THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 277, No. 34, Issue of August 23, pp. 30968–30975, 2002
Printed in U.S.A.

Rapid Mapping of Protein Structure, Interactions, and Ligand Binding by Misincorporation Proton-Alkyl Exchange*

Received for publication, April 3, 2002, and in revised form, May 17, 2002
Published, JBC Papers in Press, May 25, 2002, DOI 10.1074/jbc.M203172200

Joshua A. Silverman‡ and Pehr B. Harbury§

From the Department of Biochemistry, Stanford University, Stanford, California 94305

Understanding protein conformation, interactions, and ligand binding is essential to all biological inquiry. We report a novel biochemical technique, called misincorporation proton-alkyl exchange (MPAX), that can be used to footprint protein structure at single amino acid resolution. MPAX exploits translational misincorporation of cysteine residues to generate probes for physical analysis. We apply MPAX to the triosephosphate isomerase $(\beta/\alpha)_8$ barrel, accurately determining its substrate-binding site, a protein-protein interaction surface, the solvent-accessible protein surface, and the stability of the barrel. Because MPAX requires only microgram quantities of material and is not limited by protein size, it is ideally suited for proteins not amenable to conventional structural methods, such as membrane proteins, partially folded or insoluble proteins, and large protein complexes.

To rapidly obtain functional information for a large number of sequences, a general and efficient tool for probing protein conformation is required. In principle, protein footprinting provides a means to study protein structure, but it has been far less successful than the corresponding techniques developed for nucleic acids (1). Because of the chemical heterogeneity of the amino acid side chains, no reagent exists with the ability to cleave uniformly the protein backbone under native conditions. Furthermore, protein separation techniques such as SDS-PAGE do not possess the single monomer resolution of the urea-acrylamide gels used for the separation of nucleic acids, which complicates the analysis of observed cleavage patterns. Finally, because of the cooperative nature of protein unfolding, proteolytic cleavage at one site often leads to a global loss of structure and to increased cleavage at other sites in the same molecule, resulting in artifactual data.

Footprinting by chemical modification of amino acid side chains represents a different approach to the problem. Modification of side chains is carried out under native conditions, whereas detection of modifications can be performed under arbitrary conditions. The susceptibility of each side chain to modification reports its solvent accessibility. Acylation of lysine residues (2, 3), oxidation of methionine residues (4), and alkylation of cysteine residues (5, 6) have been used previously to footprint protein structures. These studies have been limited in scope, however, because they examine only a few naturally occurring residues or require extensive site-directed mutagenesis to introduce additional structural probes. Here we describe a new experimental approach, termed misincorporation proton-alkyl exchange (MPAX),[1] that overcomes many of the limitations of traditional protein footprinting and facilitates high throughput analysis.

MPAX utilizes the sulfhydryl groups of cysteine residues as reactive sites that can be modified specifically under native conditions. Because cysteine residues occur naturally at a low frequency, auxiliary cysteines must be introduced into a protein of interest to act as structural probes. An expedient strategy for introducing cysteine residues is to subvert the protein translational machinery. Recently, mutant bacterial cysteine tRNAs were created by substituting the cysteine anticodon triplet with an isoleucine or methionine anti-codon triplet (7). These tRNAs were demonstrated to misincorporate cysteine at isoleucine and methionine codons complementary to the mutant triplets. Based on this finding, we engineered anticodon variants of the cysteine tRNA that substitute cysteine residues for a variety of other amino acids. These misincorporator tRNAs are expressed from autonomous plasmids that can be co-transformed with arbitrary protein expression vectors.

The MPAX strategy is outlined in Fig. 1. Co-expression of a cysteine misincorporator tRNA with a protein of interest in *Escherichia coli* results in cysteine being misincorporated at low frequency for one type of amino acid. The substituted proteins are exposed to a thiol-specific alkylating reagent. The rate of alkylation at each misincorporated cysteine provides a measurement of its solvent accessibility. Cysteine alkylation is measured by mass spectrometry or by protection from backbone cleavage by the cysteine-specific cutting reagent 2-nitro-5-thiocyanobenzoic acid (NTCB) (8). Because misincorporation occurs at a single amino acid type, cleavage sites are generally distant in the sequence, and cleavage fragments can be separated using SDS-PAGE. Combining data from separate gel lanes for misincorporation at different amino acids overcomes the intrinsic limitations of these gels and can provide structural information at single amino acid resolution. Because typical misincorporation frequencies are ~1%/site, a very small fraction of proteins is expected to contain more than a single cysteine substitution. Rates of modification at slowly reacting sites can be measured accurately despite the existence of rapidly reacting sites, because the misincorporated cysteines are located in different molecules. The method is not limited by protein size or solubility and requires only microgram quantities of material.

* This work was supported by a grant from the Chicago Community Trust (to P. B. H.) and a Terman fellowship (to P. B. H.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ Supported by the Paul and Mildred Berg Stanford Graduate Fellowship.

§ To whom correspondence should be addressed: Dept. of Biochemistry, Stanford University, 279 Campus Dr. W., Stanford, CA 94305. E-mail: harbury@cmgm.stanford.edu.

[1] The abbreviations used are: MPAX, misincorporation proton-alkyl exchange; GdmCl, guanidinium chloride; ICAT, isotope coded affinity tag; NTCB, 2-nitro-5-thiocyanobenzoic acid; TIM, triosephosphate isomerase; Tricine, *N*-[2-hydroxy-1,1-bis(hydroxymethyl)ethyl]glycine.

