Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry

Yuen Ho*, Albrecht Gruhler*, Adrian Heilbut*, Gary D. Bader†‡, Lynda Moore*, Sally-Lin Adams*, Anna Millar*, Paul Taylor*, Keiryn Bennett*, Kelly Boutilier*, Lingyun Yang*, Cheryl Wolting*, Ian Donaldson*, Søren Schandorff*, Juanita Shewnarane*, Mai Vo*†, Joanne Taggart*†, Marilyn Goudreault*†, Brenda Muskat*, Cris Alfarano*, Danielle Dewar†, Zhen Lin†, Katerina Michalickova†‡, Andrew R. Willems†§, Holly Sassi†, Peter A. Nielsen*, Karina J. Rasmussen*, Jens R. Andersen*, Lene E. Johansen*, Lykke H. Hansen*, Hans Jespersen*, Alexandre Podtelejnikov*, Eva Nielsen*, Janne Crawford*, Vibeke Poulsen*, Birgitte D. Sørensen*, Jesper Matthiesen*, Ronald C. Hendrickson*, Frank Gleeson*, Tony Pawson†§, Michael F. Moran*, Daniel Durocher†§, Matthias Mann*, Christopher W. V. Hogue*†‡, Daniel Figeys* & Mike Tyers†§

** MDS Proteomics, 251 Attwell Drive, Toronto, Canada M9W 7H4, and Staermosegaardsvej 6, DK-5230 Odense M, Denmark*
† Programme in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto, Canada, M5G 1X5
‡ Department of Biochemistry, University of Toronto, 1 Kings College Circle, Toronto, Canada M5S 1A8
§ Department of Medical Genetics and Microbiology, University of Toronto, 1 Kings College Circle, Toronto, Canada M5S 1A8

The recent abundance of genome sequence data has brought an urgent need for systematic proteomics to decipher the encoded protein networks that dictate cellular function[1]. To date, generation of large-scale protein–protein interaction maps has relied on the yeast two-hybrid system, which detects binary interactions through activation of reporter gene expression[2–4]. With the advent of ultrasensitive mass spectrometric protein identification methods, it is feasible to identify directly protein complexes on a proteome-wide scale[5,6]. Here we report, using the budding yeast *Saccharomyces cerevisiae* as a test case, an example of this approach, which we term high-throughput mass spectrometric protein complex identification (HMS-PCI). Beginning with 10% of predicted yeast proteins as baits, we detected 3,617 associated proteins covering 25% of the yeast proteome. Numerous protein complexes were identified, including many new interactions in various signalling pathways and in the DNA damage response. Comparison of the HMS-PCI data set with interactions reported in the literature revealed an average threefold higher success rate in detection of known complexes compared with large-scale two-hybrid studies[3,4]. Given the high degree of connectivity observed in this study, even partial HMS-PCI coverage of complex proteomes, including that of humans, should allow comprehensive identification of cellular networks.

To survey the yeast proteome, we chose an initial set of 725 bait proteins representative of a variety of different functional classes, including 100 protein kinases, 36 phosphatases and regulatory subunits, and 86 proteins implicated in the DNA damage response (DDR). A small scale, one-step immuno-affinity purification based on the Flag epitope tag was used to capture bait proteins, which were transiently overexpressed from the heterologous *GAL1* or *tet* promoters. Proteins from 1,558 individual immunoprecipitations were resolved by SDS–polyacrylamide gel electrophoresis (PAGE), visualized by colloidal Coomassie stain, excised from the gel and subjected to tryptic digestion before mass spectrometric analysis (Supplementary Information Fig. 1). As our isolation procedure often yielded multiple proteins from single excised bands, which cannot be resolved by peptide-mass-fingerprinting alone, we used tandem mass spectrometry (MS/MS) fragmentation to identify unambiguously proteins in each gel slice[6]. A total of 15,683 gel slices were processed, yielding approximately 940,000 MS/MS spectra that matched sequences in the protein sequence database. Over 35,000 protein identifications were made, corresponding to 8,118 potential interactions with a set of 600 bait proteins that were expressed at detectable levels (Supplementary Information Table 1). Ubiquitous, nonspecifically binding proteins, defined empirically on the basis of frequency of occurrence (see Supplementary Information), were subtracted from the raw data set to yield 3,617 interactions with 493 baits—for further discussion of filtering criteria and mass spectrometry methodology, see Supplementary Information. This filtered data set contained 1,578 different interacting proteins representing 25% of the yeast proteome (Supplementary Information Table 2). In a preliminary direct validation of the HMS-PCI data set, 64 out of 86 interactions (74%) in a random set of new associations detected by HMS-PCI were recapitulated in immunoprecipitation-immunoblot experiments (data not shown). The HMS-PCI method was able to identify known complexes from a variety of subcellular compartments, including the cytoplasm, cytoskeleton, nucleus, nucleolus, plasma membrane, mitochondrion and vacuole (see Supplementary Information). Of all the proteins identified, 531 corresponded to hypothetical uncharacterized proteins predicted from the yeast genome sequence (Supplementary Information Table 3).

