Real-Time Measurement of Multiple Intramolecular Distances during Protein Folding Reactions: A Multisite Stopped-Flow Fluorescence Energy-Transfer Study of Yeast Phosphoglycerate Kinase†

M. Pilar Lillo

Departamento de Biofísica, Instituto de Química-Física, CSIC, Serrrano 119, E-28006 Madrid, Spain

Barbara K. Szpikowska and Maria T. Mas*

Division of Biology, Physical Biochemistry Section, Beckman Research Institute of the City of Hope, Duarte, California 91010

Jason D. Sutin and Joseph M. Beechem*

Department of Molecular Physiology and Biophysics, Vanderbilt University, Nashville, Tennessee 37232

Received April 7, 1997; Revised Manuscript Received June 23, 1997⊗

ABSTRACT: Understanding the set of rules which dictate how the primary amino acid sequence determines tertiary structure is an unsolved problem in biophysics. If it were possible to simultaneously measure all of the intramolecular distances in a protein (in real time) during a folding reaction, the "second" genetic code problem would be solved. Regrettably, no such technique currently exists. As a first step toward this goal, an optical distance assay system has been developed for a two-domain protein, yeast phosphoglycerate kinase (PGK), using Förster resonance energy transfer [Lillo, M. P., et al. (1997) *Biochemistry 36*, 11261−11272]. In this study, real-time stopped-flow distance changes are measured using six unique pairs of donor/acceptor fluorescent labels strategically placed throughout the tertiary structure of PGK. These multiple donor/acceptor sites were genetically engineered into PGK by cysteine substitution mutagenesis followed by extrinsic labeling with fluorescent probes, 5-[[[(2-iodoacetyl)amino]-ethyl]amino]naphthalenesulfonic acid (as a donor) and 5-iodoacetamidofluorescein (acceptor). The unfolding of PGK is found to be a sequential multistep process (native → I_1 → I_2 → unfolded) with rate constants of 0.30, 0.16, and 0.052 s^{-1}, respectively (from native to unfolded). Unique to this unfolding study, six intramolecular distance vectors have been resolved for both the I_1 and I_2 states. With this distance information, it is shown that the transition from the native to I_1 state can be modeled as a large hinge-bending motion, in which both domains "swing away" from each other by about 15 Å. As the domains move apart, the carboxyl-terminal domain rotates almost 90° about the hinge region connecting the two domains. It is also shown that the amino-terminal domain remains intact during the native → I_1 transition, consistent with our previous site-specific tryptophan fluorescence anisotropy stopped-flow study [Beechem, J. M., et al. (1995) *Biochemistry 34*, 13943−13948]. Future experiments are proposed which will attempt to resolve in detail the unfolding/refolding transitions in this protein with a resolution of approximately 5−10 Å.

Real-time measurement of all of the intramolecular distances during a protein folding reaction would allow one to directly "map" the mechanism by which the primary sequence is translated into tertiary structure. Regrettably, no such measurement technology currently exists, although stopped-flow quench NMR studies have made important contributions in this area (Roder et al., 1988; Udgaonkar & Baldwin, 1988; Dyson & Wright, 1996). In this study, a more realistic goal of the real-time measurement of six intramolecular distances was established. Using six fluorescence donor/acceptor pairs genetically engineered in yeast phosphoglycerate kinase (Lillo et al., 1997), stopped-flow fluorescence energy-transfer measurements were performed during the time course of an unfolding reaction. In this manner, changes in distance along six distinct intramolecular vectors within the protein three-dimensional structure were measured in real time as the bidomain enzyme PGK[1] unfolds. The three-dimensional locations of the six donor/acceptor sites are shown in Figure 1 in the preceding paper (Lillo et al., 1997). These six locations were chosen in order to highlight several aspects of the mechanism associated with

† This work was supported by NIH Grants GM41360 (M.T.M.) and GM45990 (J.M.B.). M.P.L. acknowledges support from DGICYT (PB93-126) and CSIC (Spain).

⊗ Abstract published in *Advance ACS Abstracts,* August 15, 1997.

[1] Abbreviations: D-PGK, donor-labeled protein; D-PGK-A, donor- and acceptor-labeled protein; EDTA, ethylenediaminetetraacetic acid; FRET, Förster resonance energy transfer; IAEDANS, 5-[[[2-iodoacetyl)-amino]ethyl]amino]naphthalenesulfonic acid ; AEDANS, reacted IAEDANS covalently attached to protein; IAF, 5-iodoacetamidofluorescein; AF, reacted IAF covalently attached to protein; GuHCl, guanidine hydrochloride; MOPS, 3-(*N*-morpholino)propanesulfonic acid; DTT, DL-dithiothreitol; PGK, phosphoglycerate kinase from *Saccharomyces cerevisiae*; Trp, tryptophan; N, I_1, I_2, and U, native, first intermediate, second intermediate, and unfolded state of PGK, respectively.

S0006-2960(97)00789-7 CCC: $14.00

the folding properties of PGK as a bidomain enzyme. Residues 1−184 form the amino-terminal domain (N-domain) which is connected to the carboxyl-terminal domain (C-domain, residues 200−401) through a flexible hinge region (residues 185−199). The C-terminal peptide (residues 401−415) folds back onto the amino-terminal domain. Position 412 is situated close to the C-terminal end, but in the three-dimensional sense belongs to the N-domain, and was chosen as a common reference point to follow (a) unfolding of the N-domain by monitoring changes in two intradomain distances, 135 ↔ 412 and 412 → 75, and (b) changes in the position of the C terminus relative to two different points in the C-domain, one in close proximity and one further apart in the folded structure, 202 ↔ 412 and 290 ↔ 412, respectively. Two additional interdomain distances, independent of the C-terminal end, were also measured to monitor domain−domain uncoupling (135 ↔ 290 and 75 ↔ 290).

In our previous study (Lillo et al., 1997), we were able to show that the donor fluorescence intensities were dominated by Förster energy-transfer mechanisms for all six donor/acceptor pairs. In this study, methodologies are developed which will reveal that changes in all six of these intramolecular distances can be monitored in real time during GuHCl-induced unfolding reactions. A unique experimental strategy was developed, which utilizes intrinsic tryptophan fluorescence signals as an "internal control" to experimentally determine the effects of the extrinsic donor and/or acceptor label on the overall protein unfolding kinetics. It was found that the unfolding kinetics, as observed by the intrinsic tryptophan fluorescence, for both the singly labeled and doubly labeled proteins were identical, allowing a direct experimental determination of the changes in fluorescence energy-transfer efficiencies during the time course of the unfolding reaction. The observed changes in energy-transfer efficiencies allow the reconstruction of a minimal "structural model" which describes the unfolding reaction along these six unique intramolecular distance vectors. Global analysis algorithms were utilized to simultaneously analyze all six intramolecular energy-transfer efficiency vectors in terms of a single internally consistent population of unfolding intermediates. It will be shown that the very first structural transition which occurs in yeast PGK during unfolding is a large hinge-bending and twist motion, which allows the amino-terminal domain and carboxyl-terminal domain to move away from each other by approximately 15 Å. To our knowledge, this is the first real-time long-range structural distance measurement, with enough detail to be interpreted in terms of a three-dimensional transition away from the native state.

