

CD-Tagging: A New Approach to Gene and Protein Discovery and Analysis

BioTechniques 20:896-904 (May 1996)

J.W. Jarvik^{1,2}, S.A. Adler¹, C.A. Telmer¹, V. Subramaniam¹ and A.J. Lopez¹

¹Carnegie Mellon University, Pittsburgh, PA, and ²Vyrex Corporation, La Jolla, CA, USA

ABSTRACT

We describe a new method for gene discovery and analysis, CD-tagging, that puts specific molecular tags on a gene, its transcript and its protein product. The method has been successfully tested in two organisms, the haploid unicellular alga *Chlamydomonas reinhardtii* and the metazoan *Drosophila melanogaster*. The method utilizes a specially designed DNA molecule, the CD-cassette, that contains splice acceptor and donor sites surrounding a short open reading frame. Insertion of the CD-cassette into an intron in a target gene introduces a new exon, represented by the open reading frame of the CD-cassette, surrounded by two functional hybrid introns. As a result (i) the gene is tagged by a specific nucleotide sequence, (ii) the mRNA is tagged by a specific nucleotide sequence and (iii) the protein is tagged by a specific peptide sequence. Because these tags are unique, specific nucleotide or antibody probes can be used to obtain and/or analyze the gene, transcript or protein. As a gene discovery technology, CD-tagging has two unique advantages: 1) Genes can be identified through a primary screen at the protein level, and so the very process by which a gene is identified provides specific empirical information about its biological function. 2) The cassette arms, which are spliced out of the transcript of the target gene, are available to carry a wide variety of DNA sequences, such as genes encoding drug resistance that can be used to select for the presence of the CD-cassette in the genome.

INTRODUCTION

A fundamental concern in genetics and molecular biology is establishing the correspondence between gene and protein. Thus, when a new gene is discovered it is usually desirable to identify the protein or proteins it encodes; conversely, when a new protein is identified it is usually desirable to identify the gene that encodes it. In this communication we describe a new molecular-genetic method that adds specific tags to a gene, mRNA and protein in a single recombinational event, and we use the method to tag cloned genes from *Chlamydomonas reinhardtii* and *Drosophila melanogaster*. With this method, establishing the correspondence between gene and protein in gene discovery is dramatically simplified, indeed obviated, by the fact that the gene and the gene product are discovered simultaneously. We call the method "CD-tagging" because it tags each of the three classes of molecule (DNA, RNA and protein) referred to in the central dogma of molecular biology.

CD-tagging employs a specially designed DNA molecule that we call a CD-cassette. Within the cassette is an open reading frame flanked by consensus sites for RNA splicing. As shown in Figure 1, when the CD-cassette is introduced into an intron in a target gene, the result is two chimeric introns surrounding a new "guest exon". When the tagged gene is transcribed, spliced and translated, a protein containing a "guest peptide" insert will be produced. If the guest peptide includes an epitope recognized by an existing antibody, as in the examples presented in this paper, the result will be an epitope-tagged protein. For CD-tagging to work as depicted in Figure 1, the chimeric introns surrounding the guest exon must be

functional. On the basis of numerous cases in the literature in which introns with 5' splice sites from one source and 3' splice sites from another were observed to function appropriately (3,9, 11,14), we expected that in most cases the chimeric introns generated by CD-tagging would be functional as well. In this report we show, by tagging known introns in known genes, that our expectation was met.

MATERIALS AND METHODS

Strains and Culture Conditions

Chlamydomonas strains CC1861 (*arg-7 mt-*) and CC1032 (*pf14- mt+*) were obtained from the Chlamydomonas Genetics Center, Duke University. An *arg7- pf14-* strain was constructed by standard genetic methods (15). Cells were grown on HSA medium (15) supplemented, when appropriate, with 0.005%–0.01% arginine. *Drosophila* cell culture followed standard methods (7).

Synthesis of the CD-1 DNA

The CD-1 DNA was constructed by the Midland Certified Reagent Company (Midland, TX, USA) through the chemical synthesis of four overlapping polynucleotides that were annealed and inserted into the *Xba*I site of pUC118 to produce plasmid pJJ225. The nucleotide sequence of the entire CD-1 insert in pJJ225 was confirmed by dideoxy sequencing.

