Redesigning the Quaternary Structure of R67 Dihydrofolate Reductase

CREATION OF AN ACTIVE MONOMER FROM A TETRAMERIC PROTEIN BY QUADRUPLICATION OF THE GENE*

(Received for publication, June 7, 1996, and in revised form, August 20, 1996)

Thomas D. Bradrick‡, Chris Shattuck‡, Michael B. Strader‡, Carol Wicker§, Edward Eisenstein‡, and Elizabeth E. Howell***

From the ‡Department of Biochemistry, Cell and Molecular Biology, The University of Tennessee, Knoxville, Tennessee 37996-0840, §Mycogen Corp., San Diego, California 92121, the ¶Center for Advanced Research in Biotechnology, Rockville, Maryland 20850, and the †Department of Chemistry and Biochemistry, University of Maryland, Baltimore County, Baltimore, Maryland 21229

R67 dihydrofolate reductase (DHFR) provides resistance to the antibacterial drug trimethoprim. This R-plasmid-encoded enzyme does not share any homology with chromosomal DHFR. A recent crystal structure of active, homotetrameric R67 DHFR (Narayana, N., Matthews, D. A., Howell, E. E., and Xuong, N.-H. (1995) Nat. Struct. Biol. 2, 1018–1025) indicates that a single active site pore traverses the length of the molecule. Since the center of the pore possesses exact 222 symmetry, site-directed mutagenesis of residues in the pore will produce four mutations/active site. To break this inevitable symmetry, four copies of the gene have been linked in frame to create an active monomer possessing the essential tertiary structure of native tetrameric R67 DHFR. The protein product, quadruple R67 DHFR, is 4 times the molecular mass of native R67 DHFR in SDS-polyacrylamide gel electrophoresis and is monomeric under nondenaturing conditions as measured by sedimentation equilibrium experiments. The catalytic activity of quadruple R67 DHFR is decreased only slightly when compared with native R67 DHFR. Folding of quadruple R67 DHFR is completely reversible at pH 5. However, at pH 8, folding is not fully reversible; this is likely due to a competition between productive intramolecular versus nonproductive intermolecular domain association. The production of a fully active, monomeric R67 DHFR variant will enable the design of more meaningful site-directed mutants where single substitutions per active site pore can be generated.

Recent protein engineering studies of oligomeric proteins have attempted to shift the assembly equilibrium toward the protomer by the introduction of destabilizing mutations at subunit interfaces (Refs. 1–3 and references therein). These studies allow the functional role of protein-protein interactions at subunit interfaces to be probed. This strategy assumes that most oligomeric proteins possessing n protomers possess n active sites and is less useful for enzymes that have active sites to which several subunits contribute.

In contrast to the above approach that forms monomers by dissociation of the native oligomeric species, we have constructed an active, stable monomeric form of R67 dihydrofolate reductase (DHFR)† by association. This strategy involves the use of genetic engineering techniques to covalently link four protomers together. This approach is necessarily different, since R67 DHFR is a homotetramer with a single active site pore, and stabilization of the 78-amino acid monomer would yield only a partial active site.

In R67 DHFR, a 222 symmetry element occurs at the center of the pore, and crystallography and binding studies indicate that two molecules of either substrate (dihydrofolate/folate) or cofactor (NADPH) can bind to both sides of the active site pore under binary complex conditions (4, 5). Under ternary complex conditions, a total of two ligands can be bound; the possibilities include two NADPH molecules, two folate molecules, or one NADPH and one folate bound on opposite sides of the pore. The latter is the productive ternary complex. Cofactor inhibition is not observed due to negative cooperativity between both bound NADPH molecules (microscopic $K_d$ values of 5 and 48 $\mu$M), and substrate inhibition is not observed due to positive cooperativity between two bound dihydrofolate molecules (microscopic $K_d$ values of 250 and 4.4 $\mu$M; Ref. 5). This model for catalysis predicts that either dihydrofolate or NADPH can bind to each half-pore, although presumably with different orientations.

Because of the 222 symmetry in the active site of tetrameric R67 DHFR, mutagenesis of residues designed to explore substrate binding will also affect cofactor binding. To break the inevitable symmetry of this active site and make asymmetric mutations, we have engineered tandem arrays of the gene. As a first step toward this goal, we duplicated the gene and characterized the active dimer that resulted (6). We have now quadruplicated the gene and characterized the active, monomeric gene product. In addition, a byproduct of the cloning step was a gene triplication event. These genes and methods will allow us to break the symmetry of the active site and introduce a prescribed number of mutations in specific positions. It will also let us ask whether covalent linkage of four identical subunits introduces competition during folding reactions and re-

*This work was supported by National Institutes of Health Grant GM35308 (to E. E. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked “advertisement” in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

†The abbreviations used are: DHFR, dihydrofolate reductase; GdnHCl, guanidine hydrochloride; CD, circular dichroism; MES, 4-morpholinoethanesulfonic acid; quadruple R67 DHFR, a monomer created by quadruplication of the R67 DHFR gene that possesses the essential tertiary structure of native, tetrameric R67 DHFR; triple R67 DHFR, the gene product from the triplicated gene; double R67 DHFR, the gene product from the duplicated gene.
sults in nonproductive protein association and/or altered folding topologies.