MATERIALS AND METHODS

For protein purification and labeling procedures, consult the preceding paper (Lillo et al., 1997).

Stopped-Flow System and Fluorescence Detection

Unfolding reactions were initiated by a SFM-3 stopped-flow unit (Bio-Logic/ Molecular Kinetics, Pullman, WA) with a 50 μL FC.15 fluorescence cuvette and three independently driven stepper-motor syringes. Fluorescence

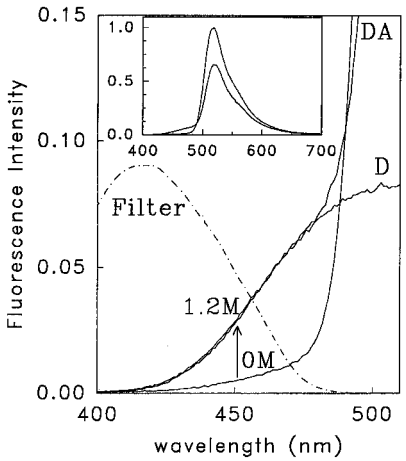


FIGURE 1: Expanded view of the emission wavelength region utilized for the quantitation of the donor-side stopped-flow energy-transfer kinetics. Corrected emission spectra ($\lambda_{ex} = 340$ nm) of the donor-only (D) and donor/acceptor (DA) PGK (202 ↔ 412 sample) under native (0 M GuHCl) and unfolded (1.23 M GuHCl) conditions. MOPS buffer, pH 7.5, $T = 25$ °C. Filter (−•−) is the transmittance spectra (45% transmitance at $\lambda_{max} = 418$ nm) of the donor-side filter combination (see Materials and Methods). (Inset) Full emission wavelength scan of the donor/acceptor sample at 0.0 M (lower curve at 480 nm, upper at 520 nm) and 1.23 M GuHCl.

detection utilized a home-built T-format steady-state stopped-flow fluorimeter assembled using the following components. The excitation beam was provided by a 450 W xenon arc lamp coupled to a 0.22 Spex monochromator (Edison, NJ). The monochromator output was focused onto a fused-silica fiber optic (10 × 100 μm fiber bundle, circle to slit configuration) and directed into the 50 μL stopped-flow observation cuvette (excitation and emission path lengths of 2 mm). Each emission channel consisted of a collimating lens, filter holder, and single-photon counting PMT assembly. The filters utilized in this study were as follows: 340 nm interference filter for tryptophan emission (P/N 58630 Oriel Co., Stratford, CT), the combination of a 400 nm cut-on type (L-40, Hoya Optics, Fremont, CA) and a band-pass filter (58814 Oriel Co.) for IAEDANS emission (donor side) (see Figure 1 for filter shape profile), and a 520 nm interference filter (54351 Oriel Co., 58% transmitance at 519 nm, acceptor side) for AF emission. Signals from each PMT (Hamamatsu R928) simultaneously recorded the fluorescence from both the donor- and acceptor-side emission. The PMT outputs were amplified by a SR455 DC-300 MHz amplifier (Stanford Research) and detected with a two-channel photon-counting unit (Stanford Research SR-400). The output of the discriminators from the SR-400 were sent to two MCS-II multichannel scaler cards (Tennelec Nucleus, Oak Ridge, TN) interfaced to two microcomputers (Intel 486). The MCS-II cards were synchronized with the SFM-3 stopped-flow unit through an external synch-output pulse from a Molecular Kinetic stepper motor controling unit. Data were collected using dwell times from 10 to 25 ms in 8000 total channels.

Stopped-flow unfolding transitions were initiated by mixing equal volumes (100 μL each) of a solution of labeled PGK (donor-only or donor/acceptor sample) and the denaturing buffer (2.46 M GuHCl in 50 mM MOPS, 100 mM NaCl, 1 mM EDTA, and 1 mM DTT at pH 7.5). All reactions were performed at 25 ± 0.2 °C, and reaction mixtures

contained a final concentration of 1 μM PGK in 1.23 M GuHCl/MOPS buffer. Flow rates for unfolding were between 2 and 6 mL/s (2 mL/s most of the experiments). Background measurements were also collected at the beginning of each run. Solutions of labeled PGK in MOPS buffer, diluting buffer (MOPS buffer), and denaturing buffer (2.46 M GuHCl in MOPS buffer) were loaded into separate, independently controlled syringes.

A typical stopped-flow energy-transfer run would collect (within 8000 channels) the entire reaction sequence below: (1) buffer background, 400 μL of MOPS buffer (1 mL/s); (2) folded protein control, 100 μL of MOPS buffer + 100 μL of PGK (2 mL/s, 2 s total); (3) GuHCl background, 100 μL of MOPS buffer + 100 μL of 2.46 M GuHCl (2 mL/s, 2 s); and (4) protein unfolding reaction, GuHCl jump from 0 to 1.23 M, 100 μL of PGK + 100 μL of 2.46 M GuHCl (2 mL/s).

For all experiments, solutions were extensively degassed.

Stopped-Flow Reaction Protocol for Energy-Transfer Measurements

Stopped-flow energy-transfer experiments require an extended reaction sequence protocol, wherein experiments are sequentially performed on the donor-only protein (D-PGK) followed by experiments on the donor- and acceptor-labeled proteins (D-PGK-A). The donor-only sample (D-PGK) is examined first. Multiple stopped-flow unfolding runs (10−20) are collected for the tryptophan fluorescence ($\lambda_{ex} = 295$ nm). These stopped-flow intensity changes serve as an "internal control" for the energy-transfer kinetics (see Results and Discussion). The excitation wavelength is then shifted to 340 nm (donor absorbance band), and the fluorescence emission is simultaneously collected at the donor side and the acceptor side. Multiple kinetics runs are summed (30−40), in order to obtain adequate signal-to-noise ratios. The donor/acceptor sample (D-PGK-A) is then examined at the same concentration as the donor alone experiments. Again, multiple runs for tryptophan emission ($\lambda_{ex} = 295$ nm) and donor emission (now the donor is in presence of the acceptor, $\lambda_{ex} = 340$ nm) and additional experiments with excitation directly into the acceptor at 430 nm (no donor absorption, no energy transfer) were carried out. Filter combinations can be assembled which allow absolutely pure collection of the tryptophan emission and the donor AEDANS fluorescence (see Figure 1). Due to the severe emission spectral overlap of AEDANS and AF, there is no method for collecting absolutely pure acceptor emission when exciting into the donor absorption band. Pure acceptor emission can, however, be obtained by excitation above the donor absorbance band.