To begin to assess cellular signalling events on a proteome-wide level, we used most of the protein kinases and phosphatases encoded in the yeast genome to capture associated components. As an example, HMS-PCI analysis of the mitogen-activated protein kinase (MAPK) Kss1 identified many known components of the mating/filamentous growth pathway[7], including Ste11, Ste7, and four known downstream targets, the transcriptional regulators, Ste12, Tec1, Dig1/Rst1 and Dig2/Rst2 (Fig. 1a). We identified other Kss1 interactions of potential biological significance, including Bem3, which is a GTPase-activating protein that may help attenuate the upstream Cdc42 Rho-type GTPase needed for mating. Relevant interactions were also detected between the cell wall integrity MAPK Slt2 and its upstream activators Bck1 and Mkk2, and between the osmotic stress response MAPK Hog1 and Rck2, a downstream target kinase[7] (Supplementary Information Table 2).

Numerous kinases and phosphatases formed discernable complexes, often with regulatory factors that serve to localize or control activity[1]. An extensive network was assembled around Cdc28, the primary cyclin-dependent kinase (CDK) for cell division control[8], including interactions with its known cyclin partners Cln1, Cln2, Clb2, Clb3 and Clb5, as well as with the small CDK-binding subunit Cks1 (Fig. 1b). In many cases the detected interactions are bridged by intermediary partners; for example crystallographic evidence indicates that Cks1–cyclin interactions occur through the CDK

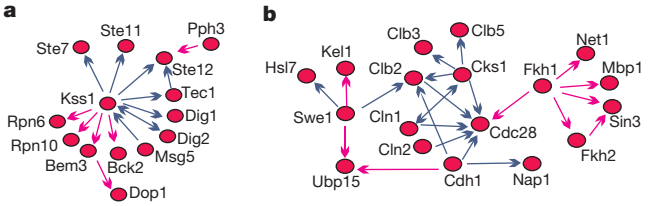

Figure 1 Two kinase-based signalling networks. **a**, Interaction diagram for Kss1 complexes. **b**, Interaction diagram for Cdc28 and Fkh1/2 complexes. Arrows point from the bait protein to the interaction partner. Blue arrows indicate known interactions; red arrows indicate new interactions. Not all detected interactions with all components are shown, nor are connections necessarily direct (see Supplementary Information Table 2).

subunit[8]. However, in the absence of additional evidence, there is no a priori means to elucidate connectivity of the interactions, and so we have chosen to represent each interaction as a direct pair (see below). The dual-specificity kinase Swe1, which inhibits Cdc28, was associated with both Clb2 and Hsl7, a known negative regulator of Swe1 (ref. 9). An interaction between Swe1 and Kel1, a protein that is involved in cell fusion and cell polarity[10], might reflect the link between polarized growth and Swe1 activity. Cdc28 also associated with Fkh1, one of the transcription factors that drives expression of the mitotic cyclins *CLB1/2* and other G2/M-regulated genes, thereby providing direct physical closure of the known Clb1/2-Cdc28 positive feedback loop[11]. Other Fkh1/2 interactions were consistent with roles in transcriptional activation and repression. Another local network was detected around the mitotic exit network (MEN) kinases Cdc5, Cdc15, Dbf2 and Dbf20, including the cohesin complex recently implicated as a Cdc5 target for activation of sister chromatid separation[12]. Finally, numerous interactions between protein phosphatases, regulatory subunits and substrates were found (Supplementary Table S2). For example, the dual specificity phosphatase Msg5 was associated with both Fus3 and Slt2, consistent with its genetic role in attenuating MAPK signalling[7].

