Provided with this examination are two highlight papers that discuss two research papers (also provided) on chemical biology methodology development. On each combination (highlight and underlying research papers) four questions are asked. In answering these you may use all papers discussed in the 15 chapters of the course, your notes, and also your imagination. In answering the questions it is advised to use chemical structures where appropriate, and written text to explain the rationale behind your answers.

As explained during the course, printed manuscripts and hand-written notes are allowed at the examination. Electronic devices are not allowed.

1) Profiling of functional cysteines
Highlight  
_Hypersensitive response to over-reactive cysteines_
Hoogendoorn, Angew. Chem. Int. Ed. 2011, 50, 5434

Paper _Quantitative reactivity profiling predicts functional cysteines in proteomes_
Weerapana, Nature 2010, 468, 790 (this paper was discussed in the course)

A) Describe the general aim of the Weerapana paper.

B) Motivate the use of iodoacetamide-based cysteine-reactive groups (instead of, for instance, fluoroacetamide). Provide an alternative cysteine-reactive group for this purpose.

C) Design a chemically cleavable linker as an alternative for the enzymatic cleavable linker used in the work.

D) The highlight mentions possible extention of the method to activity-based protein profiling. Give an example (activity-based probe, target enzymes, strategy).

2) Activity-based glycosidase profiling
Highlight  
_Getting lucky in the lysosome_
Goddard-Borger, Nat. Chem. Biol. 2010, 6, 881

Paper _Ultrasensitive in situ visualization of active glucocerebrosidase molecules_
Witte, Nat. Chem. Biol. 2010, 6, 907 (new paper, related papers were discussed)

A) Describe the general aim of the Witte paper.

B) Describe why this paper was, according to Goddard-Borger, a 'lucky shot'.

C) Describe a strategy (based on discussion papers from the course) to label more retaining beta-glucosidases in a single experiment.

D) Labeling in the Witte paper takes place in lysosomes. Design an alternative probe that contains a fluorophore that is visible exclusively in lysosomes.
Over the last few years, the field of proteomics has made considerable progress thanks to the development of new techniques for the analysis of protein expression, function, and activity in complex biological samples. These approaches enable not only the identification and characterization of proteins but also the accurate quantification of protein levels and activity in native proteomes. Quantitative proteomics techniques make use of either stable-isotope labeling (e.g. ICAT, SILAC)\cite{1} or label-free techniques based on mass spectrometry (MS)\cite{2} to identify proteins and quantitatively analyze expression levels (Figure 1a), which may differ significantly between different cell types and tissues, and under various conditions. In addition, most proteins are subject to posttranslational modifications, such as oxidative processes, that influence their function and stability. These modifications as well as the differences between different proteomes, for example, healthy versus diseased, can be analyzed by quantitative proteomics.

In the case of proteins with catalytic activity, protein abundance studies are not sufficient to provide insight into the enzymatic activity, which is usually subject to many levels of regulation. Hence, methods are required that directly monitor protein activity irrespective of protein expression levels. Activity-based protein profiling (ABPP) enables the direct quantification of enzymatic activity with the use of small-molecule activity-based probes (ABPs).\cite{3} These interact specifically with the catalytically active form of target enzymes and may be equipped with an affinity tag for isolation and identification by mass spectrometry. While traditional ABPP experiments are aimed at identification of either the complement of the labeled proteins or the active-site fragments targeted by the probe, a tandem orthogonal proteolysis/activity-based protein profiling (TOP-ABPP) method was recently developed to simultaneously identify tagged proteins and sites of modification (Figure 1b).\cite{4}

In general, the reactivity of amino acid side chains, being either catalytic activity or susceptibility to posttranslational modification, is largely dependent on the local protein microenvironment. However, no consensus sequences are known that systematically identify highly reactive amino acid residues and distinguish them from their nonreactive counterparts. This complicates the global identification of reactive sites in the proteome as well as the annotation of newly discovered proteins. ABPP and quantitative proteomics techniques both target a specific subset of reactive amino acid residues, but none of the approaches fully covers the total “reactivity profile” of the proteome.

Of all naturally occurring amino acid residues the free thiol group of cysteine is considered the most reactive group, since it is highly nucleophilic and very sensitive to oxidative modification. Various thiol-specific labeling reagents are available to study cysteine residues. Among these, iodoacetamide (IA) is frequently used in quantitative proteomics, where cysteine residues in two different proteomes are labeled with a stoichiometric amount of light or heavy probe and the differences analyzed (Figure 1a).\cite{1}

Alternatively, ABPP experiments often make use of ABPs that exclusively target catalytically active cysteine residues in a specific class or subclass of enzymes. Selectivity is accomplished by tuning the reactivity of the ABP in such a way that it reacts only at specific sites in the proteome and leaves other functionalities unaffected. An example hereof is the activity-based probe DCG-04 which selectively labels the cysteine protease cathepsins.\cite{5,6} In addition, the nucleophilicity of particular cysteine residues is determined by measuring pKₐ values or the rate of alkylation by specific electrophiles, but this is possible only with purified proteins,\cite{7} which presents an obvious limitation.

Recently, Weerapana et al.\cite{8} designed a strategy for the direct quantification of amino acid side chain reactivity, in particular that of cysteine residues, on a global scale in native biological samples. This approach combines the advantages of TOP-ABPP and quantitative proteomics in the search for functional cysteines in complex proteomes. In this method, termed isoTOP-ABPP (isotopic TOP-ABPP, Figure 1c), the alkyn-iodoacetamide (IA probe) is “clicked” to either a heavy or light azido-TEV-biotin tag. Based on the hypothesis that functionality is reflected by the nucleophilicity and therefore the “hyperreactivity” of cysteine residues, it was reasoned that hyperreactive cysteine residues are labeled to completion with low concentrations of IA probe, whereas the less-reactive cysteine residues are labeled in a concentration-dependent manner. Hence, treatment of a proteome with a low concentration of the heavy IA probe and increasing concentrations of the light IA probe, followed by enrichment of tagged proteins by streptavidin pull-down, trypsin and
TEV digest, and LC-MS/MS analysis gives a light/heavy ratio for each labeled cysteine-containing peptide. Very reactive cysteine residues are expected to have isoTOP-ABPP ratios $R_{\text{light}}:R_{\text{heavy}}$ of approximately 1, and for less-reactive cysteine residues $R_{\text{light}}:R_{\text{heavy}} < 1$. Since this ratio is dependent on reactivity rather than protein abundance and the IA probe is small and cell-permeable, this technique can be used to study functional cysteine residues in native proteomes.

Weerapana et al. show that isoTOP-ABPP ratios for individual cysteine residues are indeed independent of protein abundance and tissue origin. Furthermore, the group of peptides that showed a ratio of $R_{\text{light}}:R_{\text{heavy}} < 2$ (for an IA-probe concentration of 10:1; with a tenfold excess of IA probe compared with a heavy IA probe) is enriched in cysteine residues that are annotated as being catalytically active, part of an active site, or subject to posttranslational (oxidative) modifications. A low isoTOP-ABPP ratio of cysteine residues in uncharacterized proteins provides valuable information that may lead to elucidation of their function. Additionally, this method can be used to predict the functionality of computationally designed proteins in a complex mixture.

The results of this study show that with an elegant concentration-based design cysteine reactivity can be correlated to functionality. Instead of lowering the reactivity of the probe in order to achieve selectivity (as in ABPP), a very reactive probe is now used in different concentrations to discriminate between residues of different reactivity. At low probe concentrations cysteine residues compete for reaction with the probe, resulting in labeling differences. A similar approach is used in organic chemistry, for instance to determine the relative reactivity of glycosyl donors by competition against a common activator.[9] However, since labeling of cysteine residues is also largely dependent on abundance, Weerapana et al. make use of the ratio between high and low concentrations of light and heavy probes rather than absolute labeling quantification.[8]

A shortcoming of the method is that a subset of (functional) cysteine residues might not be labeled by the probe due to sterical reasons. Other cysteine residues might be dependent on cofactors for their activity or have a different mode of action. Therefore, the absence of hyperreactivity does not exclude functionality and only its presence can be used as a strong indication that the residue is involved in a functional process. One could envisage that with a different set of electrophilic probes, targeting a different fraction of the proteome, the scope of this method would be greatly enhanced. Currently, however, very few general reagents for functional groups other than thiols are at hand, so the challenge lies in the development of these. Alternatively, this technique could be expanded to specific ABPs, like DCG-04. Stoichiometric amounts of probe would label the family of enzymes as a whole, whereas lower concentrations of probe could be used to gain insight in the activity of individual members of the family under various conditions. This would for instance give an entry to assess relative cathepsin activities in a setting where their concerted action in the processing of MHC class II antigenic peptides is the subject of study.[10] Future research will show if modest changes in isoTOP-ABPP ratios can be used to quantify even subtle differences in activity.


Cysteine is the most intrinsically nucleophilic amino acid in proteins, where its reactivity is tuned to perform diverse biochemical functions. The absence of a consensus sequence that defines functional cysteines in proteins has hindered their discovery and characterization. Here we describe a proteomics method to profile quantitatively the intrinsic reactivity of cysteine residues en masse directly in native biological systems. Hyper-reactivity was a rare feature among cysteines and it was found to specify a wide range of activities, including nucleophilic and reductive catalysis and sites of oxidative modification. Hyper-reactive cysteines were identified in several proteins of uncharacterized function, including a residue conserved across eukaryotic phylogeny that we show is required for yeast viability and is involved in iron–sulphur protein biogenesis. We also demonstrate that quantitative reactivity profiling can form the basis for screening and functional assignment of cysteines in computationally designed proteins, where it discriminated catalytically active from inactive cysteine hydrolase designs.

Large-scale scientific endeavours such as genome sequencing and structural genomics are providing a wealth of new information on the full complement of proteins present in eukaryotic and prokaryotic organisms. Many of these proteins, however, remain partly or completely unannotated with respect to their biochemical activities. New methods are therefore needed to characterize protein function on a global scale. Much effort is currently devoted to the characterization of post-translational modification events because these covalent adducts can have profound and dynamic effects on protein activity. Another frequently overlooked parameter that defines functional ‘hotspots’ in the proteome is amino acid side-chain reactivity, which can vary by several orders of magnitude for a given residue depending on local protein microenvironment. Methods to measure side-chain reactivity en masse directly in complex biological systems have not yet been described, and as such, the reactive landscape of the proteome remains largely unexplored.

Among the protein-coding amino acids, cysteine is unique owing to its intrinsically high nucleophilicity and sensitivity to oxidative modification. The \( pK_a \) of the free cysteine thiol is between 8 and 9, meaning that only slight perturbations in the local protein microenvironment can result in ionized thiolate groups with enhanced reactivity at physiological pH. Diverse families of enzymes use cysteine-dependent chemical transformations, including proteases, oxidoreductases and acyltransferases. In addition to its role in catalysis, cysteine is subject to several forms of oxidative post-translational modification, including sulphenation (SOH), sulphination (SO\(_2\)H), nitrosylation (SNO), disulphide formation and glutathionylation, which endow it with the ability to serve as a regulatory switch on structural genomics are providing a wealth of new information on the full complement of proteins present in eukaryotic and prokaryotic organisms. Many of these proteins, however, remain partly or completely unannotated with respect to their biochemical activities. New methods are therefore needed to characterize protein function on a global scale. Much effort is currently devoted to the characterization of post-translational modification events because these covalent adducts can have profound and dynamic effects on protein activity. Another frequently overlooked parameter that defines functional ‘hotspots’ in the proteome is amino acid side-chain reactivity, which can vary by several orders of magnitude for a given residue depending on local protein microenvironment. Methods to measure side-chain reactivity en masse directly in complex biological systems have not yet been described, and as such, the reactive landscape of the proteome remains largely unexplored.

Among the protein-coding amino acids, cysteine is unique owing to its intrinsically high nucleophilicity and sensitivity to oxidative modification. The \( pK_a \) of the free cysteine thiol is between 8 and 9, meaning that only slight perturbations in the local protein microenvironment can result in ionized thiolate groups with enhanced reactivity at physiological pH. Diverse families of enzymes use cysteine-dependent chemical transformations, including proteases, oxidoreductases and acyltransferases. In addition to its role in catalysis, cysteine is subject to several forms of oxidative post-translational modification, including sulphenation (SOH), sulphination (SO\(_2\)H), nitrosylation (SNO), disulphide formation and glutathionylation, which endow it with the ability to serve as a regulatory switch on proteins that is responsive to the cellular redox state.

