

15. M. F. Bietenholz *et al.*, *Astrophys. J.* **457**, 604 (1996); M. F. Bietenholz, N. Bartel, M. P. Rupen, *Astrophys. J.*, in press.
16. R. Bandiera, F. Pacini, M. Salvati, *Astrophys. J.* **285**, 134 (1984).
17. W. Kundt, *Fundam. Cosmic Phys.* **20**, 1 (1998).
18. All uncertainties in this paper are standard errors. The standard errors of fit parameters were calculated to reflect a χ^2_ν of unity in case the measured $\chi^2_\nu > 1$. The standard errors of individual data points had statistical and estimated systematic contributions added in quadrature.
19. A. H. Rots and W. W. Shane, *Astron. Astrophys.* **45**, 25 (1975).
20. S. R. Trammell, D. C. Hines, J. C. Wheeler, *Astrophys. J.* **414**, L21 (1993); J. M. Lewis *et al.*, *Mon. Not. R. Astron. Soc.* **266**, L29 (1994). At later times, the maximum velocity could not be determined reliably from the H α absorption trough because of blending effects with newly developed emission lines.
21. The standard error includes the errors from the optical velocity and the rate of change of the VLBI angular radius, and the upper limit of deviations from isotropic expansion (N. Bartel *et al.*, in preparation).
22. We investigated the model dependence of our result and found that the asymmetric brightness distributions, possible shell thickness variations, and opacity effects would not alter the significance of the change of the deceleration rate by more than $\sim 1\sigma$. Our results are in conflict with a measurement of $m = 0.89 \pm 0.03$ for the late time from $t = 541$ to 1304 days (17). The significant discrepancy could be due to these authors not having used the more sensitive and convolution-free method of model fitting. Their method of convolving the source with a beam increasing in size $\propto t$ (implicitly assuming $m = 1$) and measuring the radius for a particular contour level (50%) biases the determination of m toward values of unity (as they note) and decreases the sensitivity of detecting deviations from a self-similar expansion.
23. B. M. Gaensler *et al.*, *Astrophys. J.* **479**, 845 (1997).
24. R. A. Chevalier, *Astrophys. J.* **259**, 302 (1982).
25. S. D. Van Dyk, K. W. Weiler, R. A. Sramek, M. P. Rupen, *Astrophys. J.* **432**, L115 (1995).
26. T. Suzuki and K. Nomoto, *Astrophys. J.* **455**, 658 (1995); C. Fransson, P. Lundqvist, R. A. Chevalier, *Astrophys. J.* **461**, 993 (1996).
27. C. Fransson and C.-I. Björnsson, *Astrophys. J.* **509**, 861 (1998).
28. H. Nussbaumer and R. Walder, *Astron. Astrophys.* **278**, 209 (1993).
29. P. Lundqvist, in *Circumstellar Media in the Late Stages of Stellar Evolution*, R. E. S. Clegg, I. R. Stevens, W. P. S. Meikle, Eds. (Cambridge Univ. Press, Cambridge, 1994), pp. 213–220.
30. J. C. Houck and C. Fransson, *Astrophys. J.* **456**, 811 (1996).
31. A. Sandage, *The Hubble Atlas of Galaxies* (Carnegie Institution of Washington, Washington, DC, 1961).
32. The sequence of radio images has been made into a movie of exploding star SN1993J (available at <http://aries.phys.yorku.ca/~bartel/SNmovie.html>).
33. Research at York University was partly supported by the Natural Sciences and Engineering Research Council of Canada. NRAO is a facility of NSF operated under cooperative agreement by Associated Universities, Inc. The NASA/JPL Deep Space Network is operated by JPL/California Institute of Technology, under contract with NASA.

16 June 1999; accepted 23 November 1999

Protein Interaction Mapping in *C. elegans* Using Proteins Involved in Vulval Development

Albertha J. M. Walhout,^{1*} Raffaella Sordella,¹ Xiaowei Lu,^{2†}
James L. Hartley,³ Gary F. Temple,³ Michael A. Brasch,³
Nicolas Thierry-Mieg,^{1,4} Marc Vidal^{1*‡}

Protein interaction mapping using large-scale two-hybrid analysis has been proposed as a way to functionally annotate large numbers of uncharacterized proteins predicted by complete genome sequences. This approach was examined in *Caenorhabditis elegans*, starting with 27 proteins involved in vulval development. The resulting map reveals both known and new potential interactions and provides a functional annotation for approximately 100 uncharacterized gene products. A protein interaction mapping project is now feasible for *C. elegans* on a genome-wide scale and should contribute to the understanding of molecular mechanisms in this organism and in human diseases.

Complete genome sequences are available for *Escherichia coli*, *Saccharomyces cerevisiae*, and *C. elegans*, and are expected soon for other model organisms and humans (1). In addition to facilitating the identification and cloning of genes and providing valuable insights on evolution, this information is likely to change the way biological questions are addressed. It is becoming possible to study molecular mechanisms globally in the context of complete sets

of genes, rather than analyzing genes individually. For example, DNA microarrays and chips can be used to monitor simultaneously the expression of nearly all genes of an organism (2). However, the function of most gene products predicted from sequencing projects is still completely uncharacterized, and it is widely accepted that this limitation needs to be overcome before full advantage can be taken of complete genome sequences. Functional assays aimed at characterizing the cellular localization of proteins, their spatial and temporal expression patterns, and their potential interacting partners should provide a backbone of functional annotations from which new biological questions can be formulated (3). Because the number of unannotated gene products in each model organism ranges from thousands to tens of thousands, it is important to develop standardized functional assays in which the same procedure can be applied to many proteins at a time, allowing utilization of high-throughput procedures.

Protein-protein interactions are crucial for many biological processes. Therefore, the knowledge of potential interactions involving otherwise uncharacterized proteins may provide insight into their function. The two-hybrid system, a standardized functional assay, facilitates the identification of potential protein-protein interactions and has been proposed as a method for the generation of protein interaction maps (4–7). Before the approach can be applied on a genome-wide scale, however, conceptual and technical issues need to be addressed. Conceptually, the biological information generated by two-hybrid analyses is often questioned because of the inherent artificial nature of the assay. Therefore, this method should be tested in a model organism using groups of proteins for which functional data are available. Technically, the cloning of open reading frames (ORFs) into appropriate expression vectors with the current techniques is laborious and expensive when dealing with hundreds or thousands of genes. Consequently, methods to standardize this process are required. Furthermore, it is necessary to decide upon a format in which the interaction data will be made available to the research community. Finally, it will be necessary to develop methods to determine the biological relevance of the potential interactions identified.

To address the issue of the potential biological relevance of protein interaction maps, we selected *C. elegans* as a model organism (8). The nearly complete *C. elegans* genome sequence led to the prediction of $\sim 20,000$ gene products of which approximately 700 have been functionally characterized [for example, (9)]. As a starting point, we chose to focus on *C. elegans* genes involved in the regulation of vulval development. At least four different pathways function coordinately to form a single vulva in the adult hermaphrodite, including a receptor tyrosine kinase (RTK)/Ras pathway (RTK/Ras), a Notch pathway (Notch), and two

¹Massachusetts General Hospital Cancer Center, Charlestown, MA 02129, USA. ²Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA. ³Life Technologies Inc., Rockville, MD 20850, USA. ⁴Laboratoire LSR-IMAC, 38402 St-Martin-D'Herès Cedex, France.