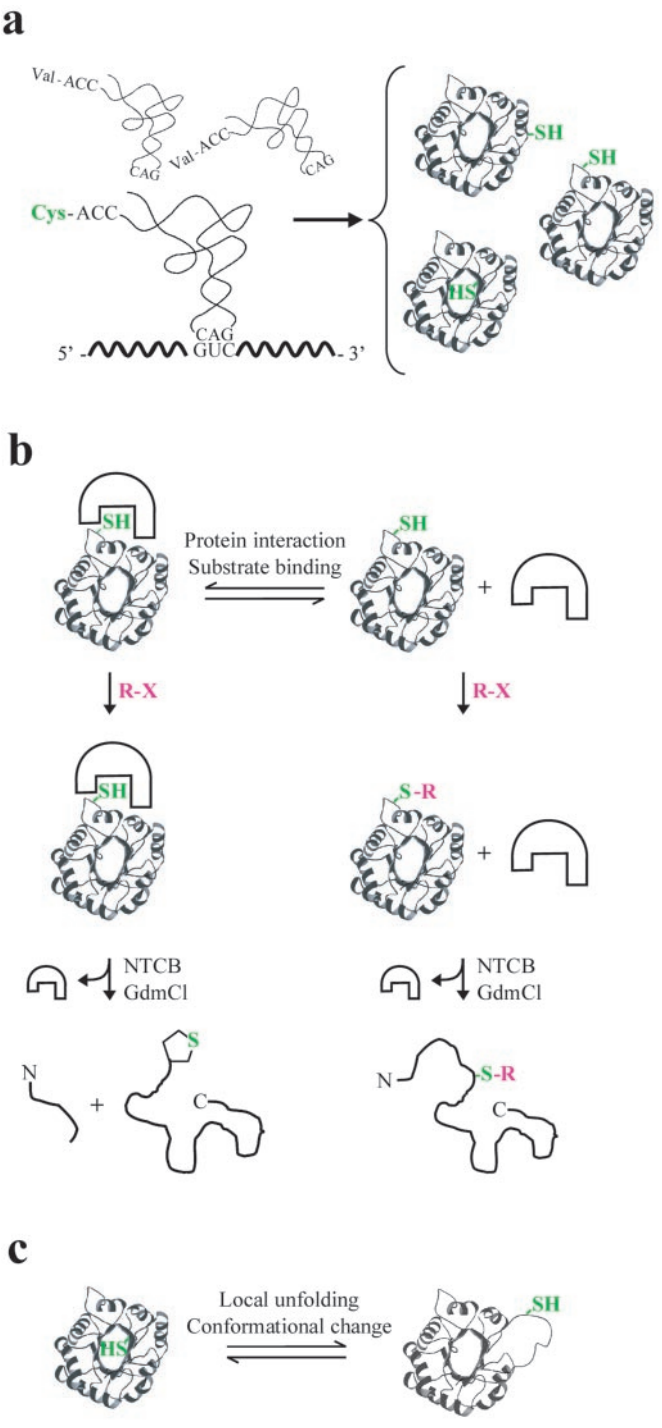

FIG. 1. **Misincorporation proton-alkyl exchange.** *a*, a cysteine tRNA with the valine anti-codon GAC competes with native valine tRNAs in the cell. In this example, cysteine is misincorporated in place of valine at a low level, resulting in an ensemble of proteins containing single cysteine substitutions. *b*, misincorporated cysteines can be used as probes of solvent accessibility. Exposed cysteines on the protein surface react with the alkylating reagent R-X to generate an alkylated sulfhydryl group. These alkylated cysteines do not promote backbone cleavage when exposed to the cysteine-specific cutting reagent NTCB. Cysteines protected from solvent by protein-protein interactions or ligand binding are not alkylated and do promote backbone cleavage when exposed to NTCB. Cleavage fragments are visualized by gel electrophoresis. *c*, changes in cysteine exposure that result from local protein unfolding or a conformational change can also be detected.

EXPERIMENTAL PROCEDURES

Misincorporation of Cysteine—The *E. coli* cysteine tRNA was cloned under the control of the T7 promoter of the pET21a plasmid (Novagen, Madison, WI), and the *E. coli* cysteinyl tRNA synthetase was cloned into the *Bgl*II and *Sph*I sites of the same plasmid to generate pMPAX. Derivative misincorporator plasmids were generated by Kunkel mutagenesis (9) of the tRNA anti-codon sequence and named pMPAXDNB(ABC), where ABC denotes the mutant anti-codon triplet. A synthetic yeast TIM gene (10) was amplified by PCR, adding a C-terminal protein kinase A tag (amino acid sequence: GRRASIY), and cloned into the *Eco*RI and *Hin*dIII sites of pET28a (Novagen). The two native cysteines were substituted by Kunkel mutagenesis to generate the C41V/C126A double mutant designated pH6_TIM_PKA. BL21(DE3) cells were co-transformed with pH6_TIM_PKA and the pMPAX derivative of interest. The cultures were grown in M63 medium (11) containing 50 μg/ml ampicillin and 20 μg/ml kanamycin and induced overnight by the addition of isopropyl-1-thio-β-D-galactopyranoside to 1 mM.

NTCB Cleavage—Ten μg of purified protein was radioactively labeled by incubation with [γ-^{33}P]ATP and protein kinase A (Sigma). Alkylation of cysteines was accomplished by incubating the protein in 50 mM sodium bicine, pH 8.6, 10 mM iodoacetamide for 2 min at room temperature. The reaction was quenched by the addition of an equal volume of 20 mM β-mercaptoethanol, 0.1 mg/ml bovine serum albumin. NTCB cleavage at cysteine residues was performed as described (8, 12) with several modifications. An equal volume of 8 M guanidinium chloride, 300 mM sodium bicine, pH 8.6, 100 mM NTCB was added to the protein sample and incubated for 5 min. The cyanylated proteins were trichloroacetic acid-precipitated (13), and the pellet was resuspended in 10 μl of 8 M urea, 0.1 M NH_4OH. After a 1-h incubation, 5 μl of 500 mM Tris chloride, pH 6.5, 25% glycerol, 5% SDS, 0.001% Coomassie Brilliant Blue G250 was added to quench the cleavage, and the resulting solution was loaded onto a Tricine gel (14) with an additional comb gel layer (15). This protocol resulted in an average cleavage efficiency of 84%. Molecular weight markers were constructed by creating a series of single cysteine mutants of TIM by Kunkel mutagenesis. The pooled proteins were cleaved with NTCB to generate a ladder of defined peptides. The gels were run at 140–160 volts overnight, transferred onto Whatman 3MM Chr paper, and dried. The gels were exposed on image plates and quantitated on a PhosphorImager (Molecular Dynamics, Sunnyvale, CA).

Substrate Binding—Labeled protein misincorporated at leucine or valine was incubated with or without 50 mM sodium glyceraldehyde-3-phosphate in 200 mM sodium bicine, pH 8.6, 10 mM iodoacetamide for 2 min before quenching and cleavage. Illustrations were generated using MOLSCRIPT (16) and the 7TIM structure (17).

Antibody Binding—The Myc epitope sequence (EQKLISEEDL) was inserted at position 132 in the yeast TIM sequence by Kunkel mutagenesis. The denatured protein was diluted from 5 M urea into a reaction containing a final concentration of 1 M urea, 50 mM sodium bicine, pH 8.6, 150 mM NaCl with or without 1 μl of polyclonal anti-Myc serum and incubated with 10 mM iodoacetamide for 2 min.

Measurement of Protection Factors—Labeled protein was prepared and incubated for 24 h with 50 mM sodium bicine, pH 8.6, 10 mM iodoacetamide. The samples were taken at various time points, acid-quenched, and stored at −20 °C. The protection factor at each position was defined as the ratio of the observed rate of alkylation in the unfolded state to the observed rate of alkylation in the folded protein. The results are the averages of three independent measurements. Alkylation rates in the unfolded state were measured to be 3.3 ± 0.1 $M^{-1}\,s^{-1}$ in 4 M GdmCl, 50 mM sodium bicine, pH 8.6, 25 °C for all probes. This value was corrected to 1.5 $M^{-1}\,s^{-1}$ for denaturant-free buffer based on the effect of GdmCl on the alkylation rate of glutathione. The fractional burial of each side chain was calculated using X-PLOR (18) and the 7TIM structure. To determine the stability of TIM, labeled protein at a final concentration of 2 μM was incubated with 50 mM sodium bicine, pH 8.6, 10 mM iodoacetamide for 5 h in the presence of 0.7–1.1 M GdmCl (in 0.1 M increments). The values of RT*ln(protection factor) for valine 91 were plotted against denaturant concentration and extrapolated linearly to zero. The unfolding free energy of TIM was evaluated as the *y* intercept.