Total-intensity data from D-PGK and D-PGK-A were background subtracted. The average FRET efficiency, $\langle E(t)\rangle$, was determined from the ratio of the two total intensity signals from the D-PGK-A and D-PGK kinetic curves (same number of runs, same protein concentration) using

$$\langle E(t)\rangle = 1 - \frac{I_{DA}(t)}{I_{D}(t)} \qquad (1)$$

where $I_{DA}(t)$ and $I_{D}(t)$ represent the measured donor/acceptor and donor-only time-dependent stopped-flow intensities, respectively.

Data Analysis

(a) Simultaneous Analysis of the Stopped-Flow Energy-Transfer Efficiencies in Terms of Internally Consistent Populations and Unique Distances. Global analysis was performed on all 12 D-PGK and D-PGK-A stopped-flow fluorescence time curves using the Globals Unlimited (Urbana, IL) (Beechem et al., 1991) general purpose nonlinear analysis program. Additional subroutines were developed (described in Results and Discussion) and added to the global analysis, which allowed for the simultaneous fitting of all of the stopped-flow energy-transfer data (donor only, donor/acceptor, and all six energy-transfer pairs) in terms of internally consistent sets of kinetic populations and distances. The final result of this analysis consists of the resolution of an internally consistent population of PGK intermediates during unfolding (constrained to be identical for all of the mutant proteins) and a distance/transfer efficiency unique for each donor/acceptor pair.

(b) Global Analysis of the First Folding Intermediate in PGK in Terms of a Simple Structural Transition of the Native State. The recovered six intramolecular distances associated with the first unfolding intermediate in PGK (obtained in part a above) were further examined in terms of a simple structural transition of the native state. Analysis was performed by modeling PGK as two domains connected by a flexible hinge. All six intramolecular distances obtained for the first intermediate unfolding state were then modeled as a simple hard-body motion of the two domains using (as fitting parameters) the Euler angles (α, γ, and ϕ). The only additional unknown fitting parameter in this system is the origin (center of locus) of the hard-body rotation. A grid pattern (with lattice points spaced every 1 Å) was scanned throughout the entire central helix region connecting the amino-terminal and carboxyl-terminal domains. At each grid location, a full nonlinear minimization was performed on each Euler angle, minimizing the difference betweeen the six measured intramolecular distances and that predicted by a simple rotation centered at that locus.

RESULTS AND DISCUSSION

The stopped-flow total-intensity changes were simultaneously measured on both the donor and acceptor sides of the fluorescence emission. Measurements of the donor emission were performed using a combination of filters (see Materials and Methods), which allowed examination of the AEDANS intensity in donor/acceptor samples with no contribution of the AF (acceptor) emission (Figure 1). In this study, the average fluorescence resonance energy-transfer (FRET) efficiency was obtained as a function of time from signals obtained from the donor-side-measured stopped-flow intensity curves (using eq 1). The simultaneous measurements obtained from the acceptor-side fluorescence emission were completely consistent with the donor-side measurements. However, the signal-to-noise ratio which can be obtained from the acceptor side data was vastly inferior with respect to that obtained from the donor-side emission. Hence, the acceptor-side stopped-flow experiments were utilized only as an additional internal consistency check concerning the overall magnitude of the transfer efficiencies observed in the measurements of the donor-side emission. It was found that the magnitudes of the change in energy-transfer efficiencies as observed by the acceptor emission

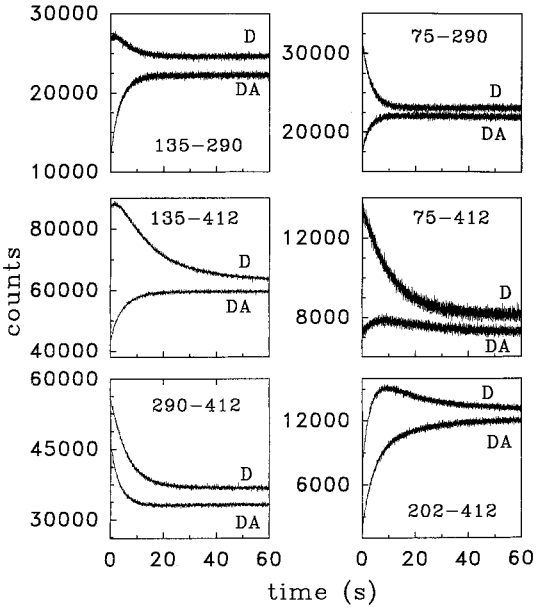


FIGURE 2: Fluorescence stopped-flow unfolding data obtained (using the donor-side detection) for all six energy-transfer pairs. At time zero, an unfolding transition was initiated in PGK by a GuHCl concentration jump from 0 to 1.23 M. $T = 25$ °C, and $\lambda_{exc} = 340$ nm, with a 4 nm excitation band width. Results from the global simultaneous analysis of all of the donor-only (D) and donor/acceptor (DA) experiments can be found in Table 1.

signals were all within ±5% of that determined from the donor-side emission. This result also supports [together with the time-resolved results (Lillo et al., 1997)] the fact that the observed changes in donor intensities are dominated by true energy-transfer terms and not some other photophysical event.

Figure 2 represents the background-subtracted, fluorescence stopped-flow signals obtained from all six donor-only (D-PGK) and donor/acceptor (D-PGK-A) pairs during an unfolding reaction. These curves correspond to the sum of multiple unfolding kinetic runs (30−40 runs). The complete experimental traces (data regions not shown in Figure 2) contained signals for native protein in MOPS buffer, as well as backgrounds for both the MOPS buffer and 1.23 M GuHCl. All of the subsequent global analyses described below simultaneously fit these 12 data sets (e.g., Figure 4).

Figure 3 shows a representative internal control tryptophan stopped-flow intensity change for a D-PGK-Cys and D-PGK-A sample. The internal tryptophan signals represent an independent measurement of the unfolding kinetics of the extrinsically labeled proteins. If the presence of the extrinsic labels was altering the overall unfolding kinetics of PGK, one would have expected that the donor-only and donor/acceptor proteins would have altered tryptophan fluorescence unfolding kinetics (the lack of energy transfer between the intrinsic PGK tryptophans and IAEDANS was ruled out in a previous study; Lillo et al., 1997). At each specific location, the donor- and donor- and acceptor-labeled tryptophan fluorescence stopped-flow unfolding signals overlap (to within a few percent) (Figure 3). This surprising result indicates that the unfolding kinetics of the D-PGK and D-PGK-A are essentially identical. This result provides an experimental justification of the use of eq 1 (which explicitly utilizes the ratio of signals between the singly and doubly labeled proteins) to determine the fluorescence energy-transfer efficiencies during the unfolding reaction. If the

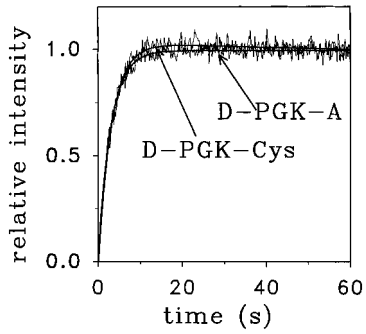


FIGURE 3: Absolute overlap of the stopped-flow unfolding kinetics for both singly labeled (D-PGK-Cys) and doubly labeled (D-PGK-A) PGK using intrinsic tryptophan fluorescence as an internal control. Tryptophan stopped-flow unfolding data were obtained in parallel with and on the same samples as the energy-transfer unfolding data in Figure 2 (see Materials and Methods). At time zero, an unfolding transition was initiated by a GuHCl concentration jump from 0 to 1.23 M. $T = 25$ °C, and $\lambda_{exc} = 295$ nm, with a 4 nm excitation band width; emission was collected through a 340 nm interference filter. Representative data on the 135 ↔ 290 mutant are shown; data on all other mutants revealed a similar overlap.

