state model of unfolding in R67 DHFR. Several lines of evidence support this hypothesis. First, the transition zone for unfolding is coincident when monitored by both fluorescence and circular dichroism techniques at identical protein concentrations. Second, the unfolding curve is not biphasic. Third, a consistent ΔG_{H_2O} value (13.9 kcal mol^{-1}) is calculated over a 78-fold change in protein concentration. These results all suggest folded monomer is not a populated species during unfolding and that folding and dimerization are tightly coupled.

In examining the dimeric R67 DHFR crystal structure, Matthews et al. (1986) noted that β-strands B–D from one monomer associate with β-strands B–D from the second monomer, forming a third β-barrel structure. As the third β-barrel occurs at the monomer–monomer interface, it may play a role in stabilizing the structure of the dimer and/or loss of this structure may destabilize folded monomer. We also note that numerous residues at the monomer–monomer interface are hydrophobic and if the interface is disrupted by GdnHCl, folded monomers may be destabilized upon exposure of the hydrophobic residues to solvent.

Since Trp45 occurs at the monomer–monomer interface, we may be monitoring changes in its environment in our fluorescence and absorbance unfolding studies. A second tryptophan (38) occurs per monomer and from the dimeric crystal structure appears to be exposed to solvent. However, in the crystal structure of tetrameric R67 DHFR, Trp38 is buried at the dimer–dimer interface, and we can describe a tetramer to dimer equilibrium by monitoring the environment of this tryptophan (Nichols et al., 1991, and unpublished results; also see footnote 2).

Why is truncated R67 DHFR less stable than native R67 DHFR? Several factors may be involved. First, the position of the new N-terminus with a positively charged group may place it too close to a hydrophobic pocket at the monomer–monomer interface, destabilizing the dimer. Second, the disorder of residues 1–16 seen in the crystal structure could be described by a totally random orientation of the N-terminus, or it could also be described by rapid interconversion of several (>3) stable conformations. The multiple stable conformations could increase the stability of native over truncated R67 DHFR. However, if this were true, one might expect M_G describing the unfolding transition to be different for native and truncated R67 DHFR. Shortle (1989) defines M_G as a measure of the extent of unfolding. Since M_G is the same (within error) for native and truncated R67 DHFR, this suggests the N-terminus is not ordered in native R67 DHFR; i.e., if the N-terminus were ordered, native R67 DHFR would unfold to a greater extent than truncated DHFR, and this would be reflected in an altered M_G value. However, if only a *short* segment of the N-terminus is ordered (i.e., several rapidly interconverting stable conformations for possibly Asn15-Phe16), it is likely that changes in M_G from unfolding curves would be minimal and within our standard deviations. At present, we cannot differentiate between this last possibility and any destabilizing effect associated with moving the N-terminal amino group closer to the β-barrel structure.

Why is truncated R67 DHFR not found in vivo? Several factors may be involved; for example, mRNA stability could be different for full-length and truncated R67 DHFRs, or refolding rates for truncated R67 DHFR could be altered such that aggregation (or proteolysis) now readily occurs (Krueger et al., 1990; Mitraki et al., 1991; Lee et al., 1988; Chin et al., 1988). Also, differences in the N-terminal sequence have been found to affect the protein half-life in eucaryotes (Bachmair et al., 1986; Arfin & Bradshaw, 1988). Perhaps similar mechanisms exist for affecting protein turnover in procaryotes.

From our equilibrium unfolding studies, one likely factor involved in nonexpression of truncated R67 DHFR is alteration of unfolding and/or refolding rates such that rapid proteolysis occurs in vivo. Our preliminary kinetic unfolding studies indicate a slow rate of unfolding for native R67 DHFR with a relaxation time (τ) of 80 s (0–3 M GdnHCl transition, 2.3 μM R67 DHFR as monomer, data not shown). Under similar conditions, unfolding of truncated DHFR has a τ of 2 s, an approximately 40-fold difference in relaxation times. Refolding rates are currently being investigated. Whether the combination of decreased stability and an approximately 40-

fold change in unfolding rates is sufficient to explain why truncated R67 DHFR is not found in vivo is not known.

A specific example where small changes in refolding rates drastically affect protein expression/exportation has been seen for precursor and mature forms of exported proteins such as maltose binding protein and ribose binding protein (Park et al., 1988); 2–3-fold slower refolding rates for these precursor proteins have been proposed to play a physiological role, i.e., retardation of folding to allow exposure of a binding site for the chaperonin SecB (Liu et al., 1989; Randall et al., 1990). In this case, a small change in refolding rates drastically alters SecB binding which consequently affects exportation of these proteins. Unfolding rates were not altered.