*Present address: Dana Farber Cancer Institute, Department of Genetics, Harvard Medical School, Boston, MA 02115, U.S.A.

†Present address: Department of Anatomy, University of California San Francisco, San Francisco, CA 94143, USA.

‡To whom correspondence should be addressed. E-mail: Marc_Vidal@DFCI.Harvard.edu

REPORTS

functionally redundant synthetic multivulva pathways: synMuv class A (synMuv A) and B (synMuv B) (10). Because many protein-protein interactions have been reported to be important for this process, the rate of false negatives in two-hybrid analyses could be estimated. However, the relationships between the products of many other genes involved in vulval development still remain to be determined. For

example, among 15 synMuv gene products (11), four have been characterized in more detail so far: LIN-35[retinoblastoma protein (pRB)], its associated proteins LIN-53(RbAp48) and HDA-1 [histone deacetylase (HDAC)] (12), and LIN-36 (13). This provided an opportunity to determine whether protein interaction maps might be helpful to point to novel functional relationships between the

products of less characterized (or uncharacterized) genes.

To address the technical problem of cloning multiple genes simultaneously, we took advantage of a novel method, "recombinational cloning" (RC) (Fig. 1A). RC is based on the recombination reactions involved in phage lambda integration into, and excision from, the *E. coli* genome. This method allows both the direction-

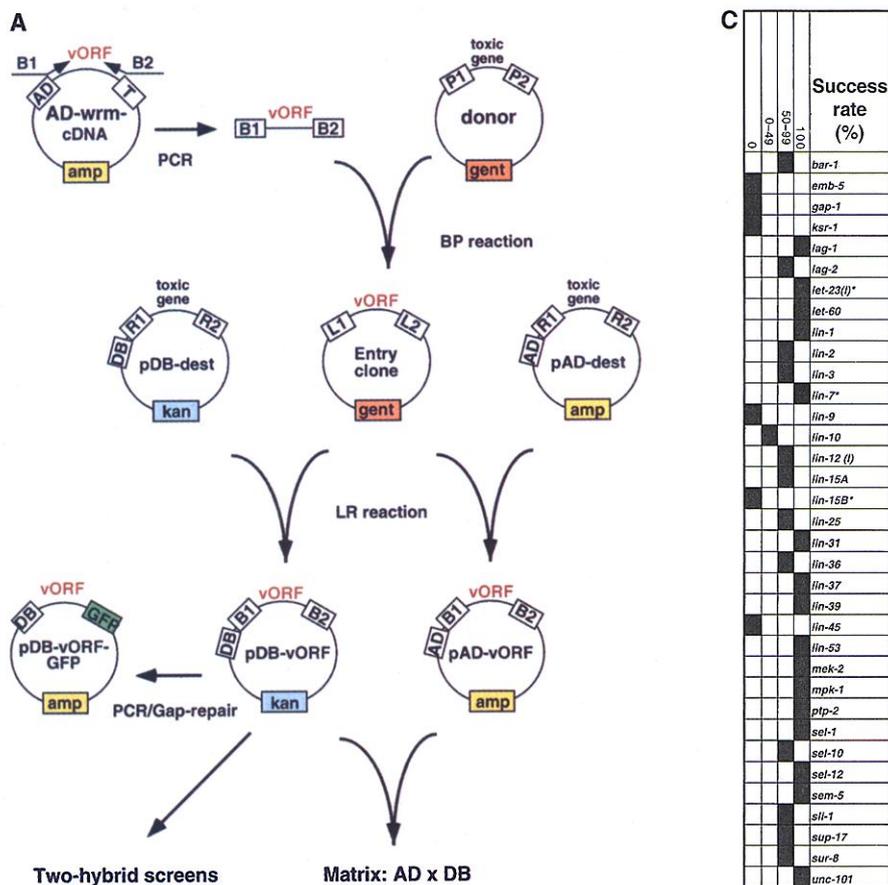


Fig. 1. Cloning of ORFs of genes involved in vulval development (vORFs). (A) Recombinational cloning (RC) (14). RC is based on the recombination reactions that mediate the integration and excision of phage λ into and from the *E. coli* genome, respectively. The integration involves recombination of the attP site of the phage DNA within the attB site located in the bacterial genome (BP reaction) and generates an integrated phage genome flanked by attL and attR sites. The excision recombines attL and attR sites back to attP and attB sites (LR reaction). The integration reaction requires two enzymes [the phage protein Integrase (Int) and the bacterial protein integration host factor (IHF)] (BP clonease). The excision reaction requires Int, IHF, and an additional phage enzyme, Excisionase (Xis) (LR clonease). Artificial derivatives of the 25-bp bacterial attB recombination site, referred to as B1 and B2, were added to the 5' end of the primers used in PCR reactions to amplify the vORFs (Fig. 1B). The resulting products were BP cloned into a "Donor vector" containing complementary derivatives of the phage attP recombination site (P1 and P2) using BP clonease. The resulting "Entry clones" contain vORFs flanked by derivatives of the attL site (L1 and L2) and were subcloned into two-hybrid "destination vectors" which contain derivatives of the attL-compatible attR sites (R1 and R2) using LR clonease. This resulted in "expression clones" in which vORFs are flanked by B1 and B2 and fused in frame to the DNA-binding domain (DB) or the activation domain (AD) of Gal4p. To ensure that both NH₂- and COOH-terminal fusion proteins can be generated, the B1 and B2 sequences were designed to be in frame with the vORF sequences. Note that different RC vectors harbor different selectable markers. In addition, both Entry and Destination vectors contain a toxic gene which prevents growth of most commonly used *E. coli* strains. This allows a genetic selection for the desired end products of each reaction. In addition to R1 and R2 sites DB-dest and AD-dest vectors contain yeast ARS and CEN sequences and *LEU2* or *TRP1* selectable marker, respectively. Because protein immunoblotting techniques are not compatible with high-throughput experiments, full-length vORF expression was tested using COOH-terminal fusions to GFP. However, no pDB-GFP destination vector is available at this point. Thus, vORFs were shuffled by PCR-Gap repair (17). (B) PCR amplification of vORFs. A summary of the functional information available for the vORFs shown here can be found on WormPD (23). A high-quality poly-dT primed cDNA library (AD-wrmcDNA) generated using mRNAs derived from all stages of development (15) was used as template DNA and the resulting vORF PCR products were analyzed on an

agarose gel. PCR reactions were considered successful when a single band of the expected size was observed (M = DNA size-markers). Twenty-nine vORFs were successfully cloned (16) (four are not shown here) by RC into an Entry vector and subsequently into DB-dest and AD-dest vectors. The design and sequence of the PCR primers will be described elsewhere (15). (C) Recombinational cloning of vORFs. The success rate of the first RC cloning step was measured by PCR (15). Briefly, the size of the insert of Entry clones was verified and in most cases more than 50% of the colonies contained a correct size insert. For a subset of clones, the fidelity of the second RC step was verified and in all cases an insert of the correct size was observed. Five vORFs could not be cloned because of unsuccessful PCR reactions. For *lin-15B*, a correct-size PCR product was obtained but could not be cloned. (I), intracellular domain; *, PCR product obtained from plasmid template DNA.

al cloning of PCR products into a reference vector and the subsequent transfer of the resulting DNA inserts into many different expression vectors *in vitro* (14). Importantly, restriction enzymes and ligase are not required for any of these steps. The ORFs corresponding to the genes involved in vulval development (vORFs) were introduced into two-hybrid vectors using this technique (15) and the resulting clones were subsequently verified using PCR analysis (Fig. 1, B and C) (16).