**MATERIALS AND METHODS**

**DNA Sequencing**—An ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit from Perkin Elmer was used to sequence the quadruplicated R67 DHFR gene. The reactions were run on an ABI373A Stretch Automated Sequencer (Perkin Elmer) using Well-to-Read plates. A Sequenase kit (U.S. Biochemical Corp.) was used to read the triplicated gene sequence.

**Protein Purification**—Quadruple and triple R67 DHFRs were purified using the same protocol as employed for native R67 DHFR (6). Stellar2 Escherichia coli cells (F' mcrA Δ(mcrBC-hsdRMS-mpR) recA endA1 gyrA96 thi supE44 relA1 λ- Δ(lac-proAB); Life Technologies, Inc.) were utilized to minimize/eliminate recombination between gene copies. Cells were grown at 30 °C. Protein concentrations were determined by biuret assays, and all proteins were homogeneous according to SDS-polyacrylamide gel electrophoresis.

**Steady State Kinetics and Ligand Binding**—Steady state kinetic data were obtained using a Perkin-Elmer λ3a spectrophotometer interfaced with an IBM PS2 (7). The buffer was 50 mM MES, 100 mM Tris, 50 mM acetic acid, and 10 mM β-mercaptoethanol (MTA). This polybuffer maintains a constant ionic strength between pH 4.5 and 9.5 (8).

Dissociation constants describing binding of NADPH to 2–6 μM protein were determined at 25 °C by fluorescence quenching using a Perkin-Elmer LS-5B fluorimeter with excitation and emission wavelengths of 282 nm for DHF (9) and 6220 M-1 cm-1 (11). The molar extinction coefficient used to calculate the amount of DHF was 12,300 M-1 cm-1 (11).

**pH Dependence of Oligomerization State**—To monitor a pH-dependent equilibrium between tetramer and two dimers in native R67 DHFR or an equivalent dimer, 2 monomers transition in double R67 DHFR, or an equivalent monomer as altered monomer conformational change in quadruple R67 DHFR, tryptophan fluorescence was monitored as a function of pH (6, 12). The emission spectra for the DHFRs (excitation, 295 nm) were measured from 300 to 450 nm at each pH during the titrations. The intensity-averaged emission wavelength (λ) for each emission spectrum was calculated using the equation,

\[
\lambda = \frac{\sum (I_\lambda) \cdot \lambda}{\sum I_\lambda}
\]  

(Eq. 2)

where I is intensity and λ is the wavelength (13). The intensity-averaged emission wavelength is an integral measurement and is less sensitive to noise. Data for native R67 DHFR were fit to Equation 5 in Nichols et al. (12), and data for double R67 DHFR were fit to the following,

\[
F_{\text{flu}} = \frac{F_{\text{flu,max}} - F_{\text{flu,min}}}{\sum K_{\text{on,all}}P_{\text{M}}/4K_{\text{on,all}}P_{\text{M}}} \bigg(1 + \left(1 + \frac{8K_{\text{on,all}}[H^+]^{3/2}}{F_{\text{flu,max}} - F_{\text{flu,min}}}\right)^{3/2} + F_{\text{flu}}\bigg) \bigg)
\]  

(Eq. 3)

where \(K_{\text{on,all}} = (\text{dimers} [H^+])^n/\text{(monomers}[H^+]_n)^2\) in units of M, M^n, or M^n for n = 1, 1.5, or 2, respectively; \(F_{\text{flu,max}}\) is the observed fluorescence; \(F_{\text{flu,min}}\) and \(F_{\text{flu}}\) are the calculated limits for monomer and dimer fluorescence at low and high pH values; and \(P_{\text{M}}\) is the total protein concentration. The quadruple R67 DHFR data were fit to a simple ionization equation (Equations 5–7 from Ref. 14), where the number of protons was also a variable. To facilitate comparison, all the data were normalized by fixing \(F_{\text{flu, max}}\) to 1, \(F_{\text{flu, min}}\) to 1, \(F_{\text{flu}}\) to 1, and \(Y_{\text{flu, max}}\) to 1. If not, \(Y_{\text{flu, max}}\) was a fractional value between 0 and 1 and \(Y_{\text{flu, min}}\) and \(Y_{\text{flu}}\) are the optical values associated with the observed pH and with the pH limits of 8 and 4, respectively. Within error, a plot of the fluorescence intensity at 350 nm yields the same fit values, although the pre- and posttransition slope is more substantial for the intensity data (15).