ACKNOWLEDGMENTS

We thank Dave Matthews for sharing his crystal structure data, Bob Matthews, Bob Villafane, and Ira Ropson for helpful discussions, Tom Bradrick for help in stopped-flow data acquisition and analysis, and Linda Musick, Carol Booth, Paul Foster, and Charles Linn for their excellent technical assistance.

REFERENCES

Amyes, S. G. B. (1989) *J. Med. Microbiol. 28*, 73–83.
Amyes, S. G. B., & Smith, J. T. (1974) *Biochem. Biophys. Res. Commun. 58*, 412–418.
Amyes, S. G. B., & Smith, J. T. (1976) *Eur. J. Biochem. 61*, 597–603.
Amyes, S. G. B., Towner, K. J., Carter, G. I., Thomson, C. J., & Young, H. K. (1989) *J. Antimicrob. Chemother. 24*, 111–119.
Arfin, S. M., & Bradshaw, R. A. (1988) *Biochemistry 27*, 7979–7984.
Arnone, A. (1972) *Nature 247*, 146–147.
Baccanari, D., Phillips, A., Smith, S., Sinski, D., & Burchall, J. (1975) *Biochemistry 14*, 5267–5273.
Bachmair, A., Finley, D., & Varshavsky, A. (1986) *Science 234*, 179–186.
Blakley, R. L. (1960) *Nature (London) 40*, 1684–1685.
Bowie, J. U., & Sauer, R. T. (1989) *Biochemistry 28*, 7139–7143.
Brisson, N., & Hohn, T. (1984) *Gene 28*, 271–275.
Brito, R. M. M., Reddick, R., Bennett, G. N., Rudolph, F. B., & Rosevear, P. R. (1990) *Biochemistry 29*, 9825–9831.
Brito, R. M. M., Rudolph, F. B., & Rosevear, P. R. (1991) *Biochemistry 30*, 1461–1469.
Chin, D. T., Fogg, S. A., Webster, T., Smithe, T., & Goldberg, A. L. (1988) *J. Biol. Chem. 263*, 11718–11728.
De Francesco, R., Pastore, A., Vecchio, G., & Cortese, R. (1991) *Biochemistry 30*, 143–147.
Ellis, K. J., & Morrison, J. F. (1982) *Methods Enzymol. 87*, 405–426.
Fishel, L. A., Villafranca, J. E., Mauro, J. M., & Kraut, J. (1987) *Biochemistry 26*, 351–360.
Flensburg, F., & Steen, R. (1986) *Nucleic Acids Res. 14*, 5933.
Fling, M. E., & Elwell, L. P. (1980) *J. Bacteriol. 141*, 779–785.
Fling, M. E., Kopf, J., & Richards, C. (1988) *Plasmid 19*, 30–38.
Gittleman, M. S., & Matthews, C. R. (1990) *Biochemistry 29*, 7011–7020.
Gornall, A. G., Bardawill, C. J., & David, M. M. (1949) *J. Biol. Chem. 177*, 751–766.
Herold, M., & Kirschner, K. (1990) *Biochemistry 29*, 1907–1913.
Howell, E. E., Booth, C., Farnum, M., Kraut, J., & Warren, M. S. (1990) *Biochemistry 29*, 8561–8569.
Huovinen, P. (1987) *Antimicrob. Agents Chemother. 31*, 1451–1456.
Joyner, S. S., Fling, M. E., Stone, D., & Baccanari, D. P. (1984) *J. Biol. Chem. 259*, 5851–5856.
Kilmartin, J. V. (1976) *Br. Med. Bull. 32*, 209–212.
Kraut, J., & Matthews, D. A. (1987) *Biological Macromolecules and Assemblies*, (Jurnak, F. A., & Mcpherson, A., Eds.) Vol. 3, pp 1–71, Wiley, New York.
Krueger, J. K., Stock, A. M., Schutt, C. E., & Stock, J. B. (1990) *Protein Folding: Deciphering the Second Half of the Genetic Code* (Gierasch, L. M., & King, J., Eds.) pp 136–142, American Association for the Advancement of Science, Washington, D.C.
Lee, Y. S., Park, S. C., Goldberg, A. L., & Chung, C. H. (1988) *J. Biol. Chem. 263*, 6643–6646.
Liu, G., Topping, T. B., & Randall, L. L. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 9213–9217.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Matthews, D. A., Smith, S. L., Baccanari, D. P., Burchall, J. J., Oatley, S. J., & Kraut, J. (1986) *Biochemistry 25*, 4194–4204.
Mitraki, A., Fane, B., Hasse-Pettingell, C., Sturtevant, J., & King, J. (1991) *Science 253*, 54–58.
Morrison, J. F., & Sneddon, M. K. (1990) *Chemistry and Biology of Pteridines 1989*, pp 728–733, Walter de Gruyter, Berlin.
Motulsky, H. J., & Ransnas, L. A. (1987) *FASEB J. 1*, 365–374.
Nichols, R., Weaver, D., Reece, L., & Howell, E. (1991) *FASEB J. 5*, A1521.
Pace, C. N., Shirley, B. A., & Thomson, J. A. (1990) in *Protein Structure: A Practical Approach* (Creighton, T. E., Ed.) Chapter 13, IRL Press, Oxford.
Park, S., Liu, G., Topping, T. B., Cover, W. H., & Randall, L. L. (1988) *Science 239*, 1033–1035.
Pattishall, K. H., Acar, J., Burchall, J. J., Goldstein, F. W., & Harvey, R. J. (1977) *J. Biol. Chem. 252*, 2319–2323.
Penner, M., & Frieden, C. (1985) *J. Biol. Chem. 260*, 5366–5369.
Randall, L. L., Topping, T. B., & Hardy, S. J. S. (1990) *Science 248*, 860–862.
Ropson, I. J., Fordon, J. I., & Freiden, C. (1990) *Biochemistry 29*, 9591–9599.
Santoro, M. M., & Bolen, D. W. (1988) *Biochemistry 27*, 8063–8068.
Shortle, D. (1989) *J. Biol. Chem. 264*, 5315–5318.
Skold, O., & Widh, A. (1974) *J. Biol. Chem. 249*, 4324–4325.
Smith, D. R., & Calvo, J. M. (1980) *Nucleic Acids Res. 8*, 2255–2274.
Smith, D. R., & Calvo, J. M. (1982) *Mol. Gen. Genet. 187*, 72–78.
Smith, S., & Burchall, J. J. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 4619–4623.
Smith, S., Stone, D., Novak, P., Baccanari, D. P., & Burchall, J. J. (1979) *J. Biol. Chem. 254*, 6222–6225.
Stone, D., & Smith, S. (1979) *J. Biol. Chem. 254*, 10857–10861.
Stone, D., Paterson, S. J., Raper, J. H., & Phillips, A. W. (1979) *J. Biol. Chem. 254*, 480–488.
Stone, S. R., & Morrison, J. F. (1986) *Biochim. Biophys. Acta 869*, 275–285.