The vORF two-hybrid clones were used in two versions of the two-hybrid system. First, a matrix experiment was performed with 29 vORF-encoded proteins to determine the percentage of recovery of previously reported interactions. For each DB-vORF/AD-vORF pairwise combination a diploid yeast strain was generated by mating (17, 18) and tested for protein-protein interactions by scoring two-hybrid phenotypes (Fig. 2A). At least 50% (6 of 11) of the interactions reported in the literature, either in *C. elegans* or in other model organisms, were detected. In most cases, failure to detect interactions can be explained by the inherent restrictions of the two-hybrid assay or the physiology of yeast cells (19). Interestingly, two novel potential interactions, LIN-10/LIN-10 and LIN-53(RbAp48)/LIN-37 were identified using this approach.

Second, a more extensive protein interaction map was generated by exhaustive two-hybrid screens (20) of an AD-Y worm cDNA library with 27 DB-vORFs (two DB-vORFs were removed from the assay because they strongly activate reporter-gene expression in the absence of any interacting protein) (Fig. 2B). These two-hybrid selections identified 992 AD-Y encoding sequences that were subsequently amplified by PCR directly from yeast colonies and sequenced to generate "interaction sequence tags" (ISTs). Such ISTs corresponded to a total of 148 interactions involving 124 different potential interactors, of which 15 have been pre-

viously identified genetically and 109 were predicted from the *C. elegans* genome sequence. The number of ISTs identified for each DB-vORF bait varied between 0 and 29 (Fig. 2B). We systematically verified that the interacting sequence expressed in frame with AD corresponded to ORFs predicted by the genome sequencing project, rather than out-of-frame sequences encoding short irrelevant peptides (6).

To make the IST data publicly available, we took advantage of the ACeDB (*a C. elegans database*) database management system (21). ACeDB stores the genetic and physical maps and the nearly complete genome sequence along with its predicted ORFs. In addition, for many of the predicted ORFs, expressed sequence tags (ESTs) are available. The IST information was introduced into a local version of ACeDB (22) with the goal of connecting ORFs on the basis of a functional parameter rather than a genetic or physical link. Through ISTs, several ORFs are now linked by virtue of the ability of their products to interact in the context of a yeast two-hybrid assay. The information can be found by querying ACeDB for vORFs (Fig. 2C). In addition, we have introduced the IST data on a Web page (22) with hyperlinks to "WormPD" (23). WormPD, a recently released database, outlines published functional information on *C. elegans* genes in a standard format similar to that of YPD, a database of yeast functional annotation. In the future, the IST hyperlinks to WormPD should allow the integration of the protein interaction map with other worm functional genomics projects [for example, (24)].

Because the two-hybrid system is an artificial assay, the IST data should first be integrated with other information to evaluate the likelihood of biological relevance of each potential interaction. This in turn should allow the formulation of meaningful hypotheses. Therefore, we classified the potential interactions according to two-hybrid criteria and/or known biolog-

ical information. We explored the possibility that the knowledge of interactions conserved in other organisms might represent useful biological information (X/Y conserved interactions are referred to here as worm "interologs" of X'/Y' interactions in other species if X' and Y' are orthologs of X and Y, respectively). Hence, for each partner of worm DB-X/AD-Y potential interactions, we performed BLAST searches to identify X' and Y' orthologs and concentrated on those that have been reported to interact in other species. The first class observed consisted of interactions previously reported both in *C. elegans* and in at least one other model organism (Fig. 3A, class I). For example, the LET-60(Ras)/SUR-8 interaction is an interolog of the human Ras/hsSUR-8 interaction (25). The second class represents novel potential interologs. These interactions have been shown in other model organisms but have not been reported previously in *C. elegans* (Fig. 3A, class II). We propose that such potential interologs point to new hypotheses of function for the corresponding *C. elegans* proteins. For example, the SEL-10(Cdc4p)/SKP-1(Skp1p) potential interaction is a probable interolog of the yeast Cdc4p/Skp1p interaction (26). Because Cdc4p/Skp1p plays a role in protein degradation in yeast and SEL-10 interacts physically with LIN-12(Notch) and SEL-12(Presenilin) (Fig. 2A), it is possible that worm SKP-1 is involved in the degradation of components of the Notch pathway. Similarly, in the Ras pathway, LET-60(Ras)/F28B4.2(RalGDS) is a potential interolog of the human Ras/RalGDS interaction (27), suggesting that F28B4.2(RalGDS) might modulate LET-60(Ras) activity in *C. elegans*. Finally, LIN-53(RbAp48)/EGR-1[Metastasis Associated Protein (MTA1)] (28) might be an interolog of a human interaction because human MTA1 was found in the NURD complex that also contains RbAp48 (29).

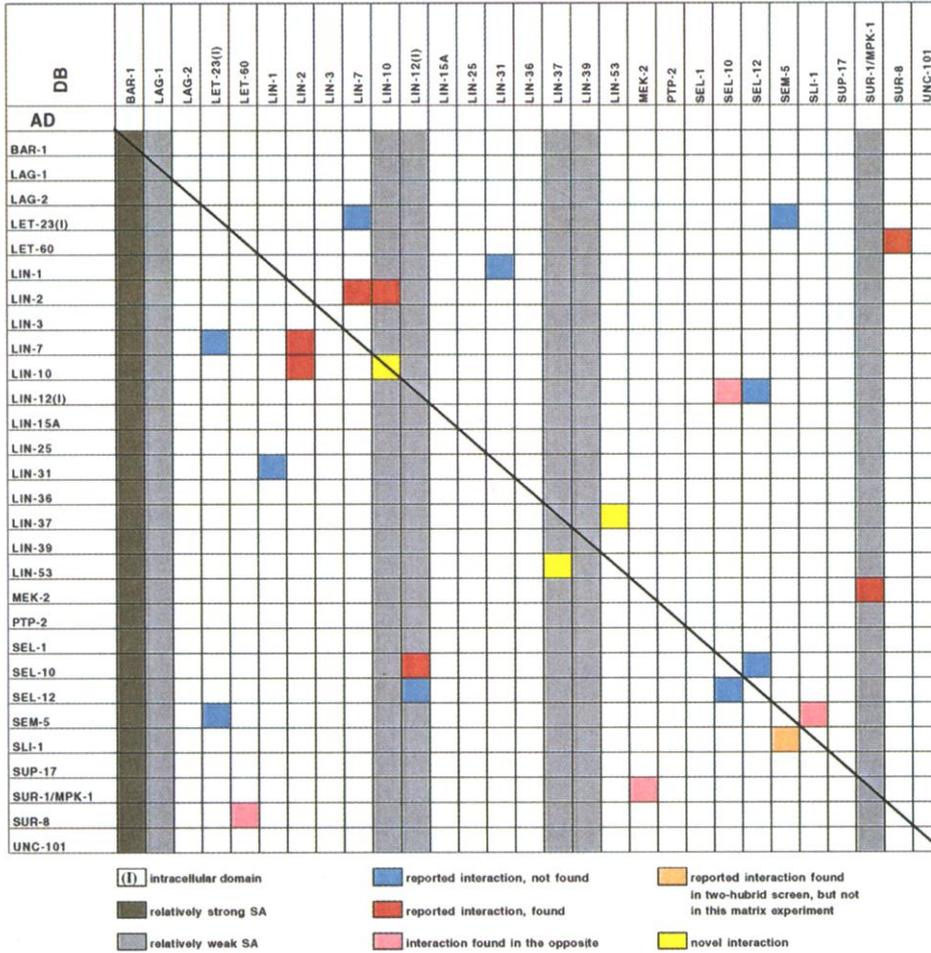
Although potentially powerful, applying the

