Solubilizing Buried Domains of Proteins: A Self-Assembling Interface Domain from Glutathione Reductase†

Bernd Leistler‡ and Richard N. Perham*

Cambridge Centre for Molecular Recognition, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, U.K.

Received September 28, 1993; Revised Manuscript Received November 15, 1993®

ABSTRACT: In the dimeric glutathione reductase (GR) from *Escherichia coli*, the interface domain is largely surrounded by the other three domains in each subunit of the protein. Subgenes encoding three forms of the interface domain have been expressed in *E. coli* and the products purified from inclusion bodies: INT is the excised interface domain, as it is found in native GR; INT^N and INT^{FN} are variants carrying exchanges of surface residues in what would have been hydrophobic contact regions with other neighboring domains. The isolated INT domain was found to be a soluble and folded protein, but it was isolated as a mixture of the dimer and at least two species of higher molecular weight. The latter were believed to arise by further association of the dimer via the newly exposed and unsatisfied hydrophobic contact regions. In the variant INT^N, three hydrophobic residues normally involved in the contact with the NADPH-binding domain in GR were replaced. This partly suppressed the further aggregation of the dimers. However, continued aggregation at high protein concentrations suggested that at least one further site of unwanted aggregation was still present. After four additional amino acid replacements in the region normally in contact with the FAD-binding domain, the resulting variant INT^{FN} exhibited no unspecific aggregation, even at concentrations as high as 3.2 mg/mL. The conformational stability of INT^N and INT^{FN} was not affected by these exchanges, as judged by superimposable cooperative GuCl-induced equilibrium unfolding transitions at 1.0 M GuCl. Conditions were also found for GR and for INT^{FN} under which they dissociate into compact monomers. The effects on dimerization of mutations introduced at the subunit interface can therefore now be studied in both proteins. An equilibrium dissociation constant of 25 μM was estimated for INT^{FN}. This work shows that it is possible to generate a soluble, folded form of a normally buried protein domain by the rational introduction of a small number of amino acid exchanges at its newly exposed hydrophobic surfaces, without interfering with its intrinsic structure and stability. The repertoire of protein folds accessible to design and redesign is thus substantially increased.

Recognition between the subunits of an oligomeric protein must be highly specific to guard against adventitious assembly with the subunits of other, even homologous, proteins in the same cell and sufficiently strong to maintain assembly. Although the amino acid composition of subunit interfaces frequently resembles that of the interior of proteins and protein domains (Chothia & Janin, 1975; Janin et al., 1988; Miller, 1989), these processes of molecular recognition often impose on them properties that are somewhat different. Thus, whereas proteins and functional domains within proteins have evolved to maintain a delicate balance between stability and flexibility, thereby optimizing their biological activity in a particular environment (Jaenicke, 1991a,b), subunit interfaces are generally densely packed and rigid parts of the protein.

The homodimeric enzyme glutathione reductase (GR)[1] is an interesting example. Crystallographic studies of the human (Karplus & Schulz, 1987) and homologous *Escherichia coli* (Ermler & Schulz, 1991) enzymes have revealed that each subunit consists of four well-delineated domains: an FAD-binding domain and an NADPH-binding domain (both $\beta\alpha\beta\alpha\beta$ Rossmann folds), followed by a smaller central domain and an interface domain (Figure 1). The two identical active sites are located at the dimer interface across a 2-fold axis and consist of amino acid residues contributed by both subunits. The subunit interface itself comprises two distinct parts, which are structurally dissimilar and separated by a solvent-filled cavity. The major ("upper") part is generated exclusively by the interface domain, which is the most rigid and best conserved region of the enzyme. The second ("lower") part is made up of extensions from the FAD-binding domain: this part is the most flexible region of the molecule and is relatively poorly conserved. It appears from the crystal structure that it makes little or no contribution to dimer stability (Karplus & Schulz, 1987); in keeping with this conclusion, neither the introduction of an intersubunit disulfide bridge at position 75 in this region of the *E. coli* enzyme (Scrutton et al., 1988) nor the introduction of mutually repelling electrostatic charges in the same position (Scrutton et al., unpublished results) has any detectable effect on the thermal stability of the *E. coli* enzyme. In contrast, the exchange of Gly418 for a bulky tryptophan (G418W) in the upper part of the dimer interface causes a significant disruption of the protein structure, insufficient to prevent assembly to the dimeric enzyme but substantially raising the susceptibility to thermal denaturation. This distortion is accompanied by the acquisition of cooperative

† B.L. was an Overseas Fellow of The Wellcome Trust.

* To whom correspondence should be addressed (telephone, +44 223 333663; Fax, +44 223 333345).

‡ Present address: elias Entwicklungslabor für Immunoassays GmbH, Obere Hardtstrasse 18, D-79114 Freiburg im Breisgau, Germany (telephone, +49 761 478 0563; Fax, +49 761 478 0520).

® Abstract published in *Advance ACS Abstracts,* February 1, 1994.

[1] Abbreviations: CD, circular dichroism; DTT, dithiothreitol; EDTA, ethylenediaminetetraacetate; GR, glutathione reductase (EC 1.6.4.2); GuCl, guanidinium chloride; INT, interface domain of *Escherichia coli* GR, comprising residues 334–450; INT^N, INT variant I339E, V343S, and I349S; INT^{FN}, INT variant I339E, V343S, I349S, M378E, V382S, T383S, and T384S.

