indirect immunofluorescence using an anti-cytochrome *c* antibody (65972A, PharMingen, 1:1,000), or by cell fractionation into cytosol and membrane components and immunoblotting (cytochrome *c* antibody 7H8.2C12, PharMingen, 1:1,000 dilution).

1. Chin, L., Merlino, G. & DePinho, R. A. Malignant melanoma: modern black plague and genetic black box. *Genes Dev.* **12**, 3467–3481 (1998).
2. Serrone, L. & Hersey, P. The chemoresistance of human malignant melanoma: an update. *Melanoma Res.* **9**, 51–58 (1999).
3. Soengas, M. S. *et al.* Apaf-1 and caspase-9 in p53-dependent apoptosis and tumor inhibition. *Science* **284**, 156–159 (1999).
4. Schmitt, C. A. & Lowe, S. W. Apoptosis and therapy. *J. Pathol.* **187**, 127–137 (1999).
5. Lowe, S. W., Ruley, H. E., Jacks, T. & Housman, D. E. p53-dependent apoptosis modulates the cytotoxicity of anticancer agents. *Cell* **74**, 957–967 (1993).
6. Hakem, R. *et al.* Differential requirement for caspase 9 in apoptotic pathways *in vivo. Cell* **94**, 339–352 (1998).
7. Fearnhead, H. O. *et al.* Oncogene-dependent apoptosis is mediated by caspase-9. *Proc. Natl Acad. Sci. USA* **95**, 13664–13669 (1998).
8. Baylin, S. B. & Herman, J. G. DNA hypermethylation in tumorigenesis: epigenetics joins genetics. *Trends Genet.* **16**, 168–174 (2000).
9. Robertson, K. D. & Jones, P. A. DNA methylation: past, present and future directions. *Carcinogenesis* **21**, 461–467 (2000).
10. Bird, A. P. & Wolffe, A. P. Methylation-induced repression—belts, braces, and chromatin. *Cell* **99**, 451–454 (1999).
11. Hark, A. T. *et al.* CTCF mediates methylation-sensitive enhancer-blocking activity at the H19/Igf2 locus. *Nature* **405**, 486–489 (2000).
12. Bell, A. C. & Felsenfeld, G. Methylation of a CTCF-dependent boundary controls imprinted expression of the Igf2 gene. *Nature* **405**, 482–485 (2000).
13. Lowe, S. W. & Ruley, H. E. Stabilization of the p53 tumor suppressor is induced by adenovirus 5 E1A and accompanies apoptosis. *Genes Dev.* **7**, 535–545 (1993).
14. Green, D. R. & Reed, J. C. Mitochondria and apoptosis. *Science* **281**, 1309–1312 (1998).
15. Rodriguez, J. & Lazebnik, Y. Caspase-9 and APAF-1 form an active holoenzyme. *Genes Dev.* **13**, 3179–3184 (1999).
16. Zou, H., Li, Y., Liu, X. & Wang, X. An APAF-1 cytochrome c multimeric complex is a functional apoptosome that activates procaspase-9. *J. Biol. Chem.* **274**, 11549–11556 (1999).
17. Hu, Y., Benedict, M. A., Ding, L. & Nunez, G. Role of cytochrome c and dATP/ATP hydrolysis in Apaf-1-mediated caspase- 9 activation and apoptosis. *EMBO J.* **18**, 3586–3595 (1999).
18. Teitz, T. *et al.* Caspase 8 is deleted or silenced preferentially in childhood neuroblastomas with amplification of MYCN. *Nature Med.* **6**, 529–535 (2000).
19. Yamamoto, H., Gil, J., Schwartz, S. Jr & Perucho, M. Frameshift mutations in Fas, Apaf-1, and Bcl-10 in gastro-intestinal cancer of the microsatellite mutator phenotype. *Cell Death Differ.* **7**, 238–239 (2000).
20. Schmitt, C. A., McCurrach, M. E., de Stanchina, E., Wallace-Brodeur, R. R. & Lowe, S. W. INK4a/ARF mutations accelerate lymphomagenesis and promote chemoresistance by disabling p53. *Genes Dev.* **13**, 2670–2677 (1999).
21. Chin, L. *et al.* Cooperative effects of INK4a and ras in melanoma susceptibility *in vivo. Genes Dev.* **11**, 2822–2834 (1997).
22. Walker, G. J. *et al.* Virtually 100% of melanoma cell lines harbor alterations at the DNA level within CDKN2A, CDKN2B, or one of their downstream targets. *Genes Chromosom. Cancer* **22**, 157–163 (1998).
23. Grossman, D., McNiff, J. M., Li, F. & Altieri, D. C. Expression and targeting of the apoptosis inhibitor, survivin, in human melanoma. *J. Invest. Dermatol.* **113**, 1076–1081 (1999).
24. Irmler, M. *et al.* Inhibition of death receptor signals by cellular FLIP. *Nature* **388**, 190–195 (1997).
25. Selzer, E. *et al.* Expression of Bcl-2 family members in human melanocytes, in melanoma metastases and in melanoma cell lines. *Melanoma Res.* **8**, 197–203 (1998).
26. Mora, J. *et al.* Clinical categories of neuroblastoma are associated with different patterns of loss of heterozygosity on chromosome Arm 1p. *J. Mol. Diag.* **2**, 1–9 (2000).
27. Woo, M. *et al.* Essential contribution of caspase 3/CPP32 to apoptosis and its associated nuclear changes. *Genes Dev.* **12**, 806–819 (1998).
28. Marcellus, R. C., Teodoro, J. G., Charbonneau, R., Shore, G. C. & Branton, P. E. Expression of p53 in Saos-2 osteosarcoma cells induces apoptosis which can be inhibited by Bcl-2 or the adenovirus E1B-55 kDa protein. *Cell Growth Differ.* **7**, 1643–1650 (1996).
29. Capodieci, P., Magi-Galluzzi, C., Moreira, G. Jr Zeheb, R. & Loda, M. Automated *in situ* hybridization: diagnostic and research applications. *Diagn. Mol. Pathol.* **7**, 69–75 (1998).