**Sedimentation analyses of the R67 DHFR variants were performed with a Beckman XL-A Optima analytical ultracentrifuge using a four-hole An-55 rotor. Equilibrium experiments were performed at 25 °C and a rotor speed of 25,000 rpm, whereas velocity experiments were performed at 20 °C and a rotor speed of 50,000 rpm. The concentration distribution of the samples at sedimentation equilibrium was acquired as an average of 25 measurements of absorbance at each radial position, with nominal spacing of 0.001 cm between radial positions. Sedimentation velocity data were collected every 2 min after attaining speed in a single measurement with nominal spacing of 0.003 cm between radial positions.

**Equilibrium data were analyzed as described previously (12) to determine the homogeneity and number of components of the R67 DHFR variants.** Sedimentation velocity data were analyzed to yield sedimentation coefficients by using the time derivative of the boundary, g'(s), as described by Stafford (16) using instrumentally supplied software. The hydrodynamic shape of R67 DHFR under various conditions was estimated by determining the frictional coefficient of the molecule from the sedimentation coefficient using the Sedverd equation and by assuming a level of hydration of 0.4 g of H2O/g of protein (17). These values were then compared with the frictional coefficients calculated from the semianalytical ellipsoid, a sphere, a prolate ellipsoid (18), or a cylinder with dimensions corresponding to those of the R67 DHFR structure.

**Equilibrium Unfolding**—Equilibrium unfolding curves at pH 5.0 were monitored as a function of guanidine HCl (GdnHCl) concentration (6, 19). The emission spectra for the DHFRs (excitation, 295 nm) were measured from 300 to 450 nm at each denaturant concentration (between 0 and 5 M GdnHCl). The intensity-averaged emission wavelength for each emission spectrum was calculated and normalized as described above. The Gibbs free energy change, ΔG, between folded and unfolded states can then be calculated by nonlinear regression of the following equation for both double and quadruple R67 DHFR data,

\[
Y_{\text{flu,max}} = \left(\frac{Y_{\text{flu,min}} + M_p[G\text{dnHCl}]}{Y_{\text{flu,min}} + M_p[G\text{dnHCl}] + M_d[G\text{dnHCl}] + M_t[G\text{dnHCl}] + M_z[H_2O]}\right)^{1/2} \left(1 + 1 + \exp\left(-14\Delta G_{\text{H2O}} + M_p[G\text{dnHCl}] + M_d[G\text{dnHCl}] + M_z[H_2O]\right)\right)^{1/2}
\]  

(Eq. 4)

where R is the gas constant, T is the temperature, ΔG_{H2O} is the free energy change between native and unfolded protein in the absence of GdnHCl, \(M_p\) is the slope describing the dependence of ΔG on denaturant concentration, \(Y_{\text{flu,min}}\) and \(Y_{\text{flu,max}}\) are the concentration-independent optical values, and \(M_d\) and \(M_t\) are the slopes of the pre- and post-transition regions, respectively (6). The native R67 DHFR equilibrium folding data were fit to Equation 4 as in Beneo et al. (19). Within error, a plot of the fluorescence intensity at 340 nm yields the same fitted values, although the pre- and post-transitional slopes are more substantial for the intensity data (15).

**Circular Dichroism**—Circular dichroism (CD) spectra of 5 μM protein in MTA buffer without β-mercaptoethanol were obtained using a Jasco J-710 spectropolarimeter at 22 °C. For each sample, 10 spectra were obtained using 1-nm steps and 2-s integrations, and an averaged spectrum was calculated. A buffer base-line scan was subtracted from the averaged protein scan. The mean residue ellipticity calculated for the CD data used 108 g/mol as the mean residue molecular weight.

**RESULTS**

Rationale and Modeling of Linkers to Connect the N and C termini in Tetrameric R67 DHFR—Previous results were helpful in designing a gene coding for monomeric R67 DHFR. First, a crystal structure of the dimeric form of R67 DHFR indicates 16–18 N-terminal amino acids are disordered in each monomer (20). In accord with this observation, removal of 16 N-terminal residues by chymotrypsin treatment does not affect activity (19). Also the addition of 27 amino acids to the C terminus of R67 DHFR had no effect on activity (21). Finally, gene duplications successfully produced active dimer that mimicked the essential tertiary structure of native tetramer (6). These observations indicate that the functional core of R67 DHFR consists of the β-barrel structure formed by residues 27–77 and that...
extending and/or fusing the N and C termini of adjacent monomers might reasonably be expected not to impair function.