0006-2960/94/0433-2773$04.50/0

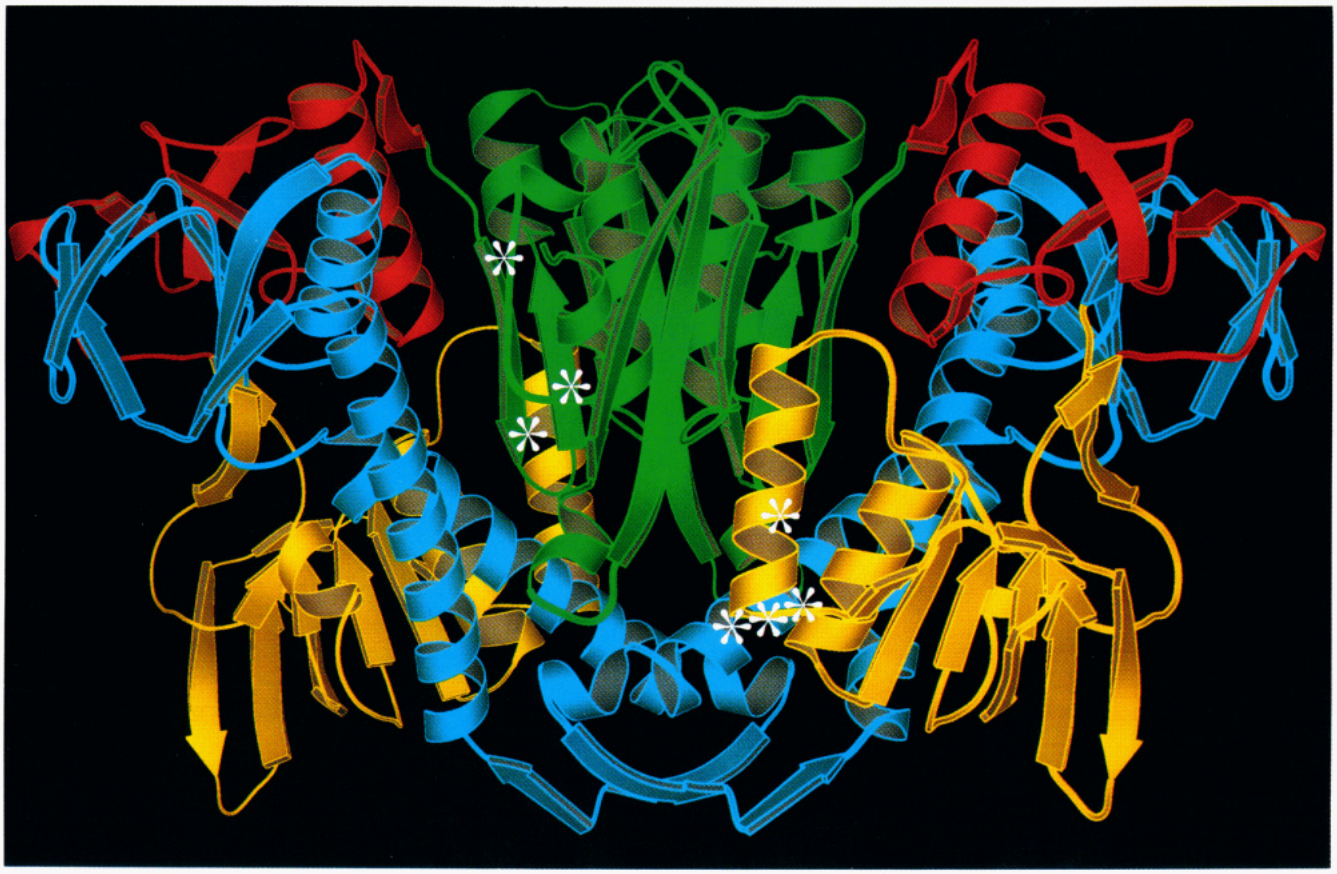

FIGURE 1: Domain structure of *E. coli* glutathione reductase. View perpendicular to the 2-fold symmetry axis of the GR dimer. The interface domain (containing the upper interface region) is depicted in green, the FAD-binding domain (providing the lower interface) is depicted in blue, the NADPH-binding domain is depicted in yellow, and the central domain is depicted in red. The sites of the amino acid exchanges (Figure 2) are indicated by asterisks. The structure was generated by the program MOLSCRIPT (Kraulis, 1991).

kinetic behavior with respect to the substrate glutathione (Scrutton et al., 1992).

In order to obtain a better idea of the real contribution of the upper interface region to dimer stability, we have attempted to create an independent interface domain (INT) of *E. coli* GR. This interface domain (molecular mass 12.8 kDa) is stabilized by a central 5-fold β-pleated sheet, and in the intact enzyme it is buried to a large extent by the other three domains (Figure 1). These contacts with other domains leave behind unsatisfied hydrophobic areas on the surface of the interface domain when it is generated in an independent form. To prevent nonspecific aggregation of the domain, we have found it necessary to identify these contact regions and to amend their hydrophobic nature by the suitable introduction of charged and hydrophilic amino acid residues. The product is a soluble domain whose general folding appears undisturbed by the changes wrought in it and which is still capable of specific dimerization.

These experiments indicate that it should be possible to create soluble forms of other folding domains from globular proteins and open the way to a study of protein folding and evolution at the level of individual folding units.

MATERIALS AND METHODS

Materials. *E. coli* GR was purified as described by Scrutton et al. (1987). $VENT_R$ DNA polymerase and the restriction enzymes *Nco*I and *Hin*dIII were obtained from New England Biolabs. The plasmid pET11d and the host *E. coli* strain BL21(DE3)plysS were purchased from AMS Biotechnology. Complex bacteriological media were from Difco Laboratories, and all media were prepared as directed in Sambrook et al. (1989). [^{35}S]dATP-αS triethylammonium salt (400 Ci/mmol) for DNA sequencing was supplied by Amersham International. All reagents used were of the highest purity available.

Recombinant DNA Techniques. DNA fragments (0.56 kb) containing the coding sequences of the INT domain and of its three variants were amplified by polymerase chain reaction (PCR) from the plasmid pKK223-3(C42A) described by Deonarain et al. (1990). To amplify the coding sequence of the INT domain, we used the following terminal primers (mismatches underlined): (a) 5′-CCGGATGAGCCCATG-GATTACAGC-3′, which primes on the coding strand and contains an *Nco*I cleavage site to provide a new in-frame start codon, and (b) 5′-GGCTGAAAATCTTCTCT-3′, which primes on the noncoding strand downstream of the single *Hin*dIII restriction site which in turn is situated downstream of the termination signal of the *gor* gene. To amplify the coding sequence of the INT^N variant, primer a was replaced by a primer with sequence c: 5′-GGACACTTTGCCATG-GATTACAGCAACGAACCGACCGTGTCCTTCAGCC-ATCCGCCGAGTGGTACTGTT-3′. It contains an *Nco*I cleavage site and codes for the three desired amino acid exchanges in the region contacting the NADPH-binding domain. This oligonucleotide was designed for use with a different plasmid but nonetheless worked well for this reaction. The coding sequence for INT^{FN} was obtained in two steps, using two overlapping mutagenesis primers (Higuchi, 1989). First, two 0.23-kb DNA fragments were amplified using primer c and primer d, 5′-GAGGAGGAGGCGGTATATTCCGCG-GTGAAAGAGGA-3′, which primes on the noncoding strand and contains the mutations to code for the desired amino acid exchanges in the region contacting the FAD domain. A 0.31-

