Proteins have long been known to be affected by UV irradiation in processes ranging from amino acid side chain degradation, the reduction of disulfide bridges, aggregation, and even to consequent loss of biological activity. However, natural examples of clean, site-selective photocleavage are rare. To the best of our knowledge, the only reported examples are fluorescent proteins Kaede and EosFP that experience cleavage (between Phe61, His62 at 400 nm) and require prior formation of a specific chromophore.

A more general, clean, photoinduced cleavage mechanism, with potential biotechnological utility in protein preparation, is therefore also of fundamental interest. Indeed, while a number of site-selective photolytic methods requiring the addition of extraneous cleavage agents have been investigated, these are to date poorly selective and low yielding. Use of unnatural amino acids, such as α-nitro-Phe, only allows photocleavage in up to ~30%. Unexpectedly, while investigating the photochemical modification of extraneous cleavage agents have been investigated, these are to date poorly selective and low yielding. Use of unnatural amino acids, such as α-nitro-Phe, only allows photocleavage in up to ~30%. Unexpectedly, while investigating the photochemical modification of proteins, we have discovered a clean, photocleavage reaction localized to TIM-barrel proteins from family 1 of glycosylhydrolases (GH-1).

UV irradiation of archetypal GH-1 protein from the hyperthermo-

A more general, clean, photoinduced cleavage mechanism, with potential biotechnological utility in protein preparation, is therefore also of fundamental interest. Indeed, while a number of site-selective photolytic methods requiring the addition of extraneous cleavage agents have been investigated, these are to date poorly selective and low yielding. Use of unnatural amino acids, such as α-nitro-Phe, only allows photocleavage in up to ~30%. Unexpectedly, while investigating the photochemical modification of proteins, we have discovered a clean, photocleavage reaction localized to TIM-barrel proteins from family 1 of glycosylhydrolases (GH-1).

UV irradiation of archetypal GH-1 protein from the hyperthermo-

Bioinformatic analysis reveals that the His-Trp (HW) diad is widespread and, indeed, is repeated at positions His424Trp425 in SsβG. Since this site did not dissociate on exposure to UV light, we considered that an extended series of residues bordering His150 and Trp151 must contribute to the remarkable regioselectivity of this cleavage reaction. To probe the underlying molecular contributions, short peptide fragments of SsβG, expected to have little inherent conformation but correct primary sequence, were synthesized: H-HWP-NH$_2$ and H-LNMYHWPLPL-NH$_2$, representing residues 150–152 and 146–155 of SsβG, respectively. Neither fragmented under UV light.

These results implicated conformation and secondary/tertiary structure as key determinants in fragmentation. Ala-scanning was used to probe a 3.5 Å sphere of residues around the cleavage site: SsβG mutations Y149A, P152A, F222A, and W433A prevented fragmentation, while alteration of more remote residues (Q18A, R79A, N81A, L153A, N205A, E206A, V210A, and E387A) had no effect. Modeling of the cleavage site (Figure 1) reveals that essential residue groups Y149, F222, W433, and P152 cradle residue W151 in an hydrophobic, π-rich environment.

Figure 1. SsβG’s photocleavage site (blue) and associated essential (red) and nonessential (green) residues. Model based on PDB 1gow.

1 CRL, University of Oxford.
2 University of Nottingham.
3 PTCL, University of Oxford.
4 Computational Biology, University of Oxford.
5 Plant Sciences, University of Oxford.
Digestion-MS/MS analysis allowed accurate mass and formulae determination of the C-terminal amino acid of the 18.2 kDa fragment and N-terminal amino acid of the 39.1 kDa fragment giving results consistent with His-amine 1 and enamine 2 (Scheme 2). Lack of an N-terminal amine in 2 is also consistent with its lack of reactivity under Edman degradation. MS analyses of the b1 ions of 2 and 3-indole-acrylic acid were identical; both showed characteristic loss of CO followed by loss of HCN from the indole ring. The presence of indolylamidine in 2 was also confirmed by UV spectroscopy ($\lambda_{\text{max}}$ 340 nm).20-22

Evidence to suggest the involvement of singlet oxygen in the cleavage reaction was not found. Exclusion of oxygen did not alter reaction rate. After bubbling with argon or addition of scavenging reagent trolox,23 protein cleavage occurred at the same rate had oxygen not been removed. Reaction rate was also unaffected by D2O.24 Furthermore, one-photon photosensitized production of singlet oxygen (340 nm).20 Failure to observe measurable induction of cleavage by either 1-pyrenebutanoic hydrazide, or fluorescein hydrazide.21-22

The general theory of the photocleaved His-Trp motif was investigated through various methods. Using the motif’s atomic coordinates, all proteins with known 3D structure containing similarly spatially arranged HWP-triads were identified (using SPASM29). The most similar the arylsulfatase from *Medicago sativa*,30 the chalcone reductase from *CtG*,31 and the β-glycosidase from *SsG*,32 were expressed and purified; only the closest *TajG* (c.05 Å rmsd, 54% sequence similarity) was cleaved by UV, whereas other proteins were degraded by UV radiation.33 Daughter fragments (17.2 and 38.1 kDa) revealed a similar, the arylsulfatase from *Medicago sativa*,30 the chalcone reductase, arylsulfatase, and *SsG*, respectively; Dr. S. Kengen for *S. setosa* samples, and no cleavage activity (*HHFD/FHWD*, e.g., *HHFD/FHWD*) or photocleavage activity. Together these analyses correctly predicted to explore transplantation of the photocleavable motif, we created a GFP-3x5G (His$_{16}$tag fusion protein; Ni-Chromatography allowed affinity purification and then “photoelease” of GFP (see Supporting Information (SI)).

In conclusion, we have discovered an efficient photocleavage reaction present in a natural motif of GH1 proteins that can be used to create a photocleavable tag. The biological function, if any, of this photocatalytic post-translational processing is not clear. However, early phylogenetic analyses revealed a cluster of photocleavable proteins that are products of isolated genes or genes in operons associated with peptide transport.36

**Acknowledgment.** We thank Drs. J. McCullagh, R. Procter for technical assistance; Prof. P. Ogilby for helpful discussions; Prof. J. Noel, Drs. M. Kertesz, M. Moracci for plasmids encoding chalcone reductase, arylsulfatase, and *TajG*, respectively; Dr. S. Kengen for *PjG*, Prozomix for Ref*βG*, Cte*βG*, Spj*βG*, and the International AIDS Vaccine Initiative (IAVI) for funding.

**Supporting Information Available:** Experimental procedures, bioinformatics, and mass spectrometry data. This material is available free of charge via the Internet at http://pubs.acs.org.

**References**

19. Such electrophile-rich indolylamidines are poor conjugate electrophiles (see ref. 35) and the 39.1 kDa fragment was not observed with ethanethiol, thiophenol, or diethanolamine.
26. The reaction was unaffected by metals, EDTA, or substrate.
31. Degradation of noncleavable over 10 h. Optimal cleavable reaction times 2–4 h give background degradation <<20% (see SI); LASER light allows cleaner 15 min reactions.

JA9026105