Sundstrom, L., Vinayagamoorthy, T., & Skold, O. (1987) *Antimicrob. Agents Chemother. 31*, 60–66.
Sundstrom, L., Radstrom, P., Swedberg, G., & Skold, O. (1988) *Mol. Gen. Genet. 213*, 191–201.
Swift, G., McCarthy, B. J., & Heffron, F. (1981) *Mol. Gen. Genet. 181*, 441–447.
Tartof, K. D., & Hobbs, C. A. (1987) *BRL Focus 9*, 12.
Vermersch, P. S., & Bennett, G. N. (1988) *DNA 7*, 243–251.
Vermersch, P. S., Klass, M. R., & Bennett, G. N. (1986) *Gene 41*, 289–297.
Villafranca, J. E., Howell, E. E., Voet, D. H., Strobel, M. S., Ogden, R. C., Abelson, J. N., & Kraut, J. (1983) *Science 222*, 782–788.
Villafranca, J. E., Howell, E. E., Oatley, S. J., Xuong, N., & Kraut, J. (1987) *Biochemistry 26*, 2182–2189.
Weaver, C. D., Crombie, B., Stacey, G., & Roberts, D. M. (1991) *Plant Physiol. 95*, 222–227.
Wlodawer, A., Miller, M., Jaskolshki, M., Sathyanarayana, B. K., Baldwin, E., Weber, I. T., Selk, L. M., Clawson, L., Schneider, J., & Kent, S. B. H. (1989) *Science 245*, 616–621.
Wylie, B. A., Amyes, S. G. B., Young, H. K., & Koornhof, H. J. (1988) *J. Antimicrob. Chemother. 22*, 429–435.
Yanish-Perron, C., Viera, J., & Messing, J. (1985) *Gene 33*, 103–119.
Zieg, J., Maples, V. F., & Kushner, S. R. (1978) *J. Bacteriol. 134*, 958–966.
Zolg, J. W., & Hanggi, U. J. (1981) *Nucleic Acids Res. 9*, 697–710.