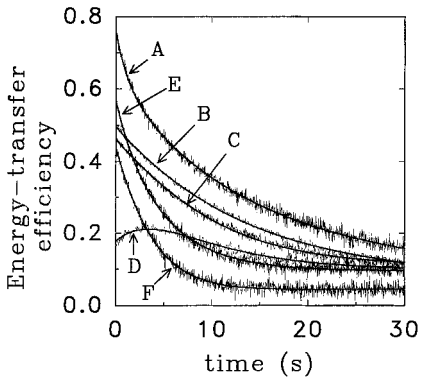


FIGURE 4: Observed and calculated fluorescence energy-transfer efficiency kinetics for all six energy-transfer pairs (A, 202 ↔ 412; B, 135 ↔ 412; C, 412 → 75; D, 290 ↔ 412; E, 135 ↔ 290; F, 75 ↔ 290). The solid lines represent the best global analysis fit to the $N \rightarrow I_1 \rightarrow I_2 \rightarrow U$ model. The recovered rate constants are as follows: $k(N \rightarrow I_1) = 0.30$ (0.276−0.328) s^{-1}, $k(I_1 \rightarrow I_2) = 0.16$ (0.10−0.174) s^{-1}, and $k(I_2 \rightarrow U) = 0.052$ (0.045−0.061) s^{-1}. The recovered energy-transfer efficiencies for each state from this analysis can be found in Table 1.

D-PGK and D-PGK-A tryptophan fluorescence samples did not superimpose, then one could not utilize eq 1 to calculate the energy-transfer efficiencies. The time scale associated with the tryptophan unfolding reaction (between all of the different mutant proteins) is also essentially identical. Only the 202 ↔ 412 mutant appears to have slightly slower unfolding kinetics.

Given the measured changes in stopped-flow signal intensities for the donor and donor/acceptor pairs (Figure 2) and the fact that the unfolding kinetics as observed by the intrinsic tryptophan kinetics were identical within each individual donor or donor/acceptor protein, one can utilize eq 1 to generate an experimental measure of the energy-transfer efficiency as a function of time for each site (Figure 4). These measured energy-transfer efficiencies are essentially "model-independent", in that they are simply experimental determinations of the ratio of fluorescence for the donor-only and donor- and acceptor-labeled species. In all of the sites examined, the observed transfer efficiencies decrease except for the transient rise in transfer efficiency which occurs at the 290−412 position early during the

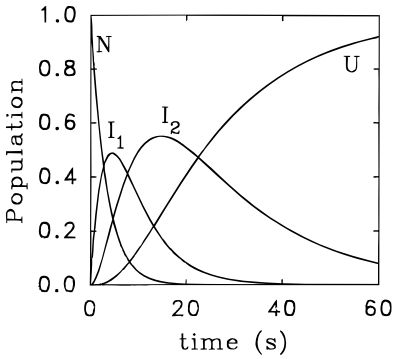


FIGURE 5: Recovered internally consistent change in PGK populations upon GuHCl-induced unfolding (0.0 → 1.23 M GuHCl) using the kinetic model native (N) → I_1 → I_2 → unfolded (U). This population change was linked for all six energy-transfer pairs which were analyzed. The six intramolecular distances associated with each intermediate population can be found in Figures 6 and 7.

unfolding reaction. Contrary to the intrinsic tryptophan fluorescence kinetics (Figure 3), the observed energy-transfer efficiency unfolding kinetics associated with each distinct donor/acceptor pair are unique (Figure 4). This is to be expected, as each donor/acceptor pair is reporting on distance changes occurring over distinct intramolecular vectors within the three-dimensional structure of the protein.

Before going into a detailed numerical analysis of the change in energy-transfer efficiencies, quite a large amount of information can be obtained by simple visual inspection of the data. Note that the fastest and largest changes in transfer efficiency are found in curves E and F (Figure 4). Curves E and F correspond to changes in distance along the intramolecular vector connecting residues 75 ↔ 290 and 135 ↔ 290. Both of these distances correspond to "looking" across from the amino-terminal domain to the carboxyl-terminal domain over the interdomain "gap" (use Figure 1 in the preceding paper as a reference). The change in distance along these vectors therefore represents the speed in which the two domains "swing away" from each other during the unfolding reaction. Curve D is rather unusual, in that the measured transfer efficiency actually increases initially, followed by a slow decrease. As will become apparent in the more detailed analysis (given below), this transfer efficiency increase corresponds to the two domains swinging away from one another. As the domains separate, they actually "twist" around a hinge that brings the net distance between 290 and 412 closer (fast phase), before each individual domain begins to "melt" (slow phase). We have double-checked curve D now on three completely different protein preparations, all revealing identical results. Examination of the acceptor-side stopped-flow profiles of the 290 ↔ 412 experiment (curve D), reveals an identical change in transfer efficiency (i.e., the transient decrease in donor fluorescence is exactly paralleled by a transient increase in acceptor fluorescence), indicating that the observed rise and/or fall is due to a true energy-transfer transient and not due to some other complex quenching/enhancement process. This "visual-inspection" analysis will now be performed with a firm mathematical basis, in order to further quantitate the observed unfolding transitions. Before the detailed global numerical analyses of the energy-transfer efficiencies are described (smooth fit to the data shown in Figure 4), it is important to examine the overall fitting strategy which has been utilized for these data.