Functional cysteines, regardless of whether they are catalytic residues or sites of post-translational modification, do not conform to a canonical sequence motif, which complicates their systematic identification and characterization. \( pK_a \) measurements can identify cysteine residues with heightened nucleophilicity (or ‘hyper-reactive’ cysteines), but this requires purified protein and detailed kinetic and mutagenic experiments that cannot be performed on a proteome-wide scale. Additional methods have been introduced to computationally predict redox-active cysteines, identify cysteines with specific modifications, and qualitatively inventory electrophile-modified cysteines in proteomes. Some of these studies have provided suggestive evidence that nucleophilic cysteines may possess a variety of important functions, although the non-quantitative methods used in each case precluded a robust and systematic evaluation of this potential relationship. We adopted a different strategy to globally characterize cysteine functionality in proteomes based on quantitative reactivity profiling with isotopically labelled, small-molecule electrophiles.

**Quantifying cysteine reactivity in proteomes**

Our approach, termed isoTOP-ABPP (isotopic tandem orthogonal proteolysis–activity-based protein profiling), has four features to enable quantitative analysis of native cysteine reactivity (Fig. 1a): (1) an electrophilic iodoacetamide (IA) probe, to label cysteine residues in proteins, that also has (2) an alkyne handle for ‘click chemistry’ conjugation of probe-labelled proteins to (3) an azide-functionalized TEV-protease recognition peptide containing a biotin group for streptavidin enrichment of probe-labelled proteins to (4) an isotopically labelled valine for quantitative mass spectrometry (MS) measurements of IA-labelled peptides across multiple proteomes (Supplementary Fig. 1). After tandem on-bead proteolytic digestions with trypsin and TEV protease, probe-labelled peptides attached to isotopic tags are released and analysed by liquid-chromatography-high-resolution MS to identify IA-modified cysteines and quantify their extent of labelling based on MS2 and MS1 profiles, respectively. An isoTOP-ABPP ratio, \( R \), is generated for each identified cysteine that reflects the difference in signal intensity between light and heavy tag-conjugated proteomes.

We first verified the accuracy of isoTOP-ABPP by labelling varying amounts of a mouse liver proteome \((1 \times, 2 \times, 4 \times)\) with the IA probe.
isoTOP-ABPP involves proteome labelling, click-chemistry-based incorporation of isotopically labelled cleavable tags, and sequential on-bead protease digestions to provide probe-labelled peptides for MS analysis. The IA probe is shown in the inset. LC-MS/MS, liquid-chromatography-MS/MS. 

Measured isoTOP-ABPP ratios for cysteines from MCF7 cells labelled with four pairwise IA probe concentrations (10:10 μM, 20:10 μM, 50:10 μM, 100:10 μM). The blue box highlights peptides with low isoTOP-ABPP ratios (R < 2.0). Chromatographs for creatine kinase B (CKB; low ratio) and plastin 2 (LCP1; high ratio) are shown, with elution profiles for heavy- and light-labelled peptides in blue and red, respectively, and green lines depicting peak boundaries used for quantification. Isotopic envelopes are shown for light- and heavy-labelled peptides with green lines representing predicted values. Sequences are shown for tryptic peptides containing IA-probe-labelled cysteines (marked by asterisks) in CKB and LCP1. RT, retention time.

Additional chromatographs from isoTOP-ABPP experiments are in Supplementary Table 7.

Figure 1 | A quantitative approach to globally profile cysteine reactivity in proteomes. a, isoTOP-ABPP involves proteome labelling, click-chemistry-based incorporation of isotopically labelled cleavable tags, and sequential on-bead protease digestions to provide probe-labelled peptides for MS analysis. The IA probe is shown in the inset. LC-MS/MS, liquid-chromatography-MS/MS. 

b, Measured isoTOP-ABPP ratios for peptides from MCF7 cells labelled with four pairwise IA probe concentrations (10:10 μM, 20:10 μM, 50:10 μM, 100:10 μM). The blue box highlights peptides with low isoTOP-ABPP ratios (R < 2.0). Chromatographs for creatine kinase B (CKB; low ratio) and plastin 2 (LCP1; high ratio) are shown, with elution profiles for heavy- and light-labelled peptides in blue and red, respectively, and green lines depicting peak boundaries used for quantification. Isotopic envelopes are shown for light- and heavy-labelled peptides with green lines representing predicted values. Sequences are shown for tryptic peptides containing IA-probe-labelled cysteines (marked by asterisks) in CKB and LCP1. RT, retention time. Additional chromatographs from isoTOP-ABPP experiments are in Supplementary Table 7.

In contrast to traditional cysteine-alkylating protocols for proteomics that use millimolar concentrations of IA to stoichiometrically modify all cysteines in denatured proteins, we proposed that, by applying low (micromolar) concentrations of the IA probe to native proteomes, differences in the extent of alkylation would reflect differences in cysteine reactivity, rather than abundance. This hypothesis predicts that the reactivity of cysteines can be measured on a proteome-scale in isoTOP-ABPP experiments that compare low versus high concentrations of IA probe, where hyper-reactive cysteines would be expected to label to completion at low probe concentrations (generating isoTOP-ABPP ratios with R^{[\text{high}]/[\text{low}] < 1} and less reactive cysteines should show concentration-dependent increases in IA-probe labelling (generating isoTOP-ABPP ratios with R^{[\text{high}]/[\text{low}] > 1}) (Supplementary Fig. 2). We tested this idea by performing four parallel isoTOP-ABPP experiments with the soluble proteome of the human breast cancer cell line MCF7 using pair-wise IA-probe concentrations of 10:10 μM, 20:10 μM, 50:10 μM and 100:10 μM (light/heavy). More than 800 probe-labelled cysteines were identified on 522 proteins, the vast majority of which exhibited escalating isoTOP-ABPP ratios (Fig. 1b) expected for reactions that did not reach completion over the tested probe concentration range. In contrast, a small subset of cysteines (<10%) showed nearly identical ratios at all probe concentrations tested (R^{[\text{1:1}] ≈ R^{[\text{2:1}] ≈ R^{[\text{5:1}] ≈ R^{[10:1]} ≈ 1}}, Fig. 1b, shaded blue box). An expanded analysis of multiple human cancer line (Supplementary Fig. 5 and Supplementary Table 1) and mouse tissue (Supplementary Fig. 6 and Supplementary Table 2) proteomes treated with low (10 μM) and high (100 μM) IA-probe concentrations revealed consistent isoTOP-ABPP ratios for individual cysteine residues, indicating that the propensity of a cysteine to display high IA reactivity is an intrinsic property of the residue (and presumably its local protein environment), and not, in general, contingent on features specific to a particular cell or tissue. Additionally, isoTOP-ABPP ratios showed no correlation with either protein abundance or peptide ion intensity (Supplementary Fig. 7), indicating that they were independent of potential MS-based ionization sources for saturation. Finally, we confirmed that similar isoTOP-ABPP ratios were obtained for cysteines in reactions where time rather than the concentration of probe was varied (Supplementary Fig. 8 and Supplementary Table 3), confirming that low isoTOP-ABPP ratios reflect rapid reaction kinetics (hyper-reactivity), rather than saturable binding interactions (see Supplementary Discussion).

Hyper-reactivity predicts cysteine functionality

We next sought to assess the functional ramifications of the special subset of cysteines that showed hyper-reactivity in isoTOP-ABPP experiments. We first noted that multiple sites of IA-probe labelling on the same protein often showed markedly different isoTOP-ABPP ratios. For example, the glutathione S-transferase GSTO1 was labelled on four cysteine residues, three of which showed high ratios (C90, C192 and C237 had ratios of R^{[10:1]} = 5.6, 7, and 5.4, respectively), whereas the fourth (C32) showed a low ratio of R^{[10:1]} = 0.9 (Fig. 2a). Interestingly, C32 is the active-site nucleophile of GSTO1 (ref. 22). Acetyl-CoA acetyltransferase-1 (ACAT1) was also labelled on four cysteines and three showed high ratios (C119, C196 and C413 showed ratios of R^{[10:1]} = 8.8, 8.2 and 4, respectively), whereas the fourth, the active site nucleophile C126 (ref. 23), yielded a low ratio of R^{[10:1]} = 1.1 (Fig. 2a).

The aforementioned findings indicated that heightened IA reactivity might be a good predictor of cysteine functionality in proteins. To examine this premise more systematically, we queried the Universal Protein Resource (UniProt) database to retrieve functional annotations for the 1,082 cysteine residues labelled by the IA probe. This analysis revealed that the most hyper-reactive cysteines were remarkably enriched in functional residues, with 35% of the cysteines with R^{[10:1]} < 2 being annotated as active-site nucleophiles or redox-active disulphides compared to 0.2% for all cysteine residues in the UniProt database (Fig. 2b, c, Supplementary Fig. 9 and Supplementary Tables 4 and 5). Hyper-reactive cysteines were also, as a group, more conserved across eukaryotic evolution (Supplementary Fig. 10). A broader survey of hyper-reactive cysteines identified several that have been ascribed functional properties in the literature despite lacking annotation in UniProt (Supplementary Fig. 11). For example, a single hyper-reactive cysteine C108 (R^{[10:1]} = 1.0) was identified in the uncharacterized protein D15Ws75e. This protein and its orthologues are predicted...
Function of the hyper-reactive cysteine in FAM96B

Intrigued by the diverse functional properties showed by hyper-reactive cysteines, we reasoned that critical activities might be inferred for such residues in hitherto uncharacterized proteins. A survey of the cysteines to be cysteine proteases based on conservation of a prototypical Cys-His catalytic dyad24. Interestingly, C108 corresponds to the putative active-site nucleophile of this catalytic motif and a recent crystal structure confirms the proximity of C108 to a conserved histidine (H38) (Supplementary Fig. 12). Thus, quantitative reactivity profiling supports structural predictions that D15Ws75e is a functional cysteine protease.

Hyper-reactive cysteines also corresponded to sites for post-translational modification. For instance, C101 (R\(_{10:1} = 1.92\)) in the protein arginine methyltransferase PRMT1 has been identified as a site of modification by the endogenous oxidative product 4-hydroxy-2-nonenal (HNE)25. This cysteine, although nonessential for catalytic function, is an active site residue that makes direct contact with the S-adenosylmethionine cofactor26 (Fig. 3a). Interestingly, we found that HNE inhibited both the IA-labelling (Fig. 3b) and catalytic activity (Fig. 3c) of wild-type PRMT1. A C101A mutant of PRMT1 showed substantially reduced IA-labelling (Fig. 3b) and HNE sensitivity (Fig. 3c). These data indicate that PRMT1 may be regulated by oxidative stress pathways through selective HNE modification of its hyper-reactive, active-site C101 residue. Additional hyper-reactive cysteines represented sites for glutathionylation27 (CLIC1 (C24), CLIC3 (C25) and CLIC4 (C35); R\(_{10:1} = 2.02, 1.07 and 1.45\), respectively) and nitrosylation28 (RTN3; C42, R\(_{10:1} = 0.78\)). These data, taken together, indicate that heightened reactivity is not only a feature of catalytic cysteines, but also of ‘non-catalytic’, active-site cysteines, as well as those that undergo various forms of oxidative modification.

Function of the hyper-reactive cysteine in FAM96B

Figure 2 | Hyper-reactive cysteines are highly enriched in functional residues. a, Chromatographs from an isoTOP-ABPP experiment using 100:10 μM IA probe are shown for peptides from GSTO1 (top) and ACAT1 (bottom). The cysteine nucleophiles (asterisks) show low ratios (R\(_{10:1} \approx 1\)), whereas other cysteines show high ratios (R\(_{10:1} \gg 4\)). b, Pie charts illustrating the percentage of functionally annotated cysteines for three isoTOP-ABPP ratio ranges, including an average derived from all cysteines in the UniProt database. c, Correlation of isoTOP-ABPP ratios with functional annotations from the UniProt database where active-site nucleophiles or redox-active disulphides are shown in red, and all other cysteines in black. A moving average (window of 50) of functional residues is shown as a dashed blue line, demonstrating a profound enrichment within R\(_{10:1} < 2.0\). Data are from experiments in three human cancer cell lines (MCF7, MDA-MB-231 and Jurkat).