The protein–protein interaction map of *Helicobacter pylori*

Jean-Christophe Rain*, Luc Selig*, Hilde De Reuse†, Véronique Battaglia*, Céline Reverdy*, Stéphane Simon*, Gerlinde Lenzen*, Fabien Petel*, Jérôme Wojcik*, Vincent Schächter*, Y. Chemama*, Agnès Labigne† & Pierre Legrain*

* Hybrigenics SA, 180 avenue Daumesnil, Paris 75012, France
† Unité de Pathogénie Bactérienne des Muqueuses, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France
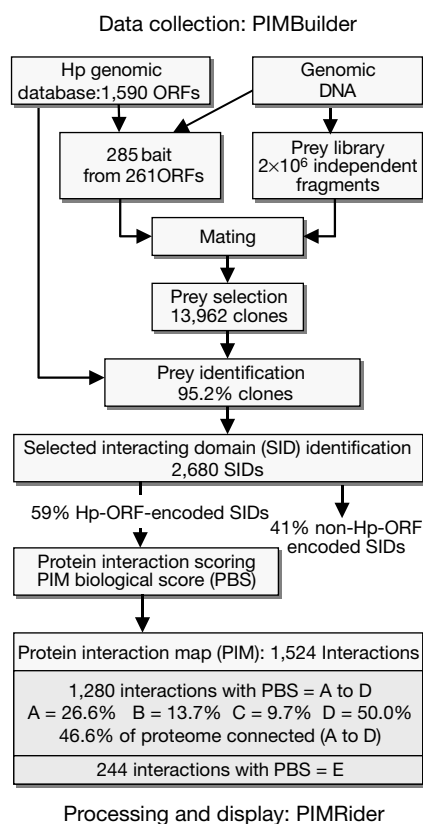
With the availability of complete DNA sequences for many prokaryotic and eukaryotic genomes, and soon for the human genome itself, it is important to develop reliable proteome-wide approaches for a better understanding of protein function[1]. As elementary constituents of cellular protein complexes and pathways, protein–protein interactions are key determinants of protein function. Here we have built a large-scale protein–protein interaction map of the human gastric pathogen *Helicobacter pylori*. We have used a high-throughput strategy of the yeast two-hybrid assay to screen 261 *H. pylori* proteins against a highly complex library of genome-encoded polypeptides[2]. Over 1,200 interactions were identified between *H. pylori* proteins, connecting 46.6% of the proteome. The determination of a reliability score for every single protein–protein interaction and the identification of the actual interacting domains permitted the assignment of unannotated proteins to biological pathways.

During the past few years, interaction maps have been proposed for viral[3–5] and eukaryotic (*Saccharomyces cerevisiae*[2,6,7] and *Caenorhabditis elegans*[8], for example) genomes. Here we describe the first procaryotic interaction map, for which the strategy used (Fig. 1) is also a variation of the two-hybrid assay[2]. It differs considerably by both the type of results generated and the throughput of the experimental procedures. We constructed a library of random genomic fragments of the *H. pylori* strain 26695 that had been previously sequenced[9]. A high complexity library was first obtained in *Escherichia coli* (over ten million clones). Ninety-seven per cent of the plasmids contained a single genomic insert (mean size 1,000 ± 550 nucleotides). This library was then introduced into yeast by transformation. Two million independent yeast colonies were collected, pooled and stored at −80 °C as equivalent aliquot fractions of the same library.

In parallel, a large set of bait plasmids was constructed in a bait vector designed to decrease the level of transcriptional auto-activation[2]. Bait constructs were specifically adapted for interaction screens, yielding 'validated baits'; for example, hydrophobic putative trans-membrane domains were discarded to avoid any non-nuclear localization of the bait protein in the yeast cell. In some cases, specific open reading frame (ORF) domains were selected for the bait design. For every single bait construct, we performed a preliminary small-scale screening experiment that determines the selective pressure (that is, modifying the selective medium, see Methods) to be applied to obtain a small number of independent positive clones per million interactions tested (usually less than 10). All positive colonies were then picked, and prey fragments were individually identified by sequence analysis and comparison with the genomic database through a dedicated integrated laboratory production management system (Fig. 1).