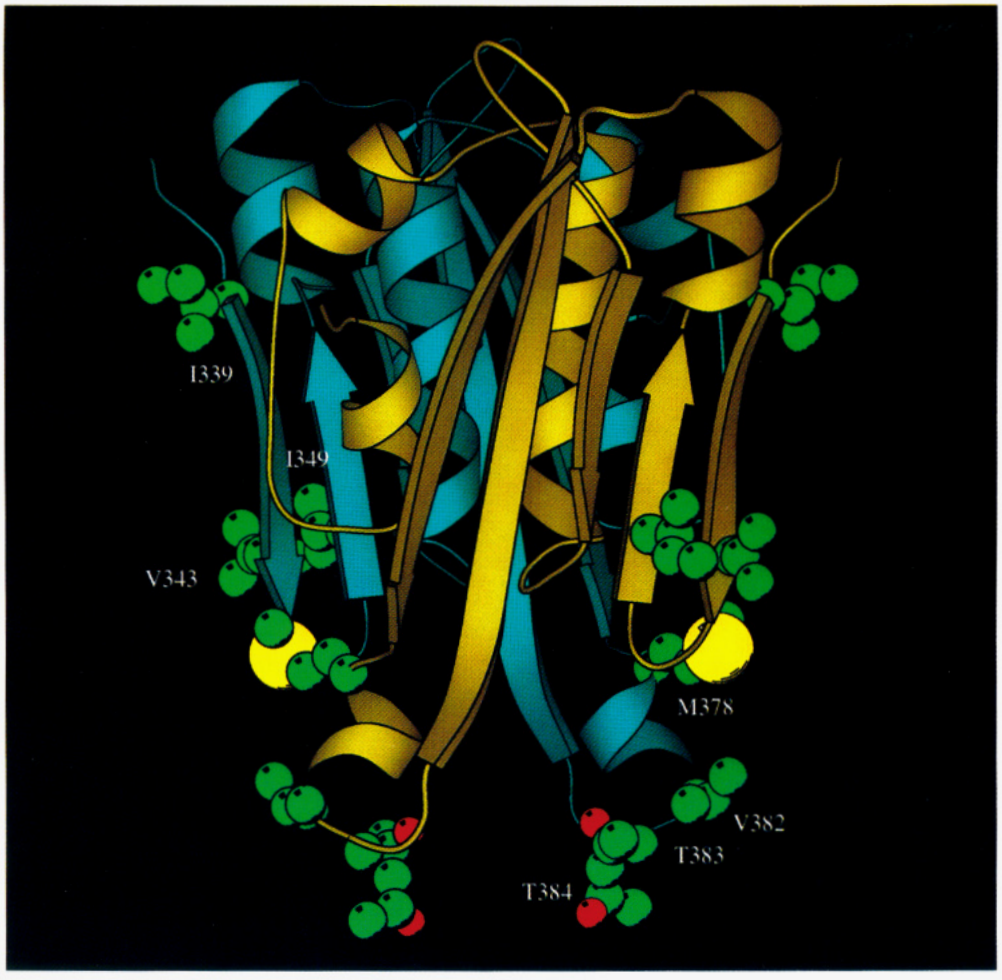

FIGURE 2: Location of the amino acid exchanges in the structure of the interface domain. The dimer form of the interface domain is taken from the structure of intact GR, viewed as in Figure 1 . The two contributing domains are depicted in pale green and blue. The seven amino acid residues exchanged [three (Ile339, Ile349, and Val343) at the contact site with the NADPH-binding domain, followed by four (Met378, Val382, Thr383, and Thr384) at the contact site with the FAD-binding domain] to give the INT^N and INT^{FN} variants, respectively, are indicated (labeled for the blue domain only). The structure was generated by the program MOLSCRIPT (Kraulis, 1991).

kb DNA fragment was amplified using primer b and primer e, 5′-GGAATATACCGCCTCCTCCTCTCACCGCCA-GCCGT-3′, which primes on the coding strand and hybridizes with primer d at 21 bases around the sites of the mutations. After purification, the fragments were used as templates for primer b and primer c to amplify a 0.56-kb fragment containing the 0.35-kb coding sequence of INT^{FN}, which also contained the *Nco*I and *Hin*dIII restriction sites.

The polymerase chain reactions were carried out under mineral oil in a volume of 100 μL. The reaction buffer contained 200 mM dNTPs, 2 mM $MgCl_2$, 100 pmol of each sense and antisense primer, 0.025 pmol of plasmid or 0.23- and 0.31-kb fragments, and 2 units of $VENT_R$ DNA polymerase. The reactions were carried out on a programmable TECHNE PHC-3 thermal cycler. The cycling parameters with heating and cooling rates of 1 °C/s were as follows: 2 min at 95 °C (1 cycle); 1 min at 94 °C, 2 min at 50 °C, and 3 min at 72 °C (25 cycles); 10 min at 72 °C (1 cycle). After digestion of the purified 0.56-kb fragments with *Nco*I and *Hin*dIII, the resulting 0.53-kb DNA fragments were subcloned into the pET11d expression vector (Studier et al., 1990). The genes were sequenced by the dideoxy method (Sanger et al., 1977) using the Pharmacia T7 sequencing kit according to the manufacturer's instructions. The plasmids carrying the coding sequences for INT and the two variants were transformed into BL21(DE3)plysS cells (Studier et al., 1990). All other recombinant DNA techniques followed standard protocols (Sambrook et al., 1989).

Choice of Amino Acid Exchanges. To predict the sites of unspecific aggregation at the surface of INT, the side chains of the interface domain in a refined X-ray structure of *E. coli* GR (G. E. Schulz, unpublished data) were screened for the number of van der Waals contacts they make with residues of the other three domains. The search was done automatically using the protein structure analysis program NAOMI (Brocklehurst & Perham, 1993). The two clusters of nonpolar contacts thus detected were assumed to be exposed in INT and therefore potential sites of nonspecific aggregation. The residues with the most prominent side chains in each cluster were then replaced (Figure 2). Serine residues were chosen for most of the exchanges because the serine side chain is polar but should not interact electrostatically with other side chains (other than by hydrogen bonding), and it confers no susceptibility to proteinases. One glutamic acid residue was introduced in each cluster in order to increase the net negative charge of INT, which is only −2. The glutamic acid residue introduced in the FAD-domain contact site had the additional purpose of maintaining a hydrophobic contact of the side chain of the exchanged residue, Met378, with the core of the domain.

Purification of Protein Domains. All proteins were purified as follows. An overnight culture (100 mL) of transformed *E. coli* cells was grown at 37 °C in 2 × TY medium (Sambrook et al., 1989) containing 100 μg/mL ampicillin and 50 μg/mL chloramphenicol. This was used to inoculate 1 L of the same medium. The cells were grown at 37 °C to an A_{600} of 0.5–0.7,

and expression of protein was then induced with 1 mM isopropyl thiogalactoside. After growth for a further 4–6 h, the cells were harvested by centrifugation and resuspended in 100 mM potassium phosphate buffer, pH 7.8, containing 5 mM DTT, 2 mM EDTA, 0.5M NaCl, and 1% Triton TX-100 for disruption in a French press. The insoluble material was recovered by centrifugation and was washed and centrifuged four times in the same buffer. It was then dissolved and stirred for 60–90 min in approximately 20 mL of 100 mM potassium phosphate buffer, pH 7.8, containing 0.1 M DTT, 2 mM EDTA, and 6 M GuCl. The solution was clarified by centrifugation (30 min at 20000*g*) and diluted to a final volume of 250–300 mL by addition of 100 mM potassium phosphate buffer, pH 7.8, containing 1 mM DTT, 1 mM EDTA, and 4 mM GuCl, after which the GuCl was removed by dialysis (three changes) against a large volume (at least 10-fold) of refolding buffer (100 mM potassium phosphate, pH 7.8, containing 1 mM DTT, 1 mM EDTA, 100 mM KCl, and 10% glycerol) at 4 °C. Further purification was achieved by anion-exchange chromatography on a HiLoad Q Sepharose column (Pharmacia). After overnight dialysis at 4 °C against 0.05 M NH_4HCO_3, pH 7.7, containing 1 mM DTT and 1 mM EDTA, the proteins were loaded onto the column and eluted with a linear gradient (100 mL) of the same buffer (0.05–1.0 M NH_4HCO_3) at the same temperature and a flow rate of 2 mL/min. All purified proteins were stable on ice for at least several days, as judged by gel filtration. For long-term storage, the proteins were dripped into liquid nitrogen and the frozen pellets kept at −80 °C.