Selection for Misincorporation—The *Helicobacter pylori* amidase AmiE was amplified from genomic DNA by PCR and cloned between the *Bgl*II and *Sph*I sites of pET24a (Novagen) to generate pAmiE. Mutations at Cys^{165} were created by Kunkel mutagenesis. Mutant pMPAX libraries were generated by treating 10 μg of pMPAX(GUG) or pMPAX(GCU) DNA with ultraviolet light in a Stratalinker (Stratagene, La Jolla, CA) for 30 s. The libraries were transformed into BL21(DE3) Tuner cells (Novagen) containing pAmiE with either a C165S or a C165H mutation, resulting in 10^5–10^6 transformants. pMPAX plasmids from cells that grew faster than the parent plasmid

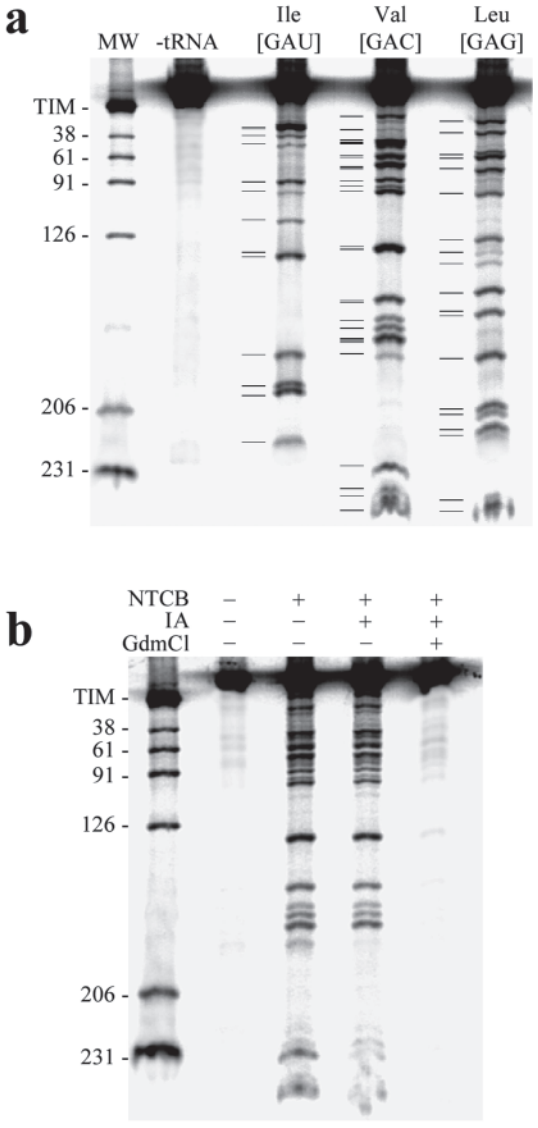

FIG. 2. **Specific cysteine misincorporation.** *a*, TIM was expressed in the presence of the indicated misincorporator tRNA and labeled at its C terminus with radioactive phosphate. The labeled protein was then cleaved with NTCB. The expected pattern of cleavage is shown as *black lines* to the *left* of each *lane*. *MW* denotes the molecular weight markers, which are labeled according to the residue cleaved in the TIM protein. The position of the full-length protein is designated as *TIM*. *b*, TIM with cysteine misincorporated at valine positions was treated with 10 mM iodoacetamide (*IA*) for 2 min in the presence or absence of 4 M GdmCl and cleaved with NTCB.

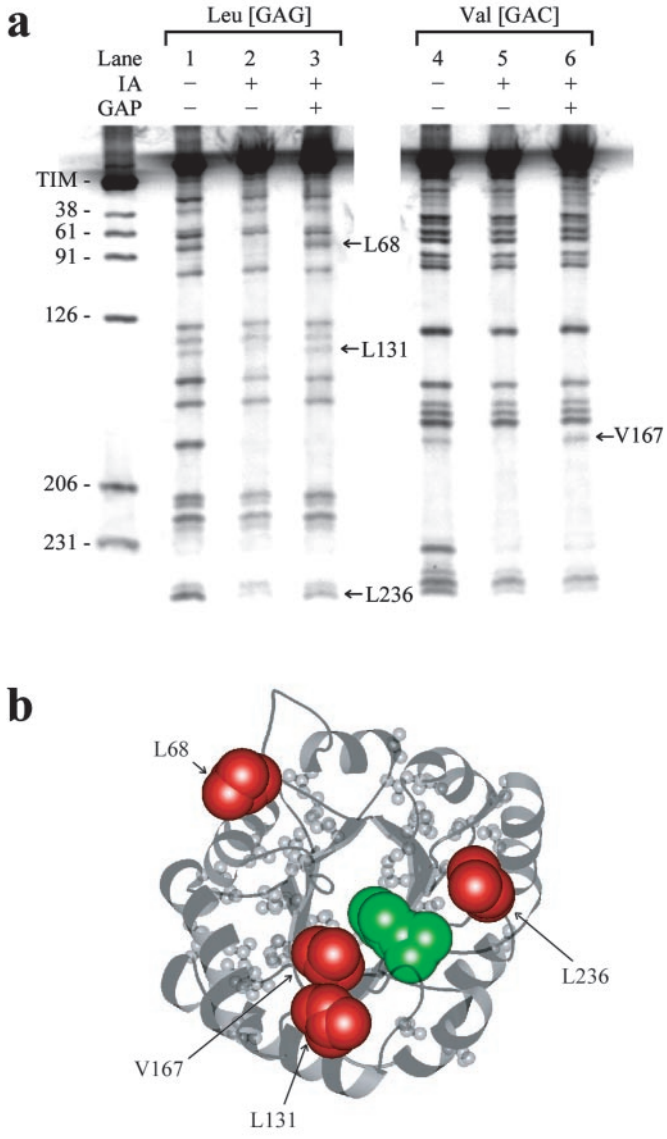

FIG. 3. **Mapping ligand binding.** *a*, TIM protein was treated with 10 mM iodoacetamide (*IA*) for 2 min in the presence (*lanes 3* and *6*) or absence (*lanes 2* and *5*) of 50 mM glyceraldehyde-3-phosphate. The protein was then cleaved with NTCB. The K_m of TIM for glyceraldehyde-3-phosphate is 0.5 mM (30). Unalkylated protein (*lanes 1* and *4*) is shown for comparison. *b*, the locations of amino acids protected from alkylation by glyceraldehyde-3-phosphate are illustrated in the yeast TIM crystal structure. The positions protected by the substrate are shown as *red* van der Waals' surfaces, whereas those that show no protection are shown in *gray*. The substrate analog is shown as a *green* van der Waals' surface.

in acetamide medium (50 mM potassium phosphate, pH 7.8, 10 mM glucose, 1 mM $MgCl_2$, 100 μM citric acid, 50 μM $FeCl_3$, 25 μM $MnCl_2$, 25 μM $CaCl_2$, 100 mM acetamide, 25 μM isopropyl-1-thio-β-D-galactopyranoside, 50 μg/ml ampicillin, and 20 μg/ml kanamycin) were isolated, co-transformed with pH6_TIM_PKA, and screened for misincorporation efficiency.