The DDR includes DNA repair processes and checkpoint pathways that dictate cell cycle progression, transcription, protein degradation and DNA repair itself[13]. The global DDR network revealed by HMS-PCI contained many known interactions as well as many new interactions of probable biological significance (Fig. 2a; see also Supplementary Information Table 2). Most of the interactions identified did not depend on treatment with exogenous DNA-damaging agents, perhaps reflecting the fact that low level DNA damage normally occurs during replication[13]. Examples of known interactions include: the replication factor C complex (RFC, Rfc1–5) and the RFC^{Rad24} subcomplex, as well as the PCNA-like (PCNAL) Mec3–Rad17–Ddc1 complex, both of which transduce DNA damage signals; part of the Mms2–Ubc13–Rad18 post-replicative repair (PRR) complex; and the Mre11–Rad50–Xrs2 (MRX) complex that mediates double-strand-break repair by homologous and non-homologous mechanisms[13]. We also recovered nearly all known nucleotide excision repair (NER) factors in their dedicated subcomplexes[14]: Rad1–Rad10–Rad14 (NEF1); Rad3–TFB3–Kin28–Ccl1 (NEF3/TFIIH); and Rad7–Rad16 (NEF4). The Rad4–Rad23 interaction (NEF2) was not found, but we nevertheless detected an association between Rad4 and NEF1, a known interaction among NER factors.

The comprehensive coverage of DDR proteins readily identified pathway and network connections. For example, we recovered Rfc4 in Ddc1 complexes, consistent with the hypothesis that the PCNAL complex might be loaded onto DNA by the RFC^{Rad24}complex[15]. In terms of new interactions, the HMS-PCI approach revealed that Met18 can associate with Rad3, a component of the transcription factor TFIIH complex needed for both RNA Pol II-dependent transcription and NER. An association between Rad23 and the ubiquitin chain assembly factor Ufd2 (ref. 16) is corroborated by genetic interactions that suggest that *RAD23* and *UFD2* act antagonistically[17]. Ufd2 also interacted with a second ubiquitin-like (UBL) domain-containing protein, Dsk2. Another NER component, Rad7, bound the yeast elongin C homologue, Elc1, for which a function remains to be assigned. In mammalian cells, elongin C associates with elongin B, the cullin Cul2, the RING-H2 domain protein Rbx1 and substrate recruitment factors called SOCS box proteins to form E3 ubiquitin ligase complexes[18]. Consistent with the Elc1–Rad7 interaction, sequence alignments revealed a divergent SOCS box motif in Rad7, suggesting that it may mediate substrate ubiquitination during excision repair (Fig. 2b, c).

The Rad53 kinase, which corresponds to Chk2 in mammals and Cds1 in fission yeast, is a conserved mediator of DDR signals[13]. In addition to the known Rad53 interaction with Asf1 (refs 19, 20), HMS-PCI revealed several new, associated proteins of probable biological significance, including the protein phosphatase 2C (PP2C) enzyme Ptc2, which is genetically implicated as a negative regulator of Rad53 (ref. 21). Furthermore, the uncharacterized gene