Protein–protein interaction maps are built on experimental data that ideally yield a heuristic value for each connection. Our procedure involves several steps of processing of raw two-hybrid results (Fig. 1). First, positive prey fragments are clustered into families of overlapping fragments. The common sequence shared by

these fragments is referred to as the selected interacting domain (SID). Second, SIDs that do not code for part of an *H. pylori* ORF are discarded. Third, for every remaining SID, a PIM biological score (PBS) is computed. The PBS is based on a statistical model of the competition for bait-binding between fragments. It is computed like a classical expected value (*E* value), and ranges from 0 (specific interaction) to 1 (probable artefact). For practical use, the scores were divided into four categories, from A (score very close to 0) to D (close to 1). A fifth category, E, was added to distinguish interactions involving only highly connected prey domains (SIDs found as prey with frequency greater than a fixed threshold). These are most probably two-hybrid artefacts. Although they may have some biological significance, they add little specific information to the interaction map. It should be emphasized that domains, rather than whole proteins, are tagged. For example, the carboxy-terminal region of protein HP0705/UvrA was found as a prey in 16 different interactions that were scored in the E category, but the same protein was selected through another domain as a specific interacting prey. Because global connectivity is taken into account, the PBS is

computed incrementally over the whole PIM and its discriminatory power increases as screening results accumulate.

We carried out 285 screens on 261 *H. pylori* bait ORFs chosen as follows: first, a core set of 50 proteins known to be involved in complexes and/or in pathogenicity was used to validate the approach; and second, 211 baits were then picked randomly, with a slight bias toward regions of the proteome that were still unexplored. Positive colonies (13,962) were selected (Table 1), and more than 95% of them were identified by sequencing the prey insert. From these prey fragments, 2,680 independent SIDs were defined, out of which 1,100 fell into non *H. pylori* ORF coding regions and were discarded (Fig. 1). Thirty-one SIDs were classified in the E category. The remaining SIDs define 1,280 interactions including 62 homo-oligomeric interactions. In total, 46.6% *H. pylori* proteins of the proteome were connected, which corresponds to an average connectivity of 3.36 partners per connected protein (without counting homodimeric connections). Only 14 screens out of 285 yielded no positive clones or interactions with nonsignificant score values, illustrating the fact that our technology reduces the rate of false negatives.

Given that little information is available on protein interactions in *H. pylori*, we used a data set of known interactions in *E. coli* to validate further our experimental strategy, and to evaluate the correlation between the PBS and the actual biological significance of interactions. For each *H. pylori* protein present in the interaction map, significant *E. coli* orthologues (FastA score < 0.01) were selected and their annotations in the SwissProt database (release 38)[10] were verified manually for known interactions shown by various biochemical means. The resulting *E. coli* interaction list was compared with interactions found in *H. pylori* (Fig. 2). Among these *E. coli* interacting pairs, 53% of the homodimers and 67% of the heterodimers were found. Among heterodimers, five out of six negative pairs were tested only in one direction, suggesting that performing the reciprocal screens would decrease this number. Most interactions that were described for orthologous proteins in *E. coli* fell into the high-scoring interaction category (A) according to our PBS calculation (7/10 homodimers and 9/12 heterodimers) confirming the heuristic value of the classification. The interaction map was also analysed according to the classification of *H. pylori* proteins into 14 functional categories previously proposed[9]. For 10 categories out of 14, more intra-category protein–protein interactions were observed than expected from a random theoretical distribution (*Z*-scores ranging from 2 to 50; in all cases, $P < 0.05$), suggesting the existence of a significant correlation between functional grouping and detection of interactions.

To display and analyse the interaction data, we developed a software platform composed of a database, a web-based graphical interface layer and various query and analysis tools (the PIMRider, Fig. 3; see also http://pim.hybrigenics.com). Starting from a gene name (or an ORF name), the PIMRider draws an automatic layout of the neighbourhood of this protein in the protein interaction map (Fig. 3a). Paths connecting two proteins can also be queried.

Figure 1 Outline of the strategy for building an *H. pylori* (Hp) proteome-wide interaction map. A production database (the PIMBuilder) was built that contains information related to the genomic sequence of *H. pylori*, which codes for 1,590 putative proteins or ORFs. It was populated with raw data from screening experiments. The PIMBuilder tracks all biotechnological or bioinformatics operations performed during the production processes, stores information about all biological objects produced during experiments, and interfaces with robots and bioinformatics modules. It also implements the procedure used to construct PIMs from raw experimental data. After identification of almost all positive clones, overlapping prey fragments were clustered into families to define SIDs. Those families that had no biological coding capability (antisense or intergenic region, out of frame fragments occurring in a single frame) were discarded. A PIM biological score (PBS; see Methods) was then calculated for *H. pylori* ORF-encoded SIDs. Interactions were grouped into categories A to D (from high to low heuristic values). The global connectivity of the PIM was also analysed to detect highly connected prey polypeptides. Those interactions were grouped in the E category. Processing of data and visualization of interactions were performed by an in-house bioinformatic platform (PIMRider).