Fig. 2. (Opposite) Protein interaction mapping. (A) Matrix of two-hybrid interactions between vORF-encoded proteins. The 29 vORFs cloned into pDB-dest and pAD-dest (Fig. 1) were transformed into yeast cells of opposite mating types (MaV103 and MaV203, respectively) (17). Diploids for every pairwise combination were generated by mating and tested for two-hybrid phenotypes. Color coding is as follows. Dark gray squares: selfactivation (SA) levels that are too high for two-hybrid screens (SA occurs from the ability of a DB-bait protein to up-regulate two-hybrid reporter gene expression in the absence of any AD interactor); light gray squares: intermediate SA levels which are compatible with two-hybrid screening using higher concentrations of 3-aminotriazole (3AT) (17); blue squares: interactions previously reported either in *C. elegans* or in other model organisms (potential interologs, Fig. 3A) and undetected in either the DB-X/AD-Y or the AD-X/DB-Y orientation (false negatives); red squares: interactions previously reported and detected in the Matrix assay; pink squares: interactions previously reported and detected in the Matrix assay in the opposite configuration only; orange square: interaction not found in the Matrix but uncovered in the screens described in Fig. 2B; yellow squares: novel potential interactions between the products of vORFs. **(B)** Exhaustive two-hybrid screens using 27 DB-vORFs as baits. DB-vORF baits were tested for SA on plates containing different concentrations of 3AT. SA ranged between levels for which concentrations of 100 mM 3AT were not sufficient to prevent growth

(####) to levels only detectable on X-Gal (#). Different concentrations of 3AT were used ("3AT") depending on the level of selfactivation of the corresponding bait. After transformation of the AD-wrmcDNA library (15), yeast colonies expressing potential interactions were selected on appropriate 3AT plates. The number of colonies screened varied between 0.8 and 4.2x10⁶ (# Colonies). Other abbreviations are as follows: (3AT)^R: number of 3AT resistant colonies; (3AT)^{R*}: number of 3AT^R clones that exhibited at least one additional two-hybrid phenotype [growth on plates lacking uracil or no growth on plates containing 5-Fluorouracil (5-FOA), or expression of β-Galactosidase activity on X-Gal-containing plates (17)]; ISTs (interaction sequence tags): number of distinct genes isolated as potential interaction partners (inserts of the AD-Y interacting clones were amplified by PCR directly from yeast colonies and subsequently sequenced using an ABI protocol). **(C)** ISTs in ACeDB. For each of the vORF that corresponds to a bait screened in this project, a window can be opened to retrieve ISTs. An example is shown for DB-LIN-53(RbAp48). Genomic and functional information on each IST can be retrieved. In the example shown, 13 independent cDNAs were selected for the LIN-53(RbAp48)/LIN-37 IST. The intron-exon junctions are shown in a format similar to ESTs currently available in ACeDB. If an interactor has itself been used as a bait, for example LIN-37 (thin, vertical red box in the center of the screen), a new window can be opened and the process reiterated.

REPORTS

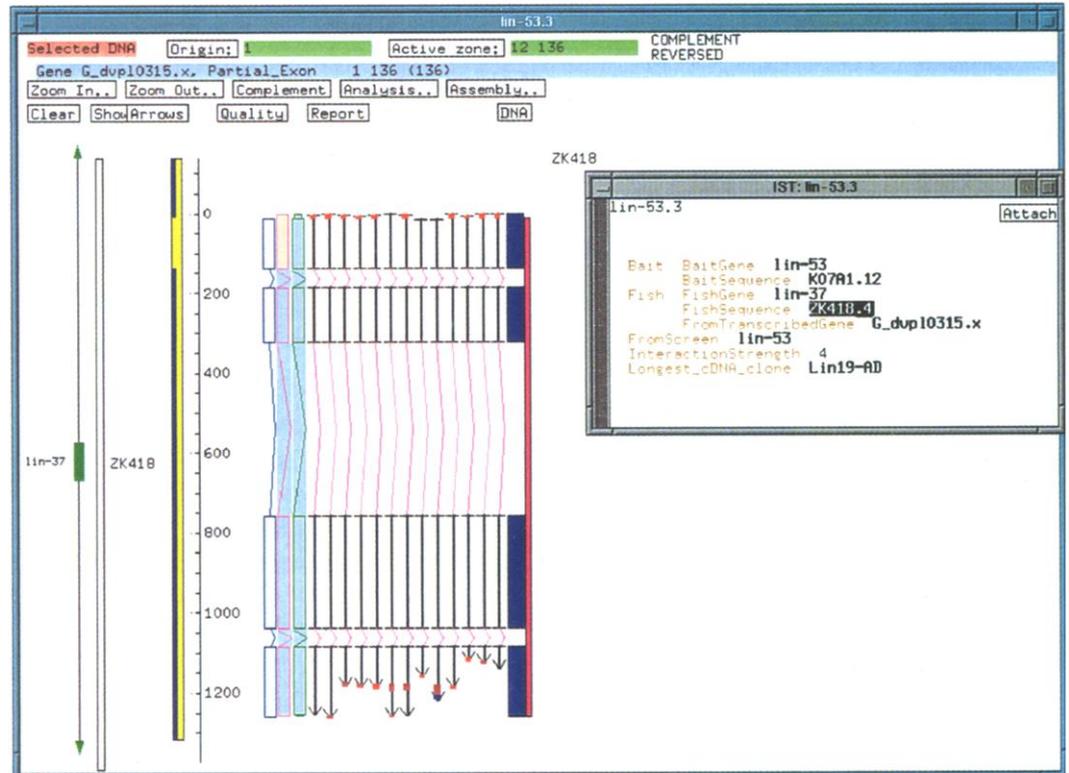
A



B

Gene	SA	[3AT]	# Colonies	# (3AT)R	# (3AT)R*	ISTs
<i>bar-1</i>	###					
<i>lag-1</i>	###	100	3.5	155	85	8
<i>lag-2</i>		20	1.3	41	3	0
<i>let-23(l)</i>		20	2.3	25	1	0
<i>let-60</i>		20	2	33	15	4
<i>lin-1</i>		20	2.8	16	1	0
<i>lin-2</i>		20	2.8	83	76	17
<i>lin-3</i>		20	4.1	20	5	0
<i>lin-7</i>		20	2.5	55	4	2
<i>lin-10</i>	###					
<i>lin-12(l)</i>	###	100	1.8	97	45	0
<i>lin-15A</i>		20	2.4	86	65	13
<i>lin-25</i>		20	2.1	12	0	
<i>lin-31</i>		20	2.8	46	28	1
<i>lin-36</i>		20	1.1	228	185	29
<i>lin-37</i>	##	60	2.3	152	122	14
<i>lin-39</i>	##	60	2.4	39	27	3
<i>lin-53</i>		20	1.8	131	29	3
<i>mek-2</i>		20	1.4	8	8	2
<i>mpk-1</i>	#	20	4.2	227	153	27
<i>ptp-2</i>		20	2.4	23	0	
<i>sel-1</i>		20	1.4	13	0	
<i>sel-10</i>		20	1.5	51	38	4
<i>sel-12</i>		20	3.2	24	0	
<i>sem-5</i>		20	3.2	74	40	11
<i>sli-1</i>		20	2.7	21	17	1
<i>sup-17</i>		20	0.8	15	0	
<i>sur-8</i>		20	2.1	38	6	2
<i>unc-101</i>		20	1.5	49	39	7
				1762	992	148