Figure 3 | Functional characterization of the hyper-reactive cysteines in PRMT1. a, Crystal structure of rat PRMT126 (green, PDB accession code 1ORJ) showing the hyper-reactive cysteine C101 in contact with an S-adenosylhomocysteine (SAH) cofactor (cyan). b, Wild-type (WT) and C101A mutant of human PRMT1 were labelled with the IA probe, followed by click chemistry to incorporate a fluorescent rhodamine tag. In-gel fluorescence demonstrates robust labelling of the wild-type but not C101A mutant of PRMT1, and shows that IA-probe labelling of wild-type PRMT1 is inhibited by HNE (upper panel). Lower panel shows Coomassie blue staining for treated protein samples. c, Catalytic activity of purified wild-type, but not C101A mutant of PRMT1 was measured by monitoring transfer of \(^3\)H-methyl from \(^3\)H-S-adenosylmethionine (SAM) to a histone 4 substrate.

Table: IsoTOP-ABPP Ratios

<table>
<thead>
<tr>
<th>Cysteine</th>
<th>Ratio</th>
<th>Functional Annotation</th>
</tr>
</thead>
<tbody>
<tr>
<td>C2</td>
<td>0.9</td>
<td><em>Annotated active site + redox-active disulphide</em></td>
</tr>
<tr>
<td>C90</td>
<td>1.1</td>
<td><em>Annotated active site + redox-active disulphide</em></td>
</tr>
<tr>
<td>C112</td>
<td>1.1</td>
<td><em>Annotated active site + redox-active disulphide</em></td>
</tr>
<tr>
<td>C192</td>
<td>1.45</td>
<td><em>Annotated active site + redox-active disulphide</em></td>
</tr>
<tr>
<td>C237</td>
<td>4.0</td>
<td><em>Annotated active site + redox-active disulphide</em></td>
</tr>
</tbody>
</table>

Moving average: 0.78, 1.07, 1.45.
Predicting functional cysteines in designed proteins

The marked correlation between cysteine hyper-reactivity and functionality observed in native proteomes led us to ask whether this relationship would extend to de novo designed proteins. We compared the IA labelling of twelve proteins that were computationally designed to act as cysteine hydrolases. These proteins originated from structurally distinct scaffolds and were all designed to contain cysteine-histidine dyads within an active site cavity (see Supplementary Methods for more details). Two of the designed proteins, ECH13 and ECH19, showed significant hydrolytic activity using a fluorogenic ester substrate, whereas the other ten designs were inactive (Fig. 5a and Supplementary Fig. 11), where they may assist in the transfer of assembled FeS clusters to client proteins. Consistent with this premise, queries of the Saccharomyces genome databank (SGD) revealed that YHR122W has been found in several large-scale protein interaction studies to bind to proteins involved in cytosolic iron-sulphur (FeS) cluster assembly, namely Nar1 and Cia1 (ref. 30; Fig. 4b). We found that the activity of the FeS-client protein isopropylmalate isomerase (Leu1) was markedly reduced in YHR122W (ref. 30; Fig. 4b). We found that the activity of the non-FeS enzyme alcohol dehydrogenase (ADH) was rescued by expression of wild-type YHR122W, thus indicating the importance of C161 for the activity of the non-FeS enzyme alcohol dehydrogenase (ADH). Error bars represent standard deviation, n = 3. ***p < 0.001, Student’s t-test.

**Figure 4** | Functional characterization of YHR122W/FAM96B. 

a. Expression of wild type and a C161A mutant of YHR122W in a yeast strain with a doxycycline (dox)-repressible YHR122W gene demonstrated a dominant-negative phenotype on induction of the C161A mutant expression (−dox/+gal, middle panel) and rescue of viability by expression of wild type, but not the C161A mutant of YHR122W (+dox/+gal, right panel). b. The cytosolic FeS cluster assembly pathway contains multiple proteins with hyper-reactive cysteines (in red). YHR122W/FAM96B (YHR) is a putative member of the FAM96B family. 

We also observed that expression of the C161A mutant of YHR122W caused a severe growth defect in non-suppressive media indicative of a dominant-negative phenotype (Fig. 4a and Supplementary Fig. 14). These data confirm the importance of C161 for the in vivo function of YHR122W and, by extension, other members of the FAM96B family.

We first evaluated IA labelling of protein designs using a clickable, fluorescent reporter tag and SDS–polyacrylamide gel electrophoresis (SDS–PAGE) analysis, where similar amounts of each protein were tested in a homogeneous background proteome representing a mix of Escherichia coli and human (MCF7 cell line) proteins. The two active protein designs ECH13 and ECH19 showed strong IA-labelling signals compared to inactive designs (Fig. 5a), and, in both cases, mutation of the active-site cysteine to alanine abolished labelling (Fig. 5b) and hydrolytic activity (data not shown). We next combined the proteomes containing all twelve protein designs, diluted them into a background human cell proteome, and analysed the mixture by isoTOP-ABPP. Notably, both ECH13 and ECH19 showed isoTOP-ABPP ratios that were equivalent to the most hyper-reactive cysteines in human and E. coli proteomes (R_{10:1} = 0.92 and 1.27, respectively), whereas the remaining inactive protein designs all showed higher ratios ranging from 1.88–6.11 (Fig. 5c and Supplementary Fig. 15b, c). These data thus reveal a strong correlation between cysteine hyper-reactivity and hydrolytic activity across a diverse panel of protein designs and designate heightened cysteine nucleophilicity as a key feature of successful cysteine hydrolase designs.

**Conclusions**

Here, we have described a quantitative method to profile the intrinsic reactivity of cysteine residues in native proteomes. Measurement of the rate of alkylation by IA (or other carbon electrophiles) has been used by enzymologists to assess the nucleophilicity of cysteine residues in individual, purified proteins. With isoTOP-ABPP, these studies can now be extended to quantitative, proteome-wide surveys of cysteine reactivity in complex biological systems. A key advantage of isoTOP-ABPP over more traditional proteomic methods that target cysteine-containing peptides is the use of an alkylated IA probe in place of more bulky biotinylated reagents, which have shown an impaired ability to label cysteines in native proteins. Alkynylated IA probes, owing to their cell permeability, also afford the opportunity to perform cysteine reactivity profiling in living systems. In pilot experiments, we have found that a large fraction of hyper-reactive cysteines are labelled by the IA probe in living cells (Supplementary Fig. 16). Furthermore, isoTOP-ABPP selectively targets probe-accessible cysteines in native proteins. In this way, structural cysteines engaged in disulphide bonds or buried within the body of a protein are avoided to provide preferential access to a specific fraction of cysteines that are profoundly enriched in functionality (the...
IA probe labelled 1,082 out of a total of 8,910 cysteines present on the 890 human proteins detected in this study). Projecting forward, it is possible that, by varying the nature of the electrophile, isoTOP-ABPP probes can be created that profile the reactivity of different subsets of cysteines, as well as other amino acids in proteomes, such as serine, threonine, tyrosine and glutamate/aspartate, which have also been shown to react with small-molecule probes16,18,33–35.

We discovered that hyper-reactivity can predict cysteine function in both native and designed proteins. The fact that hyper-reactivity was strongly correlated with catalytic activity in de novo designed cysteine hydrolases is interesting from the principles of both enzyme engineering and assay development, as it indicates that heightened cysteine nucleophilicity is a key feature of active catalysts and, accordingly, electrophile reactivity could serve as an effective primary screen for novel cysteine-dependent enzymes. We show that these screens can be performed directly in complex proteomes using either gel or MS (isoTOP-ABPP) detection platforms, thus offering a versatile and relatively high-throughput way to evaluate many protein designs in parallel. The isoTOP-ABPP platform has the additional advantage of reading out the relative cysteine reactivity of designs independent of their expression levels against a ‘background’ of native, hyper-reactive cysteines for comparison. isoTOP-ABPP might also offer a complementary way to perform cysteine reactivity/accessibility experiments that monitor protein stability and ligand interactions36,37.

The relationship between cysteine reactivity and functionality extends beyond nucleophilic catalysis to include other enzymatic activities (oxidative/reductive), as well as sites of electrophilic and oxidative modification. Quantitative reactivity profiling thus distinguishes itself as a complementary and perhaps more inclusive strategy to survey cysteine function compared to previous computational11,14,17 and experimental11–13,17 methods that focus on specific cysteine-based activities or modification events. Considering further that hyper-reactive cysteines corresponded to sites for glutathionylation27, nitrosylation34 and HNE-modification35, we speculate that cysteine nucleophilicity is a property that may have been selected for during evolution to offer points of protein control by oxidative stress pathways. Determining how the reactivity of cysteine residues is honed will require further investigation, but we anticipate that quantitative proteomic data, when integrated with the output of ongoing structural genomics programs, may eventually uncover unifying mechanistic principles that explain cysteine reactivity in proteins. In this regard, it is interesting to note that, although hyper-reactive cysteines did not conform to any obvious consensus sequence motifs, many of these residues were found at the N termini of α-helices (Supplementary Fig. 17). This finding is consistent with literature reports ascribing a role for α-helix dipoles in the stabilization of cysteine thiolate anions38.

Finally, it is important to stress that some functional cysteines may be inherently reactive, but inaccessible to our IA probe for steric reasons. Other cysteine-reactive electrophilic probes16,17 may prove more suitable for such cysteine residues. Also, hyper-reactivity is not necessarily a defining feature for all functional cysteines. Some enzymes with catalytic cysteines may, for instance, show reduced reactivity until they bind their physiological substrates or may rely more on substrate recognition than inherent catalytic power for function. This may be the case with the E1-activating and E2-conjugating enzymes, which recognize a specific class of ubiquitinated substrates and possess active-site cysteines that showed only moderate levels of electrophile reactivity (Supplementary Fig. 18). Other cysteines may have activities that are not dependent on their nucleophilicity. Our data do indicate, however, that those cysteines that are hyper-reactive in proteomes probably perform important catalytic and/or regulatory functions for their parent proteins. The large number of newly discovered residues that fall into this category foretell a broad role for hyper-reactive cysteines in mammalian biology.

**METHODS SUMMARY**

**Probes and tags.** The IA probe and the light and heavy variants of the azide-TEV-biotin tags were synthesized as previously described39,40.

**Sample preparation, mass spectrometry and data analysis.** For concentration-dependent experiments, proteome samples in PBS were probe labelled with the desired probe concentration for 1 h. Click chemistry was performed with either the light or heavy variants of the azide-TEV-biotin tags and the samples were mixed and subjected to streptavidin enrichment and subsequent trypsin and TEV digestion. The resulting TEV digests were analysed by Multidimensional Protein Identification Technology (MudPIT) on an LTQ-Orbitrap instrument. The resulting tandem MS data were searched using the SEQUEST algorithm40 using the following criteria: E1, E2 and E3; catalytic cysteines in ECH13 and ECH19 show low isoTOP-ABPP ratios (red) compared with other designs (blue). Chromatographs are shown for peptides from the nine designs identified in this experiment (bottom panel), in the same order as shown in the top panel.
a concatenated target/decoder variant of the human, mouse and E. coli protein sequence databases. Quantification of light/heavy ratios (isoTOP-ABPP ratios, R) was performed using in-house software. Detailed information on sample preparation, mass spectrometry methods and data analysis is presented in Methods.

**Complementation of S. cerevisiae YHR122W deletion mutant.** Complementary DNA encoding wild-type YHR122W was subcloned into the pESC-Leu vector (Stratagene). The YHR122W/CIC161A mutant was generated using the Quickchange procedure (Stratagene). These constructs were introduced into the yeast Tet promoter Hughes (yTHC) strain harbouring a conditional (doxycycline-dependent) disruption in the YHR122W gene (Open Biosystems). Growth of these transformed cell lines on -gal/-doc media was monitored for 3 days. These cell lines were also used to monitor Leu and alcohol dehydrogenase (ADH) activity. Detailed information on the protocols used to subclone, transform and monitor the growth of the yeast strains and measure enzyme activity is available in Methods.

**Full Methods** and any associated references are available in the online version of the paper at www.nature.com/nature.
METHODS

All compounds and reagents were purchased from Novabiochem, Sigma or Fisher, except where noted.

Preparation of mouse proteomes. Mouse tissues (heart and liver) were harvested and immediately frozen in liquid nitrogen. The tissues were then Dounce homogenized in 1× PBS, pH 7.4. Centrifugation at 100,000 × g (45 min) provided soluble fractions (supernatant) and membrane fractions (pellet). Protein concentrations for each proteome were obtained using the Bio-Rad DC protein assay and stored at −80 °C till use.