Mass Spectrometry. Samples of the pure proteins were dialyzed against 0.05 M NH_4HCO_3 and lyophilized from water twice. Samples were redissolved in 69% water, 20% acetonitrile, 10% 2-propanol, and 1% acetic acid at a protein concentration of 10–20 pmol/μL. Electrospray mass spectrometry was performed on a Hewlett-Packard MS-Engine-5989A mass spectrometer using myoglobin as the calibration standard.

Determination of Protein Concentration. INT, INT^N, and INT^{FN} each contain four tyrosine and no tryptophan residues. Each domain also contains six phenylalanine residues, but their absorbance (maximum 248 nm; Levine & Federici, 1982) does not overlap with that of the tyrosine residues (maximum 276 nm). According to Levine and Federici (1982), the molar absorbance of the three proteins at 276 nm in a solution of 6 M GuCl should therefore be 5320 M^{-1} cm^{-1}. The value in 50 mM potassium phosphate buffer, pH 7.8, containing 1 mM DTT, 1 mM EDTA, and 150 mM NaCl was measured to be 4993 M^{-1} cm^{-1}. This is lower by 15% than that determined by the method of Bradford (1976). The spectra were measured on a Hewlett-Packard 8452A diode array spectrophotometer with 2-nm resolution.

Analysis of Assembly and Unfolding by Gel Filtration. Analytical gel filtration was performed in 50 mM potassium phosphate, pH 7.8, containing 1 mM DTT, 1 mM EDTA, and 150 mM NaCl on a Superose 12 HR 10/30 column (Pharmacia) at a flow rate of 0.4 mL/min and on a Superdex 75 HR 10/30 column (Pharmacia) at 0.5 mL/min. Protein was detected at 230 nm. The Superose 12 column was used for the investigation of aggregation and unfolding and the Superdex 75 column for the determination of molecular weight. The columns were attached to a Pharmacia FPLC chromatography system equipped with an LKB 2141 variable wavelength detector detecting absorbance at 230 or 276 nm. Protein unfolding was monitored at 23 °C by following the change in elution volume from a Superose 12 gel filtration column with increasing concentrations of GuCl. Samples of 12–50 μg (2–8 μM) in a volume of 100 or 500 μL were injected automatically after each 20 mL of a linear gradient of GuCl in 50 mM potassium phosphate, pH 7.8, containing 1 mM DTT, 1 mM EDTA, and 150 mM NaCl. The samples had been equilibrated overnight on ice with the respective concentration of GuCl and brought to 23 °C 1 h before injection. The data were normalized by extrapolating the linear change in the elution volume between 1.0 and 4.0 M GuCl (peak of compact GR monomer), above 3.0 M GuCl (peak of unfolded GR), and above 2.4 GuCl (unfolded INT and INT variants) to zero GuCl concentration. The two separate peaks observed in the unfolding of GR between 2.4 M and 4.2 M GuCl were weighted by their areas, and the position of an imaginary intermediary peak was calculated between the two baselines. The normalized relative change of elution volume is the apparent fractional deviation (F_{app}) between both baselines (GR) or from the single baseline (INT and INT variants). Dissociation of INT^{FN} was monitored at 23 °C by following the change in elution volume from the Superdex 75 column after progressive dilution of a stock protein solution (3.2 mg/mL; 250 μM) with 50 mM potassium phosphate, pH 7.8, containing 1 mM DTT, 1 mM EDTA, and 150 mM NaCl. Samples of 100 μL were injected onto the column.

CD Spectroscopy. The far-ultraviolet CD spectra were recorded in 50 mM potassium phosphate buffer, pH 7.8, containing 0.1 mM DTT, 0.1 mM EDTA, and 20 mM NaCl on a JASCO J-720 instrument at 20 °C using a cuvette of 0.1-cm path length. The protein concentration was 12 μM. Five repetitive scans were recorded at a scan speed of 50 nm/min (response time 1 s) and subsequently averaged. The buffer baseline was subtracted.

SDS–Polyacrylamide Gel Electrophoresis. Discontinuous SDS–polyacrylamide gel electrophoresis was carried out by the method of Laemmli (1975), with the following acrylamide concentrations: stacking gel 5%; resolving gel 20%.

Measurement of Catalytic Activity. The catalytic activity of GR was assayed in the direction of NADPH oxidation at 25 °C under saturating substrate conditions (Scrutton et al., 1987). The effect of GuCl was tested by assaying after equilibration of the protein with increasing concentrations of GuCl in 50 mM potassium phosphate, pH 7.8, containing 1 mM DTT, 1 mM EDTA, and 150 mM NaCl.

RESULTS

Gene Expression and Domain Purification. The subgene encoding the C-terminal interface domain (INT, residues 334–450, Figure 2) of *E. coli* GR was expressed from the pET11d vector as described in Materials and Methods. The overexpressed protein was found exclusively in the insoluble fraction of the disrupted cells and amounted to more than 50% of the total cell protein. The same was true later for the INT^N and INT^{FN} domains (Figure 2), and they were all purified by the same method. The insoluble fraction of the cells was washed and dissolved in 6 M GuCl, and the GuCl was removed by dialysis; at this stage the domains were more than 90% pure, as judged by SDS–polyacrylamide gel electrophoresis (Figure 3). A 1-L culture typically yielded 2.5 g of wet cells and 80 mg of renatured and soluble protein. Higher purity was achieved by means of chromatography on a HiLoad Q Sepharose anion-exchange column. When 20 mg of INT or INT^N domains was loaded, the proteins were eluted in two successive peaks, at 0.20 and 0.47 M NH_4HCO_3, respectively, corresponding to two different species of high molecular mass (cf. Figure 5). A sample (50 mg) of INT^{FN} was eluted in

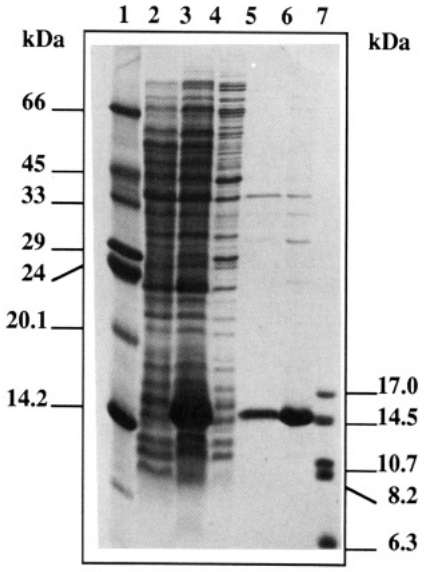

FIGURE 3: Expression and refolding of the interface domain. Samples of the cell-free extracts of *E. coli* cells expressing the INT domain were submitted to SDS–polyacrylamide gel electrophoresis at various stages of the purification process. Lanes: 1 and 7, marker proteins (molecular mass in kDa); 2, noninduced cells; 3, cells grown for 4 h after induction with 1 mM IPTG; 4, soluble fraction after cell disruption; 5, insoluble fraction after cell disruption; 6, redissolved inclusion bodies, after being washed, dissolved in 6 M GuCl, and dialyzed to permit refolding. Proteins were visualized by staining with Coomassie Brilliant Blue R250. For other details, see the text.

small part at the same salt concentrations, but the majority was eluted in a new peak at 0.33 M NH_4HCO_3.