Read-out by Mass Spectrometry—Details of the synthesis and characterization of the ICAT reagents are provided as supplementary material upon request. Wild-type TIM was incubated with 10 mM ^{13}C-ICAT, 50 mM sodium bicine, pH 8.6, at room temperature for 10, 300, or 1440 min. ^{12}C-ICAT was added to 100 mM, followed by guanidinium chloride to 4 M, and the reactions were incubated for 45 min. Two samples of wild-type TIM were treated respectively with 100 mM of ^{13}C-ICAT or ^{12}C-ICAT in 100 mM sodium bicine, pH 8.6, 4 M guanidinium chloride for 1 h and then pooled to generate a control sample with a 1:1 ratio of ^{12}C:^{13}C at each misincorporated cysteine. The samples were trichloroacetic acid-precipitated to remove excess alkylating reagents and resuspended in 5 M urea. The urea was diluted to 0.5 M in 100 mM Tris chloride, pH 8.0, 25 μg/ml sequencing-grade trypsin (Roche Molecular Biochemicals), and the protein was digested overnight at 37 °C. The samples were clarified by centrifugation, and the supernatant was added to 250 μl of phenylboronate acrylamide beads (Pierce) pre-equilibrated in binding buffer (50 mM sodium HEPES, pH 9.0, 500 mM NaF, 10% acetonitrile). The mixture was rotated for 30 min and then washed for three 30-min incubations with 1 ml of binding buffer. Modified peptides were eluted by two 15-min incubations with 250 μl of elution buffer (10 mM Tris chloride, pH 8.0, 100 mM sorbitol, 10% acetonitrile). The eluates were pooled, concentrated under vacuum, and injected onto a 1 × 50-mm C18 column (Michrom, Auburn, CA) in buffer A (0.025% trifluoroacetic acid, 0.1% formic acid in water) and eluted in a linear gradient of 5–70% buffer B (0.022% trifluoroacetic acid, 0.085% formic acid in acetonitrile) over 50 min. Eluting peptides were analyzed by tandem mass spectrometry on an LCQ ion trap mass spectrometer (ThermoFinnigan, San Jose, CA) at the Stanford Mass Spectrometry Facility essentially as described in Ref. 19. No difference in the elution times of the ^{13}C-ICAT or ^{12}C-ICAT modified peptides was observed. The entire peak area was integrated for calculation of the mass ratio.

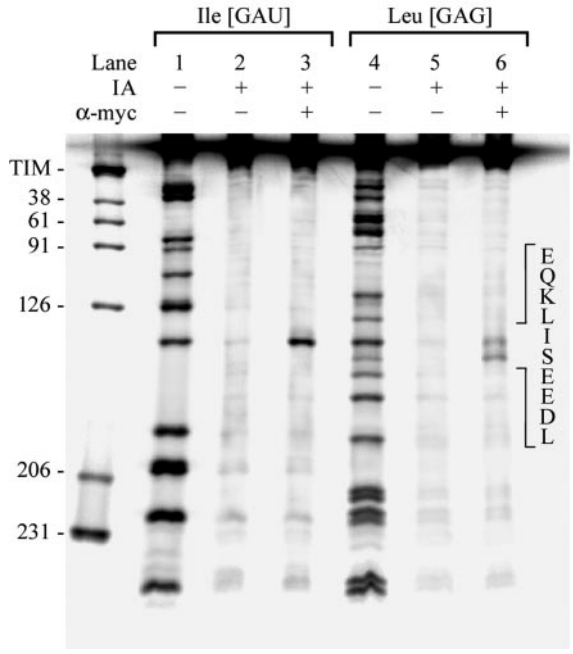

FIG. 4. **Mapping a protein-protein interaction.** A Myc epitope tag sequence was inserted into loop 5 of TIM (shown at *right*). The tagged protein was treated with 10 mM iodoacetamide (*IA*) for 2 min in the presence (*lanes 3* and *6*) or absence (*lanes 2* and *5*) of polyclonal rabbit serum inoculated against the Myc tag (*α-myc*). The protein was then cleaved with NTCB. Unalkylated protein (*lanes 1* and *4*) is shown for comparison. The three protected bands correspond to the leucine and isoleucine residues present in the epitope sequence.

RESULTS

Demonstration of Misincorporation—A cysteine-free variant of yeast TIM was used as a model system to investigate the MPAX technique. TIM is a dimeric $(\beta/\alpha)_8$ barrel protein. The TIM construct used here includes a C-terminal protein kinase A tag to allow labeling with radioactive phosphate, and an N-terminal His_6 tag on a 30-amino acid linker for purification. This linker shifts the full-length, 290-residue protein away from the fainter cleavage products during gel electrophoresis.

We first verified that cysteine could be misincorporated at specific amino acid positions by expressing TIM in the presence or absence of a series of cysteine misincorporator tRNAs. Cleavage of TIM by NTCB is observed only when TIM is co-expressed with a misincorporator tRNA (Fig. 2*a*). The observed ladder of cleavage bands depends on the anti-codon sequence of the tRNA and corresponds to the pattern expected based on the amino acid sequence of TIM (Fig. 2*a*). Amino acid analysis confirmed the presence of low levels of cysteine (~0.3 cysteines/protein) in the purified proteins. Treatment of the protein with 10 mM iodoacetamide for 2 min under denaturing conditions completely blocked cleavage at misincorporated cysteines (Fig. 2*b*). However, treatment of the protein with iodoacetamide under native conditions did not block cleavage at most sites. The data suggest that protein structure protects these residues from alkylation.

Mapping Binding Sites—Strategies for chemically mapping protein interaction sites fall into three broad categories: interference, cross-linking, and protection (20). Interference is based on scanning mutagenesis of a protein. If a residue substitution interferes with a physical interaction, it is inferred that the residue plays a role in binding. In the chemical cross-linking approach, residues that participate in cross-links are identified, and the existence of the cross-link implies spatial proximity to the binding partner. Finally, protection methods are based on chemical modification of a protein in the presence and absence of a binding partner. Binding sites are identified as residues protected from modification by the presence of the partner. The three techniques provide distinct and complementary information. For example, protection and interference experiments identify regions of a protein that undergo a conformational change upon binding, whereas cross-linking only maps direct contact surfaces. Each of the chemical mapping techniques can be carried out in conjunction with parallel misincorporation of cysteine residues. Because the protection strategy is most general with respect to multiple proteins and multiple binding partners, we describe a protection experiment here.

To footprint the ligand-binding site of triosephosphate isomerase, TIM protein with misincorporated cysteines was alkylated with 10 mM iodoacetamide for 2 min in the presence and absence of 50 mM glyceraldehyde-3-phosphate, its substrate. Binding of substrate was found to specifically protect a subset of the fastest alkylating residue positions (Fig. 3*a*). Three of these are located in close proximity to the crystallographically determined substrate-binding site (Fig. 3*b*) and correspond to the only solvent-accessible (less than 99% buried) valine and leucine residues within 15 Å of the substrate. A fourth residue protected by the substrate is located in the dimerization loop, which makes contacts to the substrate in the opposite monomer.

We also investigated the utility of MPAX for mapping a protein-protein interaction surface. Because native triosephosphate isomerase is a homodimer that does not interact with other protein partners, we created tagged variants of TIM with Myc epitope insertions before or after the fifth helix of the barrel. We were not able to refold either tagged protein, so the antibody binding experiment was performed with denatured TIM. Anti-Myc antibodies specifically protected from alkylation only cysteine residues misincorporated within the Myc epitope, providing a direct mapping of the antibody-binding site (Fig. 4). Alkylation at single amino acid resolution was observed by separating leucine and isoleucine bands in different gel lanes. The data demonstrate the utility of MPAX for examining partially folded proteins. Importantly, a protein-binding interface was revealed in a single experiment.

Mapping the Protein Surface—Current *de novo* protein structure prediction algorithms yield multiple reasonable structural models given an input sequence. The inclusion of sparse experimental NMR data in the prediction process significantly improves the accuracy and convergence of the computed models (21). To address the possibility that MPAX data might also be useful for guiding computational structure prediction, we investigated whether MPAX could be used to map a protein surface.