Preparation of human cancer cell line proteomes. MDA-MB-231 cells were grown in L15 media supplemented with 10% fetal bovine serum at 37 °C in a CO2-free incubator. Jurkat cells and MCF7 cells were grown in RPMI-1640 supplemented with 10% fetal bovine serum at 37 °C with 5% CO2. For in vitro labelling experiments, cells were grown to 100% confluency, washed three times with PBS and scraped in cold PBS. Cell pellets were isolated by centrifugation at 1,400g for 3 min, and the cell pellets stored at −80 °C until further use. For in situ labelling of MDA-MB-231 and MCF7 cells, the cells were grown to 90% confluency, the media was removed and replaced with fresh media containing 10 μM IA probe. The cells were incubated at 37 °C for 1 h and harvested as detailed above. The harvested cell pellets were lysed by sonication and fractionated by centrifugation (100,000 × g, 45 min) to yield soluble and membrane proteomes. The proteomes were diluted to 2 mg/mL and stored at −80 °C until use.

Protein labelling and click chemistry. Proteome samples were diluted to 2 mg protein/ml solution in PBS. Each sample (2× 0.5 mL aliquots) was treated with 10, 20, 50, or 100 μM of IA probe using 5 μL of a 1, 2, 5, or 10 mM stock in DMSO. The labelling reactions were incubated at room temperature (25 °C) for 1 h. Click chemistry was performed by the addition of 150 μM of either the light TEV tag or heavy TEV tag (15 μL of a 5 mM stock, 1 mM tris(2-carboxyethyl)phosphine (TCEP; fresh 50× stock in water), 100 μM ligation (17× stock in DMSO:tol-butanol 1:4) and 1 mM CuSO4 (50× stock in water). Samples were allowed to react at room temperature for 1 h. After the click chemistry step, the light- and heavy-labelled samples were mixed together and centrifuged (5,900g, 4 min, 4 °C) to pellet the precipitated proteins. The pellets were washed twice in cold MeOH, after which the pellet was solubilized in PBS containing 1.2% SDS via sonication and heating (5 min, 80 °C).

For time course experiments, proteome samples were labelled with 100 μM of IA probe (using 5 μL of a 10 mM stock in DMSO). After 6 min of probe labelling, an aliquot of the reaction was quenched by passing the sample through a NAP-5 column (GE Healthcare) to remove excess, unreacted protein. After 60 min of probe labelling, the other sample was quenched as before and click chemistry was performed as described earlier.

Streptavidin enrichment of probe-labelled proteins. The SDS-solubilized, probe-labelled proteome samples were diluted with 5 mL of PBS for a final SDS concentration of 0.2%. The solutions were then incubated with 100 μL of streptavidin-agarose beads (Pierce) for 3 h at room temperature. The beads were washed with 10 mL 0.2% SDS/PBS, 3× 10 mL PBS and 3× 10 mL H2O and the beads were pelleted by centrifugation (1,300g, 2 min) and resuspended in 200 μL of 2 M urea/PBS, 1 mM CaCl2 (100× stock in H2O), and trypsin (2 μg). The digestion was allowed to proceed overnight at 37 °C. The digest was separated from the beads using a Micro BioSpin column and the beads were then washed with 3× 500 μL PBS, 3× 500 μL H2O, and 1× 150 μL of TEV digest buffer. The washed beads were then resuspended in 150 μL of TEV digest buffer with AcTEV Protease (Invtrogen, 5 μL) for 12 h at 29 °C. The eluted peptides were separated from the beads using a Micro BioSpin column and the beads washed with H2O (2× 75 μL). Formic acid (15 μL) was added to the sample, which was stored at −20 °C until MS analysis.

Liquid-chromatography-mass-spectrometry (LC-MS) analysis. LC-MS/MS analysis was performed on an LTQ-Orbitrap mass spectrometer (ThermoFisher) coupled to an Agilent 1100 series high-performance liquid chromatography system. TEV digests were pressure loaded onto a 250 μm fused silica desalting column packed with 4 cm of Aqua C18 reverse phase resin (Phenomenex). The peptides were then eluted on a biphasic column (1,300 g μm fused silica with a 5 μm tip, packed with 10 cm C18 and 3 cm Partisphere strong exchange resin (SCX, Whatman) using a gradient 5–100% buffer B in buffer A (buffer A: 95% water, 5% acetonitrile, 0.1% formic acid; buffer B: 20% water, 80% acetonitrile, 0.1% formic acid). The peptides were then eluted from the SCX onto the C18 resin and into the mass spectrometer using four salt steps as previously described.20,21 The flow rate through the column was set to 0.25 μL min−1 and the spray voltage was set to 2.75 kV. One full MS scan (FTMS) (400–1,800 MW) was followed by 18 data dependent scans (ITMS) of the most intense ions with dynamic exclusion disabled.

Peptide identification. The tandem MS data were searched using the SEQUEST algorithm44 using a concatenated target/decoy variant of the human and mouse International Protein Index databases. A static modification of +57.02146 on cysteine was specified to account for iodoacetamide alkylation and differential modifications of +464.28596 (light probe modification) and +470.29977 (heavy probe modification) were specified on cysteine to account for probe modifications with the either light or heavy variants of the IA-probe-TEV adduct. SEQUEST output files were filtered using DTASelect 2.0. Reported peptides were required to be fully tryptic and contain the desired probe modification and discriminant analyses were performed to achieve a peptide false-positive rate below 5%. The actual false-positive rate was assessed at this stage according to established guidelines43 and found to be ~3.5%. Additional assessments of the false-positive rate were performed following the application of additional filters (described later) resulting in a final false-positive rate below 0.05%.

Ratio quantification. Quantification of light/heavy ratios (isoTOP-ABPP ratios, R) was performed using in-house software written in the R programming language that utilizes routines from the open-source XCMS package45 for MS data analysis to read in raw chromatographic data in the mzXML format46. Each experiment consisted of two LC/MS/MS runs: light:heavy 10 μM:10 μM, and light:heavy 100 μM:10 μM IA-probe concentration. Both runs were searched using SEQUEST and filtered with DTASelect as described earlier. Because the mass spectrometer was configured for data-dependent fragmentation, peptides are not always identified in every run. As such, peptides were identified in either 1) only the 10 μM:10 μM run, 2) only the 100 μM:10 μM run, or 3) both runs. In the case of peptides that were sequenced in both runs, identification of the corresponding peaks was made by choosing peaks that co-elute with the peptide identification. In the case of probe-modified peptides that were sequenced in one, but not the other run, an algorithm was developed to identify the corresponding peak in the run without the SEQUEST identification. To accomplish this, the retention time of the ‘reference’ peptide is used to position a retention time window (±10 min) across the run lacking a peptide identification. Extracted ion chromatograms (±10 p.p.m.) of the target peptide m/z with both ‘light’ and ‘heavy’ modifications are generated within that window. The program then searches for candidate co-eluting pairs of light-heavy MS1 peaks, and for each candidate pair calculates the ratio of integrated peak area between the light and heavy peaks. Several filters are used to ensure that the correct peak pair is identified. First, the extent of co-elution for each peak pair is quantified using a Pearson correlation, an established method to gauge elution profile similarity47. Second, the predicted pattern of the isotopic envelope of the target peptide is generated and compared to the observed high-resolution MS1 spectrum. This comparison generates an ‘envelope correlation score’ (Env) that also enables confirmation of the monoisotopic mass and charge state of each candidate peak. Peak pairs that have poor co-elution scores, or that have the incorrect monoisotopic mass or charge, or whose isotopic envelopes are not well correlated with the predicted envelope are eliminated from consideration. After application of these filters, in the rare case that multiple candidates still exist, then no peak is chosen and a ratio is not recorded. Usually, however, application of these filters results in a single candidate peak pair and the ratio for this peak pair is recorded for the peptide in the corresponding run. In this way, each experiment yields two ratios, one for the 10 μM:10 μM run and one for the 100 μM:10 μM run. Following application of these filters, the false-positive rate was reassessed, and found to be less than 0.05% in all cases.

AFTER ratios for unique peptide entries are calculated for each experiment, overlapping peptides with the same labelled cysteine (for example, same local sequence around the labelled cysteines but different charge states, MudPIT segments, or tryptic termini) are grouped together, and the median ratio from each group is reported as the final ratio (R). All of these values can be found in Supplementary Tables 1, 2 and 3 and representative chromatographs can be seen in Supplementary Table 7. Raw result files of peptide identification using SEQUEST can be found in Supplementary Table 9.

Functional annotation of labelled cysteines. For automated functional analyses, custom perl-scripts were developed to query the UniProtKB/Swiss-Prot Protein Knowledgebase release 57.4 (current as of 16 June 2009). Sequence annotation in the (Features) section of the relevant UniProt entry was mined and any annotation corresponding to the labelled residue was collected. This functional annotation was then summarized across the four tables 4 and 5.

Recombinant PRMT1 protein expression and purification. Full-length cDNA encoding human PRMT1 in pOTB7 was purchased from Open BioSystems and subcloned into PET-45B (+) (Novagen). BL21(DE3) E. coli containing this vector was grown in LB media containing 75 μg mL−1 carbenicillin with shaking at 37 °C.
to an OD<sub>600nm</sub> of 0.5. The cells were then induced with 1 mM isopropl-β-D-thiogalactoside (IPTG) and harvested 4 h later by centrifugation. Cells were lysed by stirring for 20 min at 4 °C in 50 mM Tris-HCl (pH 8.0) with 150 mM NaCl and supplemented with 1 mg/ml lysozyme and 1 mg/ml DNaSE. The lysate was then sonicated and centrifuged at 10,000g for 10 min. Talocon cobalt affinity resin (Clontech; 400 µl of slurry per gram of cell paste) was added to the supernatant, and the mixture was rotated at 25 °C for 30 min. Beads were collected by centrifugation at 700g for 3 min, washed twice with Tris buffer, and applied to a 1-cm column. The column was washed twice with Tris buffer (10 ml per 400 µl of resin slurry) and Tris buffer with 500 mM NaCl once. The bound protein was eluted by the addition of 100 mM imidazole (2 ml per 400 µl of resin). Imidazole was removed by passage over a Sephadex G-25M column (GE Healthcare), and the eluate was concentrated using an Amicon centrifugal filter device (Millipore). Protein concentration was determined using the Bio-Rad DC protein assay kit.

These conditions yielded PRMT1 at approximately 0.5 mg/ml of culture. A C101A mutation was introduced into the pET-45b(+) construct described earlier using the Quickchange Site-Directed Mutagenesis Kit (Stratagene), and the resulting mutant protein was expressed identically and isolated with a similar yield.

**In-gel fluorescence characterization of PRMT1.** Thirteen micrograms of recombinant PRMT1 (wild type or C101A mutant) was pre-incubated with 0, 25 or 50 µM HNE (Calbiochem, 50 mM stock in ethanol) for 1 h at room temperature and was then labelled with 100 nM of the IA probe (5 µM stock in DMSO) and the reactions incubated for 1 h at room temperature. Click chemistry was performed with 20 µM rhodamine-azide, 1 mM TCEP, 100 mM TBTA ligand and 1 mM CuSO<sub>4</sub>. The reaction was allowed to proceed at room temperature for 1 h before quenching with 50 µl of 2X SDS–PAGE loading buffer (reducing). Quenched reactions were separated by SDS–PAGE (30 µl of sample/lane) and visualized in-gel using a Hitachi FMBio II flatbed laser-induced fluorescence scanner (MiraBio).

**PRMT1 in vitro methylation assays.** Five-hundred nanograms of recombinant human PRMT1 (wild type or C101A mutant) was pre-incubated with HNE (Calbiochem) for 30 min and methylation activity was monitored after addition of 1 mg of recombinant histone 4 (M250S; NEB) and SAM (2 µCi) in methylation buffer (20 mM Tris, pH 8.0, 200 mM NaCl, 0.4 mM EDTA). Reactions were incubated for 90 min at 30 °C and stopped with SDS sample buffer. SDS–PAGE gels were fixed with 10% acetic acid/10% methanol v/v, washed, and incubated with Amplify reagent (Amersham) before exposing at −80 °C.