Table 1 General features of the *H. pylori* interaction screens

Screening data	Total or mean value	Comments
Number of screens	285	261 different *H. pylori* ORFs
Combination of bait/prey polypeptides assayed	5.6×10^9	
Number of prey polypeptides assayed per bait	2×10^7	
Number of selected colonies per bait (1st reporter)	916	From 0 to 15,000
Number of selected clones per bait (2 reporters)	49	From 0 to 139
Number of selected colonies per million pairs tested	3.8	From 0 to 25
Number of selected clones	13,962	
Number of sequenced clones	13,296	95.2 %

Connections are displayed with their PBS scores, and can be filtered according to score categories. A graphical summary of information describing all interaction domains within a given protein can be displayed (Fig. 3b). Raw data on every interaction can be retrieved. Finally, the PIMRider supplies a description of each gene with functional and genomic information, and includes links to significant bibliographic references and to relevant external databases, such as PyloriGene (http://genolist.pasteur.fr/PyloriGene/). PyloriGene is a manually annotated database of the two *H. pylori* sequenced genomes[9,11], that integrates all publicly available information on genes and proteins and has been elaborated with a structure similar to that of the *Bacillus subtilis* SubtiList database[12].

Exploring the protein–protein interaction map reveals biological pathways and allows prediction of protein function. A first example concerns chemotaxis (see Fig. 3a). The *H. pylori* genome reveals three homologues of *E. coli* proteins that are involved in the chemotactic pathway (CheA, CheW and CheY) and proteins such as TlpA similar to chemotaxis receptors (MCPs). CheA was found to interact with CheY and CheW, and distinct interacting domains were precisely identified (Fig. 3b). The domain of CheA which binds to CheY precisely overlaps with the interacting domain assigned by a structural study in *E. coli*[13]. The TlpA-binding site for CheW was localized in a domain known, in *E. coli*, to be methylated and implicated in the transduction of the chemotactic signal.

The urease complex was also examined. Urease activity is essential for *H. pylori* pathogenicity and its synthesis requires two structural subunits, UreA and UreB, and the product of four accessory genes: *ureE, ureF, ureG, ureH*[14]. Complexes between accessory proteins and their role in nickel incorporation at the urease active site have been described for orthologues, but little information is available for *H. pylori* (for review, see ref. 15). The
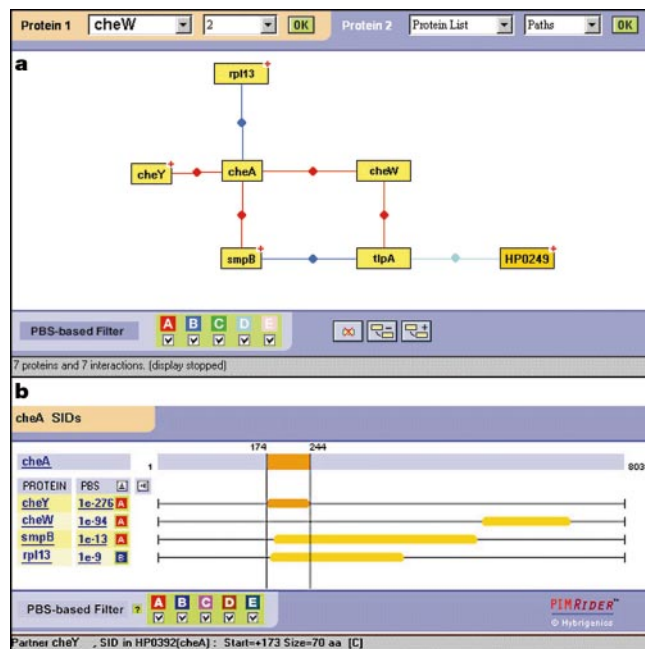
protein interaction map revealed the connection between UreA and UreB, and one of the two expected homo-oligomeric interactions of structural subunits (UreA); the UreB homodimer could not be detected. A connection between accessory proteins and the structural subunits occurs via UreH and UreA, which is consistent with the presumed chaperone role of UreH[15]. A new structural link was found between UreG and UreE. The UreF and UreH proteins were connected, but no connection between UreG and UreF or UreH was detected. In addition to the accessory proteins, the urease operon codes for an inner-membrane protein, UreI, essential for resistance to acidity[16] and recently described as a H^+-gated urea channel[17]. The third cytoplasmic domain of this protein reveals a potential interaction with the ExbD protein which is involved in transmission of PMF (proton motor force) energy to outer-membrane receptors.