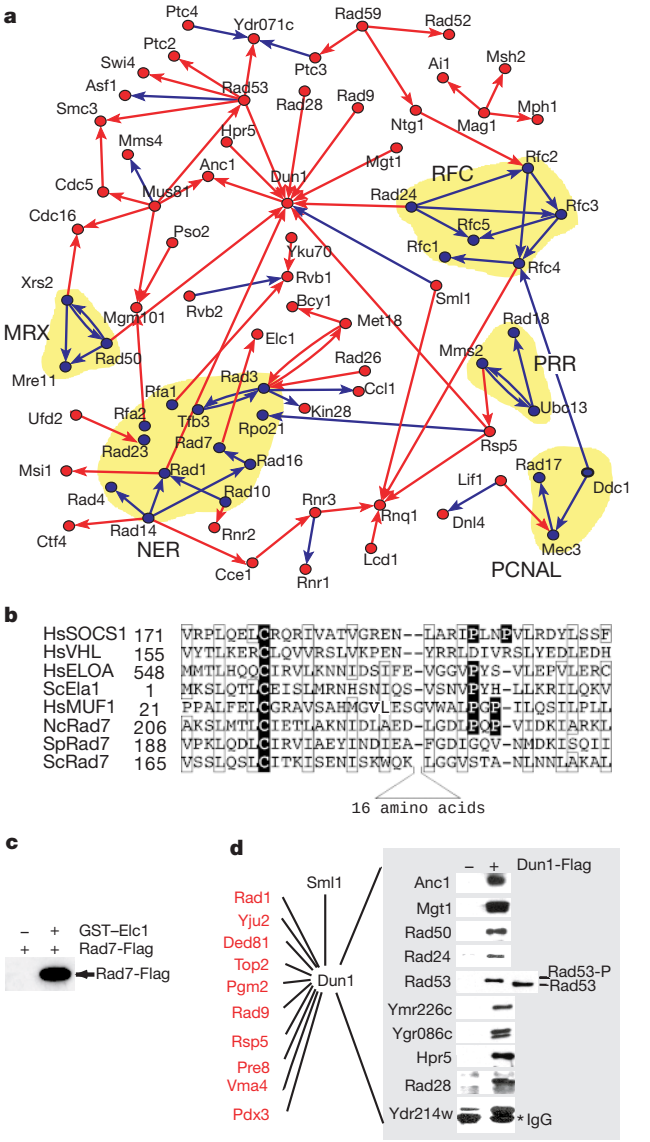

Figure 2 The DNA damage response network. **a**, Global connections in the network represented by a subset of known and new interactions. Interactions were initially nucleated from 86 proteins implicated in the DDR. Blue nodes indicate known interactions within dedicated complexes, highlighted in yellow. Blue arrows indicate known interactions; red arrows indicate new interactions. Not all detected interactions with all components are shown (see Supplementary Information Table 2), nor are connections necessarily direct. **b**, Alignment of SOCS box motif in Rad7 with known SOCS box proteins, including human MUF1, which like Rad7 contains carboxy-terminal leucine-rich repeats. Conserved hydrophobic residues are boxed; other diagnostic residues are shaded. Hs, *Homo sapiens*; Sc, *Saccharomyces cerevisiae*; Nc, *Neurospora crassa*; Sp, *Schizosaccharomyces pombe*. **c**, Rad7 interacts with Elc1. Flag-tagged Rad7 and glutathione *S*-transferase (GST)–Elc1 or GST alone were expressed in yeast, followed by capture on glutathione resin and detection of the Flag epitope. **d**, Complete set of Dun1 interactions detected by HMS-PCI. Co-immunoprecipitation tests of a subset of new Dun1 interactions are shown on the right; untested interactions are shown on the left. Note that the phosphorylated form of Rad53 interacts specifically with Dun1. The colour scheme is the same as in **a**. IgG, immunoglobulin-γ. Heavy chain; asterisk.

product Ydr071c was detected with both Rad53 and the PP2C family members Ptc3 and Ptc4, suggesting that Ydr071c may be a DDR-specific regulatory factor of PP2C-type phosphatases. We found consistently a genetic interaction between *YDR071C* and *RAD53* (R. Woolstencroft and D.D., unpublished data). With regard to Rad53 substrates[13], the putative targets Swi4 and Cdc5 were linked directly or indirectly to Rad53 by HMS-PCI.

The Dun1 protein kinase has a similar overall structure to Rad53, most notably through the presence of phosphothreonine-binding modules termed forkhead-associated (FHA) domains[22]. Dun1 is implicated in many aspects of the DDR, yet the identities of its upstream regulators and downstream effectors remain largely unknown. HMS-PCI analysis of Dun1 revealed potential upstream regulators Rad9, Rad53, Rad24, Hpr5/Srs2 and Rad50 (Fig. 2d). Dun1 interacted with a probable substrate, the ribonucleotide reductase inhibitor Sml1, which is degraded on *DUN1*-dependent phosphorylation[23]. Interestingly, Dun1 also associated with Rsp5, an E3 ubiquitin ligase that targets the RNA polymerase II large subunit Rpo21 for ubiquitin-mediated degradation after DNA damage[24]. Rsp5 is thus a candidate E3 ubiquitin ligase for Sml1. Of the interactions tested by co-immunoprecipitation, 10 out of the 10 that involved Dun1 were confirmed. Of particular significance are the interactions between Dun1 with Rad24, Rad50, Rad53, Hpr5/Srs2, Rad28 and Mgt1, as all of these proteins are known to participate in the DDR. We note that Dun1 was able to retrieve selectively the hyperphosphorylated form of Rad53, probably through the FHA domain of Dun1 (Fig. 2d). This observation suggests that the DNA damage signal is transmitted directly from Rad53 to Dun1. Finally, Dun1 also interacts with Ymr226c and Ygr086c, two proteins of unknown function that are induced on general cell stress, perhaps indicating that Dun1 may affect processes other than DNA damage.