The crystal structure of chymotrypsin-truncated, tetrameric R67 DHFR is shown in Fig. 1A, and the N and C termini are indicated.\(^3\) In designing a link between monomers, we simply connect the C terminus (Asn78) of monomer A and the N terminus (Val117) of monomer B would span 28 Å, straight line distance) and only one link crosses the dimer-dimer interface (AB to CD; 42 Å). Alternatively, the domains could fold with connectivities of ACDB or ADCB with two covalent links crossing the dimer-dimer interfaces and only one link spanning the monomer-monomer interface. A third possibility is a connectivity of ADBC or ACBD, where the domains could fold with two covalent links crossing the dimer-dimer interfaces and the N and C termini introduced by gene quadruplication are depicted by arrows. The arrows connecting the ellipsoids are not meant to describe the exact position of the tethers, just the general topological linkage possibilities.

The amino acids in the first monomer are labeled 1–78, those in the second monomer 101–178, those in the third monomer 201–278, and those in the fourth monomer 301–378. For brevity, when a single residue is mentioned here, all four residues are implied. This nomenclature is retained when double and quadruple R67 DHFR are discussed although residue 78 is followed in frame by residue 101, etc.

The amino acids in the first monomer are labeled 1–78, those in the second monomer 101–178, those in the third monomer 201–278, and those in the fourth monomer 301–378. For brevity, when a single residue is mentioned here, all four residues are implied. This nomenclature is retained when double and quadruple R67 DHFR are discussed although residue 78 is followed in frame by residue 101, etc.

**Fig. 1.** A (top), the α-carbon backbone of tetrameric native R67 DHFR (4). 16 amino acids were removed by chymotrypsin treatment to yield this crystal form in space group D2. The ordered amino acids in monomer A are numbered 17–78 (lower left), those in monomer B, 117–178 (upper left), those in monomer C, 217–278 (upper right), and those in monomer D, 317–378 (lower right). The active site pore occurs in the center of the figure. B (bottom), six different topological connections for quadruple R67 DHFR. Each ellipse describes a native monomer. The AB and CD borders are equivalent to the native monomer-monomer interactions, and the AD and BC overlaps correspond to the native dimer-dimer interfaces. ~ identifies the dimer-dimer interfaces where His\(^{76}\), His\(^{122}\), His\(^{192}\), and His\(^{262}\) and Trp\(^{138}\), Trp\(^{248}\), and Trp\(^{318}\) occur. The active site pore occurs in the center of each diagram. The covariant connections between the N and C termini introduced by gene quadruplication are depicted by arrows. The arrows connecting the ellipsoids are not meant to describe the exact position of the tethers, just the general topological linkage possibilities.

---

\(^3\)The amino acids in the first monomer are labeled 1–78, those in the second monomer 101–178, those in the third monomer 201–278, and those in the fourth monomer 301–378. For brevity, when a single residue is mentioned here, all four residues are implied. This nomenclature is retained when double and quadruple R67 DHFR are discussed although residue 78 is followed in frame by residue 101, etc.
gonucleotides that are complementary to the coding sequence were not used, since they could bind to any/all of the gene copies, and a unique sequence would not be determined. No extraneous mutations were found.

Both proteins resulting from gene triplication and quadruplication were homogenous. As shown in Fig. 3, triple and quadruple R67 DHFR migrate in SDS-polyacrylamide gel electrophoresis with 3–4 times the apparent molecular mass of the native monomer. The duplicated gene product is additionally shown for comparison.

**pH-dependent Protein Characterization**—Native R67 DHFR is a tetramer at high pH (>8.0) that undergoes a pH-dependent, reversible dissociation to form dimers at low pH (≤5.0; Ref. 12). Dissociation is linked to ionization of His62 and its symmetry-related partners, His162, His262, and His362. In a similar fashion, double R67 DHFR is a dimer at pH 8.0 that dissociates into monomers at pH 5.0. Quadruple R67 DHFR would not be expected to dissociate, since the covalent linkages introduced at the gene level would transform the native 78-amino acid polypeptide into domains. However, quadruple R67 DHFR may undergo a conformational change from a “closed” to a “more extended” form upon histidine ionization as the pH decreases. Therefore, the ability of quadruple R67 DHFR to fold into a native-like structure was evaluated using several techniques at low and high pH.