The apparent M_r values inferred from SDS–polyacrylamide gel electrophoresis deviated somewhat from those expected. The INT domain migrated with an apparent molecular mass of 15 kDa (Figure 3) instead of 12.8 kDa calculated from its amino acid sequence, and the INT^N and INT^{FN} domains migrated with an apparent molecular mass of 17 kDa (data not shown). However, electrospray mass spectrometry gave molecular mass values which deviated insignificantly from the expected values: the measured values were 12 875.16 Da (+1.28 Da) for INT, 12 851.99 Da (+0.28 Da) for INT^N, and 12 810.60 Da (+1.53 Da) for INT^{FN} (the deviations from the molecular mass values calculated from the amino acid sequences are appended in parentheses). The reasons for the discrepancies in the SDS–polyacrylamide gel electrophoresis values are not clear, but the results of electrospray mass spectrometry are convincing evidence of the correctness of the intended constructs.

Circular Dichroism. The far-UV CD spectra of the purified domains between 195 and 260 nm are shown in Figure 4. That of the INT domain has a minimum at 216 nm, which is indicative of a high content of β-sheet. The baseline is intersected at 204 nm, which indicates a low content of random structure. The INT^N and INT^{FN} domains have superimposable spectra that differ significantly from that of INT. The negative ellipticity at 216 nm is 125% that of INT, and at 209 nm it is 161%. The baseline intersection is shifted slightly to 202 nm. An increased negative ellipticity generally indicates a higher content of secondary structure.

Suppression of Nonspecific Aggregation. After being refolded at 0.06 mg/mL (4.6 μM) in 100 mM potassium phosphate buffer, pH 7.8, containing 1 mM DTT, 1 mM EDTA, 100 mM KCl, and 10% glycerol, the INT and INT^N proteins were analyzed by gel filtration on a Superose 12 column. INT exhibited three main peaks (Figure 5A), with apparent molecular mass values of >1000, 115, and 20 kDa,

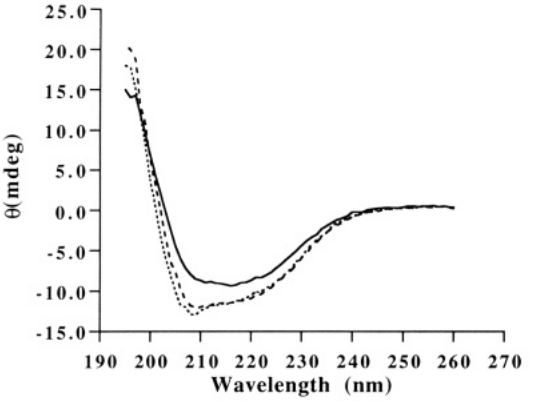

FIGURE 4: Far-UV circular dichroism spectra of the purified INT, INT^N, and INT^{FN} domains. Average of five scans each in a cell of 0.1-cm path length (protein concentration 12 μM). The buffer baselines are subtracted. Curves: —, INT; - - -, INT^N; ···, INT^{FN}.

respectively, whereas INT^N eluted predominantly as a 20-kDa protein (Figure 5B). These data indicate that INT aggregated easily beyond the expected dimer to perhaps an octameric species and even higher aggregates. The surface exchanges (I339E, V343S, and I349S on the region normally contacting the NADPH-binding domain) made to create the INT^N domain (Figure 2) appeared to have suppressed this nonspecific aggregation. However, at higher protein concentrations, INT^N also aggregated further. Figure 5C shows the elution profile of INT^N at a concentration of 0.66 mg/mL (52 μM). The putative octamers partially dissociated after addition of 0.4 M GuCl (data not shown), a concentration which leaves the dimer intact (cf. Figure 8), and they dissociated almost completely into dimers after addition of 0.4 M GuCl and dilution to a protein concentration of 0.07 mg/mL (5.4 μM). The aggregates of INT did not dissociate under these conditions.

To combat this residual aggregation, further amino acid replacements (M378E, V382S, T383S, and T384S) were introduced into INT^N to increase the hydrophilicity of the region normally contacting the FAD-binding domain in the native enzyme dimer (Figure 2). Thr383 and Thr384 were included in the exchanges because of the hydrophobic contacts made with the FAD-binding domain by the two side-chain methyl groups. The properties of the resulting INT^{FN} domain suggested a drastic reduction of the residual nonspecific aggregation, as desired. Even at a protein concentration of 3.2 mg/mL (250 μM), gel filtration indicated a single species with an apparent molecular mass of 26 kDa and revealed no evidence of any significant aggregation of the INT^{FN} dimers (Figure 6).

Dimer–Monomer Equilibrium. The expected molecular mass of the dimeric INT, INT^N, and INT^{FN} domain is 25.7 kDa. Thus an apparent molecular mass of 20 kDa measured by gel filtration is either anomalous or, more likely, corresponds to a dimer–monomer equilibrium with only 50% in the dimer form. Since the dimers of INT^N aggregated at protein concentrations higher than 8 μM, the amount of dimer appeared never to increase beyond this percentage. However, INT^{FN} did not aggregate nonspecifically beyond the dimer, and it was possible, therefore, to determine its quaternary structure over a wide range of concentrations. Figure 7 shows that the apparent molecular mass decreased when diluted samples of INT^{FN} were gel filtered through a Superdex 75 column. When the concentration of INT^{FN} applied to the column was decreased from 250 to 0.25 μM, the apparent molecular mass fell steadily from 26 to 16 kDa. This suggests