Resonance Raman Spectra of Plastocyanin and Pseudoazurin: Evidence for Conserved Cysteine Ligand Conformations in Cupredoxins (Blue Copper Proteins)†

Jane Han,‡ Elinor T. Adman,§ Teruhiko Beppu,∥ Rachel Codd,⊥ Hans C. Freeman,⊥ Laila Huq,⊥ Thomas M. Loehr,‡ and Joann Sanders-Loehr*,‡

Department of Chemical and Biological Sciences, Oregon Graduate Institute of Science and Technology, Beaverton, Oregon 97006-1999, Department of Biological Structure, University of Washington, Seattle, Washington 98195, Department of Agricultural Chemistry, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan, and Department of Chemistry, University of Sydney, Sydney 2006, New South Wales, Australia

Received June 13, 1991; Revised Manuscript Received August 23, 1991

ABSTRACT: New resonance Raman (RR) spectra at 15 K are reported for poplar (*Populus nigra*) and oleander (*Oleander nerium*) plastocyanins and for *Alcaligenes faecalis* pseudoazurin. The spectra are compared with those of other blue copper proteins (cupredoxins). In all cases, nine or more vibrational modes between 330 and 460 cm^{-1} can be assigned to a coupling of the Cu–S(Cys) stretch with Cys ligand deformations. The fact that these vibrations occur at a relatively constant set of *frequencies* is testimony to the highly conserved ground-state structure of the Cu–Cys moiety. Shifts of the vibrational modes by 1–3 cm^{-1} upon deuterium exchange can be correlated with N–H···S hydrogen bonds from the protein backbone to the sulfur of the Cys ligand. There is marked variability in the *intensities* of these Cys-related vibrations, such that each class of cupredoxin has its own pattern of RR intensities. For example, plastocyanins from poplar, oleander, French bean, and spinach have their most intense feature at ~425 cm^{-1}; azurins show greatest intensity at ~410 cm^{-1}, stellacyanin and ascorbate oxidase at ~385 cm^{-1}, and nitrite reductase at ~360 cm^{-1}. These variable intensity patterns are related to differences in the electronic excited-state structures. We propose that they have a basis in the protein environment of the copper–cysteinate chromophore. A further insight into the vibrational spectra is provided by the structures of the six cupredoxins for which crystallographic refinements at high resolution are available (plastocyanins from *P. nigra*, *O. nerium*, and *Enteromorpha prolifera*, pseudoazurin from *A. faecalis*, azurin from *Alcaligenes denitrificans*, and cucumber basic blue protein). The average of the Cu–S(Cys) bond lengths is 2.12 ± 0.05 Å. Since the observed range of bond lengths falls within the precision of the determinations, this variation is considered insignificant. The Cys ligand dihedral angles are also highly conserved. Cu–S_γ–C_β–C_α is always near −170° and S_γ–C_β–C_α–N near 170°. As a result, the Cu–S_γ bond is coplanar with the Cys side-chain atoms and part of the polypeptide backbone. The coplanarity accounts for the extensive coupling of Cu–S stretching and Cys deformation modes as seen in the RR spectrum. The conservation of this copper–cysteinate conformation in cupredoxins may indicate a favored pathway for electron transfer.

The blue copper proteins have been the subject of intensive spectroscopic and structural studies (Adman, 1985, 1991; Solomon et al., 1986). Particularly noteworthy are their very intense (ϵ = 3000–5000 M^{-1} cm^{-1}) absorption bands in the visible spectrum (590–625 nm) arising from cysteinate S → Cu(II) charge transfer. In addition, they exhibit distinctive resonance Raman (RR) spectra, axial or rhombic EPR spectra with abnormally low copper hyperfine splitting constants, and relatively high redox potentials of 200–700 mV (compared to $E'_0 \simeq 160$ mV for the Cu^{2+}/Cu^{1+} pair). Blue copper centers are found in both mononuclear and multinuclear copper proteins. Where the biological function has been established,

† This research was supported by Grants GM 18865 (T.M.L., J.S.-L.) and GM 31770 (E.T.A.) from the National Institutes of Health and by the Australian Research Council, Grant A 28930307 (H.C.F.).
* To whom correspondence should be addressed.
‡ Oregon Graduate Institute of Science and Technology.
§ University of Washington.
∥ University of Tokyo.
⊥ University of Sydney.

0006-2960/91/0430-10904$02.50/0