**Complementation of S. cerevisiae YHR122W deletion mutant.** A cDNA encoding YHR122W was purchased as a full-length expressed sequence tag (Open Biosystems). The construct for subcloning into the yeast epitope tagging vector pESC-Leu (Stratagene) was generated by polymerase chain reaction (PCR) from the corresponding cDNA using the following primers: sense primer, 5'-GAAGCGGCCGCAATGTCTGAGTTTTTGAATGA-3'; antisense primer, 5'-CCGACTAGTGCCTTACAAGTCACTAACATCTTAG-3'.

The PCR product was digested with NotI-SpeI and subcloned into a NotI-SpeI-digested pESC-Leu vector and sequenced. The YHR122W(C161A) mutant was generated using the Quickchange procedure (Stratagene). The mutant cDNA was sequenced and found to contain only the desired mutation.

Constructs containing wild-type and C161A mutant YHR122W were introduced into the yTHC strain YSC1180-7428770 (Open Biosystems) using the reagents provided in the Yeastmaker Yeast Transformation System 2 (Clontech). The yeast was grown in synthetic dextrose minimal medium (−Leu) and spot assays were performed in either synthetic dextrose minimal medium (−Leu) or synthetic galactose minimal medium (−Leu) + agar plates ± 50 µg ml<sup>−1</sup> doxycycline. The plates were cultured at 30 °C for 3 days.

**Isopropylmalate isomerase (Leu1) assay.** Yeast strains harbouring either an empty vector or wild-type YHR122W (see earlier section) were cultured in synthetic dextrose minimal medium (−Leu) to an OD<sub>600nm</sub> of 1.0 and transferred into synthetic galactose minimal medium (−Leu) ± 50 µg ml<sup>−1</sup> doxycycline for 12 h. Yeast were lysed and Leu1 semi-purified by ammonium sulphate precipitation (40–70%). The activity assays were performed using N-threo-3-isopropylmalic acid as the substrate and product formation was measured by monitoring absorbance at 235 nm for 10 min.

**ADH assay.** Yeast cell lysates in 0.1 M sodium pyrophosphate buffer (pH 9.2, 1.5 ml) were treated with 2 mM ethanol (0.5 ml) and 0.025 M NAD (1.0 ml) and ADH activity was measured by absorbance increase at 340 nm for 3 min.

**De novo designs of cysteine hydrolases and hydrolysis activity assays.** We used the Rosetta computational enzyme design methodology<sup>46</sup> to search a set of protein scaffolds for constellations of backbones capable of supporting an idealized transition state for ester hydrolysis derived from the geometries and mechanisms of natural cysteine hydrolases<sup>47</sup>. The idealized active-site models feature a nucleophilic cysteine, a general base/acid histidine and at least one side-chain or backbone hydrogen bond donor as the oxanion hole. The sequence of residues surrounding the putative active sites was optimized using the Rosetta design algorithm to maximize transition state stabilization<sup>48</sup>. A set of 12 designed proteins in 10 distinct scaffolds was chosen for experimental characterization. For each designed protein, synthetic genes were obtained and protein expression and purification was performed in E. coli as previously described<sup>49</sup>. Activity was measured with the substrate by following the initial (<5% substrate conversion) increase in fluorescence due to the appearance of the product coumarin. A protein concentration of 20 µM and substrate concentration of 100 µM were used in 25 mM HEPES buffer, 150 mM NaCl, 1 mM TCEP, pH 7.5. The background rate was measured under identical conditions but without the protein. Kunkel mutagenesis was used for creating point mutations in the active-site residues. A detailed description of the design and characterization of the cysteine hydrolases will be presented elsewhere. Amino acid sequences of the 12 designs can be found in Supplementary Information.

**In-gel fluorescence and isoTOP-ABPP characterization of designed proteins.** For in-gel fluorescence studies, E. coli lysates overexpressing the designed proteins were diluted to 2 mg protein/ml in PBS. Each sample (25 µl) was mixed with 25 µl of MCF7 human cell soluble proteome (2 mg ml<sup>−1</sup>) and was labelled with 100 nM of the IA probe (5 µM stock in DMSO) and the reactions incubated for 1 h at room temperature. Click chemistry, SDS–PAGE separation and in-gel fluorescence visualization were performed as described in previous sections.

For isoTOP-ABPP studies, 10 µl of each of the E. coli lysates (2 mg protein/ml) overexpressing the designed constructs were mixed together and the total volume was brought to 1 ml by the addition of 2 mg ml<sup>−1</sup> of MCF7 soluble proteome. Time-dependent and concentration-dependent labelling with the IA probe, click chemistry, on-bead trypsin and TEV digestions, LC-MS runs and MS data analysis were performed as described in previous sections.

that spontaneously reactivates after the destruction of incorrect structural elements3.

By exploiting the unique attributes of the F-T luciferase substrate and the idiosyncrasies of the DnaK functional cycle (specifically, the need for DnaJ and ATP hydrolysis for high-affinity substrate binding and of GrpE for ADP and substrate release), the authors reach three major conclusions that had so far been out of experimental reach. First, substoichiometric amounts of DnaK are particularly effective at promoting F-T luciferase refolding, which implies that the binding of multiple DnaK molecules to a single misfolded protein is not always required for—and could in fact be detrimental to—substrate renaturation. Second, ATP- and DnaJ-mediated locking of DnaK onto F-T luciferase drives local unfolding, allowing at least some of the substrate molecules to reach a native conformation after GrpE-mediated release from the chaperone. Thus, in this context, DnaK is better classified as an unfoldase rather than a foldase. Finally, by capitalizing on their ability to conduct saturation kinetic experiments, the authors determine that one molecule of F-T luciferase can be converted to the native state at an ATP cost that is three orders of magnitude lower than that needed for the complete degradation and resynthesis of F-T luciferase.

The results of Sharma and coworkers suggest that a re-evaluation of the role of DnaK in cell physiology might be in order. As pointed out by the authors, if DnaK operates as such an energy-efficient unfoldase, it could play an important role in the quality control of the proteome of unstressed cells by inexpensively reactivating misfolded proteins whose structures are proximal to the native state. In essence, by clamping onto hydrophobic segments exposed to the solvent by an improperly folded substrate subdomain, the chaperone would melt the surrounding incorrect structure, allowing for proper subdomain refolding and repacking following release (Fig. 1b). This would be very much in line with the documented ability of DnaK (and DnaJ) to alter the conformation of native proteins, such as the heat shock sigma factor σ32, upon binding3. Whether the unfoldase activity and the one chaperone, one substrate paradigm are equally significant in stressed cells is more of an open question. DnaK’s collaboration with the ClpB-Hsp104 disaggregate, which, like AAA+ proteases, uses a motor-threading activity to unfold proteins, provides an interesting starting point for considering this issue. There is strong evidence that stress-aggregated proteins are remodeled by DnaK-DnaJ-GrpE before their transfer to the central channel of ClpB8, a process that is fully consistent with DnaK functioning as an unfoldase. However, the Hsp70 system is also believed to operate downstream of ClpB, helping largely unfolded proteins emerging from the ClpB pore reach a native conformation. Here, DnaK’s holdase function might dominate, allowing for the formation of folding nuclei next to the clamped position(s) while simultaneously reducing the concentration of aggregation-prone species. Considering the conformational diversity of folding intermediates, the two mechanisms are likely not exclusive and could explain why members of the Hsp70 family are so adept at performing a broad range of cellular functions.

François Baneyx and Brent L. Nannenga are in the Department of Chemical Engineering, University of Washington, Seattle.

e-mail: baneyx@uw.edu

References

Competing financial interests
The authors declare no competing financial interests.

MOLECULAR PROBES

Getting lucky in the lysosome

Conjugation of a known, mechanism-based glycosidase inhibitor to sensitive fluorophores yielded unexpectedly potent and selective probes for quantifying active lysosomal glucocerebrosidase. These conjugates could prove to be invaluable tools for diagnosing and studying Gaucher disease.

Ethan D Goddard-Borger, Tom Wennekes & Stephen G Withers

Gaucher disease, the most prevalent of the lysosomal storage disorders and the most common type of pediatric neurodegenerative disease, is caused by insufficient glucocerebrosidase (GBA) activity, a direct result of simple mutations in the encoding gene, GBA. As a consequence of this reduction in activity, there is a substantial accumulation of its substrate, glucosylceramide, in the lysosomes, which leads to the various pathologies of Gaucher disease. Diagnosis and treatment of Gaucher disease requires regular quantification of the active GBA in a patient’s tissues, not just the total GBA concentration or total β-glucosidase activity. To date, this has been accomplished by a cumbersome combination of techniques involving antibodies and low-pH activity assays on cell lysates. However, recent work from the laboratories of Overkleeft and Aerts is set to change this3. These authors report new, surprisingly potent activity-based probes (ABPs) that can selectively label active GBA, even in a whole-cell context, by covalent attachment of a fluorophore, enabling simple and sensitive quantification of this enzyme by SDS-PAGE (Fig. 1a).

The majority of Gaucher disease–causing mutations in GBA impair intracellular trafficking of the enzyme, rather than affecting its inherent activity. These mutations compromise the ability of the enzyme to obtain and retain its native fold within the endoplasmic reticulum (ER), and so most of the mutant enzyme undergoes ER-associated degradation (ERAD) instead of being trafficked to its intended destination, the lysosome. Interestingly, individuals with just 10–15% of normal GBA activity levels are asymptomatic; thus, if the flux of the GBA mutant to the lysosome can be increased by even a small amount, the Gaucher disease pathologies may be ameliorated. Enzyme enhancement therapy (EET)
news & views

seeks to do just this by using small-molecule chaperones, such as isofagomine, to bind and stabilize correctly folded mutant GBA within the ER, thereby elevating the steady-state concentration of folded enzyme within this organelle and increasing the amount of enzyme trafficked to lysosomes (Fig. 1b)²,³. A simple means to determine the amount of active GBA within cells would be of great value not only as a diagnostic tool but also in the development of such new therapies. How were these appropriately sensitive and selective ABPs for active GBA contrived? The epoxide-containing natural product cyclophellitol, a mechanism-based glycosidase inactivator that has previously been used to generate Gaucher disease models in mice⁴, was modified with an azide at a position corresponding to C6 of glucose. This compound was then conjugated to red and green borondipyrromethene (BODIPY) fluorophores using the ever-popular Cu(i)-catalyzed azide-alkyne cycloaddition ‘click’ reaction. Astonishingly, given the size and location of these modifications, the products proved to be much better inactivators of GBA than the parent compound. Fortuitously, these probes also reacted

Figure 1 | Getting a grip on GBA. (a) An activity-based probe (ABP) is able to selectively modify a target enzyme within a complex sample or living cell. Its structure generally consists of three parts: a tag, a linker and an active-site binding motif incorporating a reactive group (sometimes called a warhead). For these GBA ABPs, the tag is a BODIPY fluorophore (green) that is linked (black) to a cyclophellitol derivative (orange), which contains a reactive epoxide group (blue). The mechanism for hydrolysis of glucosylceramide by GBA involves a Glu340-bound covalent intermediate (left box). The same residue reacts with the epoxide of cyclophellitol to inactivate GBA (right box). (b) General overview of protein folding in the ER as affected by calnexin or calreticulin complexes and controlled by N-linked glycan processing. On the right, the EET concept is illustrated for GBA by the action of the small-molecule chaperone isofagomine (purple).
Cellulose squeezes through

The glucose-based polymer cellulose is of great biological and economical importance; however, little is known about how cellulose is synthesized. Now, structural estimates of one of the cellulose-synthesizing subunits in the bacterium Acetobacter xylinum help to explain the extrusion of the newly synthesized glucan chains.

Anne Endler, Clara Sánchez-Rodríguez & Staffan Persson

Cellulose, the major terrestrial biopolymer, consists of hydrogen-bonded, β-1,4-linked glucose units. Apart from having a considerable biological role, cellulose also has a substantial economic impact. For instance, cellulose provides the raw material for textile and paper products, and it is anticipated to become a key precursor for glucose-derived ethanol. The bulk of cellulose is obtained from plant resources that use photosynthetically derived nucleotide sugar UDP-glucose (UDP-Glc) to synthesize cellulose. Nevertheless, many nonphotosynthetic organisms, including bacterial, fungal and animal species, also produce cellulose. Among these, the common vinegar bacterium Acetobacter xylinum, or Gluconacetobacter xylinus, converts UDP-Glc into cellulose. However, the means by which A. xylinum and other bacteria synthesize cellulose is not fully understood. Hu et al. have now resolved the structure of one of the subunits responsible for cellulose synthesis in A. xylinum, and they propose a model for how the glucose-based chains are extruded from these Gram-negative (double-membranated) bacteria.