We compared the HMS-PCI data set, represented as hypothetical direct interactions between bait and associated proteins, with comprehensive high-throughput yeast two-hybrid (HTP-Y2H) data sets[3,4] using interactions reported in the literature as a benchmark. Interaction data sets were entered into the Biomolecular Interaction Network Database (BIND)—a standardized repository for all forms of biological interaction data, including protein–protein interactions[25]. To systematically compile a set of published interactions, we used a search engine called PreBIND (J. Martin *et al.*, in preparation), which is a support vector machine and natural language-processing-based algorithm designed to identify abstracts that describe protein–protein interactions. Beginning with all 600 bait proteins used in this study, PreBIND was used to collect a non-exhaustive set of 697 protein interactions from the literature. The PreBIND data was combined with 545 interactions derived from the MIPS protein interaction table[26] to create a literature-based set of 1,003 interactions that involve the HMS-PCI bait set (see Supplementary Information). When compared against this literature benchmark, the HMS-PCI data set contained 2.6- to 3.4-fold more literature-derived interactions per bait than each large-scale HTP-Y2H data set, and 1.9-fold more interactions when compared with the combination of both comprehensive HTP-Y2H data sets (Fig. 3a, b and Table 1)[3,4]. In addition to published interactions, a number of new interactions were shared by the HMS-PCI and HTP-Y2H data sets (Fig. 3c). Functional annotation of the HMS-PCI data set indicated that 275 of the detected complexes contained two or more interaction partners within the same gene ontology (GO) biological process (see Supplementary Information). Finally, we note that the connectivity distribution of the HMS-PCI data set follows a power law, as observed for other large-scale biological networks (see Supplementary Information Fig. 3 for a Pajek representation of the data set)[27].

Proteome-wide analysis of native protein complexes by highly sensitive mass spectrometric methods allows the detection of complex cellular networks that might otherwise elude more focused approaches (G.D.B. *et al.*, manuscript in preparation). Given that approximately 40% of yeast proteins are conserved through eukaryotic evolution[28], the global yeast protein interaction map will provide a partial framework for understanding more complex proteomes. Imminent technical advances, such as gel-free analysis