C



REPORTS

interolog concept alone in estimating the significance of potential interactions precludes the finding of novel connections previously unidentified in other model organisms. As an alternative method to classify the IST information (Fig. 3B). This approach is modeled after a method developed by Lipman and colleagues (30). We established contiguous connections between vORF- encoded proteins as follows: X interacts with Y, which interacts with Z, which interacts with W, and so on (X/Y/Z/W/...). Because two-hybrid screens are not random, we reasoned that clusters formed by contiguous connections that form closed loops (such as X/Y/Z/W/X) might increase the likelihood of

biological relevance for the corresponding potential interactions. Such two-hybrid clusters have been identified for known proteins in both macromolecular complexes and signal transduction pathways (7, 31). After removal of promiscuous interactors (32), we searched the ISTs of each vORF bait (X) for potential interactors that are identical to another bait (Y), or an interactor of another bait (Y'). For each hit, we then searched the IST list of Y or Y' and reiterated the process for several (n) cycles until the starting bait (X) was recovered, leading to clusters of contiguous connections in closed loops (for example, X/Y/Z/...n...X).

The clustering analysis was most informative in the case of ISTs identified for synMuv

gene products (Fig. 3B). A cluster centered around LIN-35(pRB) has already been reported [LIN-35(pRB)/LIN-53(RbAp48)/HDA-1(HDAC)/LIN-35(pRB)] (12). We identified several additional synMuv IST clusters: for example, LIN-53(RbAp48)/LIN-37/F10G8.8/LIN-36/EGR-1(MTA1)/LIN-53(RbAp48), and LIN-36/T05E7.5/LIN-15A/Y54E2A.3/LIN-36 (Fig. 3B). Many ISTs belong to more than one cluster, and thus, most clusters could be organized as overlapping sets: for example, LIN-53(RbAp48)/LIN-37/F10G8.8/LIN-36/EGR-1(MTA1)/LIN-53(RbAp48), and LIN-37/F10G8.8/LIN-36/Y54E2A.3/LIN-37. It is also possible that a higher likelihood of biological significance can be assigned to ISTs that correspond to interologs and belong to clusters. For example, the potential DB-LIN-53(RbAp48)/EGR-1(MTA1) interaction is both a potential interolog of a human interaction in the NURD complex (Fig. 3A) and a partner in IST clusters related to LIN-35(pRB) and HDA-1(HDAC) (Fig. 3B).

Hypotheses resulting from protein interaction maps are most useful in the context of other functional data. At this point, such information is not available on a genome-wide scale. However, in addition to LIN-35(pRB), LIN-53(RbAp48), and HDA-1(HDAC) such data are available for LIN-15A, LIN-36, and LIN-37, which are involved in the IST clusters described above (12, 13, 33). Most importantly, loss-of-function mutations in *lin-36*, *lin-37*, *lin-53*, and *lin-35* confer identical phenotypes in a synMuv class A background. In addition, the LIN-36 and LIN-37 gene products have been detected in the nuclei of vulval precursor cells (VPCs) at developmental stages consistent with a role in vulval development. This expression pattern overlaps with that of LIN-35(pRB) and LIN-53(RbAp48). Finally, although previous mosaic analysis suggested that LIN-15A functions in *hyp7*, this analysis did not exclude the possibility that this protein also acts in VPC cells (34). Thus, synMuv IST clusters, together with the existence of potential interologs and available functional data, suggest that LIN-37, LIN-36, EGR-1(MTA1), and LIN-15A may interact physically and perhaps belong to a single complex. Nevertheless, it is important to note that subsets of potential interactions might take place in different cells or at different times.

A powerful approach to directly test the biological relevance of potential interactions is to correlate loss-of-function mutations with loss of interaction. Single amino acid changes that confer a synMuv phenotype *in vivo* without grossly affecting the structure of the corresponding protein may provide useful tools to address whether loss-of-function correlates with loss of interaction. We initiated such analysis with LIN-53(RbAp48)/LIN-37 since the likelihood of biological relevance for this interaction is relatively high (35). The previously described *lin-53(n833)* allele is an excellent