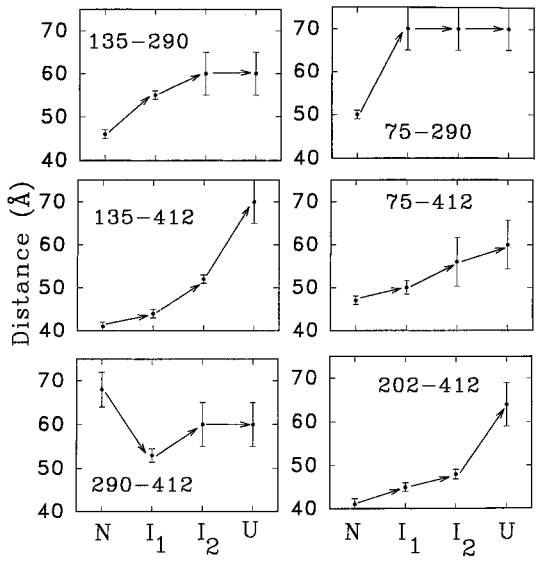
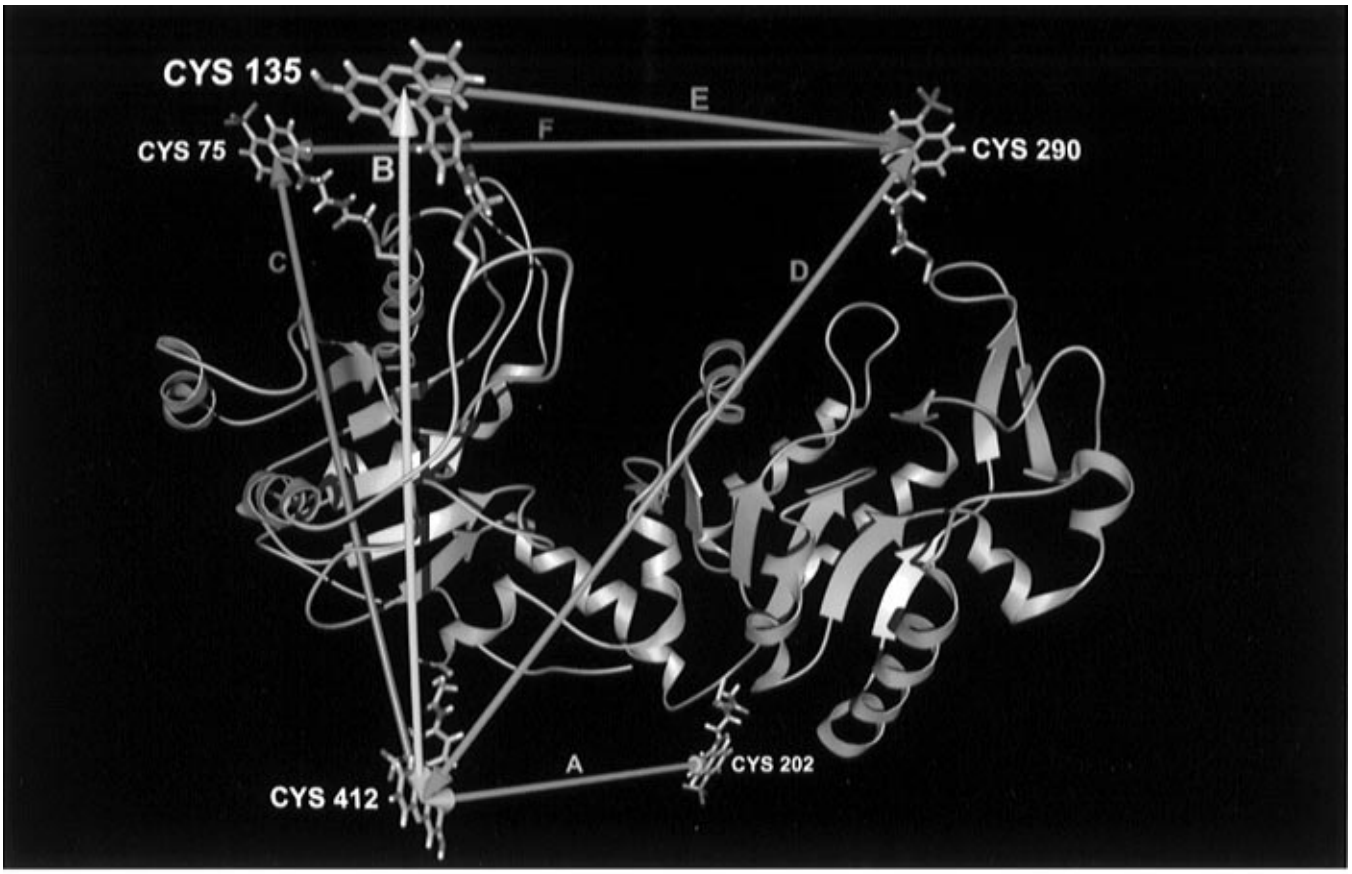
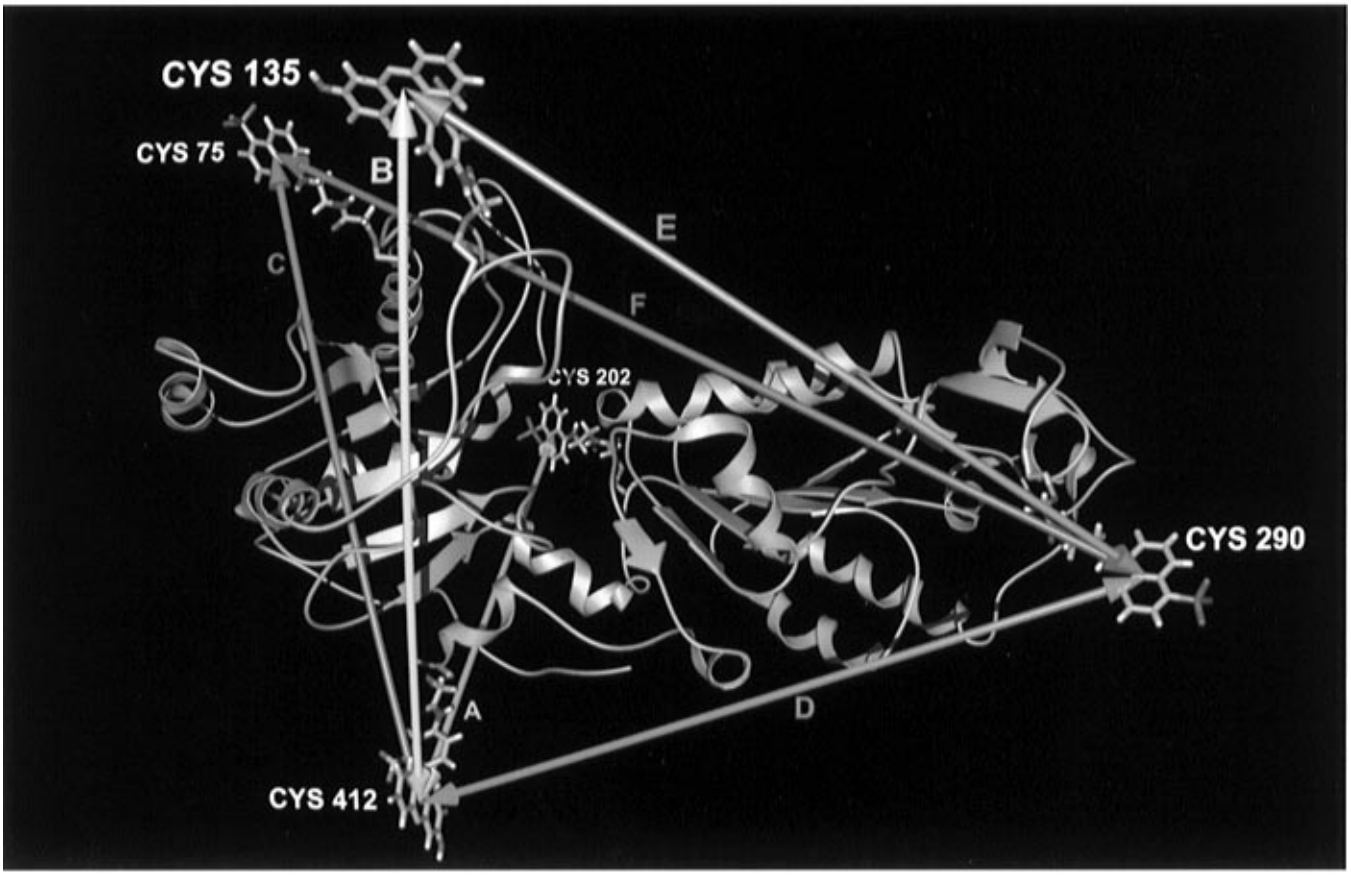