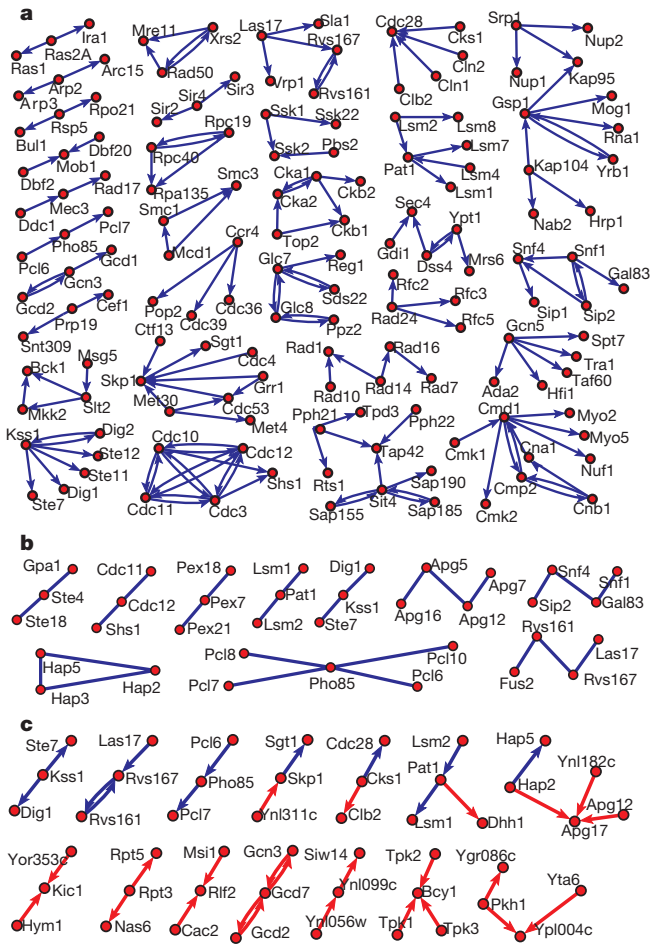

Figure 3 Comparison of large-scale protein interaction networks with interactions reported in the literature. **a**, Overlap of HMS-PCI data set and PreBIND + MIPS data set. **b**, Overlap of a comprehensive HTP-Y2H data set[4] and PreBIND + MIPS data set. **c**, Overlap of HMS-PCI and the HTP-Y2H data set[4]. Blue edges are literature-derived interactions; red edges are new interactions detected by HTP approaches. For clarity, single binary interactions are not shown. The following number of single interactions were removed from the indicated panel: **a**, 37; **b**, 23; **c**, 31.

Table 1 Literature-derived interactions in HMS-PCI and HTP-Y2H interaction data sets

	Literature data set*				
HTP data set	PreBIND	MIPS	PreBIND+MIPS	MIPS two-hybrid	MIPS biochemical
HMS-PCI	113	119	166	55	81
Uetz	42	53	63	37	22
Ito full	37	39	49	27	18
Ito core	25	26	32	19	9
Ito full+Uetz	60	71	86	47	37

To address possible methodological bias in the literature benchmark, the MIPS data set was sorted into two-hybrid interactions (MIPS two hybrid) and interactions based on biochemical purification (MIPS biochemical). As these two sets are of roughly equal size, the MIPS benchmark is impartial in this aspect. Two large-scale HTP-Y2H data sets were used (Uetz, P. *et al.* (ref. 3) and Ito, T. *et al.* (ref. 4)). 'Ito full' refers to the complete set of 4549 interactions; 'Ito core' refers to the core set of the 841 most reproducible interactions, as defined by ref. 4.
*Number of interactions reported in the literature. See Supplementary Information for details on interaction curation and comparisons.

of protein complexes, higher sensitivity mass spectrometers, systematic analysis of post-translational modifications, and protein microarrays, will undoubtedly extend the reach of the approach described here[6,29]. As the set of proteins nominally encoded by the human genome is approximately 5-fold greater than the total number of yeast proteins, comprehensive analysis of the human proteome is feasible with current technology. □

Methods

Recombination-based cloning, yeast culture and isolation of protein complexes were carried out using standard methods (see Supplementary Information for full details of methods). Cultures of strain BY4742 (*MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0 pep4 Δ :: KANR*) bearing bait plasmids were induced by either addition of galactose or doxycyclin before collection at mid-logarithmic phase (A_{600} = 1.2–1.5). After immunoprecipitation of bait complexes, protein bands were visualized by colloidal Coomassie stain, excised from polyacrylamide gels, reduced and *S*-alkylated, then subjected to trypsin hydrolysis[30]. To achieve high throughput, we constructed an automated proteomics network of mass spectrometers based on nano high-performance liquid chromatography (HPLC)-electrospray ionization-MS/MS capable of continuous operation. Liquid chromatography tandem mass spectrometry (LC-MS/MS) analysis was performed on a Finnigan LCQ Deca ion trap mass spectrometer (Thermo Finnigan) fitted with a Nanospray source (MDS Proteomics). Chromatographic separation was conducted using a Famos autosampler and an Ultimate gradient system (LC Packings) over Zorbax SB-C18 reverse phase resin (Agilent) packed into 75 μM ID PicoFrit columns (New Objective). Protein identifications were made using the commercially available search engines Mascot (Matrix Sciences), Sonar (ProteoMetrics), Sequest (ThermoFinnigan) and PepSea (MDS Proteomics). Both the raw and filtered data sets generated in this study are available at http://www.mdsp.com/yeast. The filtered data set has been deposited in BIND[25] and can be viewed at http://www.bind.ca/.

Received 27 September; accepted 5 December 2001.