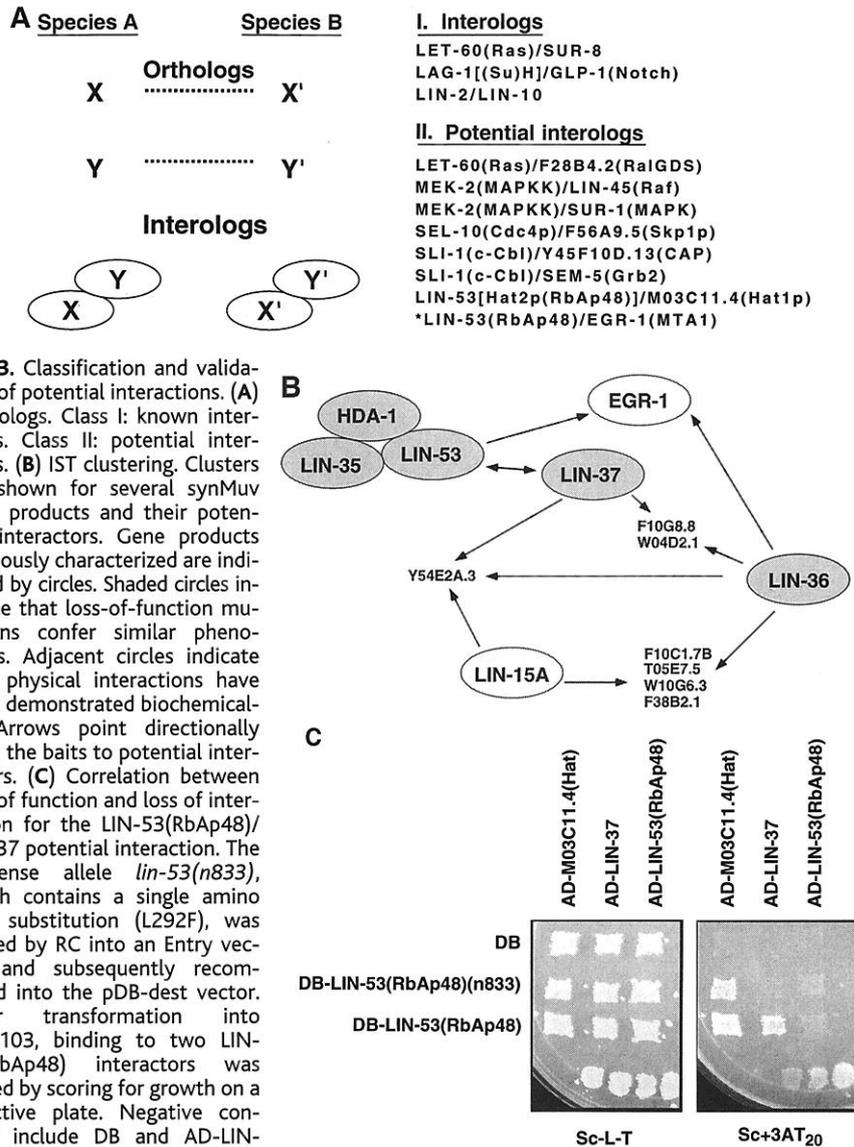


Fig. 3. Classification and validation of potential interactions. **(A)** Interologs. Class I: known interologs. Class II: potential interologs. **(B)** IST clustering. Clusters are shown for several synMuv gene products and their potential interactors. Gene products previously characterized are indicated by circles. Shaded circles indicate that loss-of-function mutations confer similar phenotypes. Adjacent circles indicate that physical interactions have been demonstrated biochemically. Arrows point directionally from the baits to potential interactors. **(C)** Correlation between loss of function and loss of interaction for the LIN-53(RbAp48)/LIN-37 potential interaction. The missense allele *lin-53(n833)*, which contains a single amino acid substitution (L292F), was cloned by RC into an Entry vector and subsequently recombined into the pDB-dest vector. After transformation into MaV103, binding to two LIN-53(RbAp48) interactors was tested by scoring for growth on a selective plate. Negative controls include DB and AD-LIN-53(RbAp48). Sc-L-T: synthetic complete medium lacking tryptophan and leucine (permissive plate), Sc+3AT₂₀: synthetic complete medium lacking tryptophan, leucine, and histidine and containing 20mM 3AT (selective plate). The four yeast patches at the bottom of each panel are controls for growth conditions. From left to right: 1st patch is a negative control (DB/AD), 2nd patch is a weak positive control for interaction (DB-pRB/AD-E2F1), 3rd patch is a strong positive control for interaction (DB-Fos/AD-Jun), 4th patch is the Gal4p positive control (DB-AD/AD) (17).

candidate for this approach because it confers a dominant negative phenotype and contains a single amino acid change (L292F) (12). The *lin-53(n833)* ORF was cloned by RC to test the ability of its encoded protein to interact with the potential partners of wild-type LIN-53(RbAp48), particularly LIN-37 (Fig. 3C). The interactions of DB-LIN-53(n833) with AD-M03C11.4(HAT) (Fig. 3C) and AD-EGR-1(MTA1) (36) were readily detected in the two-hybrid assay, suggesting that the structure and expression levels of the LIN-53 mutant protein are not grossly affected. Furthermore, LIN-53(n833) can also bind to LIN-35(pRB) in an *in vitro* assay (12). However, no interaction of DB-LIN-53(n833) with AD-LIN-37 was detected in the two-hybrid assay (Fig. 3C). Thus, it is tempting to speculate that LIN-53(n833) fails to interact with LIN-37 *in vivo* while retaining the ability to interact with other partners such as LIN-35(pRB). This might explain the dominant negative nature of *lin-53(n833)*. Taken together, these observations suggest that the LIN-53(RbAp48)/LIN-37 interaction is important for synMuv function *in vivo*.

We have addressed the feasibility of generating a genome-wide protein interaction map for *C. elegans*, the first animal model for which a complete genome sequence is available. According to our current throughput, we estimate that the scale of a *C. elegans* protein interaction mapping project should be on the order of one-tenth that of the genome sequencing project. We are currently generating a nearly complete set of cloned *C. elegans* ORFs using an automated version of RC (15). This should be useful for screening DB-ORFs against near complete AD-ORF arrays (Fig. 2A). By eliminating the need for sequencing ISTs, such arrays will increase the throughput substantially. We show that such genome-wide protein interaction maps can be interpreted at the biological level. In our hands, the two-hybrid method detected approximately 50% of reported interactions, which should allow a useful coverage of biologically important interactions. Alternative functional genomics projects based on other standardized assays will be useful to identify additional interactions. The inherent versatility of RC (14) should be of great value for such alternative approaches. On the other hand, although it is difficult to estimate the rate of false positives until more interpretation of the current IST data is performed, the data shown here suggest that false positives will not preclude the identification of relevant interactions. For example, three synMuv gene products (LIN-36, LIN-37, and LIN-15A), which had not been previously assigned to any particular step in the pathway, are now linked to each other and to LIN-35(pRB), LIN-53(RbAp48), and HDA-1(HDAC) by potential interactions. Perhaps classifications such as interologs and IST clustering will be useful for the interpretation of potential interactions. Importantly, this classification system is amenable to computa-

tion and thus should be applicable on a genome-wide scale. We also propose that relatively high-throughput genetic approaches can be used to directly test hypotheses resulting from IST information. For example, we show that a functionally defective allele of LIN-53 is specifically affected in its ability to bind LIN-37, a potential interactor of wild-type LIN-53. Finally, it is tempting to speculate that potential interactions identified here may help in understanding the molecular mechanisms involved in human tumorigenesis. Particularly, it is possible that human orthologs of LIN-36 and LIN-37 act in a pRB repressor complex and conversely, EGR-1, a worm MTA1 ortholog, might genetically interact with LIN-35(pRB).

References and Notes

1. F. R. Blattner *et al.*, *Science* **277**, 1453 (1997); A. Goffeau *et al.*, *Nature (suppl.)* **387**, 1 (1997); The *C. elegans* Sequencing Consortium, *Science* **282**, 2012 (1998); F. S. Collins *et al.*, *Science* **282**, 682 (1998).
2. M. Schena, D. Shalon, R. W. Davis, P. O. Brown, *Science* **270**, 467 (1995); D. J. Lockhart *et al.*, *Nature Biotechnol.* **14**, 1675 (1996); M. Chee *et al.*, *Science* **274**, 610 (1996).
3. E. S. Lander, *Science* **274**, 536 (1996); S. Fields, *Nature Genet.* **15**, 325 (1997).
4. P. L. Bartel, J. A. Roeklein, D. SenGupta, S. Fields, *Nature Genet.* **12**, 72 (1996); C. Evangelista, D. Lockshon, S. Fields, *Trends Cell Biol.* **6**, 196 (1997).
5. M. Fromont-Racine, J. C. Rain, P. Legrain, *Nature Genet.* **16**, 277 (1997).
6. M. Vidal and P. Legrain, *Nucleic Acids Res.* **27**, 919 (1999).
7. A. Flores *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 7815 (1999).
8. Because many developmental and behavioral mechanisms are relatively conserved [G. Ruvkun and O. Hobert, *Science* **282**, 2033 (1998); C. I. Bargmann, *Science* **282**, 2028 (1998)], information obtained from a *C. elegans* protein interaction map may be useful to formulate hypotheses about human biology. In *C. elegans*, the validity of potential protein-protein interactions can be tested using genetic methods, such as conventional epistatic analyses and RNA mediated interference (RNAi) [A. Fire *et al.*, *Nature* **391**, 806 (1998)]. Importantly, these analyses can be performed in the context of a developing animal with a fixed number of recognizable cells [W. B. Wood, *The Nematode Caenorhabditis elegans* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1988)].
9. A. J. M. Walhout, H. Endoh, N. Thierry-Mieg, W. Wong, M. Vidal, *Am. J. Hum. Genet.* **63**, 955 (1998).
10. K. Kornfeld, *Trends Genet.* **13**, 55 (1997); P. W. Sternberg and M. Han, *Trends Genet.* **14**, 466 (1998).
11. E. L. Ferguson and H. R. Horvitz, *Genetics* **123**, 109 (1989).
12. X. Lu and H. R. Horvitz, *Cell* **95**, 981 (1998).
13. J. H. Thomas and H. R. Horvitz, *Development* **126**, 3449 (1999).
14. J. L. Hartley, G. F. Temple, M. A. Brasch, in preparation. RC is also referred to as Gateway cloning. Recently, another recombination-based cloning system, referred to as "univector-plasmid fusion system" has been described [Q. Liu, M. Z. Li, D. Cortez, S. J. Elledge, *Curr. Biol.* **8**, 1300 (1998)]. This system is based on Cre-lox site-specific recombination. This system can also be used to conveniently transfer DNA inserts into different vectors. It is yet unclear how the standardization and versatility of this system will compare to RC. Details of the RC-two-hybrid system will be described elsewhere (M. A. Rhodes and M. A. Brasch, in preparation).
15. A. J. M. Walhout *et al.*, *Methods Enzymol.*, in press.
16. Sequencing of 10 independent clones corresponding to two ORFs (*lin-53* and *lin-37*) revealed approximately one mutation per 10,000 nucleotides, presumably originating from misincorporations during