Combination of genomic and proteomic data also permits function prediction. The *H. pylori* proteome contains a homologue of the *E. coli* HolB protein. In *E. coli,* this protein interacts with HolA to form part of the DNA polymerase core[18]. We found one high-scoring interaction between *H. pylori* HolB and an uncharacterized polypeptide, HP1247. A pairwise alignment between *E. coli* HolA and HP1247 highlighted structural homology (Fig. 4) not found by previous sequence analysis (Fig. 2). We thus assign the HolA function to the *H. pylori* HP1247 protein. The organization of bacterial genomes into operons suggests a functional relationship between the corresponding gene products that can be directly compared with our protein interaction map. Indeed, we found interactions between proteins that were likely to be expressed from a single operon (Table 2). Among these, we detected interactions between proteins known to interact in *H. pylori* (ScoA–ScoB) or in other organisms (RpsR–RpsF, MoaE–MoaD, FtsA–FtsZ), and between polypeptides involved in the same enzymatic activity and not yet described as interacting (HypE–HypF).

Selected interacting domains can also be analysed in terms of protein structure. The prokaryotic RNA polymerase, composed of a core enzyme ($\alpha_2\beta\beta'$) associated with a σ-factor, is one of the best studied multisubunit enzymes. In *H. pylori*, the β- and β'-subunits
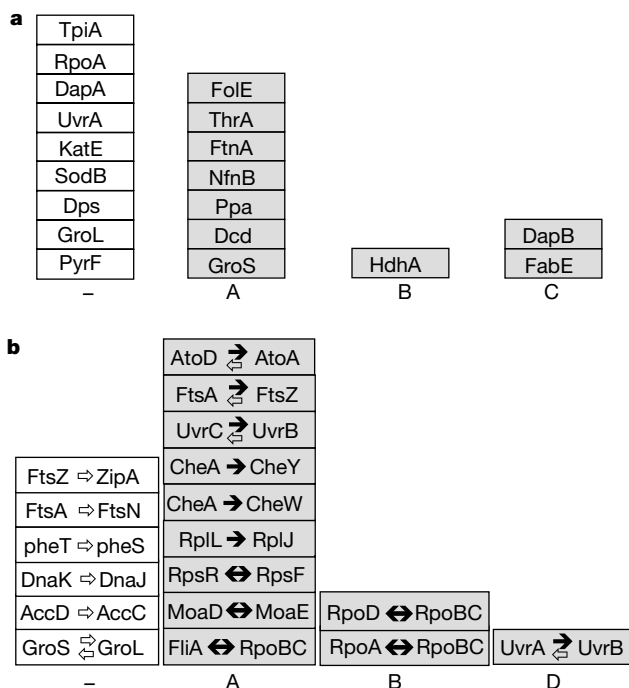
Figure 2 Sets of *E. coli* interaction data for which *H. pylori* orthologous proteins were identified and assayed in interaction screens. **a**, Homodimers; **b**, heterodimers. Names of *E. coli* proteins are boxed. Unidentified interactions between *H. pylori* orthologues are scored as '–' and shown in white boxes. Identified *H. pylori* interactions are indicated in grey boxes with their PBS score category (A, B, C or D). When several orthologues were identified, only the best scoring homologue was considered. For heterodimers, arrows indicate the direction of the screen (bait to prey) that was performed. The black (or white) colour of the arrow indicates a positive (or negative) result in the interaction screen.

Figure 3 PIMRider screen shots. **a**, The PIMViewer displays a portion of the protein interaction map around the CheA protein. Links between proteins identify connections with their colour-coded PBS score; **b**, the MultiSID Viewer exhibits the various interacting domains in the CheA protein (for details see http://pim.hybrigenics.com).

usually found in other bacteria are fused into a single polypeptide (RpoB). One of the two alternative σ-factors present in *H. pylori*, (HP1032), is similar to the *E. coli* FliA protein (sigma 28) necessary for transcription of genes involved in flagellar biosynthesis[19]. We identified a precise region of RpoB that interacts with the *H. pylori* FliA. This selected interacting domain (residues 841–959) maps exactly to a structural domain called the flexible flap[20]. The RpoB-interacting domain of FliA falls in the regions 3.2–4.2 of this σ-factor (residues 175–255). Biochemical studies suggest an interaction between the flap domain of RpoB and region 4 of sigma factors[20]. Our experiments thus characterize this interaction and support the role of the flap domain and σ-factors in the transition from an open complex to a processive elongation complex[21].