1. Pawson, T. & Nash, P. Protein–protein interactions define specificity in signal transduction. *Genes Dev.* **14,** 1027–1047 (2000).
2. Fields, S. & Song, O. A novel genetic system to detect protein–protein interactions. *Nature* **340,** 245–246 (1989).
3. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403,** 623–627 (2000).
4. Ito, T. *et al.* A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA* **98,** 4569–4574 (2001).
5. Neubauer, G. *et al.* Identification of the proteins of the yeast U1 small nuclear ribonucleoprotein complex by mass spectrometry. *Proc. Natl Acad. Sci. USA* **94,** 385–390 (1997).
6. Mann, M., Hendrickson, R. C. & Pandey, A. Analysis of proteins and proteomes by mass spectrometry. *Annu. Rev. Biochem.* **10,** 437–473 (2001).
7. Gustin, M. C., Albertyn, J., Alexander, M. & Davenport, K. MAP kinase pathways in the yeast *Saccharomyces cerevisiae*. *Microbiol. Mol. Biol. Rev.* **62,** 1264–1300 (1998).
8. Morgan, D. O. Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu. Rev. Cell. Dev. Biol.* **13,** 261–291 (1997).
9. McMillan, J. N. *et al.* The morphogenesis checkpoint in *Saccharomyces cerevisiae*: cell cycle control of Swe1p degradation by Hsl1p and Hsl7p. *Mol. Cell. Biol.* **19,** 6929–6939 (1999).
10. Philips, J. & Herskowitz, I. Identification of Kel1p, a kelch domain-containing protein involved in cell fusion and morphology in *Saccharomyces cerevisiae*. *J. Cell. Biol.* **143,** 375–389 (1998).
11. Jorgensen, P. & Tyers, M. The fork'ed path to mitosis. *Genome Biol.* **1** 1022.1–1022.4 (2000).
12. Alexandru, G., Uhlmann, F., Mechtler, K., Poupart, M. & Nasmyth, K. Phosphorylation of the cohesin subunit Scc1 by Polo/Cdc5 kinase regulates sister chromatid separation in yeast. *Cell* **105,** 459–472 (2001).
13. Zhou, B. B. & Elledge, S. J. The DNA damage response: putting checkpoints in perspective. *Nature* **408,** 433–439 (2000).
14. Prakash, S. & Prakash, L. Nucleotide excision repair in yeast. *Mutat. Res.* **451,** 13–24 (2000).
15. Thelen, M. P., Venclovas, C. & Fidelis, K. A sliding clamp model for the Rad1 family of cell cycle checkpoint proteins. *Cell* **96,** 769–770 (1999).
16. Koegl, M. *et al.* A novel ubiquitination factor, E4, is involved in multiubiquitin chain assembly. *Cell* **96,** 635–644 (1999).
17. Ortolan, T. G. *et al.* The DNA repair protein Rad23 is a negative regulator of multi-ubiquitin chain assembly. *Nature Cell Biol.* **2,** 601–608 (2000).
18. Tyers, M. & Rottapel, R. VHL: a very hip ligase. *Proc. Natl Acad. Sci. USA* **96,** 12230–12232 (1999).
19. Emili, A., Schieltz, D. M., Yates, J. R. & Hartwell, L. H. Dynamic interaction of DNA damage checkpoint protein Rad53 with chromatin assembly factor Asf1. *Mol. Cell* **7,** 13–20 (2001).
20. Hu, F., Alcasabas, A. A. & Elledge, S. J. Asf1 links Rad53 to control of chromatin assembly. *Genes Dev.* **15,** 1061–1066 (2001).
21. Marsolier, M. C., Roussel, P., Leroy, C. & Mann, C. Involvement of the PP2C-like phosphatase Ptc2p in the DNA checkpoint pathways of *Saccharomyces cerevisiae*. *Genetics* **154,** 1523–1532 (2000).
22. Durocher, D., Henckel, J., Fersht, A. R. & Jackson, S. P. The FHA domain is a modular phosphopeptide recognition motif. *Mol. Cell* **4,** 387–394 (1999).
23. Zhao, X., Chabes, A., Domkin, V., Thelander, L. & Rothstein, R. The ribonucleotide reductase inhibitor Sml1 is a new target of the Mec1/Rad53 kinase cascade during growth and in response to DNA damage. *EMBO J.* **20,** 3544–3553 (2001).
24. Beaudenon, S. L., Huacani, M. R., Wang, G., McDonnell, D. P. & Huibregtse, J. M. Rsp5 ubiquitin-protein ligase mediates DNA damage-induced degradation of the large subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* **19,** 6972–6979 (1999).
25. Bader, G. *et al.* BIND—The biomolecular interaction network database. *Nucleic Acids Res.* **29,** 242–245 (2001).
26. Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **28,** 37–40 (2000).
27. Jeong, H., Tombor, B., Albert, R., Oltvai, Z. N. & Barabasi, A. L. The large-scale organization of metabolic networks. *Nature* **407,** 651–654 (2000).
28. Chervitz, S. A. *et al.* Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* **282,** 2022–2028 (1998).
29. Zhu, H. *et al.* Global analysis of protein activities using proteome chips. *Science* **293,** 2101–2105 (2001).
30. Wilm, M. *et al.* Femtomole sequencing of proteins from polyacrylamide gels by nano-electrospray mass spectrometry. *Nature* **379,** 466–469 (1996).