Sedimentation equilibrium data indicate quadruple R67 DHFR is a single species of molecular mass 33,400 and 35,050 at pH 5 and 8, respectively. In contrast, triple R67 DHFR is a mixture of dimeric and monomeric species with a monomeric molecular mass of 25,300 at pH 5. At pH 8.0, triple R67 DHFR is a mixture of tetrameric and dimeric species with molecular weights that are 4- and 2-fold multiples of 25,000. Apparently the gene triplication product can associate into several different protein oligomerization states. It is surprising that a monomeric triple species is observed at pH 5, since it would be expected to have hydrophobic residues exposed to solvent (an intersubunit β-barrel forms at the monomer-monomer interface in native R67 DHFR). The triplication gene product was not characterized further due to the mixture of species.

At pH 8, quadruple R67 DHFR yielded a sedimentation coefficient of 2.75 S. Briefly, shape information was obtained from the sedimentation coefficients by estimating the frictional factor due solely to shape by first calculating $f/f_0$, and then correcting this value for hydration assuming 0.4 g of H$_2$O/g of protein. At pH 8, this yields an estimated value for the frictional coefficient for a toroid, $f_0$ ($4.86 \times 10^{-11}$). This value is in good agreement with the predicted value of $f_0$, for a toroid with $r = 8.5$ Å and $R = 16.5$ Å, the average values for R67 DHFR estimated from the X-ray structure, where $R$ is the distance from center of mass of the torus to the center of the ring, and $r$ is the inner toroidal radius (18). At pH 5, a sedimentation coefficient of 2.36 S was obtained for quadruple R67 DHFR, which immediately suggested a hydrodynamic swelling of the protein at low pH values. An estimate for $f_0$ of $5.652 \times 10^{-8}$ g/s at pH 5 was obtained.

The experimentally determined value for $f_0$ obtained at low pH was compared with estimates for frictional coefficients for various shapes (prolate and oblate ellipsoid, toroid, cylinder) based on approximate dimensions of the X-ray structure. This comparison suggested that at low pH, the molecule could be represented either as a “swollen” toroid with approximately a 21% increase in hydrodynamic volume or as an elongated cylinder-like particle with a width of 18 Å and a length of about 150 Å, yielding a linker length of about 58 Å. This linker length approximates that expected if the 16-amino acid linker were to adopt an extended conformation.

Circular dichroism spectra for quadruple R67 DHFR at pH 5.0 and 7.0 are shown in Fig. 4. The spectrum for quadruple R67 DHFR at pH 7 is similar to those for native and double R67 DHFRs, although subtle changes are observed. A larger difference is observed at pH 5.0. A more quantitative measure of the pH-dependent association/dissociation is afforded by alterations in protein fluorescence as Trp338 and its symmetry-related residues Trp338, Trp338, and Trp338 occur at the dimer-dimer interfaces in native R67 DHFR (or the monomer-monomer interfaces in double R67 DHFR or the domain-domain interfaces in quadruple R67 DHFR). The triplication gene product was additionally shown for comparison.

**FIG. 3.** A 14% SDS-polyacrylamide gel showing purified native R67 DHFR, double R67 DHFR, triple R67 DHFR, and quadruple R67 DHFR. Lanes A and F are molecular mass markers (10-kilodalton ladder; Life Technologies, Inc.), lane B is native R67 DHFR, M, 8,430; lane C, double R67 DHFR, M, 16,860; lane D, triple R67 DHFR; lane E, quadruple R67 DHFR. Proteins were visualized by Coomassie Blue staining.

**FIG. 4.** Circular dichroism spectra of native R67 DHFR (dashed line), double R67 DHFR (dotted line), and quadruple R67 DHFR (solid line) at equivalent concentrations (5 μM at pH 7). Data in panels A and B were obtained at pH 7.0 and 5.0, respectively. At pH 5.0, native R67 DHFR is dimeric, while double R67 DHFR and quadruple R67 DHFR are monomeric. At pH 7.0, native R67 DHFR is predominately tetrameric, double R67 DHFR is predominately dimeric, and quadruple R67 DHFR is monomeric. Data are presented as the mean residue ellipticity by taking 108 g/mol as the mean residue molecular weight. θ is the molar ellipticity expressed in degrees × cm$^2$/dmol.
the fluorescence of the more extended monomer for quadruple R67 DHFR is decreased with respect to the closed monomeric form. Clearly, the presence of 0, 2, and 3 linkers in the native, double, and quadruple R67 DHFRs has resulted in changes in the local environment of Trp38 and its symmetry-related partners. Since the natural N terminus is disordered from residues 1–17 in the dimer crystal structure, a reduction in the number of accessible conformations associated with the N terminus due to covalent tethering could result in interactions with the surface of the protein, including the normal dimer-dimer interface area where Trp38 resides.