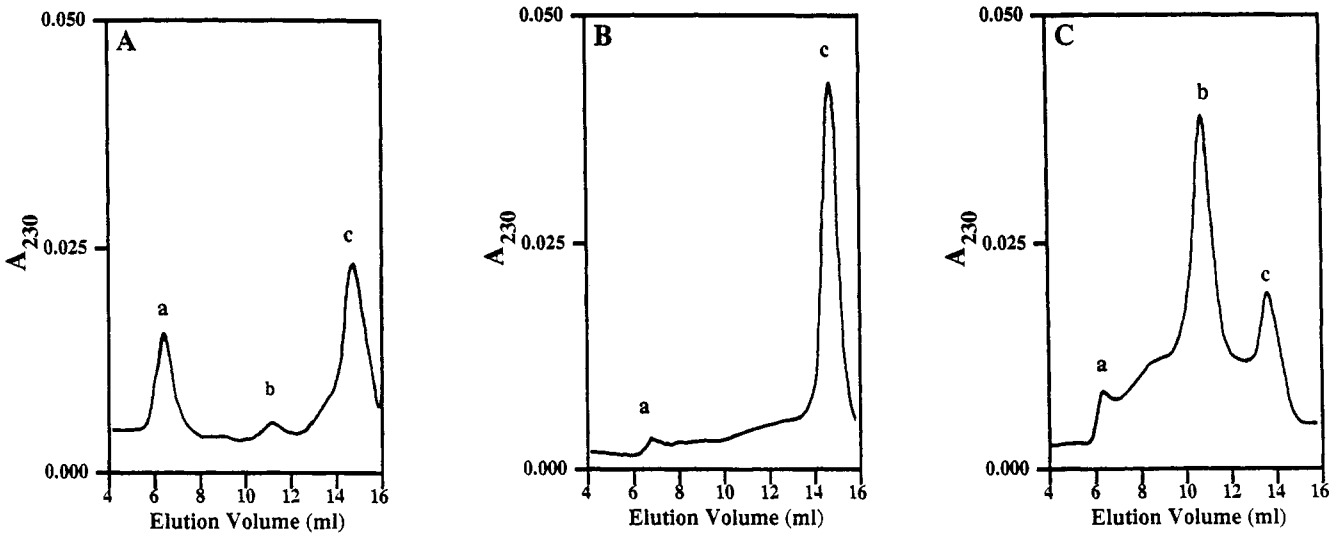

FIGURE 5: Effect of mutation of the region contacting the NADPH-binding domain. Purified domains were analyzed by gel filtration on a calibrated Superose 12 column in 50 mM potassium phosphate buffer, pH 7.8, containing 1 mM DTT, 1 mM EDTA, and 150 mM NaCl. Panels: (A) 50 μg of INT after refolding at 0.06 mg/mL (4.7 μM); (B) 50 μg of INT^N after refolding at 0.06 mg/mL; (C) 330 μg of INT^N after refolding at 0.66 mg/mL (52 μM). The elution volumes correspond to apparent molecular mass values, as follows: (a) >1000 kDa; (b) 115 kDa; (c) 20 kDa.

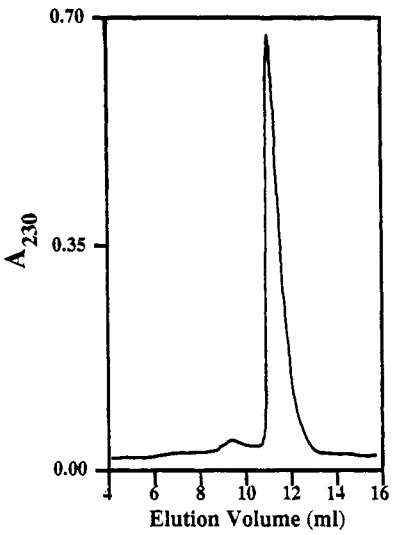

FIGURE 6: Effect of mutation of the regions contacting the NADPH- and FAD-binding domains. The purified domain (320 μg of INT^{FN}, 3.2 mg/mL, 250 μM) was analyzed by gel filtration on a calibrated Superdex 75 column in 50 mM potassium phosphate buffer, pH 7.8, containing 1 mM DTT, 1 mM EDTA, and 150 mM NaCl. The elution volume corresponds to an apparent molecular mass of 26 kDa. For other details, see the text.

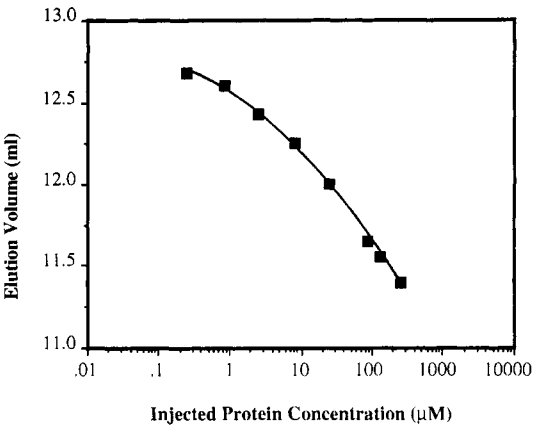

FIGURE 7: Dimer–monomer equilibrium of INT^{FN}. Samples (100 μL) of INT^{FN} diluted with 50 mM potassium phosphate buffer, pH 7.8, containing 1 mM DTT, 1 mM EDTA, and 150 mM NaCl were injected on a calibrated Superdex 75 column at the indicated protein concentrations. The column was developed with the same buffer, and the eluted protein was detected at 230 nm. For other details, see the text.

that the equilibrium dissociation constant of the dimeric species is around 25 μM, not taking into account the dilution on the column. In the transition region, the single peak was only slightly asymmetric (cf. Figure 6), indicating that the monomer and dimer are in an equilibrium that is rapid compared with the time of chromatography (23–25 min).

This experiment shows that the INT^{FN} domain can exist as a dimer or as a compact monomer, depending on the protein concentration. A similar decrease of the apparent molecular mass was obtained at a constant protein concentration by increasing the concentration of NaCl between 0.05 and 1.5 M NaCl (data not shown). The elution volumes of the standard proteins used to calibrate the column did not change under these conditions.

Stability of the Folded State. A cooperative equilibrium unfolding transition is indicative of a unique conformation and can therefore be taken as indicative of the successful design of a new protein. Similarly, the comparison of a wild-type and mutant protein is important in judging whether the amino acid replacements have affected the intrinsic stability. Analytical gel filtration is the method of choice for studying oligomeric systems in particular: dissociation into compact monomers and unfolding of the oligomers or protomers have opposite effects and are therefore distinguishable by this method (Herold & Leistler, 1991). The INT, INT^N, and INT^{FN} domains were found to have superimposable, cooperative unfolding transitions at 1.0 M GuCl (Figure 8). Therefore, the three (INT^N) or seven (INT^{FN}) amino acid exchanges did not appear to have changed the stability of the protein domain.

On the other hand, native GR unfolded with a wide transition around 3.0 M GuCl. Unlike the INT domain, GR was eluted in two distinct and symmetric peaks in the transition region (between 2.4 and 4.2 M GuCl), corresponding to the folded and unfolded forms of the protein. This indicates that the equilibrium between the folded and unfolded form of GR is slower than that of INT. The weak cooperativity of the GR transition suggests that more than a single process is involved.