Supplementary Information accompanies the paper on *Nature*'s website (http://www.nature.com).

Acknowledgements

We thank J. Chen, B. Kuehl, H. Li, V. Lay, B. Tuekam, S. Zhang, M. Patel, P. O'Donnell, I. Dutschek, U. Friedrich, M. Hansen, J. Brønd, H. Lieu, R. Woolstencroft, L. Harrington, F. Sicheri, A. Breitkreutz, C. Boone, B. Andrews and T. Hughes for discussions and/or technical assistance. This work was supported in part by grants from the Canadian Institutes of Health Research (CIHR), the Ontario Research and Development Challenge Fund and MDS-Sciex to T.P., D.D., C.H. and M.T. T.P. is a Distinguished Scientist of the CIHR; M.F.M. is a CIHR Scientist; D.D. is a Canada Research Chair in Proteomics, Bioinformatics and Functional Genomics and a Hitchings-Elion fellow of the Burroughs-Wellcome Fund; and M.T. is a Canada Research Chair in Biochemistry.

Competing interests statement

The authors declare competing financial interests: details accompany the paper on *Nature*'s website (http://www.nature.com).

Correspondence and requests for materials should be addressed to D.F. (e-mail: dfigeys@mdsp.com).

Alternative nucleotide incision repair pathway for oxidative DNA damage

Alexander A. Ischenko*† & Murat K. Saparbaev*

** Groupe "Réparation de l'ADN", UMR 8532 CNRS, LBPA-ENS Cachan, Institut Gustave Roussy, 94805 Villejuif Cedex, France*
† Novosibirsk Institute of Bioorganic Chemistry, Siberian Division of the Russian Academy of Sciences, 630090 Russia

The DNA glycosylase pathway[1], which requires the sequential action of two enzymes for the incision of DNA[2], presents a serious problem for the efficient repair of oxidative DNA damage, because it generates genotoxic intermediates such as abasic sites and/or blocking 3′-end groups that must be eliminated by additional steps before DNA repair synthesis can be initiated. Besides the logistical problems, biological evidence hints at the existence of an alternative repair pathway. Mutants of *Escherichia coli*[3] and mice (ref. 4 and M. Takao *et al.*, personal communication) that are deficient in DNA glycosylases that remove oxidized bases are not sensitive to reactive oxygen species, and the *E. coli* triple mutant *nei, nth, fpg* is more radioresistant than the wild-type strain[5]. Here we show that Nfo-like endonucleases nick DNA on the 5′ side of various oxidatively damaged bases, generating 3′-hydroxyl and 5′-phosphate termini. Nfo-like endonucleases function next to each of the modified bases that we tested, including 5,6-dihydrothymine, 5,6-dihydrouracil, 5-hydroxyuracil and 2,6-diamino-4-hydroxy-5-*N*-methylformamidopyrimidine residues. The 3′-hydroxyl terminus provides the proper end for DNA repair synthesis; the dangling damaged nucleotide on the 5′ side is then a good substrate for human flap-structure endonuclease[6] and for DNA polymerase I of *E. coli*.

Because *E. coli nfo*$^-$ and *xth*$^-$ strains[7,8] are extremely sensitive to oxidizing agents, we investigated whether the known abasic (AP) endonucleases, namely Nfo, Xth and human APEX/Ref-1/Ape1/