The alkylation rates of cysteine residues misincorporated at 61 positions in the TIM sequence were measured. Solvent-exposed residues are expected to alkylate more rapidly than buried residues, and thus alkylation rates should be useful for assigning sequence positions to interior or exterior environments. The observed alkylation rates can be interpreted using a kinetic model derived from hydrogen exchange experiments (22). In this model, the solvent accessibility of each cysteine side chain is described by an equilibrium between unfolded, solvent-accessible states and a folded, solvent-inaccessible state (Fig. 1*c*). Alkylation is assumed to occur only in the exposed states. The factor by which native protein structure slows alkylation of a cysteine residue (relative to the rate of alkylation of the same cysteine in the unfolded state) defines a protection factor for the site of misincorporation. Large protection factors indicate side chain burial, and

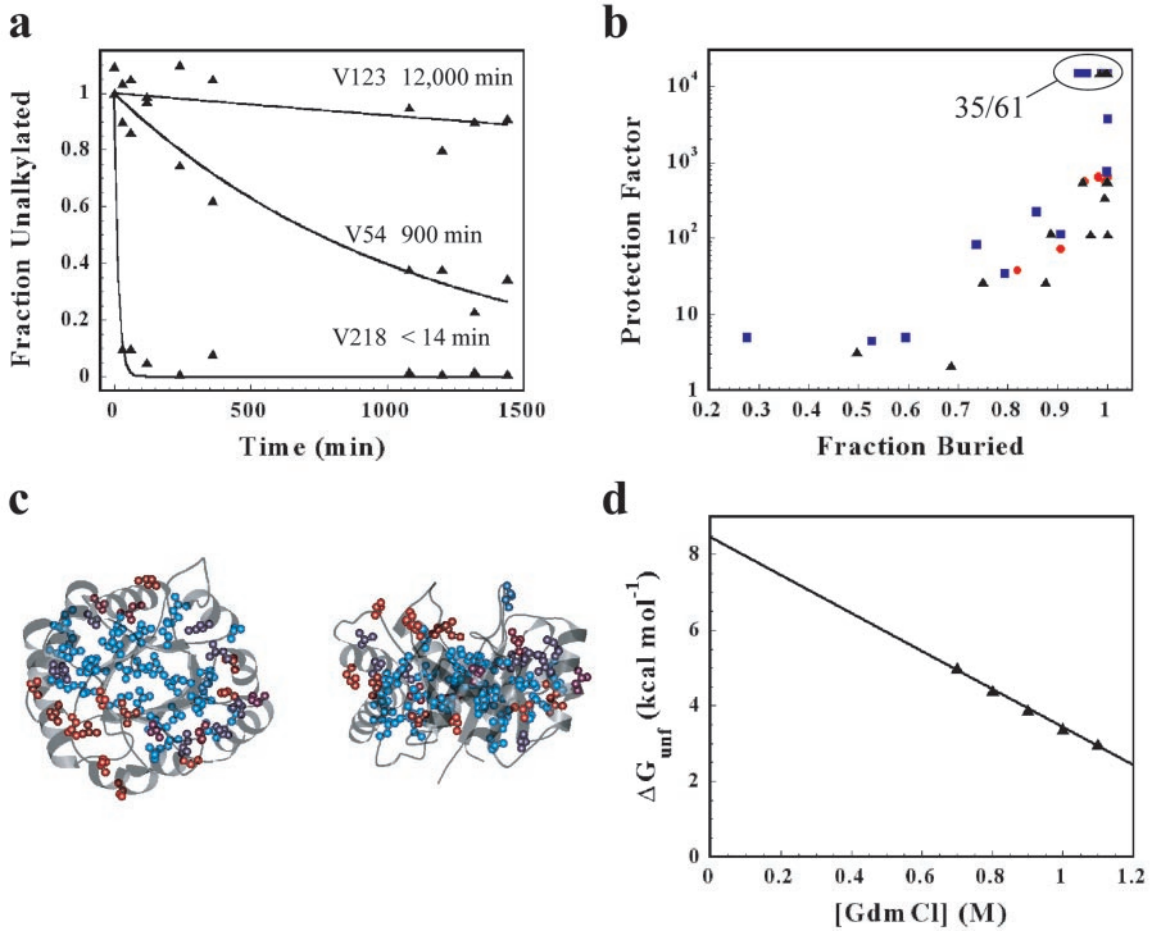

FIG. 5. **Mapping solvent accessibility and stability.** The extent to which protein structure slows the rate of alkylation at 61 positions in TIM is shown. *a*, TIM was alkylated with 10 mM iodoacetamide for variable time periods and then cleaved with NTCB. Representative data show the fractional cleavage at three valine positions (substituted by cysteine) with respect to alkylation time. The *solid lines* represent an exponential fit of the data to a first order kinetic model. The half-life for each fit is indicated. *b*, the protection factor at each misincorporation site is plotted against the fractional burial of the corresponding wild-type residue in the TIM crystal structure. 35 of the 61 probes examined exhibit protection factors of 10^4 or greater (the limit of detection in this experiment) and overlap in the *upper right-hand corner* of the plot. The data are shown for the isoleucine (●), valine (▲), and leucine (■) positions. *c*, residues in the crystal structure of yeast TIM are colored according to their protection factor. Residues with a protection factor less than 100 are *red*, whereas residues with protection factors greater than 100 are *blue*. The highly protected, apparently solvent-exposed residue at the top of the protein (*right*) is buried in the dimer interface. *d*, the unfolding free energy of the V91C TIM protein (calculated as RT*ln(protection factor)) is plotted as a function of GdmCl concentration. The *solid line* is a linear fit of the data extrapolated to zero denaturant.

the magnitude of protection increases in proportion to local protein stability.

Protection factors were measured by incubating TIM for 24 h under native conditions in the presence of 10 mM iodoacetamide. The samples were withdrawn periodically and analyzed by NTCB cleavage. This procedure allows accurate measurement of rates up to 10^4-fold slower than the intrinsic alkylation rate (a protection factor of 10^4). All of the side chains at sites displaying a protection factor greater than 10^3 are more than 94% buried in the crystal structure of the wild-type protein (Fig. 5, *b* and *c*). Conversely, all side chains at sites displaying protection factors less than 10 are at least 30% solvent-accessible. The correlation between residue burial and protection shows that MPAX can be used to assign amino acid positions to the interior or exterior of a protein, providing a constraint on the topology of the protein backbone fold.

Measuring Protein Stability—The equilibrium stability of TIM cannot be measured by spectroscopic methods because even moderate concentrations of the unfolded polypeptide aggregate. However, protection factors can provide a direct measure of protein stability under native conditions, when >99% of the protein is folded.[2] Protection factors for all of the buried probes were measured at various denaturant concentrations (data not shown). The denaturant dependence of the protection factor for the cysteine misincorporated at valine 91, one of the slowest exchanging residues in TIM, indicates that this cysteine mutant of TIM is 8.5 kcal mol^{-1} stable at 2 μM concentration (Fig. 5*d*). MPAX allows measurement of thermodynamic quantities for large and poorly folding proteins like TIM, permitting the kind of energetic analysis currently restricted to small proteins.

Misincorporation at Alternate Codons—For MPAX to be most broadly applicable, misincorporation at a variety of amino acids is required. Accordingly, the efficiency of misincorporation by 18 anti-codon mutants of the cysteine tRNA was investigated.

[2] The protection factor is equal to the equilibrium constant between folded and unfolded states when the rate of conformational closing is much faster than the rate of alkylation, a kinetic regime called EX2 (22). All of the alkylation rates were found to depend on the concentration of alkylating reagent, confirming that alkylation occurs by the EX2 mechanism.