At least four genes are involved in the production of the cellulose polymers in A. xylinum. These genes, acxesA, acxesB, acxesC and acxesD, are arranged in an operon and encode proteins that perform different roles during the production of cellulose (Fig. 1a). AxCESA and AxCESB are involved in catalyzing and regulating the polymerization of the individual glucan chains and reside in the cytoplasmic membrane. AxCESA contains amino acid motifs suggestive of processive β-glycosyltransferases and can bind UDP-Glc, which is the substrate for the cellulose glucan chains. The AxCESB subunit reversibly binds large quantities of cyclic diguanylic acid (c-di-GMP), which activates AxCESA. The two remaining subunits, AxCESC and AxCESD have been proposed to mediate glucan chain extrusion and crystallization during cellulose assembly. However, no detailed mechanism for how these two components work has been put forward. Together, these four proteins, and perhaps other components, form cellulose-producing complexes, also referred to as terminal complexes. These complexes can be viewed as pores, or crater-like structures, at the cell surface, with an approximate 150 Å diameter, as determined by electron microscopy. Terminal complexes are arranged in linear arrays along the longitudinal axis of the bacteria. These complexes extrude the nascent glucan chains (Fig. 1b), which aggregate to form twisted subelementary fibrils and later crystallize into ribbons (Fig. 1c). This

References


Competition financial interests

The authors declare no competing financial interests.

© 2010 Nature America, Inc. All rights reserved.
Ultrasensitive in situ visualization of active glucocerebrosidase molecules

Martin D Witte1, Wouter W Kallemijn2,2, Jan Aten3, Kah-Yee Li1, Anneke Strijland2, Wilma E Donker-Koopman2, Adrianus M C H van den Nieuwendijk1, Boris Bleijlevens2, Gertjan Kramer2,4, Bogdan I Florea1, Berend Hooibrink5, Carla E M Hollak6, Roelof Ottenhoff2, Rolf G Boot2, Gijsbert A van der Marel1, Herman S Overkleeft1* & Johannes M F G Aerts2*

Deficiency of glucocerebrosidase (GBA) underlies Gaucher disease, a common lysosomal storage disorder. Carriership for Gaucher disease has recently been identified as major risk for parkinsonism. Presently, no method exists to visualize active GBA molecules in situ. We here report the design, synthesis and application of two fluorescent activity-based probes allowing highly specific labeling of active GBA molecules in vitro and in cultured cells and mice in vivo. Detection of in vitro labeled recombinant GBA on slab gels after electrophoresis is in the low attomolar range. Using cell or tissue lysates, we obtained exclusive labeling of GBA molecules. We present evidence from fluorescence-activated cell sorting analysis, fluorescence microscopy and pulse-chase experiments of highly efficient labeling of GBA molecules in intact cells as well as tissues of mice. In addition, we illustrate the use of the fluorescent probes to study inhibitors and tentative chaperones in living cells.

The lysosomal hydrolase GBA hydrolyzes glucosylceramide1,2. This ubiquitously expressed enzyme is initially synthesized as a 519-residue protein that cotranslationally acquires four N-linked glycans. After the removal of its signal peptide, GBA undergoes no further post-translational proteolytic modification and does not acquire mannose-6-phosphate moieties in the Golgi apparatus. The expression of disease in individuals with a defective GBA is remarkably heterogeneous. Substantial deficiency results in Gaucher disease, and recently carrihership has been recognized as major risk for parkinsonism2,3. The manifestation of Gaucher disease is highly variable, ranging from the common non-neuronopathic (type 1) variant to more severe manifestations with lethal neurological complications (type 2 and 3 variants) and extreme cases with abnormalities in skin permeability (so-called collodion babies)2. The marked phenotypic heterogeneity is only partly explained by differences in underlining mutations in the GBA gene. The heteroallelic presence of N370S GBA, the most frequent mutation in Caucasian individuals, protects against a neuronopathic manifestation, whereas homozygosity for L444P GBA is associated with severe neurological symptoms2. Several studies have indicated that the relationship between GBA genotype and Gaucher phenotype is not very strict4. Even phenotypic heterogeneity among identical twins has been reported, suggesting that additional factors influence the in situ residual activity of GBA5.

Two treatments for Gaucher disease presently exist: enzyme replacement therapy and substrate reduction therapy. Enzyme replacement therapy is based on chronic intravenous administration of recombinant GBA (imiglucerase; trade name: Cerezyme)6,7. Substrate reduction therapy is based on chronic oral administration of N-butyldeoxynojirimycin, an inhibitor of the enzyme glucosylceramide synthase, which catalyzes the formation of glucosylceramide8,9. More recently, an alternative approach has received considerable attention, so-called chaperone therapy. Common in Gaucher patients are mutant forms of GBA that show impaired folding and retention in the endoplasmic reticulum, ultimately resulting in elimination via the ubiquitin-proteasome system, a process known as ER-associated degradation (ERAD)10,11. Studies have investigated small compounds, designated ‘chemical chaperones’, that are able to increase the amount of GBA by stabilizing and/or promoting folding of the enzyme. One extensively studied example is isofagomine (I), which is a potent competitive inhibitor interacting with the catalytic pocket16,17. Beneficial effects on the amount and lysosomal localization of mutant GBA forms in cultured cells have been reported for isofagomine, but the assays used to demonstrate increased degradative capacity have been quite artificial: cells are exposed to high concentrations of fluorogenic substrate at acidic pH (see, for example, ref. 17). It is not likely that exposing cells to low pH and millimolar concentrations of 4-methylumbelliferyl β-D-glucopyranoside for a prolonged period reflects faithfully the in situ ability of the enzyme to degrade glucosylceramide. Pharmacologic chaperones like isofagomine will only exert a positive clinical effect at a particular dose range: their concentration should be sufficiently high to promote folding of the enzyme in the endoplasmic reticulum to increase transport to lysosomes, whereas the concentration in lysosomes should also be sufficiently low to prevent marked inhibition of catalytic activity. The present lack of a suitable method for specific visualization of active GBA molecules is a major limitation in research on Gaucher disease and parkinsonism, as well as the development of new therapies. For this reason, we embarked on the development of such a method using activity-based labeling. The catalytic mechanism of GBA, a retaining glucosidase, has been elucidated in detail11,12. Briefly, unprotonated Glu340 in GBA performs the initial nucleophilic attack on the substrate, forming a covalently linked enzyme-substrate intermediate.

1 Leiden Institute of Chemistry, Leiden University, Leiden, The Netherlands. 2 Department of Medical Biochemistry, Academic Medical Center, Amsterdam, The Netherlands. 3 Department of Pathology, Academic Medical Center, Amsterdam, The Netherlands. 4 Mass Spectrometry, Swammerdam Institute for Life Sciences, Amsterdam, The Netherlands. 5 Department of Cell Biology and Histology, Academic Medical Center, Amsterdam, The Netherlands. 6 Department of Endocrinology and Metabolism, Academic Medical Center, Amsterdam, The Netherlands. 7 These authors contributed equally to this work.

*e-mail: h.s.overkleeft@chem.leidenUniv.nl or j.m.aerts@amc.uva.nl
Epoxides like conduritol B epoxide (2, CBE) and cyclophellitol (3) (Fig. 1) form first a noncovalent inhibitor–enzyme complex that then reacts with the Glu340 carboxylate to form a covalent bond, thus acting as irreversible inhibitors. Cyclophellitol resembles more closely the structure of glucoside substrates and is the more potent irreversible inhibitor of the two.\(^2\) We capitalized on this by grafting boron dipyrromethene (BODIPY) fluorophores on to the cyclophellitol core (Fig. 1). We here demonstrate highly efficient labeling of GBA in situ by these probes and reveal their use in monitoring GBA activity in Gaucher fibroblasts.

**RESULTS**

**Design and synthesis of activity-based probes**

We synthesized cyclophellitol (ref. 24) and 8-deoxy-8-azidocyclophellitol (4, KY170) (Fig. 1; for synthesis see Supplementary Results) and tested their inhibitory properties toward recombinant GBA (Genzyme). Cyclophellitol and its azido analog KY170 were found to be far more potent inhibitors of GBA than CBE. Click ligating BODIPY moieties (MDW933 to KY170) gave fluorescent inhibitors MDW933 (5) and MDW941 (6) (Fig. 1; for synthesis, see Supplementary Results). Examination of the inhibitory properties revealed that MDW933 and MDW941 were comparably potent as irreversible inhibitors, being markedly superior to CBE and even surpassing cyclophellitol and KY170. The apparent half-maximal inhibitory concentration (IC\(_{50}\)) values of both fluorescent compounds (MDW933: IC\(_{50}\) = 1.24 ± 0.04 nM; MDW941: IC\(_{50}\) = 1.94 ± 0.08 nM) were very similar, being about 100- and 1,000-fold lower than those of cyclophellitol and KY170 (IC\(_{50}\) = 0.15 ± 0.009 μM and 0.12 ± 0.004 μM, respectively) and CBE (IC\(_{50}\) = 9.49 ± 0.042 μM). We determined next the inhibition constants—the K\(_i\) (the equilibrium constant for initial binding), the rate constant (k) and the relative rate constant k/K— for CBE, cyclophellitol, KY170, MDW933 and MDW941 using a continuous substrate release assay (see Supplementary Fig. 1 for progress curves). A general trend that we observed for the equilibrium constant for initial binding was that increased hydrophobicity resulted in decreased K\(_i\) values (Table 1). Comparison of relative rate constants demonstrated that the fluorescent probes inhibited GBA 22-, 34- and 4,300-fold better than KY170, cyclophellitol and CBE.

The affinity of MDW933 and MDW941 for GBA was unexpectedly high. For a better understanding of this finding, we performed molecular docking analysis using the GBA crystal structure (PDB: 2V3E)\(^25\). The docking model revealed that at minimum free energy (−5.2 kcal mol\(^{-1}\) and −5.4 kcal mol\(^{-1}\)), the cyclitol moiety did not significantly lower for the hydrophobic fluorescent probes MDW933 and MDW941 (−8.1 kcal mol\(^{-1}\) and −8.4 kcal mol\(^{-1}\)) than for KY170 (−5.2 kcal mol\(^{-1}\)), explaining their superior inhibitory properties. Notably, when we compared the crystal structure of GBA (PDB: 2V3T) with the modeled enzyme structure (PDB: 2V3E) with CBE or MDW933 (Supplementary Fig. 2), the cyclitol moiety did not completely overlap with that of CBE covalently bound to 2V3T. The intrinsic differences in structure coordinates between the crystal structures, as well as the comparison of the positioning of CBE and MDW933 before the prenucleophilic attack with the already covalently bound state of CBE, could have caused this discrepancy. The presence of the latter was likely enough to physically alter the local protein structure.

**In vitro labeling of GBA with the fluorescent ABPs**

To examine labeling of recombinant GBA by MDW933 and MDW941, we incubated the enzyme for 30 min at 37 °C with mixtures of both probes at pH 5.2, with 0.2% (w/v) taurocholate and 0.1% (v/v) Triton X-100, an optimal condition for enzymatic activity and activity-based labeling. After the incubation, we resolved the protein preparations with SDS-PAGE and analyzed the labeled proteins by fluorescence scanning of the slab gel on a Typhoon Variable Mode Imager (Fig. 2a). Labeled recombinant GBA migrated at the expected mass of 57 kDa. At equimolar concentration of MDW933 and MDW941 (100 nM), both probes bound the enzyme equally well. Thus, labeling of GBA with both probes was comparable, as expected given their similar inhibition constants. Boiling of the samples before electrophoresis had no impact on the detection of the fluorescently labeled protein on the slab gel, indicating that the probe was firmly attached. The presence of reducing agent also did not affect the covalent binding of the probes.

We determined the sensitivity of detection of labeled GBA by incubating 2 pmol GBA with an excess of MDW933 (20 nmol at 1 mM concentration) for 1 h at 37 °C and subsequent titration of the amount applied on the gel (Fig. 2b). We could detect as little as 20 attomol GBA by fluorescence scanning. Next, we incubated equal amounts of GBA (2 pmol) for 30 min with decreasing amounts of MDW933 and applied all protein to a gel. Incubation with as little as 20 attomol of probe resulted in detectable GBA on the slab gel (Fig. 2b). Apparently, nearly all of the probe had been covalently bound to recombinant GBA, consistent with its high affinity for binding. These experiments indicated that ultrasensitive detection of GBA was feasible on slab gels following in vitro labeling with the fluorescent probes 5 and 6.