FIGURE 6: Recovered intramolecular distances associated with each intermediate formed during the GuHCl-induced unfolding of PGK. The kinetic unfolding model utilized a sequential transition from native (N) → I_1 → I_2 → unfolded (U).

The Unfolding Pathways for All of the Different Donor- and Acceptor-Labeled Proteins Are Identical (Principle 1). This assumption states that as the protein unfolds, the same sequence of protein transitions occurs for all of the different donor- and acceptor-labeled species. The experimental justification for this assumption is data from the internal control intrinsic tryptophan fluorescence kinetics (Figure 3). Nonlinear analysis of the tryptophan fluorescence kinetics revealed a biexponential decay for unfolding for all of the donor/acceptor pairs, with recovery of nearly identical rate constants for each donor- and acceptor-labeled sample. Simple graphical overlap of all of the donor/acceptor pairs reveals complete superposition of the kinetics for all of the donor/acceptor pairs except for the 202 ↔ 412 case. Global analysis of the intrinsic tryptophan unfolding data revealed that linking identical rate constants for all of the donor- and acceptor-labeled species resulted in very good global fits of the data (fitting variances nearly identical to the unlinked double-exponential analysis), with recovered rates of 0.34 (±0.05) and 0.09 s^{-1} (±0.01). Therefore, these results appear to state that the presence of the donor/acceptor extrinsic labels at the various locations does not adversely affect the unfolding kinetics of the overall protein. It should be emphasized that not all of the extrinsically labeled proteins which were initially designed for PGK revealed such ideal kinetic behavior. However, in this study, only labeled proteins which were kinetically identical (or nearly identical) were utilized.

Given principle 1 (above), it should be possible to perform a global analysis of all of the energy-transfer efficiencies in terms of an internally consistent set of unfolding populations. A global analysis was then performed with the following constraints. The protein population distributions (determined by fitting an internally consistent set of unfolding rate constants) were constrained to be identical for all donor/acceptor pairs during the analysis. Although the populations were constrained to be identical, the energy-transfer efficiency associated with each donor/acceptor pair will be unique to each pair. This is to be expected since each donor/acceptor pair is measuring a different intramolecular vector


CYS 135
CYS 75
CYS 290
CYS 412
CYS 202
A
B
C
D
E
F


CYS 135
CYS 75
CYS 290
CYS 412
CYS 202
A
B
C
D
E
F

distance. Global simultaneous analysis of all six energy-transfer pairs was then performed with these imposed constraints. It was found that all of the energy-transfer efficiencies could be simultaneously analyzed using the sequential unfolding model native (N) → intermediate 1 (I_1) → intermediate 2 (I_2) → unfolded (U). All simple three-state models (with only a single intermediate) resulted in fitting variances which were so high that they could be rejected at a >99.9% confidence level. A parallel unfolding model (native 1 → unfolded; native 2 → unfolded) could also adequately describe the stopped-flow data but was completely inconsistent with the time-resolved characterization of the native state found in the preceding paper (Lillo et al., 1997). This parallel model solution required a large population of a second native state with an energy-transfer efficiency very different from that of the first native state. From time-resolved fluorescence energy-transfer measurements of each of the six positions, however, it was apparent that no such large population with a differing energy-transfer efficiency could have existed in these samples. Therefore, the parallel state unfolding kinetic model of PGK will not be examined in any more detail at this time.

It should be emphasized that a very limited number of fitting terms were utilized to describe the stopped-flow data sets (12 different stopped-flow traces, ≈76000 data points). The unknowns in the fitting process consisted of three rate constants which were linked over all of the stopped-flow data sets. The only other fitting parameters in the global analysis were a unique energy-transfer efficiency term associated with the I_1 and I_2 states for each donor/acceptor pair. Complete resolution of all of the kinetic rate constants and energy-transfer efficiency terms of this model is revealed in rigorous confidence interval examination of the χ^2 hypersurface associated with each term (Table 1).

The recovered globally linked unfolding populations of PGK are shown in Figure 5. If one forces a model with only a single intermediate state in the analysis, the I_1 and I_2 states collapse to an almost perfect average of the two states which are recovered, with almost no change in the N and U time-dependent populations. The fitting variance associated with such fits, however, can be rejected at a >99% confidence level. Global analysis of incomplete donor/acceptor data sets (where only five instead of six pairs were examined) revealed nearly identical protein populations in time, indicating that this model is reasonably robust with respect to any single donor/acceptor pair (data not shown). It is clear, however, that the donor/acceptor pair 290 ↔ 412 contributes very strongly to the need for two intermediate unfolded states. It is impossible (within the geometry constraints of the system) to obtain a rising phase in the energy-transfer efficiency (and subsequent decrease in distance) if only a single unfolding intermediate is formed. It was for this reason that experiments on this donor/acceptor pair were repeated on three completely different donor/acceptor labeling preparations. It was found that different protein preparations yielded identical results.

If a "normal" stopped-flow fluorescence study was being performed, Figure 5 would represent the maximal amount of information which could be obtained from the data. However, in the experiments described in this work, the "optical-distance" aspect of the stopped-flow energy transfer data can now be utilized, in order to recover six independent intramolecular distances within the transiently formed unfolded states I_1 and I_2. These results are shown in Figure 6.

A very clear pattern emerges from these results. Note that, in the formation of the I_1 state, only two distances are dramatically increased (135 ↔ 290 and 75 ↔ 290). These are the two donor/acceptor pairs which are measuring the distance between the amino-terminal and carboxyl-terminal domains. This was essentially the same conclusion which was obtained by simple visual inspection of the data (Figure 4). However, with simple visual inspection of the data it would not be possible to decompose the smoothly varying transfer efficiencies (Figure 4) into an internally consistent kinetic transition between two intermediate folding states with discrete geometry associated with each state (result shown in Figure 6).