TABLE I
Misincorporation efficiency at different codons

Amino acid	Codon	Apparent cleavage efficiency[a]
		%
Cys	UGC	84.0
Tyr	UAC	1.9
Ile	AUC	0.7
Val	GUC	0.7
Leu	CUC	0.7
His (selected)	CAC	0.7
Phe	UUC	0.6
Trp	UGG	0.4
Ser (selected)	AGC	0.4
Asp	GAC	0.3
Ala	GCU	0.2
Arg	CGU	0.2
Thr	ACU	0.1
Asn	AAC	<0.1
Gln	CAA	<0.1
Glu	GAA	<0.1
Gly	GGU	<0.1
His	CAC	<0.1
Lys	AAG	<0.1
Pro	CCU	<0.1
Ser	UCG	<0.1
Ser	AGC	—[b]

[a] Average intensity of cleavage bands relative to the total protein.
[b] —, this tRNA construct is toxic in BL21 (DE3) cells.

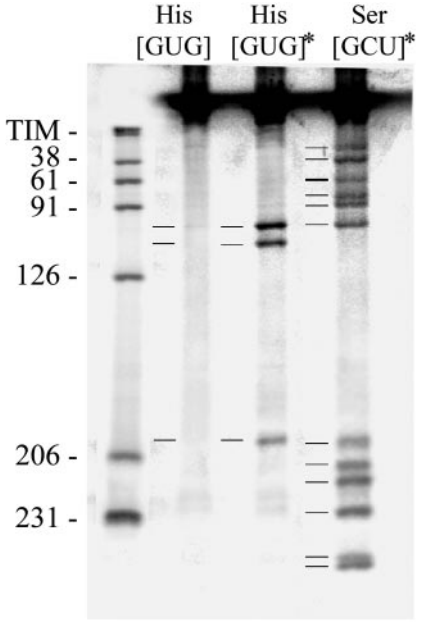

FIG. 6. **Selection for improved misincorporation.** The misincorporation induced by genetically selected misincorporator tRNAs (designated by *asterisks*) is compared with the misincorporation induced by the original histidine misincorporator tRNA. The parent serine misincorporator is not shown because it is toxic in BL21(DE3) cells. The expected cleavage pattern is shown as *black lines* to the *left* of each *lane*.

The observed misincorporation levels varied over a fairly narrow range (Table I), with the differences probably reflecting proofreading activity of the cysteinyl tRNA synthetase. Only the tRNA complementary to the serine codon AGC caused an obvious impairment in cell growth. Misincorporation appears to follow normal wobble base pairing rules (23). For example, the Val(GAC) tRNA misincorporated cysteine equally well at GUC or GUU codons, but not at GUG or GUA codons (data not shown).

To expand the number of amino acids to which MPAX can be applied, we increased the misincorporation efficiency at serine and histidine codons using a genetic selection described previously (7). The selection is based on expression of the *H. pylori* amidase AmiE in *E. coli*. AmiE function is required for growth of *E. coli* with acetamide as the sole nitrogen source. Mutation of the essential catalytic residue Cys^{165} in AmiE eliminates enzymatic function (24). Therefore, efficient misincorporation by the appropriate misincorporator tRNA construct is required to restore the function of Cys^{165} mutants. For our studies, Cys^{165} was mutated to histidine or serine and co-transformed into *E. coli* with mutant misincorporator libraries. The variants were isolated from the library based on an increased growth rate in amine-free, acetamide-containing medium. By direct biochemical measurement, some of these variants exhibit significantly higher levels of misincorporation (Fig. 6 and Table I). Isolation of mutants that efficiently misincorporate cysteine for serine and histidine residues suggests that it will be possible to select efficient misincorporator constructs for all 19 amino acids.

Mass Spectrometry Analysis—Although it is convenient to analyze MPAX data by SDS-PAGE, the gel read-out requires that the protein of interest be devoid of native cysteine residues. A native cysteine residue is present in 100% of the protein molecules and will always be cleaved by NTCB. Thus, only misincorporated cysteines located between the radioactive end label and the closest native cysteine residue are expected to produce observable cleavage bands. To overcome this limitation, we have developed a mass spectrometry method to measure the rate of alkylation at misincorporated cysteines. Our approach is based on a pulse-chase experiment that uses two ICAT reagents, *N*-iodoacetyl, *p*-chlorobenzyl-$^{12}C_6$-glucamine (^{12}C-ICAT) and *N*-iodoacetyl, *p*-chlorobenzyl-$^{13}C_6$-glucamine (^{13}C-ICAT). Except for the difference in nuclear isotope composition, the two compounds are chemically identical. The protein is incubated for a variable time under native conditions with ^{13}C-ICAT, followed by incubation under denaturing conditions with an excess of ^{12}C-ICAT (Fig. 7*a*). Solvent-exposed cysteines are modified completely by ^{13}C-ICAT during the initial pulse, whereas buried cysteines are only alkylated with ^{12}C-ICAT upon unfolding of the protein in the chase step. After alkylation, the protein is fragmented by proteolysis, and peptides containing alkylated cysteine residues are purified by boronate affinity chromatography. This affinity step takes advantage of the fact that the vicinal diols present in the glucose moiety of the ICAT reagent bind to immobilized boronate groups (25). The purified peptides are separated by reverse-phase chromatography and analyzed by mass spectrometry. Each peptide is identified by its absolute mass and by its mass fragmentation pattern. The ^{13}C-ICAT:^{12}C-ICAT isotope ratio reports the fractional alkylation of the peptide during the initial alkylation pulse (Fig. 7, *b–d*). Thus, the mass isotope ratios for cysteine-containing peptides provide a quantitative and site-specific measure of alkylation rates in the folded protein.

We demonstrate the use of the ICAT approach in measuring protection factors of misincorporated cysteines in the wild-type (cysteine-containing) TIM protein. Protection factors determined by mass spectrometry are in excellent agreement with those determined by gel methods (Fig. 7*f*). Mass spectrometry provides a high throughput read-out that does not require the removal of native cysteines from the protein. The ICAT detection method and the modular design of the misincorporation plasmids are well suited for proteome-wide studies.

DISCUSSION

We present a novel technique, MPAX, for the study of protein structure at single amino acid resolution. First, using cysteine tRNAs mutated at the anti-codon triplet, we show that we can biosynthetically misincorporate cysteine at specific amino acid sites in a bacterially expressed protein. Second, we use the pool of cysteine-misincorporated protein to map binding sites on the