Fig. 5B plots the titration of the intensity-averaged wavelength (λ) as a function of pH for native R67 DHFR, double R67 DHFR, and quadruple R67 DHFR. The curves are superimposable for the native and double proteins, while the pKα for the quadruple protein is obviously decreased. Best fit values are shown in Table I. K_{overall} can be expressed as K_{a}^{0.5}/K_{d} for the following linked equilibria,

\[
K_{a}^{0.5} = \frac{K_{d}}{K_{m}}
\]

where K_{d} is a dissociation constant; K_{a} is an ionization constant; T, D, and DH_{2} are tetramer, dimer, and protonated monomer; and n is the number of protons. For native R67 DHFR, the pKα value has been estimated as 6.84 by global fitting of three protein concentration-dependent fluorescence titration curves and ≥6.77 from NMR titrations of the C-2 proton of His62 (12). A comparison of the pKα values (5.5 versus 6.84) clearly indicates that more acid is necessary to protonate His^{62}, His^{104}, His^{202}, and His^{303} in the gene quadruplication product than in the native tetramer.

Is Quadruple R67 DHFR Active?—Quadruple R67 DHFR is almost fully active as shown by a comparison of k_{cat} and K_{m} values in Table II. In addition, fluorescence quenching was used to evaluate binding of NADPH at a low protein concentration. K_{d} values for native, double, and quadruple R67 DHFRs are 2.0 μM, 2.1 μM, and 2.0 ± 0.3 μM, while binding stoichiometries are 1.1, 1.2, and 1.1, respectively. The close identity of all these values indicates minimal perturbations in the active site of quadruple R67 DHFR due to gene quadruplication.

Equilibrium Folding/Unfolding Curves—Equilibrium unfolding of R67 DHFR at pH 5.0 was induced by GdnHCl and monitored by changes in protein fluorescence as shown in Fig. 6. Best fit values are given in Table III. Comparing equilibrium
topologies at pH 5 (Fig. 1B), where the domains would be constrained to remain near each other after His\(^{62}\) protonation. The cylindrical shape would be more consistent with an ABCD and/or ABDC topology, where the domains can move away from each other upon His\(^{62}\) protonation. For the cylindrical shape, the linker connecting the AB and CD domains at pH 5 should be flexible (i.e., two beads on a string analogy). Thus, quadruple R67 DHFR would not be expected to assume only an elongated, cylinder-like shape at pH 5.

While the topology of quadruple R67 DHFR has not been established, we note that double R67 DHFR only forms monomers where the AB domains (=CD) are linked. If double R67 DHFR monomers folded such that the AD domains (=CB) were linked, unfolding of the double dimer at pH 5 would be protein concentration-dependent, since the monomers would be held together by intersubunit β-barrel formation at the monomer-monomer interfaces. However, unfolding at pH 5 is protein concentration-independent. In addition, if the AD domains were linked (versus AB) in double R67 DHFR, pH titration of fluorescence monitoring the environment of Trp\(^{38}\) would be protein concentration-independent as the molecule changes from a toroid to a “swollen toroid” shape. However, the pH titration of fluorescence is protein concentration-dependent, consistent with dissociation of AB and CD monomers from each other upon titration of His\(^{62}\). These results are compatible with the folding pathway dictating a unique topology in double R67 DHFR and also perhaps indicate that the ≈42-Å Å link (straight line distance) across the native R67 DHFR dimer-dimer interface is more difficult to achieve.

The active conformation of quadruple R67 DHFR is more stable to pH titration than either the native tetramer or dimer, double R67 DHFR, since it takes more acid to titrate His\(^{62}\), His\(^{162}\), His\(^{362}\), and His\(^{362}\) residues as shown in Fig. 5. This occurs since the reaction being monitored is intramolecular rather than intermolecular, i.e., the effective concentration of the associating domains in quadruple R67 DHFR is greatly increased by the presence of the covalent tether.