What is also clear from the data analysis is that the two probes which are directly measuring distances spanning the amino-terminal domain (135 ↔ 412 and 412 → 75) show essentially no change between the N and I_1 states. This result is completely consistent with our previous work on stopped-flow unfolding as observed by single tryptophans engineered into the amino-terminal and carboxyl-terminal domains (Szpikowska et al., 1994; Beechem et al., 1995). This previous work revealed that, during the first phase of the unfolding, the stopped-flow anisotropy of tryptophans located in the amino-terminal domain was not altered, indicating that the rotational environment at positions W48 and W122 was completely unaffected by this first transition. The energy-transfer findings in this study now reveal that the intramolecular distances within the amino-terminal domain when measured from end to end (135 ↔ 412) and end to back (412 → 75) are also invariant during the first unfolding stage of PGK. These amino-terminal domain distance markers unfold significantly only during the I_1 → I_2 and the I_2 → U stages, also consistent with the unfolding pattern observed in the single tryptophan stopped-flow anisotropy study (Beechem et al., 1995).

The intramolecular distance 290 ↔ 412 reveals a very unique transition, in that this distance actually decreases (transiently) during the early unfolding stage (Figure 6). Our interpretation of this transient decrease (a more detailed analysis is described below) was that, as the two domains swing away from each other (as observed for 135 ↔ 290 and 75 ↔ 290), the protein pivots and twists about the hinge region, thereby bringing the 290 ↔ 412 probes closer together. In order to test this type of complex rotation model, an additional data analysis methodology was developed and utilized.

While the recovered distance plots shown in Figure 6 are very useful representations of the data, they do not reveal (in three dimensions) the type of structural transitions which are occurring in PGK as it unfolds. To further model the first unfolding transition (N → I_1), it was decided to attempt to simultaneously satisfy all of the measured intramolecular

FIGURE 7: Three-dimensional reconstructed structural transition associated with the unfolding of the PGK native state (upper) to the first unfolding intermediate (lower). The labels A−D correspond to the measured kinetic changes in energy-transfer efficiencies shown in Figure 4. The interdomain gap between the carboxyl-terminal domain (to the right) and the amino-terminal domain (to the left) expands 10 and 20 Å along the intramolecular axes E and F, respectively. The carboxyl-terminal domain also twists approximately 90° about residues 194−197 during this first stage of unfolding. AF and AEDANS are shown (to scale) at positions 135, 412 and 75, 202, and 290, respectively.

Table 1: Recovered Energy-Transfer Efficiencies from a Global Simultaneous Analysis of the Stopped-Flow Kinetics from All Six Energy-Transfer Pairs[a]

energy-transfer pair	native	I_1	I_2	unfolded	χ^2 (donor)	χ^2 (acceptor)
135 ↔ 290	0.566 (0.557−0.575)	0.20 (0.18−0.21)	0.08 (0.05−0.09)	0.10 (0.06−0.12)	1.05	1.01
75 ↔ 290	0.44 (0.424−0.443)	0.05 (0.04−0.08)	0.036 (0.03−0.06)	0.05 (0.03−0.06)	0.99	1.0
135 ↔ 412	0.50 (0.49−0.51)	0.40 (0.39−0.41)	0.19 (0.185−0.20)	0.05 (0.03−0.06)	1.03	0.98
412 → 75	0.46 (0.445−0.48)	0.355 (0.33−0.38)	0.12 (0.08−0.14)	0.10 (0.06−0.12)	1.01	1.0
290 ↔ 412	0.178 (0.153−0.20)	0.274 (0.27−0.30)	0.10 (0.07−0.12)	0.10 (0.07−0.12)	1.0	1.01
202 ↔ 412	0.76 (0.744−0.777)	0.507 (0.49−0.527)	0.265 (0.254−0.28)	0.07 (0.05−0.09)	1.0	1.01

[a] Numbers in parentheses represent the upper and lower limits of the recovered values (at the 67% confidence level) using rigorous error analysis as described in Beechem et al. (1991).

distances between the native and I_1 states as a simple hard-body motion of the two domains away from each other, with a center of locus (of rotation) somewhere within the connecting region linking the amino-terminal and carboxyl-terminal domains. In this manner, all of the individual energy-transfer distances associated with the I_1 state would be forced to be consistent with a single unfolding transition. It was found that this simple model was capable of reproducing the measured data within a tolerance level which was well under the experimental noise in the data. A graphical view of the predicted three-dimensional structure of the N → I_1 unfolding intermediate in PGK is shown in Figure 7.

Apparent in this model is the fact that the major structural effect which occurs during the first phase of the unfolding reaction is a large increase in the intramolecular distance as observed across the gap between the amino-terminal and carboxyl-terminal domains. This result is clear not only in this model but also within the model-independent distances shown in Figure 6, where both the 135 ↔ 290 and 75 ↔ 290 distances rapidly increase during the first phase of the unfolding reaction. As the domains swing away from each other, the apparent distance between 290 and 412 decreases as the carboxyl-terminal domain twists (approximately 90°) about the pivoting hinge. Multiple nonlinear analysis was performed by scanning lattice points every 1 Å through the entire hinge region, and an origin of rotation and twisting (see Materials and Methods) was found to occur somewhere between residues 194 and 197. These same residues have been found by crystallographic studies to be the location of the localized hinge bend associated with the binding of 3-phosphoglycerate to PGK (Bernstein et al., 1997). The very "natural" way in which this model is capable of reproducing the distance decrease of 290 ↔ 412, the large distance increase between the two domains, and the invariant distances within the amino-terminal domain provides some level of confidence that this simple model is capturing the essential features of the initial N → I_1 unfolding transition.

It is important, however, to keep in mind that there are several problematic aspects associated with the transition from the uniquely recovered protein populations (Figure 5) and energy-transfer efficiencies (Table 1) to absolute distances. The transfer efficiencies and protein populations can be recovered in a very robust manner. However, the conversion of the measured energy-transfer efficiencies to intramolecular distances (Figure 6) requires knowledge of the R_o term for each intermediate state. In the preceding paper (Lillo et al., 1997), a detailed time-resolved fluorescence study was performed in order to determine the R_o for each donor/acceptor pair in the native (N) and unfolded (U) states. Since these folding intermediates are only transiently formed, it is not possible to perform such a detailed spectroscopic determination. Therefore, a very difficult question was what R_o value should be utilized for the I_1 and I_2 states during the unfolding reaction?

As the initial energy-transfer data were being analyzed, however, it became apparent that the energy-transfer efficiencies associated with the amino-terminal domain were essentially constant during the first stage of the unfolding process. This was absolutely consistent with our previous stopped-flow tryptophan anisotropy of site-specific probes within the amino terminus being rotationally invariant during this first transition (Beechem et al., 1995). Therefore, in the distance determination analysis, the depolarization factors (needed for the R_o calculation) associated with the I_1 state for probes located in the amino-terminal domain were given the same values as that obtained in the native state. We feel that the probability of two independent donor/acceptor probes experiencing actual intramolecular distance changes which are both negatively compensated by changes in depolarization factors to be rather unlikely. The only other term which changes slightly for the amino-terminal domain probes during this transition in the quantum yield of the donor, but this was found to generate a maximal change of only 2% in R_o.