the PCR reactions (A. J. M. Walhout, R. Sordella, M. Vidal, data not shown). The remaining steps described here were carried out with pools of approximately 10 Entry clones per vORF. Because the average size of vORFs is about 1.5 kb, we estimate that approximately 9 of 10 Entry clones generated in the BP cloning step correspond to wild-type vORFs. For a subset of vORFs, protein immunoblot analysis showed that full-length DB-vORFs are expressed in yeast cells from the DB destination vectors. The insert size of each vORF Entry clone was verified using PCR analysis. We also demonstrated that DB-vORF Expression clones can encode full-length proteins using COOH-terminal fusions to green fluorescent protein (GFP) (Fig. 1A). It is noteworthy that both the cloning and the last two verification strategies can be performed on 96-well plates and thus are compatible with high-throughput procedures.

17. M. Vidal, in *The Yeast Two-Hybrid System*, P. Bartel and S. Fields, Eds. (Oxford Univ. Press, New York, 1997), pp. 109-147.
18. R. L. Finley Jr. and R. Brent, *Proc. Natl. Acad. Sci. U.S.A.* **91**, 12980 (1994).
19. The interaction between LET-23(RTK) and SEM-5(Grb2) is dependent upon tyrosine phosphorylation [E. J. Lowenstein *et al.*, *Cell* **70**, 431 (1992)] which is seemingly absent in yeast cells. False negatives can also occur from the fact that the two-hybrid system is based upon a transcriptional activation event that occurs in the nucleus. Hence, interactions involving HDAC, its recruiting proteins, or transcriptional repressors such as LIN-1 may result in transcriptional repression in yeast cells and thus be incompatible with transcriptional activation, for example, LIN-1(Ets-like)/LIN-31(Forkhead/HNF) [P. B. Tan, M. R. Lackner, S. K. Kim, *Cell* **93**, 569 (1998)] and LIN-35(pRB)/LIN-53(RbAp48) (72). Proteins with transmembrane domains might not be exported correctly to the yeast nucleus as DB- or AD-fusions, for example, SEL-12(presenilin)/LIN-12(Notch) and SEL-12(presenilin)/SEL-10 [W. J. Ray *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **96**, 3263 (1999); G. Wu, E. J. Hubbard, J. K. Kitajewski, I. Greenwald, *Proc. Natl. Acad. Sci. U.S.A.* **95**, 15787 (1998)]. In addition, the failure to detect an interaction can sometimes be attributed to the fact that, in the matrix setting shown here, full-length vORFs rather than discrete domains were expressed in fusion with DB and AD. For example several discrete domains of E2F family members exhibit a stronger two-hybrid read-out with pRB than the corresponding full-length proteins [M. Vidal, P. Braun, E. Chen, J. D. Boeke, E. Harlow, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 10321 (1996)]. Finally, the version of the two-hybrid system used here might affect the read-out of particular interactions. Hence, the LET-23(RTK)/LIN-7 interaction which requires the COOH-terminal end of the receptor [J. S. Simske, S. M. Kaech, S. A. Harp, S. K. Kim, *Cell* **85**, 195 (1996)] might not be detectable due to the B2-encoded peptide resulting from RC. It should be noted that the evidence for the interactions reported in the literature is sometimes limited to other versions of the yeast two-hybrid system.
20. We have developed an improved version of the two-hybrid system (17). Three reporter genes expressed from different Gal4p-inducible promoters are used to increase the selectivity of the assay. In addition, single-copy DB- and AD-vectors are used to obtain relatively low expression levels of the two-hybrid proteins, thus improving the specificity of the selection. This version gives rise to decreased rates of biologically irrelevant interactions (6) [C. Sardet *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 2403 (1995); C. Englert *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **92**, 11960 (1995); W. Du, M. Vidal, J.-E. Xie, N. Dyson, *Genes Dev.* **10**, 1206 (1996); S. S. Kim *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **93**, 15299 (1996); G. Hu *et al.*, *Genes Dev.* **11**, 2701 (1997); T. Yasugi, J. D. Benson, H. Sakai, M. Vidal, P. M. Howley, *J. Virol.* **71**, 891 (1997); T. Yasugi, M. Vidal, H. Sakai, P. M. Howley, D. Benson, *J. Virol.* **71**, 5942 (1997); A. J. M. Walhout and M. Vidal, *Genome Res.* **9**, 1128 (1999)]. We have also developed a semi-automated version of the two-hybrid system in which most steps are carried out on 96-well plates to increase the throughput of the assay (A. J. M. Walhout *et al.*, data not shown). The ISTs were retransformed into fresh DB-vORF-con-

- taining yeast cells and retested for two-hybrid readouts. This step eliminates false positives probably resulting from *cis*- and *trans*-acting mutations that can influence reporter gene activity (17).
21. R. Durbin and J. Thierry-Mieg, in *Computational Methods in Genome Research*, S. Suhai, Ed. (Plenum, New York, 1994).
 22. <http://cancerbiology.dfci.harvard.edu/cancerbiology/ResLabs/Vidal/>
 23. www.proteome.com/databases/index.htm
 24. <http://cmgm.stanford.edu/~kimlab/wmdirectorybig.html>
 25. D. S. Sieburth, Q. Sun, M. Han, *Cell* **94**, 119 (1998).
 26. C. Bai et al., *Cell* **86**, 263 (1996).
 27. A. Kikuchi, S. D. Demo, Z. H. Ye, Y. W. Chen, L. T. Williams, *Mol. Cell. Biol.* **14**, 7483 (1994).
 28. F. Solari, A. Bateman, J. Ahninger, *Development* **126**, 2483 (1999).
 29. Y. Xue et al., *Mol. Cell.* **2**, 851 (1998).
 30. R. L. Tatusov, E. V. Koonin, D. J. Lipman, *Science* **278**, 631 (1997).
 31. J. A. Printen and G. F. Sprague, *Genetics* **138**, 609 (1994); K. Y. Choi, B. Satterberg, D. M. Lyons, E. A. Elion, *Cell* **78**, 499 (1994).
 32. For yet unknown reasons, several proteins notoriously behave as promiscuous interactors in the context of the two-hybrid system. Because such promiscuous interactors would bias the clustering analysis, they should systematically be removed from the analysis. In the AD-wrmcDNA library, this is the case for AD-PAL-1, AD-UNC-15, AD-F33G12.5, and AD-F54D5.5, which were found as potential ISTs for several vORF baits but also for unrelated baits such as DB-LIN-5, DB-LIN-6, DB-MEX-3, DB-OSM-10, and DB-PCT-1 (A. J. M. Walhout, M. Lorson, S. van den Heuvel, N. Huang, C. Hunter, A. Sugimoto, A. Hart, M. Vidal, unpublished data).
 33. S. G. Clark, X. Lu, H. R. Horvitz, *Genetics* **137**, 987 (1994); L. S. Huang, P. Tzou, P. W. Sternberg, *Mol. Biol. Cell* **5**, 395 (1994); X. Lu and H. R. Horvitz, in preparation.
 34. R. K. Herman and E. M. Hedgecock, *Nature* **348**, 169 (1990).
 35. In summary, the two-hybrid readouts are excellent for LIN-53(RbAp48)/LIN-37 (interaction was detected multiple times in both DB-X/AD-Y and DB-Y/AD-X orientations, with full-length proteins in the matrix, scores positive for three reporter genes and belongs