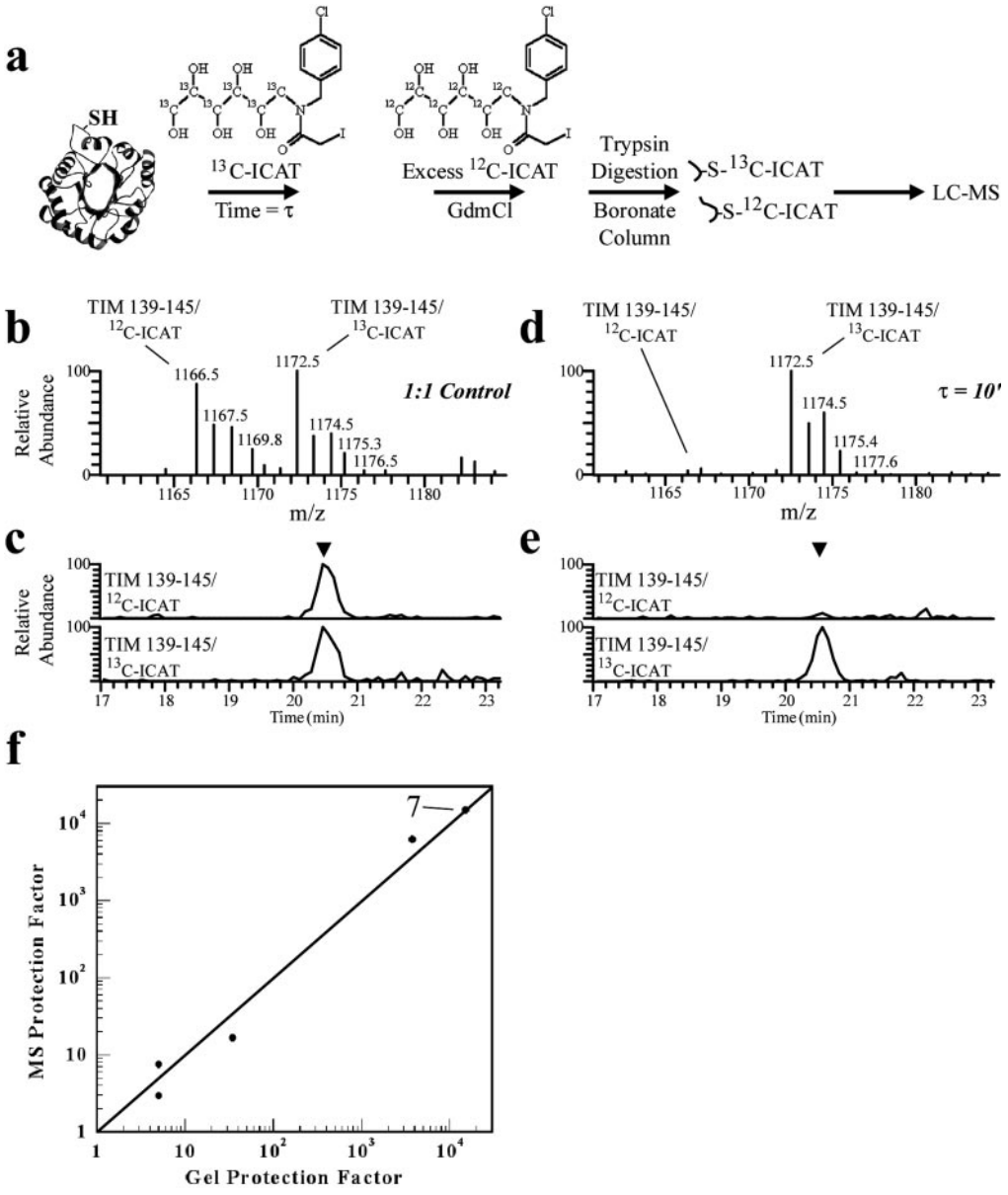

FIG. 7. **Measuring alkylation by mass spectrometry.** *a*, the ensemble of cysteine-misincorporated proteins is alkylated for a variable time τ under native conditions with ^{13}C-ICAT. Excess ^{12}C-ICAT in 4 M GdmCl is subsequently added. The proteins are digested with trypsin, and peptides containing an alkylated cysteine are purified by affinity chromatography over polyacrylamide boronate resin. The purified peptides are analyzed by reverse-phase liquid chromatography coupled with mass spectrometry (*LC-MS*). The peptides are identified by their absolute mass and by their fragmentation pattern. The ^{13}C-ICAT:^{12}C-ICAT ratio at each misincorporated cysteine determines the fractional alkylation that occurred during the initial alkylation pulse. *b*, control experiment showing the mass spectrum of a 1:1 mixture of ^{12}C-ICAT and ^{13}C-ICAT modified TIM peptide 139–145 (TLDVVER) containing cysteine misincorporated at Leu^{140}. The calculated masses for the peptide (H^+-TC^{X}DVVER), where the superior *X* indicates modification with either the ^{12}C-ICAT or ^{13}C-ICAT reagent are 1166.6 and 1172.6 Da, respectively. The additional peaks at +1, +2, and +3 mass units correspond to naturally occurring chlorine isotopes that were used to aid peak identification. *c*, mass chromatograms for the control experiment in *b*. The abundance of peptides with mass 1166–1167 (*upper panel*) or 1172–1173 (*lower panel*) is plotted *versus* elution time from a reverse-phase C18 column. The *arrowhead* indicates the time that the mass spectrum shown in *b* was taken. *d*, mass spectrum of the modified TIM peptide 139–145 (TLDVVER) prepared according to the scheme in *a* with τ = 10 min. *e*, mass chromatograms for the experiment shown in *d*. The abundance of peptides with mass 1166–1167 (*upper panel*) or 1172–1173 (*lower panel*) is plotted *versus* elution time from a reverse-phase C18 column. The *arrowhead* indicates the time that the mass spectrum shown in *d* was taken. The cysteine misincorporated at position 140 is completely alkylated by ^{13}C-ICAT in the initial alkylation pulse. The side chain of Leu^{140} is 40% solvent accessible in the native structure. *f*, plot of the protection factors measured by mass spectrometry read-out *versus* those measured by gel read-out. The data are shown for nine leucine positions and the two naturally occurring cysteines. Seven data points overlap at the limit of resolution (10^4) in the *upper right-hand corner* of the plot. The root mean square difference between the logarithms of the protection factors measured by the two methods is 0.15.

protein for ligands with K_d values ranging from millimolar (glyceraldehyde 3-phosphate binding to the TIM active site) to nanomolar (a polyclonal antibody binding to its epitope). For cysteine probes whose alkylation is completely blocked by ligand binding, probe protection increases by a factor of 1 + [ligand]/K_d in the presence of the ligand relative to its absence. Thus, in principle, MPAX could be used to determine K_d values for ligands. Third, we identify amino acids on the protein surface and buried in the protein structure using MPAX data. This type of information should be useful in predicting the structures of novel proteins. Fourth, we demonstrate that MPAX can be used to measure the protein stability of TIM under conditions that favor the native state. TIM does not reversibly unfold and refold under the conditions conven-

tionally used for stability measurements. Our data suggest that MPAX will be useful for making stability measurements on other large proteins that do not exhibit reversible refolding reactions. Fifth, we show that efficient misincorporation can be achieved at new codons by application of a simple bacterial selection scheme. Thus, it should be possible to apply the MPAX technique to all 20 amino acids. Finally, we demonstrate that MPAX data can be collected for proteins that contain native cysteine residues by using a pulse-chase experiment with an isotope-coded affinity tag detected by mass spectrometry. This approach should be useful for proteomics work.

Structural Perturbations by Cysteine Mutations—A concern regarding the use of misincorporated cysteines as a probe of protein structure is that a cysteine mutation itself could disrupt the native conformation. This concern is mitigated by the fact that the cysteine side chain is small and amphiphilic and thus a good substitute for many amino acids. To address the question directly, we measured the effects on *in vivo* folding caused by cysteine replacements at the isoleucine, leucine, and valine residues of triosephosphate isomerase. TIM was expressed in the presence of misincorporator tRNAs and was purified from both soluble and inclusion body fractions of *E. coli*. The level of misincorporation at each position was measured by NTCB cleavage. The cleavage intensity of each residue in the two protein preparations was found to differ by no more than 5-fold at any position (data not shown). None of the cysteine mutations causes a significant partitioning of triosephosphate isomerase into inclusion bodies *in vivo*, hence cysteine mutations at these positions do not appear to interfere substantially with folding.