As a next step, we analyzed the site of binding of the probe on GBA using a competition assay. Prior to labeling with MDW933, we incubated recombinant GBA with 2 mM CBE for 30 min (Fig. 2c). Preincubation with CBE, shown via crystallography to bind Glu340 (ref. 26), blocked labeling completely. Similarly, we also noticed competition with labeling by the competitive inhibitor AMP-DNM (7) (Fig. 2c)\(^27\). These results from competition experiments indicated that indeed the probe was bound in the catalytic center of GBA. We unambiguously identified the site of binding of KY170 and MDW933 by mass spectrometry. Using tryptic digestion and

---

**Table 1 | Binding constants of the inhibitors**

<table>
<thead>
<tr>
<th>Inhibitor</th>
<th>k (min(^{-1}))</th>
<th>K (μM)</th>
<th>k/K (μM(^{-1})·min(^{-1}))</th>
</tr>
</thead>
<tbody>
<tr>
<td>CBE</td>
<td>0.217 ± 0.026</td>
<td>53 ± 10.8</td>
<td>0.004</td>
</tr>
<tr>
<td>Cyclophellitol</td>
<td>0.078 ± 0.010</td>
<td>0.152 ± 0.026</td>
<td>0.514</td>
</tr>
<tr>
<td>KY170</td>
<td>0.035 ± 0.003</td>
<td>0.044 ± 0.007</td>
<td>0.794</td>
</tr>
<tr>
<td>MDW933</td>
<td>0.127 ± 0.024</td>
<td>0.077 ± 0.002</td>
<td>17.76</td>
</tr>
<tr>
<td>MDW941</td>
<td>0.208 ± 0.063</td>
<td>0.008 ± 0.003</td>
<td>25.10</td>
</tr>
</tbody>
</table>

k and K values were calculated as described in the Supplementary Results and Methods and reported with s.e.m.
Enzymatically active GBA molecules are a prerequisite for labeling with the probe, as demonstrated by the lack of labeling of GBA that had been denatured by boiling (Fig. 2c). The same conclusion could be drawn from the pH dependence of irreversible inhibition of GBA by the fluorescent probes. It exactly coincided with the pH profile of enzymatic activity toward 4-methylumbelliferyl β-D-glucopyranoside (Fig. 2d).

**Labeling of GBA in cell and tissue extracts**

To determine the labeling specificity of both fluorescent probes, we incubated homogenates of cultured cells and mouse tissues with 100 nM green fluorescent MDW933 for 30 min at 37 °C and analyzed the preparation with SDS-PAGE. In the case of homogenates of cultured RAW cells, fluorescence scanning showed exclusive labeling of GBA by MDW933. The various GBA forms, which endocytosis is blocked (Supplementary Fig. 6). Direct uptake of the probes, either by diffusion of the amphiphilic structures or by facilitated by transporters, seems most likely.

We studied *in situ* labeling of GBA in cells using fluorescence-activated cell sorting (FACS) analysis. We first preincubated cells in the absence or presence of CBE and subsequently incubated the cells with a subsaturating or an excess amount of the green fluorescent probe 5. FACs analysis revealed dose-dependent fluorescent labeling of cells and no labeling above background in CBE pretreated cells (Fig. 3b). These positive results prompted us to analyze labeling of the cells with fluorescence microscopy (Fig. 3c–f). For this purpose, we cultured fibroblasts for 2 h with 5 nM MDW941. We also detected GBA protein by indirect immunofluorescence using the specific anti-GBA monoclonal antibody 8E4 (ref. 29). Using multispectral image analysis, we could specifically distinguish the respective fluorescent emission spectra (Fig. 3d–e) from autofluorescence background (Fig. 3c). The intracellular pattern of labeling with MDW941 showed an almost complete overlap with the detection of GBA by monoclonal antibody 8E4 using this method (Fig. 3f).

As probes with a hydrophobic BODIPY moiety might nonspecifically be retained in membranes, in particular the plasma membrane, we studied this possibility more closely. First, we cultured fibroblasts obtained from a Gaucher patient homozygous for the RecNCI GBA mutation (a mutation resulting in premature degradation of GBA by

**Figure 2 | In vitro labeling of GBA with MDW933 and MDW941.** (a) Labeling of recombinant GBA (Cerezyme) with mixtures of MDW933 and MDW941 detected on slab gel. Left: labeling with the indicated amount of MDW933 in the presence of 100 nM MDW941. Right: labeling with the indicated amount of MDW941 in the presence of 100 nM MDW933. (b) Sensitivity of detection and labeling. Upper panel: Cerezyme (2 pmol) was incubated with an excess of MDW941 (1 mM) and dilutions applied on gel. Lower panel: Cerezyme (2 pmol) was incubated with a decreasing amount of MDW933. (c) Activity of Cerezyme (2 pmol) was blocked by incubating with CBE (2 mM) or AMP-DNM (2 mM) or by boiling in 1% (w/v) SDS before labeling with MDW933. (d) Effect of the pH on inhibition and enzymatic activity. The activity of GBA was determined at various pH values, normalized for the activity at pH 5.2 and plotted (circles). Inhibition by MDW933 was examined at the same pH ranges (open squares). Data represent mean values ± s.d. (e) Fluorescent labeling of GBA in homogenates of RAW cells using MDW933 (100 nM) (7.5% SDS-PAGE gel). Proteins were detected by fluorescence imaging (left lane) and Coomassie Brilliant Blue (CBB) staining (right lane). (f) Fluorescent labeling of mouse tissue lysates exposed to MDW933 (100 nM) (10% SDS-PAGE gel). Arrows indicate the molecular weight of Cerezyme. For uncut gels, see Supplementary Figure 11.
To show the versatility of the probes, we performed pulse-chase experiments using cultured cells. For this purpose, we incubated fibroblasts overnight with 10 nM red fluorescent MDW941. Subsequently, we treated the cells with 10 nM of the green fluorescent probe, harvested them at different time points (0–48 h) and subjected aliquots of cell homogenates to gel electrophoresis (see Fig. 3g for the lifecycle of GBA as visualized in this manner). It should be noted that GBA pulse labeled with the red fluorescent MDW941 disappeared gradually from the cells with an estimated half-life of about 30 h. The obtained half-life was consistent with the half-life determined previously using conventional pulse-chase labeling with radioactive methionine10. During the chase, GBA was increasingly labeled with MDW933, coinciding with formation of new GBA molecules (Fig. 3g).

We next studied the labeling of GBA by MDW941 in intact fibroblasts using time-lapse microscopy. Fibroblasts treated with the probe showed very rapid fluorescent labeling of lysosome-like structures (see Supplementary Video). With 5 nM MDW941, labeling reached a maximum within 15 min. Even after 100 h of exposure to the compound, cells did not show any signs of apoptosis or toxicity. Finally, we examined the possibility of labeling GBA in mice by intravenously administering 0.1 nmol green fluorescent MDW933 dissolved in phosphate-buffered saline to adult mice. As a control, matched mice received the buffer solution intravenously. After 2 h, we killed the animals, prepared tissue extracts, labeled them with excess red fluorescent MDW941 to visualize unlabeled GBA and subjected them to SDS-PAGE (Fig. 4 shows the outcome of a typical experiment). In most tissues—here we show lung and liver—MDW933 already labeled a considerable proportion of GBA (Fig. 4a). Consistently, in such tissues the probe also irreversibly inhibited a large proportion of GBA in the living mouse (see Fig. 4b). An exception in this respect was the brain (Fig. 4a), in which MDW933 apparently labeled almost no GBA in vivo and GBA was not inactivated. As observed earlier, intestinal fractions showed labeling of proteins of multiple molecular masses (Fig. 4a). In addition to GBA, MDW933 clearly labeled other proteins that occur in the intestine in the mouse.

Analysis of Gaucher materials

We investigated labeling of mutant GBA in fibroblasts from a normal individual and from Gaucher donors (a N370S GBA homozygote, a L444P homozygote and a RecNCI homozygote manifesting as colodion Gaucher and almost entirely lacking GBA protein) by treating cell lysates with MDW933 (10 nM) for 1 h and subjecting these to SDS-PAGE. A comparison of cells from a normal subject and from a Gaucher donor homozygous for L444P showed labeling of proteins of multiple molecular masses (Fig. 4a). In addition to GBA, MDW933 clearly labeled other proteins that could be visualized separately. For uncult gels see Supplementary Figure 12.

Figure 3 | In situ labeling of glucocerebrosidase. (a) Inactivation of GBA by MDW933 (squares) and MDW941 (circles) in situ. Fibroblasts incubated with the probes (5 nM) for the indicated time were homogenized, and residual activity was determined with 4-methylumbelliferyl β-D-glucopyranoside. Data represent mean values ± s.d. (b) FACS analysis. Cells were treated with 0 nM (red line), 2 nM (blue line) and 10 nM (green line) MDW933 for 300 min. Control cells were pretreated with CBE (0.3 mM) and incubated with 2 nM (brown line) and 10 nM (purple line) MDW933. (c–f) Representative spectral imaging micrographs of cells labeled with MDW941. (c) Autofluorescence (white). (d) MDW941 BODIPY fluorescence of GBA (red). (e) AlexaFluor488 fluorescence of GBA visualized with monoclonal Ab 8e4 (green). (f) Overlay of d and e. In all pictures, nuclei are stained with DAPI (blue). Scale bar represents 20 μm. (g) Pulse-chase experiment. Cells were incubated overnight with 10 nM MDW941 (pulse, upper panel) and then with 10 nM MDW933 for the indicated time (chase, middle panel). Lower panel: overlay of the pulse and the chase. Because of the use of low concentration of MDW933 during the first 8 h of the chase, incomplete labeling of newly formed GBA molecules was accomplished. For uncult gels see Supplementary Figure 12.

Figure 4 | Labeling of glucocerebrosidase in mice. Adult mice received intravenously 100 pmol green fluorescent MDW933 and were killed after 2 h. (a) Tissue lysates were incubated with excess (100 nM) red fluorescent MDW941 for 30 min to label unreacted GBA. In vivo labeled GBA (left panels) and in vitro labeled GBA (right panels) were visualized separately. For uncult gels, see Supplementary Figure 13. (b) Residual enzymatic activity in tissues of treated mice (black) was determined with 4-methylumbelliferyl β-D-glucopyranoside substrate and expressed as percentage of the matched control animal (gray). Data represent mean values ± s.d.
Impact of isofagomine on N370S GBA in cultured fibroblasts

It has been reported that for cells from N370S GBA homozygotes, prolonged incubation with isofagomine yields an increase in GBA activity\(^{11}\). The interaction of isofagomine with the catalytic pocket has been intensely studied, including at the level of crystals\(^{21}\). We therefore examined whether incubation of N370S GBA homozygous fibroblasts with isofagomine increased the amount of GBA that can be labeled with fluorescent probes. We cultured cells for 7 d with different concentrations of isofagomine (0, 10, 30 and 300 nM) and subsequently incubated them for 2 h with or without excess MDW941 in the presence of the original concentration of isofagomine. Determination of the activity of GBA in homogenates of cells not treated with MDW941 using 4-methylumbelliferyl β-D-glucopyranoside as substrate revealed a modest isofagomine dose–dependent increase in enzyme activity (see in Fig. 5c). Aliquots from the homogenates of cells labeled with MDW941 were subjected to gel electrophoresis, and the detected fluorescent GBA was quantified. Again we noted a modest dose-dependent increase (Fig. 5d), although less marked than the increase in in vitro enzyme activity in the homogenates. This discrepancy might be due to concomitant in situ stabilization of GBA by isofagomine in combination with competitive inhibition of enzymatic activity. We therefore investigated the effect of isofagomine on GBA activity in the intact cell using S′-pentfluorobenzoylamino-fluorescein-di-β-D-glucoside (FDG) as a substrate. Incubation of fibroblasts with various concentrations of isofagomine for 20 min, subsequent addition of FDG to the medium and quantification of the hydrolysis of FDG by FACS allowed determination of the IC\(_{50}\) value of isofagomine in intact cells\(^{22}\). The IC\(_{50}\) of isofagomine for hydrolysis of FDG was about 1 μM (Fig. 5e). Apparently, isofagomine at concentrations >1 mM completely inhibited activity of GBA in intact cells. We studied the reversibility of isofagomine competition for the fluorescent active site labeling. For this purpose, we preincubated recombinant GBA attached to monoclonal antibody BE4 immobilized to Sepharose beads for 15 min with increasing concentrations of isofagomine at pH 5.2 in the presence of taurocholate (0.2% w/v) and Triton X-100 (0.1% v/v). Prior to labeling with MDW933 (10 nM for 15 min), we either washed or did not wash the bead suspension with the same buffer. Quantification of the labeled GBA on slab gel indicated that the competition of isofagomine for the active site was fully reversible (Supplementary Fig. 10).