The DG\(_{H2O}\) values obtained from equilibrium folding/unfolding studies at pH 5 imply that native R67 DHFR is more stable than either quadruple or double R67 DHFR although more GdnHCl is required to unfold the latter two as seen in Fig. 6. This apparent discrepancy arises from comparing a bimolecular reaction with a monomolecular reaction. In other words, the unfolding/refolding of native R67 DHFR is protein concentration-dependent, since the equilibrium being considered is dimer ≈ two unfolded monomers. Although denaturation of 2.0 μM R67 DHFR requires lower GdnHCl concentrations than denaturation of 2 mM double or quadruple R67 DHFR, the unfolding curve for R67 DHFR at some higher protein concentration will overlay the quadruple and double R67 DHFR unfolding curves. At even higher concentrations of R67 DHFR, the unfolding curve will require more GdnHCl for unfolding than either double or quadruple R67 DHFR. Thus, the DG\(_{H2O}\) value for R67 DHFR is a 1 σ standard state value; at 1 μM protein, the fraction of unfolded R67 DHFR will be lower than that for

**DISCUSSION**

One requirement for creating functional protein from tandem gene arrays is that the N and C termini are close enough in three-dimensional space so that the linker can connect the termini. Another requirement is that the tether not block access to the active site. These requirements are clearly met in quadruple R67 DHFR, which is monomeric and almost fully active.

Whether quadruple R67 DHFR folds into a unique topology is not clear. Sedimentation velocity experiments indicate the conformation of quadruple R67 DHFR changes as a function of pH. The shapes most consistent with the gene quadruplication product at pH 5 are either a toroid swollen by 21% or a cylindrical shape. A “swollen” toroid conformation might approximate the shape of either ACDB, ADCB, ADBC, and/or ACBD
Asymmetric Mutations—Asymmetric-hetero-oligomeric protein species have previously been constructed by mixing native and mutant protomers (22–25). However, this approach is not appropriate for R67 DHFR, since the active tetramer dissociates into inactive dimers at low pH due to titration of four symmetry-related histidines (His62, His162, His262, and His362) that occur at the dimer-dimer interfaces (12). Thus, even if a heterotetramer were constructed, it would quickly reequilibrate into a mixture of hetero- and homospecies at pH values used to monitor catalysis. Therefore, asymmetric mutations need to be introduced into R67 DHFR by a different strategy, e.g. gene quadruplication. In this procedure, the mutant R67 DHFR gene carried in pUC8 would be utilized instead of LZ1 in Fig. 2. This would put a single mutation in either copy 1 or 4, depending on where the mutation occurs with respect to the BatEII site.

An alternate strategy for creating asymmetric mutations would be formation of a heterodimer using the gene triplication product and a single copy gene product containing a mutation. In this heterodimer, one monomer would be 3 times heavier than other monomer. To form heterodimer, both proteins would be unfolded, mixed, and refolded together at pH 5.0. Heterodimers would then be separated from homo-oligomers. Since this heterodimer would involve intermolecular β-barrel formation between monomers, it would be more stable than heterotetramers formed by mixing native and mutant dimers. The latter have no cross-links/tethers that would hold the heterospecies together across the dimer-dimer interface, when titration of His62, His162, His262, and His362 occurs. This heterodimer approach minimizes the recloning and screening steps involved in putting each mutation into a quadruplicated gene context.

Several previous studies have duplicated genes to produce active monomers from originally dimeric proteins, e.g. human immunodeficiency virus-protase and the gene V protein from bacteriophage f1 (26–29). Gene duplication in R67 DHFR, β-galactosidase, the RNA-binding protein ROF, and arc repressor (6, 30–32) have also resulted in active dimers. To our knowledge, this is the first study where a gene quadruplication has produced an active monomer.

Significance—The production of this new monomeric R67 DHFR species may be analogous to gene oligomerization events that presumably have occurred during evolution. It has been suggested that once tandem duplication has occurred, genetic drift allows the amino acid composition of the domains to change as a result of selective pressure, and alternate activities may arise. Thus, in a future study, accumulation of random mutations in the quadruplicated R67 DHFR gene product (while maintaining resistance to trimethoprim or the ability to restore fol end product prototropy to a fol minus strain of E. coli) may allow the introduction of a catalytic acid, the introduction of a specific binding site for NADPH as well as folate, and also the potential to decrease nonproductive binding modes.

Other studies of tandem gene arrays have focused on increasing binding avidity (33), introducing a high concentration of antisense sequences (34), determining whether identical domains fold independently (35), and producing polymers to increase protein stability (36). In addition, genetic drift between domains may produce homologous but not identical sequences, which in turn may facilitate the ability of the domains to fold independently (35).

Acknowledgments—We thank Dave Matthews, Narendra Narayana, and Xuong Nguyen-huu for sharing coordinates for R67 DHFR; Joe Beechem for the use of the circular dichroism spectropolarimeter; and Cynthia Peterson for helpful discussions and reading of the manuscript.