to clusters) and the available biological information on both proteins is consistent with an interaction (both proteins are required for a synMuv function and co-localize). Because no interolog has been identified yet for this interaction, it is possible that its study in *C. elegans* might reveal new features of pRB function in humans.

36. A. J. M. Walhout, H. Endoh, M. Vidal, data not shown.
37. We thank N. Dyson, F. Holstege, L. Matthews, and S. van den Heuvel for comments on the manuscript; M. Chalfie, A. Coulson, E. Harlow, H. R. Horvitz, S. Kim, E. Lander, M. Murray, R. Plasterk, J. and D. Thierry-Mieg, and S. van den Heuvel for support; J. Hitti, L. Doucette-Stamm, and B. Guild (Genome Therapeutics Corp., Waltham, MA) for their help with sequencing; H. Endoh, H. R. Horvitz, S. Kim, M. Lorson, and S. van den Heuvel for reagents or unpublished information, or both; and C. Ngwu for generous help with the design of our Web site. Supported by grants 1 R01 HG01715-01 (National Human Genome Research Institute) and 1 R21 CA81658 A 01 (National Cancer Institute) awarded to M.V.

9 August 1999; accepted 19 November 1999

A Short Fe-Fe Distance in Peroxidiferic Ferritin: Control of Fe Substrate Versus Cofactor Decay?

Jungwon Hwang,¹ Carsten Krebs,² Boi Hanh Huynh,² Dale E. Edmondson,³ Elizabeth C. Theil,⁴ James E. Penner-Hahn^{1*}

The reaction of oxygen with protein diiron sites is important in bioorganic syntheses and biomineralization. An unusually short Fe-Fe distance of 2.53 angstroms was found in the diiron (μ -1,2 peroxidiferic) intermediate that forms in the early steps of ferritin biomineralization. This distance suggests the presence of a unique triply bridged structure. The Fe-Fe distances in the μ -1,2 peroxidiferic complexes that were characterized previously are much longer (3.1 to 4.0 angstroms). The 2.53 angstrom Fe-Fe distance requires a small Fe-O-O angle ($\sim 106^\circ$ to 107°). This geometry should favor decay of the peroxidiferic complex by the release of H_2O_2 and μ -oxo or μ -hydroxo diferric biomineral precursors rather than by oxidation of the organic substrate. Geometrical differences may thus explain how diiron sites can function either as a substrate (in ferritin biomineralization) or as a cofactor (in O_2 activation).

Iron proteins mediate iron and oxygen chemistry for a variety of biological functions. Although heme proteins may be the most familiar, nonheme diiron proteins are widely distributed. Reactions of iron and oxygen at carboxylate-bridged diiron centers include reversible O_2 binding for respiration (hemerythrin), oxidation and desaturation of organic substrates [methane monooxygenase (MMOH), the R2 subunit of

ribonucleotide reductase, and the stearyl-acyl carrier protein Δ -9 desaturase (Δ 9D)], and concentration of iron in a ferric biomineral (ferritin) (1, 2). Ferritins are ubiquitous proteins that concentrate iron in cells as a mineral, storing the iron for later use; cytotoxic reactions of both Fe^{2+} and O_2 are controlled by ferritin chemistry. Ferric oxo reaction products are transported within ferritin to a large cavity in the center of the protein to form a ferric oxide hydrate. The recent detection of a peroxidiferic intermediate in the ferritin ferroxidase reaction (3, 4) firmly establishes the ferritin ferroxidase site as very similar to sites in the O_2 -activating diiron enzymes (Fig. 1). What is puzzling is how the peroxidiferic intermediate can be a catalytic cofactor in the enzymes but a substrate intermediate in the ferritins.

Similarities among the peroxidiferic intermediates of MMOH (5) and site-specific mutants of R2 (6), Δ 9D (7), and ferritin (3, 4) include a distinct blue color from a peroxo \rightarrow Fe charge transfer band (650 to 725 nm), a relatively large Mössbauer isomer shift (0.62 to 0.66 mm/s) that is characteristic of Fe^{3+} -peroxide complexes, and high-frequency (850 to 900 cm^{-1}) Raman bands that are attributable to a μ -1,2 peroxide O-O stretch (Table 1). These spectroscopic similarities stand in contrast to the reactivities of the peroxidiferic intermediates in the different proteins. Decay of the peroxo intermediate yields a high-valence Fe oxidant in the O_2 -activating enzymes but gives diferric oxo-hydroxo biomineral precursors in ferritin (Fig. 1).

To obtain further insight into the ferritin reaction, we used x-ray absorption spectroscopy (XAS), which includes x-ray absorption near edge structure (XANES) and extended x-ray absorption fine structure (EXAFS), to examine the structure of the ferritin peroxidiferic intermediate that is formed when recombinant frog M ferritin reacts with Fe^{2+} and O_2 . We observed a short Fe-Fe separation of 2.53 Å in the intermediate that allows us to define the diiron structure and to gain insight into the features responsible for modulating the peroxidiferic decay pathway.

Two intermediates in the ferritin reaction with Fe^{2+} and O_2 were trapped by rapid freeze quench at 25 ms ($+O_2^{25\text{ ms}}$) or at 1 s ($+O_2^1\text{ s}$) and were packed in delrin sample cells that are suitable for Mössbauer and XAS measurements (3, 4, 8, 9). A third sample, which was prepared in the absence of O_2 , provided an Fe^{2+} control. Mössbauer measurements (Table 2) were made both before and after x-ray irradiation to ensure sample integrity and to quantify the different iron species. In the $-O_2$ samples, most of the iron is high-spin Fe^{2+} , with a linewidth that suggests a mixture of several similar Fe^{2+} structures (denoted 1). After 25-ms reaction

¹Department of Chemistry, University of Michigan, Ann Arbor, MI 48109-1055, USA. ²Department of Physics, ³Departments of Biochemistry and Chemistry, Emory University, Atlanta, GA 30322, USA. ⁴Children's Hospital Oakland Research Institute, 5700 Martin Luther King Jr. Way, Oakland, CA 94609-1673, USA.

*To whom correspondence should be addressed. E-mail: jeph@umich.edu