DISCUSSION

The need for a method allowing visualization of active GBA molecules in situ in living cells is evident. It is of importance to understand better what the precise cell and tissue distribution of active GBA molecules is because this may render a better understanding of the pathogenesis of Gaucher disease. Moreover, demonstration of a true increase in active GBA molecules by tentative chaperones is of interest. At present the detection of GBA still relies on the use of antibodies that do not distinguish between active and inactive GBA molecules and can not label enzyme in intact cells. Our search for a suitable probe for activity-based labeling of GBA in situ has yielded the desired result. As starting point for the development of such a probe, we selected cyclophellitol, a known potent irreversible inhibitor of GBA that forms a covalent adduct. Next we linked, via a spacer, we selected cyclophellitol, a known potent irreversible inhibitor of GBA, to a fluorescein label for cell labeling. For this purpose, we preincubated recombinant GBA attached to monoclonal antibody BE4 immobilized to Sepharose beads for 15 min with increasing concentrations of isofagomine at pH 5.2 in the presence of taurocholate (0.2% w/v) and Triton X-100 (0.1% v/v). Prior to labeling with MDW933 (10 nM for 15 min), we either washed or did not wash the bead suspension with the same buffer. Quantification of the labeled GBA on slab gel indicated that the competition of isofagomine for the active site was fully reversible (Supplementary Fig. 10).
be blocked with CBE or KY170 and potent competitive inhibitors such as hydrophobic deoxynojirimycines could compete away labeling. Labeling also required the folded enzyme and occurred proportional to enzymatic activity at different pH.

The affinity of both probes for GBA is in fact quite notable: truly ultrasensitive detection of GBA molecules was obtained. The high affinity of the fluorescent probes for GBA offered the opportunity to label the enzyme very specifically. With cell or tissue lysates, exclusive labeling of GBA molecules was observed following electrophoresis. In the case of homogenates of intestine alone or tissue protein29. Another favorable feature of these probes is their ability to enter various cellular compartments. It will be of interest to determine the precise mechanism(s) more closely, although it is already clear that cellular entry seems not to depend on endocytosis. The entry of the probes into living cells allows their use in FACS analysis, and one can perform pulse-chase experiments in intact cells as we have demonstrated. Time-lapse microscopy confirmed that GBA can be labeled very efficiently in intact fibroblasts. We obtained no indications that fluorescent labeling of GBA was toxic to the cells. Initial experiments also indicated that in mice cellular GBA can be labeled with fluorescent cyclophilin-based compounds. Notably, the brain showed a different picture. Almost no brain GBA was labeled in mice upon intravenous administration of MDW933. This may suggest that MDW933 does not pass the blood-brain barrier or is actively removed from the brain by some P-glycoprotein. The fluorescence features of the green and red fluorescent probes are intrinsically suboptimal for in situ imaging of labeled GBA in tissues or whole animals.

The potential applications for the activity-based fluorescent probes MDW933 and MDW941 are substantial. They offer an alternative to antibodies, which are species-specific and can not reach compartmental GBA in intact cells. Moreover, in contrast to antibodies, our probes uniquely label active GBA molecules. One tentative area of application of the fluorescent probes may be diagnosis of Gaucher disease, in particular the demonstration of low amounts of active GBA molecules in fibroblasts or blood cells of patients. This is helpful, as low amounts of active GBA molecules are usually associated with severe, neuropathic Gaucher disease. Another area of application for these fluorescent probes may be found in the analysis of compounds for their possible inhibitory or chaperone effects. As we demonstrated, the beneficial effect of isofagomine on N370S GBA in cultured fibroblasts could be confirmed with activity-based labeling. This finding is of importance as it implies that at an optimal concentration of isofagomine, occupation of the catalytic center by the competitive inhibitor is in situ sufficiently low to allow labeling by the fluorescent probe. In other words, at an appropriate concentration of isofagomine indeed increases GBA levels and intralysosomal enzymatic capacity. This finding for isofagomine is not entirely unexpected, as it has been proposed that at the low intralysosomal pH isofagomine interacts less well with β-glucosidases than at neutral pH in the endoplasmic reticulum34. Our observations render support for the approach of chaperone therapy, although the dosing of drugs in patients to reach optimal (steady-state) concentrations in various tissues may prove to be a major challenge.

Our approach of selective detection of GBA molecules using fluorescently labeled irreversible inhibitors allows unprecedented, ultrasensitive in vivo monitoring of active enzyme molecules. It can be envisioned that the same approach is also feasible for other glycosidases, and another challenging perspective is the future use of the fluorescent probes in living animals. Two approaches can be envisioned. In the first approach the fluorescent probes are used to report on local GBA activity. In the second strategy, recombinant GBA is labeled with the fluorescent probe so that after administration, trafficking of the construct can be monitored in a strategy that is related to another recently reported strategy that uses active site labeling of recombinant and purified GBA with a radiotag35. In conclusion, the reported fluorescent activity-based probes offer very versatile research tools to visualize active GBA, ultrasensitively and specifically. This accomplishment may be not only relevant for Gaucher disease but also for Parkinsonism.

METHODS

See Supplementary Methods for the synthesis of the probes, the methods used to determine the binding constants, molecular docking studies, time-lapse microscopy and mass spectrometric analysis of GBA labeled with KY170 and MDW933.

General methods. Chemicals were obtained from Sigma-Aldrich if not otherwise indicated. Recombinant GBA was obtained from Genzyme. Monoclonal anti-human GBA antibody 8E4 was produced from hybridoma cells as described earlier. Gaucher patients were diagnosed on the basis of reduced GBA activity and demonstration of an abnormal genotype37. Fibroblasts were obtained with consent from donors. Cell lines were cultured in HAMF12 DMEM medium (Invitrogen) supplied with 10% (v/v) FBS.

Enzyme activity assays. Activity of GBA was measured at 37 °C with 4-methylumbelliferyl β-D-glucopyranoside as substrate as reported previously. To determine the IC_{50}, the inhibitors were preincubated for 30 min with the enzyme before addition of the substrate mixture. The incubation mixture contained 3 mM fluorogenic substrate, 0.2% (w/v) sodium taurocholate and 0.1% (v/v) Triton X-100 in 150 mM McIlvaine buffer, pH 5.2. After stopping the incubation with excess NaOH (final pH 10.3), we measured fluorescence with a fluorometer LS 5 (Perkin Elmer) using λ_{ex} 366 nm and λ_{em} 445 nm. Activity of lactase was quantified by measuring liberated glucose from lactose in vivo. Activity of GBA in cells was measured using FDG as substrate and FACS29.

Gel electrophoresis and fluorescence scanning. Electrophoresis in sodium dodecylsulfate containing either 7.5% or 10% polyacrylamide gels was performed as earlier described. Wet slab gels were scanned on fluorescence using the Typhoon Variable Mode Imageer (Amersham Biosciences) using λ_{ex} 488 nm and λ_{em} 520 nm (bandpass 40) for green fluorescence MDW933 and λ_{ex} 532 nm and λ_{em} 610 nm (30 nm) for red fluorescence MDW941.

Fluorescence microscopy and multispectral imaging. Fibroblasts were cultured on glass slides. Cells were incubated with MDW941 (5 nM) or control probe MDW1065 (5 nM) for 2 h. Next, cells were washed, fixed with 3% (v/v) parafomaldehyde in PBS for 15 min, washed and incubated first with 0.05% (w/v) saponine for 15 min, next with 0.1 mM NH_{4}Cl in PBS for 10 min and then with 3% (v/v) bovine serum albumin in PBS for 1 h. Next, the slides were incubated with anti-GBA monoclonal antibody 8E4 (1:500). Bound mouse monoclonal antibody was visualized with a secondary antibody conjugated with AlexaFluor488. Nuclei were stained with DAPI. Cells were examined using epifluorescence microscopy (Leica DM500B) with an HCX PL APO ×63 1.40–0.60 oil immersion objective. Filter blocks used were A4 (360/40 nm band pass excitation, 400 nm dichroic mirror, 470/40 nm band pass suppression) for DAPI, L5 (480/40 nm band pass excitation, 505 nm dichroic mirror, 527/30 nm band pass suppression) for AlexaFluor488 and N2.1 (515–560 nm band pass excitation, 580 nm dichroic mirror, 590 nm long pass suppression) for MDW941 and MDW1065. Analysis was performed with multispectral imaging using a Nuance N-MSI-20 camera with Nuance 2.10 software (Cambridge Research & Instrumentation). Data sets were acquired at 440–500 nm for A4, 500–580 nm for L5, and 580–720 nm for N2.1 filter blocks, each at 10 nm intervals. In each experiment, nonlabeled control cells were imaged to define the autofluorescence spectral library. Spectral libraries for DAPI, AlexaFluor488 and MDW941 or MDW1065, each obtained from single-stained cells, were used to unmix the triple staining patterns into the individual components and separate these from autofluorescence. Nuance software was used to construct composite images.

Fluorescence-activated cell sorting. Fibroblasts were cultured in the presence or absence of 0.3 mM CBE overnight. Next, cells were incubated with MDW933 (2 and 10 nM, for 300 min). Cells were suspended by trypsinization and analyzed by FACS using FACS Vantage (B.D. Bioscience), with the above described settings. Data sets were acquired at 440–500 nm for A4, 500–580 nm for L5, and 580–720 nm for N2.1 filter blocks, each at 10 nm intervals. In each experiment, nonlabeled control cells were imaged to define the autofluorescence spectral library. Spectral libraries for DAPI, AlexaFluor488 and MDW941 or MDW1065, each obtained from single-stained cells, were used to unmix the triple staining patterns into the individual components and separate these from autofluorescence. Nuance software was used to construct composite images.

Pulse-chase experiments. Fibroblasts were cultured overnight with MDW941 (10 nM), after which they were extensively washed with PBS and incubated with MDW933 (10 nM). Cells were harvested at different time points; homogenates were prepared and subjected to SDS-PAGE. GBA labeled with MDW933 and with MDW941 were separately visualized using the Typhoon Variable Mode Imageer with the above described settings.

NATURE CHEMICAL BIOLOGY DOI: 10.1038/nCHeMBIO.466
Labeling of GBA in live mice. Experimental procedures were all approved by the appropriate ethics committee for animal experiments. C57Bl/6j mice were obtained from Harlan and fed ad libitum commercially available lab diet (RHM-K:Hope Farms). Two Npc1 BALB/c WT (+/+ ) mice were injected intravenously via tail vein, using a restrainer with 100 μl PBS or 100 μl 100 nM MDW933 dissolved in PBS. After 2 h of administration the mice were anesthetized with FFM mix (1 ml of fentanylcitrate, 1 ml of midazalam and 2 ml of distilled water), and blood, urine and organs were collected and directly frozen in liquid nitrogen. Homogenates were made in 25 mM potassium phosphate buffer, pH 6.5, supplemented with 0.1% (v/v) Triton X-100 and labeled with MDW941 (100 nM). Homogenates were analyzed as described above.

Chaperone experiment using isofagomine. A cell line homozygous for N370S was cultured in HAMF12-DMEM medium (Invitrogen) supplied with 10% (v/v) potassium phosphate buffer, pH 6.5, supplemented with 0.1% (v/v) Triton X-100 and labeled with MDW941 (100 nM). Homogenates were described as analyzed above.

Received 10 May 2010; accepted 5 October 2010; published online 31 October 2010

References

Acknowledgments
Funding from The Netherlands Organization for Scientific Research (NWO-CW, to M.D.W., V.W.K., R.G.B., G.v.d.M., H.S.O. and J.M.F.G.A.) and The Netherlands Proteomics Centre (to B.J.F. and H.S.O.) is acknowledged.

Competing financial interests
The authors declare no competing financial interests.

Additional information
Supplementary information and chemical compound information is available online at http://www.nature.com/nchembio. Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/. Correspondence and requests for materials should be addressed to J.M.F.G.A. or R.G.B.