If a cysteine mutation at a buried position did disrupt the protein structure, the residue would likely be incorrectly classified as solvent-exposed. This type of disruption is not observed to occur in the TIM mutants that we studied. All of the residues show protection factors consistent with their solvent accessibility (Fig. 5, *b* and *c*). Alternatively, cysteine substitution in a binding interface could interfere with binding. This would produce a false negative result in a protection experiment, because no protection of the misincorporated cysteine would be observed upon addition of the binding partner. Again, this phenomenon is not observed in the TIM mutants we studied. All isoleucine, leucine, and valine residues involved in the binding sites of both a small molecule and an antibody are identified correctly (Figs. 3*b* and 4). Disruption of a binding interface should never produce a false positive result.

Comparison with Alternative Structural Approaches—MPAX is complementary to structural techniques such as x-ray crystallography and multi-dimensional NMR, because it can provide information on the dynamical behavior of proteins in complex solutions. Clearly, prior knowledge of the protein structure greatly facilitates the interpretation of MPAX data. MPAX is closely related to the amide proton-deuterium exchange technique, and essentially any type of amide proton-deuterium exchange experiment can be adapted to the MPAX format. A disadvantage of MPAX relative to amide proton-deuterium exchange is that cysteine mutations are expected to perturb protein structure to a greater extent than proton-deuteron substitutions. MPAX has the advantage relative to amide proton-deuterium exchange in that it is read out by gel methods or by mass spectrometry rather than by magnetic resonance. Thus, the MPAX technique overcomes the technical challenge of assigning a protein NMR spectrum, the requirement for specialized equipment, and the protein size limitations inherent to magnetic resonance. Finally, a number of analytical techniques are based on introduction of single cysteine substitutions into a protein by site-directed mutagenesis. Most of these approaches can be accomplished in the MPAX format, with the advantage that cysteine misincorporation provides a high throughput means to generate distributed cysteine probes.

Prospectus—MPAX can be applied to many types of experiments beyond those demonstrated here. For example, time-resolved measurements of cysteine reactivity can be used to monitor the kinetic progression of biochemical events (see Ref. 26 for analysis of *E. coli* DNA polymerase III processivity clamp assembly by the clamp loader machine and Ref. 27 for studies of apomyoglobin refolding at millisecond resolution). Alkylation of misincorporated cysteines *in vivo* would probe protein interactions in their natural environment. Presumably, the technique can be adapted to eukaryotic expression systems known to accommodate suppressor tRNAs (28, 29). Taken together, these approaches will make possible detailed structural investigations of complex and formerly inaccessible biological processes.

Acknowledgments—Anti-Myc serum and *H. pylori* genomic DNA were the generous gifts of S. Pfeffer and N. Salama, respectively. We thank A. Chien for performing the mass spectrometry analysis and for helpful discussions. We also thank R. Baldwin, S. Nautiyal, S. Pfeffer, D. Herschlag, and members of the Harbury lab for criticism and advice throughout the course of this work.

REFERENCES

1. Galas, D. J., and Schmitz, A. (1978) *Nucleic Acids Res.* **5,** 3157–3170
2. Doonan, S., and Fahmy, H. M. (1975) *Eur. J. Biochem.* **56,** 421–426
3. Hanai, R., and Wang, J. C. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91,** 11904–11908
4. de Arruda, M. V., Bazari, H., Wallek, M., and Matsudaira, P. (1992) *J. Biol. Chem.* **267,** 13079–13085
5. Doering, D. S., and Matsudaira, P. (1996) *Biochemistry* **35,** 12677–12685
6. Tu, B. P., and Wang, J. C. (1999) *Proc. Natl. Acad. Sci. U. S. A.* **96,** 4862–4867
7. Doring, V., and Marliere, P. (1998) *Genetics* **150,** 543–551
8. Jacobson, G. R., Schaffer, M. H., Stark, G. R., and Vanaman, T. C. (1973) *J. Biol. Chem.* **248,** 6583–6591
9. Kunkel, T. A., Bebenek, K., and McClary, J. (1991) *Methods Enzymol.* **204,** 125–139
10. Silverman, J. A., Balakrishnan, R., and Harbury, P. B. (2001) *Proc. Natl. Acad. Sci. U. S. A.* **98,** 3092–3097
11. Ausubel, F. M., Brent, R., Kingston, R., Moore, D. D., Seidman, J. G., Smith, J. A., Struhl, K., Albright, L. M., Coen, D. M., and Varki, A. (1995) *Current Protocols in Molecular Biology* (Janssen, K., ed) Vol. 1, p. 11, John Wiley & Sons, New York
12. Wu, J., and Watson, J. T. (1998) *Anal. Biochem.* **258,** 268–276
13. Arnold, U., and Ulbrich-Hofmann, R. (1999) *Anal. Biochem.* **271,** 197–199
14. Schagger, H., and von Jagow, G. (1987) *Anal. Biochem.* **166,** 368–379
15. Wiltfang, J., Arold, N., and Neuhoff, V. (1991) *Electrophoresis* **12,** 352–366
16. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950
17. Davenport, R. C., Bash, P. A., Seaton, B. A., Karplus, M., Petsko, G. A., and Ringe, D. (1991) *Biochemistry* **30,** 5821–5826
18. Brunger, A. (1992) *X-PLOR*, Yale University Press, New Haven, CT
19. Gygi, S. P., Rist, B., Gerber, S. A., Turecek, F., Gelb, M. H., and Aebersold, R. (1999) *Nat. Biotechnol.* **17,** 994–999
20. Creighton, T. E. (1993) *Proteins*, 2nd Ed., pp. 333–334, W. H. Freeman and Company, New York
21. Bowers, P. M., Strauss, C. E. M., and Baker, D. (2000) *J. Biomol. NMR* **18,** 311–318
22. Bai, Y., Sosnick, T. R., Mayne, L., and Englander, S. W. (1995) *Science* **269,** 192–197
23. Stryer, L. (1996) *Biochemistry*, 4 Ed., pp. 886–888, W. H. Freeman and Company, New York
24. Kobayashi, M., Yanaka, N., Nagasawa, T., and Yamada, H. (1992) *Biochemistry* **31,** 9000–9007
25. Liu, X. C., and Scouten, pp. 886–888, W. H. (2000) *Methods Mol. Biol.* **147,** 119–128
26. Turner, J., Hingorani, M. M., Kelman, Z., and O'Donnell, M. (1999) *EMBO J.* **18,** 771–783
27. Ha, J. H., and Loh, S. N. (1998) *Nat. Struct. Biol.* **5,** 730–737
28. Laski, F. A., Ganguly, S., Sharp, P. A., RajBhandary, U. L., and Rubin, G. M. (1989) *Proc. Natl. Acad. Sci. U. S. A.* **86,** 6696–6698
29. Park, H. J., and RajBhandary, U. L. (1998) *Mol. Cell. Biol.* **18,** 4418–4425
30. Hartman, F. C., LaMuraglia, G. M., Tomozawa, Y., and Wolfenden, R. (1975) *Biochemistry* **14,** 5274–5279

Rapid Mapping of Protein Structure, Interactions, and Ligand Binding by Misincorporation Proton-Alkyl Exchange

Joshua A. Silverman and Pehr B. Harbury

J. Biol. Chem. 2002, 277:30968-30975.
doi: 10.1074/jbc.M203172200 originally published online May 25, 2002