Our work provides a way to characterize proteome-wide protein interaction maps. Our results identify complexes that have been shown or postulated to exist in other organisms, such as in *E. coli*. They also complement sequence information about homologous proteins and operon prediction from the location of genes on the chromosome. Finally, identified interacting domains could be mapped on three-dimensional structures of proteins. As a whole they lead to the assignment of a functional role for many as yet uncharacterized proteins and provide tools, such as interacting domains, for further biological experiments. Our technology was designed to specifically address the main known causes for false-negative and false-positive results in two-hybrid assays. It is now clear that interactions are not necessarily detected positive by two-hybrid assays when used in reciprocal directions (Fig. 2; see also refs 4, 8, 22). Parallel screening against highly complex libraries of

fragments increases the number of 'two-hybrid-capable' candidates and reduces the rate of false negatives that arise with the classical two-hybrid matrix approach (that is, pair wise testing of a collection of proteins)[6,7]. Concerning the false positives, the specific design of selection procedures that permit a strong selectivity for all baits and the statistical analysis made possible by the experimental procedure (that is, reproducible exhaustive screening of fragment libraries) allows us to detect and tag nonspecific partners through a global scoring scheme.

Ultimate validation of biological significance should come from additional biological information, but comparison of our results with previously described interactions of *H. pylori* proteins and also of orthologous *E. coli* proteins supports the reliability of the approach. Protein interaction maps can be built at the scale of a proteome. Our technology is applicable to higher eukaryotes, for which highly complex random-primed complementary DNA libraries are screened for interacting domains. The identification of interacting domains is a direct consequence of the library screening approach and presents key advantages such as mapping of new functional domains or correlation between sequence similarity and functional homology. These interacting domains also constitute a first step towards the construction of dominant-negative mutants or the development of an assay for interaction modulation, applicable to new drug design. □

Methods

Bait cloning

Baits were constructed by PCR amplification and cloning in the pB6 plasmid derived from the original pAS2ΔΔ (ref. 2). Design of the primers was automatically proposed by a software and validated for each ORF. PCR fragments were cloned by classical enzymatic methods in a 96-well-plate format. All bait constructs used for interaction screens were fully sequenced and compared to the genome sequence; any mutant clone was discarded.

Library construction

We extracted genomic DNA from *H. pylori* 26695 as described[23], and nebulized and blunted it with a cocktail of mung bean nuclease, T4 and Klenow polymerase (NEB). Adapters containing *Sfi*1 sites were ligated to blunt DNA. The adapted DNA was cloned into pP6 plasmid derived from the original pACT2 and transformed in *E. coli* (DH10B; LifeTechnologies). Sequence analysis was performed on one hundred randomly chosen clones to establish the general characteristics of the library.

Screening procedure

The principle of the technique has been described[2]. Briefly, the screening conditions were adapted for each bait during a test screen, before performing the full-size screening experiment. The selectivity of the HIS3 reporter was modulated with 3-aminotriazole (3AT). For about 15% of the screens, diploid cells were plated on selective medium containing 3AT. In 44% of the screens, the second reporter (*lacZ*) was used directly on plates for the selection of clones. In other cases, the *lacZ* reporter was used as a second round of selection only for the selected clones. In all cases, LacZ activity was measured in a quantitative luminometric assay (Tropix).

Identification of interacting fragments

The prey fragments of the positive clones were amplified by PCR, analysed on agarose gel, partially sequenced at their 5′ junction on a PE 3700 Sequencer and mapped on the genomic sequence. Many clones were sequenced at their 3′ junction to map precisely the SID. All the steps after picking of positive clones were performed in bar-coded 96-well plates and automated with Beckman Biomek 2000 and Multimek automats. At the end of each screening experiment, the identity of the bait plasmid was controlled on a few positive clones.

PIM biological score

The PIM biological score (PBS) computation relies on two different levels of analysis: first, a local (that is, taking into account only the results of one screen) score is computed for each screen; and second, the global score is computed from the local scores by integrating results from all screens performed within the same genomic library. Local scores are thus computed only once, while global scores are recomputed each time new screens are performed. For each screen, fragments are clustered by overlap to delimit SIDs. Fragments that have no or very improbable coding capability (antisense, intergenic region, and out-of-frame fusion fragments selected in a single frame) are then eliminated from the set of prey fragments identified from positive clones. Assuming that prey fragments compete for the bait with 'equal chances', the probability p for a given fragment to be selected in an experiment is proportional to its expected number of occurrences within the library. p is
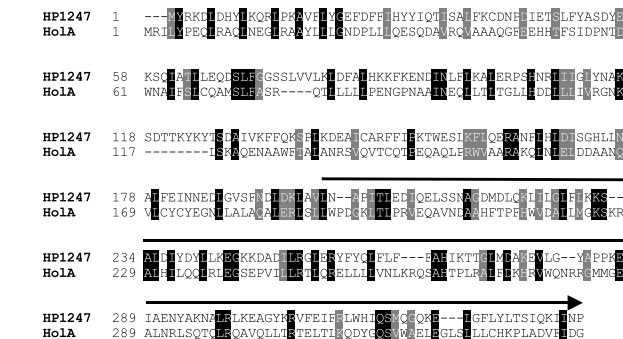

Figure 4 Alignment between *H. pylori* HP1247 protein and *E. coli* HolA. The two sequences were aligned using the FastA algorithm. Identical (black) and similar (grey) residues are outlined. The position of HP1247 C-terminal domain interacting with HolB is indicated by a line.