Using similar reasoning, the depolarization factors associated with the carboxyl-terminal domain probes (in I_1) were given a value identical to that obtained for the unfolded state, consistent with previous results with site-directed tryptophan mutants placed in this domain. A problem with this assignment arises, however, in that one can isotropically average the carboxyl-terminal domain probes, yet still have ambiguity concerning how the amino-terminal domain probes and carboxyl-domain probes are oriented relative to each other [θ_A or θ_D; notation from Dale et al. (1979)]. In Figure 6 (and subsequent analyses), this angle was set (θ_A for acceptors within the carboxyl terminus or θ_D for donors) to be identical to that of the native state. Modeling the extreme ranges associated with this assumption ($d\theta = 0$ and $d\theta = \pm 90°$), it was found that the effect on the R_o estimation [eq 4 in Lillo et al. (1997)] is approximately ±4 Å. Therefore, in the assignment of the distances associated with the intermediate states, there is a maximal uncertainty of approximately 4 Å which arises due to the unknown change in θ_A or θ_D during the unfolding process. The best estimates of the changes in R_o (in angstroms) in going from the N to the I_1 state using the above described logic are as follows: 48 → 44 Å (135 ↔ 290), 48 → 43.4 Å (75 ↔ 290), 41 → 41 Å (135 ↔ 412), 46 → 45 Å (412 → 75), 53 → 45 Å (290 ↔ 412), and 50 → 45 Å (202 ↔ 412). The I_2 and U states

all have an R_o of 41 Å ($\kappa^2 = {}^2/_3$). Uncertainties associated with the R_o values of the intermediate unfolded states represent the largest source of error in the recovery of the intramolecular distances in the intermediate folding states. Future studies wherein these experiments are performed in a "double-kinetic" manner (Beechem, 1997) would greatly decrease these uncertainties. It should be emphasized that the R_o change in the 290 ↔ 412 case actually pushes the transfer efficiency in a direction opposite to that which is measured (Figure 4D) during the rising phase. Therefore, this change is not a potential source for the observed transient increase in efficiency.

CONCLUSION

This study represents the first quantitative attempt at measuring (in real time) the full long-range distance changes that occur in a protein during unfolding. By neccessity, the spatial resolution of this study is rather limited, as only an initial set of six intramolecular distances was investigated. At the current time, additional donor/acceptor sites are being engineered into this system, and this work will be described in a future study. These additional energy-transfer pairs are being designed into the ends of the major α-helices, in order to unambiguously determine the time scale of formation/unfolding of these small structural units and their interactions. Analogous pairs are being engineered into the β-sheet regions to answer similar questions. These additional mutants will emphasize small structural subdomains, and when combined with structural data over large domains (the inter- and intradomain distances measured in this study), we may approach a point in which low-resolution maps (5−10 Å resolution) of PGK's transient folding intermediates can be obtained.

There were a variety of potential complications associated with multisite donor/acceptor energy-transfer experiments which were experimentally answered in this study. Most important was the observation that the donor-alone PGK and donor- and acceptor-labeled PGK obeyed essentially identical unfolding kinetics. This result could not have been predicted *a priori* and makes the determination of the time-dependent changes in energy-transfer efficiency much easier (no kinetic correction terms were needed in eq 1). Second, changing R_o terms associated with folding intermediates can cause errors (for this donor/acceptor pair, a maximum uncertainty of ±4 Å) in the recovered distances. Clearly, detailed time-resolved photophysical studies of each donor/acceptor site, such as that described in the preceding paper (Lillo et al., 1997), should be performed on any protein system in which multisite stopped-flow energy-transfer studies are to be performed.

The described hard-body rotational analysis of all of the energy-transfer data is essentially a way of "morphing" the native structure (N) into a proposed three-dimensional structure associated with the I_1 state. Clearly, as one follows transitions between states farther removed from the native structure, such an approach will not be possible. Luckily, in the case of PGK, our understanding of the unfolding process in this enzyme has allowed us to apply the hard-body rotational model in a manner which may be pertinent to the actual structural transition which is occurring. The invariance of the amino-terminal domain during the early stages of unfolding was clearly present in our original single tryptophan data (Beechem et al., 1995) and also clearly reveals itself in the observed energy-transfer data. The absolute invariance of the measured energy-transfer distances within the amino-terminal domain during the first stage of the unfolding transition strengthens the assertion that the measured stopped-flow parameters are faithfully recording the presence and/or absence of distance changes, and not some other photophysical effect. The hard-body rotational analysis of the carboxyl-terminal domain swinging away and twisting from the amino-terminal domain very easily captured the rather unusual observation of the 290 ↔ 412 distance transiently decreasing before increasing. No highly unusual fitting parameters were required to reproduce this result; such a result would be predicted whenever a two-domain enzyme undergoes a pivot-like motion about a central locus. With energy-transfer pairs engineered into the ends of both domains, such a transient increase in transfer efficiency (or decrease in distance) is guaranteed to occur in at least one interdomain pair. Clearly, as one becomes farther removed from the native structure (e.g., I_2), such simple analytical approaches will no longer be feasible, and one will be left with a series of distances in three dimensions "tacking" multiple residues together in space. Refolding versions of all of the above described experiments are in progress.

REFERENCES

Beechem, J. M. (1997) *Methods Enzymol. 278*, 24−49.

Beechem, J. M., Gratton, E., Ameloot, M., Knutson, J. R., & Brand, L. (1991) in *Topics in Fluorescence Spectroscopy: Principles II* (Lakowicz, J. R., Ed.) pp 241−305, Plenum Press, New York.

Beechem, J. M., Sherman, M. A., & Mas, M. T. (1995) *Biochemistry 34*, 13943−13948.

Bernstein, B. E., Michels, P. A. M., & Hol, W. G. J. (1997) *Nature 385*, 275−278.

Dale, R. E., Eisinger, J., & Blumberg, W. E. (1979) *Biophys. J. 26*, 161−194.

Dyson, H. J., & Wright, P. E. (1996) *Annu. Rev. Phys. Chem. 47*, 369−395.

Lillo, M. P., Beechem, J. M., Szpikowska, B. K., Sherman, M. A., & Mas, M. T. (1997) *Biochemistry 36*, 11261−11272.

Roder, H., Elove, G. A., & Englander, S. W. (1988) *Nature 335*, 700−704.

Szpikowska, B. K., Beechem, J. M., Sherman, M. A., & Mas, M. T. (1994) *Biochemistry 33*, 2217−2225.

Udgaonkar, J. B., & Baldwin, R. L. (1988) *Nature 335*, 694−699.

BI970789Z