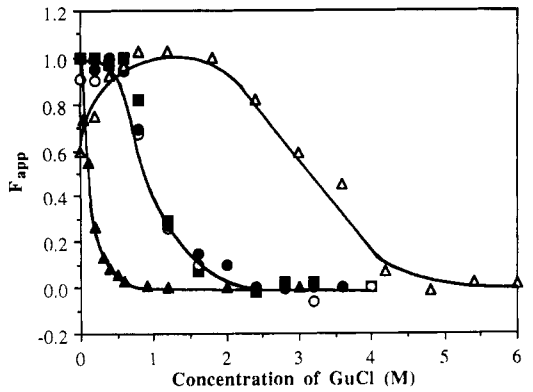

FIGURE 8: Unfolding of INT, INT^N, and INT^{FN} in the presence of GuCl. Equilibrium unfolding of (△) GR, (■) INT, (○) INT^N, and (●) INT^{FN} was monitored by following the change in elution volume from a Superose 12 gel filtration column with increasing concentrations of GuCl. The protein samples were equilibrated with the relevant concentration of the denaturant before application to the column. Protein concentrations were 7.8 μM (INT, INT^N, and INT^{FN}) and 2 μM (GR). (▲), catalytic activity of GR (1 μg/mL, 0.02 μM) measured under identical conditions. For other details, see the text.

The first transition of GR at 0.25 M GuCl monitors an apparent reduction in size, presumably reflecting dissociation into monomers. The transition is approximately matched by the loss of catalytic activity at 0.1 M GuCl. The monomeric form of the enzyme cannot be active since the catalytic site resides at the subunit interface.

Preliminary measurements (data not shown) showed that the catalytic activity of GR is unchanged at 150 mM NaCl, the salt concentration chosen for the gel filtration experiments. The first two GuCl-induced transitions of GR had no equivalent phase in the unfolding of INT or INT^N. This was probably due to the limited resolution of the Superose 12 column for proteins with molecular mass values smaller than 13 kDa and to only 35% dimerization of the INT domain under the experimental conditions (cf. Figure 7). In experiments with the Superdex 75 column, a partial GuCl-induced dissociation of INT into compact monomers at 0.05 M of the denaturant was observed.

DISCUSSION

Three different protein domains have been expressed in *E. coli* and purified from inclusion bodies. INT is the excised interface domain of glutathione reductase from *E. coli*, and INT^N and INT^{FN} are variants carrying exchanges of surface residues in what would have been contact regions with the NADPH- and FAD-binding domains, respectively (Figures 1 and 2). The isolated INT domain is completely soluble and has the properties of a folded protein. CD spectra (Figure 4) indicated the expected high content of β-sheet and a low content of random structure. Taken together with the cooperative equilibrium unfolding transition (Figure 8), this suggests that INT has a unique conformation. However, it was isolated as a mixture of at least three main species. The smallest of these was eluted from gel filtration columns with an apparent molecular mass of 20 kDa, which formally corresponds to a rapid monomer–dimer equilibrium and only 50% dimerization. The species of INT with high molecular mass were believed to be generated by nonspecific association caused by hydrophobic patches left exposed by the refolding process *in vitro* and which would normally form buried contacts with neighboring domains in the intact GR.

In the variant INT^N, three hydrophobic residues that would contact the NADPH-binding domain in GR were exchanged for more hydrophilic ones (I339E, V348S, and I349S). This largely suppressed the further aggregation of the dimers. Although high molecular mass forms occurred at high protein concentrations, they could be completely dissociated into folded protomers at low concentrations of GuCl (Figure 8). However, the residual nonspecific aggregation of INT^N suggested that at least one further crucial site of unwanted aggregation was still present. Our inference that this second aggregation site was located in what would have been the contact region with the FAD-binding domain is supported by the fact that the variant INT^{FN}, which carries additional amino acid exchanges (M378E, V382S, T383S, and T384S) in this region, did not show any nonspecific aggregation, even at protein concentrations as high as 3.2 mg/mL (Figure 6).

The conformational stability of INT^N and INT^{FN} was not affected by these exchanges. They exhibited cooperative GuCl-induced equilibrium unfolding transitions at 1.0 M GuCl that were superimposable (Figure 8). In contrast, intact GR unfolded with only weak cooperativity at 3.0 M GuCl, suggesting that this transition consists of more than a single process. It probably reflects the independent unfolding of some of the four domains of GR. Moreover, the equilibrium must be considerably slower than that of INT and the INT variants, since two separate symmetric peaks were observed in the transition region of GR unfolding, whereas INT and the INT variants were eluted as a single symmetric peak at all concentrations of GuCl.

Unfolding of GR was preceded by an apparent reduction in size at 0.25 M GuCl, suggesting that dissociation of the GR dimer, to folded monomers in the first instance, can be at least partly uncoupled from unfolding. This conclusion is reinforced by the correlation (allowing for the greater dilution at which the catalytic activity is assayed) with the concentration of GuCl at which GR becomes inactivated (Figure 8), since the catalytic site is at the dimer interface and only the dimer can be catalytically active (Karplus & Schulz, 1987, and references therein). Taking these findings together, the species that is stable between 0.8 and 1.8 M GuCl must be the folded monomer of GR. It seems a very unlikely alternative that the interface domain is completely unfolded under these conditions but that the other three domains in each polypeptide chain retain their native state, resulting in a partially native-like monomeric molecule which is more compact than the GR dimer. Corresponding dissociation without unfolding of the INT domain and its two variants was not detected in this system, probably because of the incomplete initial dimerization (approximately 35%) under the experimental conditions and the decreased resolution of the Superose 12 column for molecular mass values as low as 13 kDa. In further experiments with the Superdex 75 column, we did observe a partial GuCl-induced dissociation of INT into compact monomers at a GuCl concentration of 0.05 M. The behavior of GR and INT makes an interesting comparison with, say, the all-or-none transitions between the unfolded monomers and dimers of the trp repressor (Tasayco & Carey, 1992, and references therein).

From the dependence of the molecular mass on protein concentration, the equilibrium dissociation constant of INT^{FN} was estimated to be approximately 25 μM. This is at least 3 orders of magnitude higher than that of native GR, because GR is still active (and thus dimeric) at a protein concentration of 10 nM. It remains to be determined whether this decrease in affinity is due to the absence of interactions contributed by the lower interface region of GR, an area of 1200 $Å^2$ (Karplus & Schulz, 1987), or to conformational changes in the isolated

domain. For comparison, the upper interface region, which is present in INT and its variants, has an area of 2100 Å^2 (Karplus & Schulz, 1987). According to Chothia and Janin (1975), a loss of 1200 Å^2 of buried surface would account for a loss in hydrophobic free energy of >25 kcal/mol. However, Karplus and Schulz (1987) have doubted the applicability of such a calculation to the lower interface of GR, because it shows atypical features, such as low packing density and high backbone flexibility. It has been pointed out (Janin et al., 1988) that proteins with large subunit interfaces are unlikely to dissociate without any structural rearrangements. Thus, INT, the isolated interface domain of GR, is likely to undergo changes when it dissociates, especially with respect to the corresponding structure in GR, because it then exposes a relatively big area (2100 Å^2) derived from the upper subunit interface. The weak stability of the folded monomer of INT^{FN} in the presence of GuCl, compared with the folded monomer of GR, might therefore reflect an increased flexibility of the domain as an independent protein. A surface area of 2900 Å^2 is buried by other domains in GR (Janin & Wodak, 1983), which is even larger than the area of the dimer interface. It is probable that the interface domain is stable in the folded monomer and remains folded at higher concentrations of GuCl because it is stabilized by these additional packing interactions.