REFERENCES

1. Borchert, T. V., Abagyan, R., Jaenicke, R. & Wierenga, R. K. (1994) Proc. Natl. Acad. Sci. U.S.A. 91, 1515–1518
2. Borchert, T. V., Abagyan, R., Radha Kishan, K. V., Zeelen, J. P. & Wierenga, R. K. (1993) Structure 1, 205–213
3. Breiter, D. R., Resnik, E. & Banaszak, L. J. (1994) Protein Sci. 3, 2023–2032
4. Narayana, N., Matthews, D. A., Howell, E. E. & Xuong, N. H. (1995) Nat. Struct. Biol. 2, 1018–1025
5. Bradrick, T. D., Beechem J. M. & Howell, E. E. (1996) Biochemistry 35, 11414–11424
6. Zhuang, P., Ming, Y., Holland, J. C., Peterson, C. B. & Howell, E. E. (1993) J. Biol. Chem. 268, 26272–26279
7. Howell, E. E., Warren, M. S., Booth, C. L. J., Villafane, J. E. & Kraut, J. (1987) Biochemistry 26, 8591–8598
8. Ellis, K. J. & Morrison, J. F. (1982) Methods Enzymol. 87, 405–426
9. Blakey, R. L. (1980) Nature 281, 231–232
10. Horecker, B. L. & Kornberg, A. (1948) J. Biol. Chem. 175, 385–390
11. Baccanari, D., Phillips, A., Smith, S., Sinski, D. & Burchall, J. (1975) Biochemistry 14, 5267–5273
12. Nichols, R., Weaver, C. D., Eisinger, E., Blakley, R. L., Appleman, J., Huang, T.-H., Huang, F.-Y. & Howell, E. E. (1993) Biochemistry 32, 1695–1706
13. Royer, C. A., Mann, C. J. & Matthews, C. R. (1993) Protein Sci. 2, 1844–1852
14. Fersht, A. R. (1985) Enzyme Structure and Mechanism, 2nd Ed., pp. 155–175, W. H. Freeman and Co., New York
15. Eftek, M. R. (1994) Biophys. J. 66, 482–501
16. Stafford, W. F., III (1992) Anal. Biochem. 203, 295–301
17. Cantor, C. R. & Schimmel, P. R. (1980) Biophysical Chemical, Part II, pp. 565–570, W. H. Freeman and Co., San Francisco
18. Allison, S. A., Easterly, R. A. & Teller, D. C. (1980) Biopolymers 19, 1475–1489
19. Reece, L. J., Nichols, R., Ogden, R. C. & Howell, E. E. (1991) Biochemistry 30, 10985–10994
20. Matthews, D. A., Smith, S. L., Baccanari, D. P., Burchall, J. J., Oatley, S. J. & Kraut, J. (1986) Biochemistry 25, 4194–4204
21. Brisson, N. & Hohn, T. (1984) Gene (Amst.) 29, 271–275
22. Bradshaw, T. P. & Dunlap, R. B. (1995) Biochemistry 32, 12774–12781
23. Day, P. J., Murray, I. A. & Shaw, W. B. (1995) Biochemistry 34, 6416–6422
24. Ward, W. H. & Fersht, A. R. (1988) Biochemistry 27, 1041–1049
25. Yang, Y. R. & Schachman, H. K. (1980) Proc. Natl. Acad. Sci. U.S.A. 77, 5187–5191
26. Liang, H., Sandberg, W. S. & Terwilliger, T. C. (1993) Proc. Natl. Acad. Sci. U.S.A. 90, 7010–7014
27. Babu, L. M., Rose, J. & Craik, C. S. (1992) Protein Sci. 1, 1244–1253
28. Dilanni, C. L., Davis, L. J., Holloway, M. K., Herber, W. K., Darke, P. L., Kohl, N. E. & Dixon, R. A. F. (1990) J. Biol. Chem. 265, 17348–17354
29. Cheng, Y.-S., Yin, F. H., Foundling, S., Bloemstron, D. & Kettner, C. A. (1990) Proc. Natl. Acad. Sci. U.S.A. 87, 9660–9664
30. Kuchinke, W. & Muller-Hill, B. (1985) EMBO J. 4, 1067–1073
31. Predki, P. F. & Regan, L. (1995) Biochemistry 34, 9834–9839
32. Robinson, C. R. & Sauer, R. T. (1996) Biochemistry 35, 109–116
33. Smythe, J. A., Nardelli, B., Chatterjee, P., Gallo, R. C. & Gershoni, J. M. (1994) Protein Eng. 7, 145–147
34. Graham, G. J. & Maio, J. J. (1992) BioTechniques 13, 780–789
35. Iwakura, M. & Matthews, C. R. (1992) Protein Sci. 1, 22672–22679
36. Baccanari, D., Phillips, A., Smith, S., Sinski, D. & Burchall, J. (1975) Biochemistry 14, 5267–5273