Table 2 Interacting proteins encoded by physically related genes

Protein 1		Protein 2	
ORF	Comments	ORF	Comments
HP0047†	HypE*, hydrogenase functionally related protein	HP0048	HypF*, hydrogenase putative transcriptional regulator
HP0061	–	HP0066	ATP-binding motif
HP0064	–	HP0063	–
HP0311	–	HP0312	ATP-binding motif
HP0338	–	HP0337	–
HP0691	ScoA, 3-oxoadipate coA-transferase subunit A	HP0692	ScoB, 3-oxoadipate coA-transferase subunit B
HP0697	–	HP0696	Putative hydantoin utilization*
HP0800	MoaE*, molybdopterin converting factor, subunit 2	HP0801	MoaD*, molybdopterin converting factor, subunit 1
HP0868	–	HP0869	HypA*, hydrogenase functionally related protein
HP0874	–	HP0875	KatA, catalase
HP0978	FtsA*, cell-division protein	HP0979	FtsZ*, cell-divison protein
HP1244	RpsR, ribosomal protein	HP1246	RpsF, ribosomal protein

* Function assigned by homology.
† The nomenclature used is that of PyloriGene database.

computed as a function of the fragment length and position, and of the length and position distributions of fragments in the prey library (these distributions are calibrated using data from random sequencing).

The local score is the probability for a given SID to be obtained under the equal chance hypothesis, that is, as a result of random noise. It is deduced by combining probabilities p (using a binomial law) from each of the independent fragment defining it. A (global) PBS is computed for each protein interaction after pooling results from all screens. On the basis of an independence hypothesis, scores from different screens are combined together when the same protein domain pair is involved. The resulting PBS thus represents the probability that the protein–protein interaction is due to noise. Scores are real numbers ranging from 0 to 1, but are grouped in four categories (A, B, C and D) for practical purposes. Finally, the global connectivity of the interaction map is analysed to tag separately (category E) SIDs found as prey with frequency greater than a fixed threshold: the PBS of each protein–protein interaction involving highly connected SIDs is set to 1. Both the intercategory thresholds and the high-connectivity threshold were defined manually, taking into account the nature of the studied organism, the relevant library and the current coverage of the proteome ($A < 1^{e-10} < B < 1^{e-5} < C < 1^{e-2.5} < D$; the E category corresponds to prey SIDs selected with more than 4 baits and was arbitrarily attributed a PBS value of 1).

Bioinformatics

Several algorithms and software were implemented in the production database to facilitate experimental steps, such as a 'bait program' that designed automatically oligonucleotides for PCR amplification and sequencing of bait constructs, a 'prey program' that determined the position of each fragment in the genome and its coding capacity (such as intergenic, antisense, nucleotide position in an ORF, coding frame). The interactions were then analysed through a web-based software platform, the PIMRider developed at Hybrigenics and accessible through the web interface (http://pim.hybrigenics.com). Academic users will be granted a free licence. Other users will have to purchase a commercial licence. The *H. pylori* PIMRider platform is linked to the PyloriGene database.