It is interesting to consider the results from CD spectroscopy in this context. INT^N and INT^{FN} had identical spectra with significantly more negative ellipticity than the spectrum of INT. All three amino acid exchanges in INT^N are located in secondary structure elements, namely, in the first and second β-strands (Figure 2). The additional four exchanges in INT^{FN} are located in the long loop around residues 382–384 between strands 4 and 5. It seems reasonable to infer that the exchanges in INT^N increased the hydrophobic gradient between the surface and the interior of the domain and thus favored the formation of secondary structure in this location. The second set of exchanges, those in INT^{FN}, had no further effect because the loop in which these residues are located does not contribute to the hydrophobic core of the domain. The stabilizing effect of the INT^N mutations may have resulted in an increased stability of the folded monomers. Unfortunately, because of the problems of nonspecific aggregation, it was not possible to test this by performing the GuCl-induced unfolding experiments with INT and INT^N at the same protein concentration at which the species of folded monomers of INT^{FN} was detected.

Domains are recognized as a widespread feature of protein structure (Bränden & Tooze, 1991; Creighton, 1993), and several individual domains have been isolated before now, by either limited proteolysis or manipulation of the gene encoding the parent protein. In almost all these instances, the domains (or structural modules) have been relatively independent globular units with few direct interactions with the rest of the protein (Baron et al., 1991; Perham, 1991; and references therein). Even with phosphoglycerate kinase (Minard et al., 1989), the two domains in this monomeric enzyme have relatively little contact, but when isolated they exhibit small differences from the native structure owing to the loss of interdomain interactions and they will not complement. Similarly, the pyridoxal phosphate-binding domain of aspartate aminotransferase is much the larger of the two domains in the enzyme and provides most of the stability of the whole structure (Herold et al., 1991). In contrast, the interface domain of GR is a relatively minor part of the protein and, more importantly, is largely buried by contacts with the three other domains in the protomer. Viewed in this light, the ability to create an isolated interface domain still capable of folding and of dimerizing is an impressive demonstration of the autonomous stability of protein domains in general. It is experimental proof of the possibility of protein evolution by accretion of individual domains, possibly by a process of gene fusion and exon shuffling (Dorit et al., 1990; Doolittle, 1991). Interactions between domains would then follow with the evolutionary acquisition of additional hydrophobic surface, a reversal of the process of directed mutagenesis that we employed to create the soluble and tractable INT^{FN} domain from the starting INT form.

It appears that the conformational stability of the GR interface domain is derived predominantly from internal contacts, rather than packing with other parts of the protein. Our results suggest that, by paying special attention to exposed hydrophobic surfaces, it should be possible to tailor other domains as soluble forms, smaller than their parent proteins and more amenable to some experimental purposes, such as structural analysis by NMR spectroscopy. In particular, the excised and modified INT^{FN} domain of GR is a much simpler system to study subunit interactions and the dimerization process in detail.

ACKNOWLEDGMENT

We thank Dr. U. Ermler and Prof. G. E. Schulz for communicating the unpublished coordinates of the refined structure of *E. coli* glutathione reductase, S. M. Brocklehurst for help with the program NAOMI, Dr. Y. N. Kalia for creating the MOLSCRIPT diagrams, and Mr. C. Fuller for skilled technical assistance. We are grateful to Dr. M. Greiner and Dr. M. Herold, Hewlett-Packard Analytical Division, Waldbronn, Germany, for electrospray mass spectrometry and to Prof. K. Kirschner and Dr. R. Sterner, Biozentrum, Universität Basel, Basel, Switzerland, for use of a spectrophotometer. Other facilities were provided by the Cambridge Centre for Molecular Recognition.

REFERENCES

Baron, M., Norman, D. G., & Campbell, I. D. (1991) *Trends Biochem. Sci. 16*, 13–17.

Bradford, M. (1976) *Anal. Biochem. 72*, 248–254.

Bränden, C. I., & Tooze, J. E. (1991) *Introduction to Protein Structure*, Garland Publishing, New York.

Brocklehurst, S. M., & Perham, R. N. (1993) *Protein Sci. 2*, 626–639.

Chothia, C., & Janin, J. (1975) *Nature (London) 256*, 705–708.

Creighton, T. E. (1993) *Proteins. Structures and Molecular Properties*, 2nd ed., Freeman, New York.

Deonarain, M. P., Scrutton, N. S., Berry, A., & Perham, R. N. (1990) *Proc. R. Soc. London, B 241*, 179–186.

Doolittle, R. F. (1991) *Science 253*, 677–679.

Dorit, R. L., Schoenbach, L., & Gilbert, W. (1990) *Science 250*, 1377–1382.

Ermler, U., & Schulz, G. E. (1991) *Proteins: Struct., Funct., Genet. 9*, 174–179.

Herold, M., & Leistler, B. (1991) *J. Chromatogr. 539*, 383–391.

Herold, M., Leistler, B., Hage, A., Luger, K., & Kirschner, K. (1991) *Biochemistry 30*, 3612–3620.

Higuchi, R. (1989) in *PCR Technology. Principles and Applications for DNA Amplifications* (Erlich, H. A., Ed.) pp 61–70, M. Stockton Press, New York.

Jaenicke, R. (1991a) *Biochemistry 30*, 3147–3160.

Jaenicke, R. (1991b) *Eur. J. Biochem. 202*, 715–728.

Janin, J., & Wodak, S. J. (1983) *Prog. Biophys. Mol. Biol. 42*, 21–78.
Janin, J., Miller, S., & Chothia, C. (1988) *J. Mol. Biol. 204*, 155–164.
Karplus, P. A., & Schulz, G. E. (1987) *J. Mol. Biol. 195*, 701–729.
Kraulis, P. J. (1991) *J. Appl. Crystallogr. 24*, 946–950.
Laemmli, E. K. (1975) *Nature (London) 227*, 680–685.
Levine, R. L., & Federici, M. M. (1982) *Biochemistry 21*, 2600–2606.
Miller, S. (1989) *Protein Eng. 3*, 77–83.
Minard, P., Hall, L., Betton, J.-M., Missiakas, D., & Yon, J. M. (1989) *Protein Eng. 3*, 55–60.
Perham, R. N. (1991) *Biochemistry 30*, 8501–8512.
Sambrook, J., Fritsch, E. F., & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Sanger, F., Nicklen, S., & Coulson, A. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 5463–5467.
Scrutton, N. S., Berry, A., & Perham, R. N. (1987) *Biochem. J. 245*, 875–880.
Scrutton, N. S., Berry, A., & Perham, R. N. (1988) *FEBS Lett. 241*, 46–50.
Scrutton, N. S., Deonarain, M. P., Berry, A., & Perham, R. N. (1992) *Science 258*, 1140–1143.
Studier, F. W., Rosenberg, A. H., Dunn, J. J., & Dubendorff, J. W. (1990) *Methods Enzymol. 185*, 60–89.
Tasayco, M. L., & Carey, J. (1992) *Science 255*, 594–597.