1. Fields, S. The future is function. *Nature Genet.* **15,** 325–327 (1997).
2. Fromont-Racine, M., Rain, J. C. & Legrain, P. Toward a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nature Genet.* **16,** 277–282 (1997).
3. Bartel, P. L., Roecklein, J. A., SenGupta, D. & Fields, S. A protein linkage map of *Escherichia coli* bacteriophage T7. *Nature Genet.* **12,** 72–77 (1996).
4. Flajolet, M. *et al.* A genomic approach of the hepatitis C virus generates a protein interaction map. *Gene* **242,** 369–379 (2000).
5. McCraith, S., Holtzman, T., Moss, B. & Fields, S. Genome-wide analysis of vaccinia virus protein–protein interactions. *Proc. Natl Acad. Sci. USA* **97,** 4879–4884 (2000).
6. Ito, T. *et al.* Toward a protein–protein interaction map of the budding yeast: A comprehensive system to examine two-hybrid interactions in all possible combinations between the yeast proteins. *Proc. Natl Acad. Sci. USA* **97,** 1143–1147 (2000).
7. Uetz, P. *et al.* A comprehensive analysis of protein–protein interactions in *Saccharomyces cerevisiae*. *Nature* **403,** 623–627 (2000).
8. Walhout, A. J. M. *et al.* Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287,** 116–122 (2000).
9. Tomb, J. F. *et al.* The complete genome sequence of the gastric pathogen *Helicobacter pylori*. *Nature* **388,** 539–547 (1997).
10. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res.* **28,** 45–48 (2000).
11. Alm, R. A. *et al.* Genomic-sequence comparison of two unrelated isolates of the human gastric pathogen *Helicobacter pylori*. *Nature* **397,** 176–180 (1999).
12. Moszer, I. The complete genome of *Bacillus subtilis*: from sequence annotation to data management and analysis. *FEBS Lett.* **430,** 28–36 (1998).
13. Welch, M., Chinardet, N., Mourey, L., Birck, C. & Samama, J. P. Structure of the CheY-binding domain of histidine kinase CheA in complex with CheY. *Nature Struct. Biol.* **5,** 25–29 (1998).
14. Cussac, V., Ferrero, R. L. & Labigne, A. Expression of *Helicobacter pylori* urease genes in *Escherichia coli* grown under nitrogen-limiting conditions. *J. Bacteriol.* **174,** 2466–2473 (1992).
15. Mobley, H. L., Island, M. D. & Hausinger, R. P. Molecular biology of microbial ureases. *Microbiol. Rev.* **59,** 451–480 (1995).
16. Skouloubris, S., Thiberge, J. M., Labigne, A. & De Reuse, H. The *Helicobacter pylori* UreI protein is not involved in urease activity but is essential for bacterial survival *in vivo*. *Infect. Immun.* **66,** 4517–4521 (1998).
17. Weeks, D. L., Eskandari, S., Scott, D. R. & Sachs, G. A H$^+$-gated urea channel: the link between *Helicobacter pylori* urease and gastritic colonization. *Science* **287,** 482–485 (2000).
18. Dong, Z., Onrust, R., Skangalis, M. & O'Donnell, M. DNA polymerase III accessory proteins. I. holA and holB encoding delta and delta′. *J. Biol. Chem.* **268,** 11758–11765 (1993).
19. Liu, X. & Matsumura, P. An alternative sigma factor controls transcription of flagellar class-III operons in *Escherichia coli*: gene sequence, overproduction, purification and characterization. *Gene* **164,** 81–84 (1995).
20. Zhang, G. *et al.* Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 A resolution. *Cell* **98,** 811–824 (1999).
21. Mooney, R. A. & Landick, R. RNA polymerase unveiled. *Cell* **98,** 687–690 (1999).
22. Vidal, M. & Legrain, P. Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res.* **27,** 919–929 (1999).
23. Ferrero, R. L., Cussac, V., Courcoux, P. & Labigne, A. Construction of isogenic urease-negative mutants of *Helicobacter pylori* by allelic exchange. *J. Bacteriol.* **174,** 4212–4217 (1992).

Crystal structures of SarA, a pleiotropic regulator of virulence genes in *S. aureus*

Maria A. Schumacher*†, **Barry K. Hurlburt**‡ & **Richard G. Brennan***

* *Department of Biochemistry and Molecular Biology and* † *Vollum Institute, Oregon Health Sciences University, Portland, Oregon 97201-3098, USA*
‡ *Department of Biochemistry and Molecular Biology, University of Arkansas for Medical Sciences, Little Rock, Arkansas 72205, USA*

Staphylococcus aureus is a major human pathogen, the potency of which can be attributed to the regulated expression of an impressive array of virulence determinants. A key pleiotropic transcriptional regulator of these virulence factors is SarA, which is encoded by the *sar* (staphylococcal accessory regulator) locus[1-3]. SarA was characterized initially as an activator of a second virulence regulatory locus, *agr*, through its interaction with a series of heptad repeats (AGTTAAG) within the *agr* promoter[4]. Subsequent DNA-binding studies have revealed that SarA binds readily to multiple AT-rich sequences of variable lengths[4-11]. Here we describe the crystal structure of SarA and a SarA–DNA complex at resolutions of 2.50 Å and 2.95 Å, respectively. SarA has a fold consisting of a four-helix core region and 'inducible
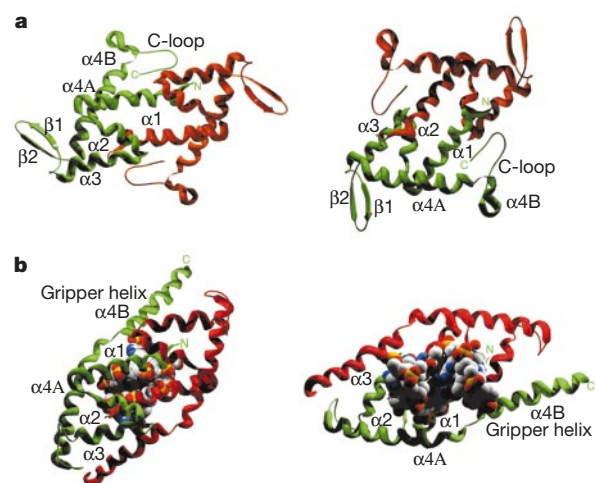
Figure 1 The SarA dimer and SarA–DNA complex. **a**, Views of SarA looking directly into the DNA-binding pocket or from the 'back' side. The secondary structures, β-hairpin and C-terminal loop of one monomer are labelled. Each monomer is red or green. **b**, The SarA–DNA complex in the identical orientation of the corresponding apo SarA directly above. The DNA duplex is shown as CPK atoms, with carbon, nitrogen, oxygen and phosphates coloured white, blue, red and yellow, respectively. The α4B gripper helix of one monomer is also labelled. The narrow and deep minor groove and major groove are seen on the left and right respectively.