Construction of CD-Tagged Genes

To tag the *Chlamydomonas reinhardtii* *pf14* gene in intron 3, the 279-bp *Nsi*I fragment from pJJ225 was ligated into the unique *Nsi*I site in plasmid

pKE-RS3, and ampicillin-resistant transformants of *E. coli* DH5 α were selected on LB ampicillin plates. Plasmid minipreps were prepared from individual colonies, digested with *Cla*I and *Hpa*I, and analyzed by agarose gel electrophoresis to determine the presence and orientation of the CD-1 insert. A plasmid with a single CD-1 insert oriented in the same direction as the *pf14* gene was retained and named pRS031-1. A plasmid with a tandem pair of CD-DNA inserts oriented in the same direction as the *pf14* gene was retained and named pRS031-2.

To add the *npt1* gene to the insert in intron 3, pRS031-1 was linearized at the unique *Xho*I site within the CD-cassette, and the 1252bp *Sal*I fragment from the *npt1*-containing plasmid pUC-4K (Pharmacia Biotech, Piscataway, NJ, USA) was inserted into the site to produce plasmid pRS031-3, which was selected on the basis of resistance to 25 μ g/mL kanamycin in *E. coli*.

To tag *pf14* in intron 1, the 203-bp *Esp*I fragment from pJJ225 was ligated into the unique *Esp*I site in plasmid pKE-RS3, and ampicillin-resistant transformants of *E. coli* DH5 α were selected on LB ampicillin plates. Plasmid minipreps were prepared from individual colonies, digested with *Cla*I and *Hpa*I and analyzed by agarose gel electrophoresis to determine the presence and orientation of the CD-1 insert. A plasmid with a single CD-1 insert oriented in the same direction as the *pf14* gene was retained and named pRS011-1. A plasmid with a two tandem CD-DNA inserts oriented in the same direction as the *pf14* gene was retained and named pRS011-2.

To tag the *Ultrabithorax* (*Ubx*) gene of *Drosophila melanogaster*, the 285-bp *Xba*I fragment from pJJ225 was inserted into the *Xba*I site within intron 3 of the *Ubx* minigene in plasmid pUMG101 (27) to produce plasmid pUMG101-CD1, which contains a single copy of the CD-1 cassette oriented in the same direction as *Ubx*. The *Ubx* minigene in pUMG101 contains all four exons and internally shortened introns, but lacks the promoter and upstream regulatory elements; it is inserted in the *Drosophila* germ line transformation vector pUAST (2), which provides transcriptional control by a

promoter that is activated by the *S. cerevisiae* *GAL4* gene product.

Transformation of *Chlamydomonas*

arg-7 and *arg7⁻ pf14⁻* strains were co-transformed with the plasmids pARG7.8 and pKE-RS3 or their CD-tagged derivatives using silicon carbide whiskers (8). Arg⁺ transformants were selected by plating on arginine-free HSA medium (15); colonies were picked into 3 mL HSA medium and assayed for motility by light microscopy. Cotransformation frequencies of *arg7⁻ pf14⁻*, as determined by the fraction of Arg⁺ transformants that were motile, ranged from 15 to 60 percent. Cotransformants of the *pf14⁺* recipient were identified on the basis of the presence of the CD-tagged gene product as assayed by immunofluorescence.

Transfection of *Drosophila* Tissue Culture Cells

Drosophila Schneider Line 2 (SL2) cells were maintained in Schneider medium (Life Technologies) supplemented with 14% fetal calf serum and 100 units/mL each penicillin and streptomycin. SL2 cells were co-transfected with CaPO₄ co-precipitates of pUMG101-CD1 and pAKG1 following standard protocols (7,18). pAKG1 (27) carries the yeast *GAL4* coding region

under control of the *Drosophila actin-5C* promoter, thus providing a constitutive source of *GAL4* protein to activate expression of the CD-tagged *Ubx* minigene in pUMG101-CD1.

Isolation of *Drosophila* Germline Transformants

pUMG101-CD1 DNA was co-injected into *w^{/w}* *Drosophila* embryos along with the P-element transposase-expressing plasmid p π 25.7, using standard methods (23). Germ-line transformants showing stable expression of the *white* gene from pUMG101-CD1 were identified among the progeny of the injected animals on the basis of their pigmented eyes.

Immunofluorescence Microscopy

Chlamydomonas cells were attached to polyethylenimine-coated wells in Teflon[®]-coated glass slides (Carlson Scientific, Peotone, IL, USA). The slides were incubated in 1% Nonidet P-40 in NB minus Ca⁺⁺ for 1.5 h prior to immunostaining. Washes, blocking reactions and treatment with fluorescein-labeled secondary antibody (Cappel, Durham, NC, USA) were as previously described (31). Primary antibody incubations were overnight at room temperature. Anti-HA mouse monoclonal antibody 12CA5 (Boehringer

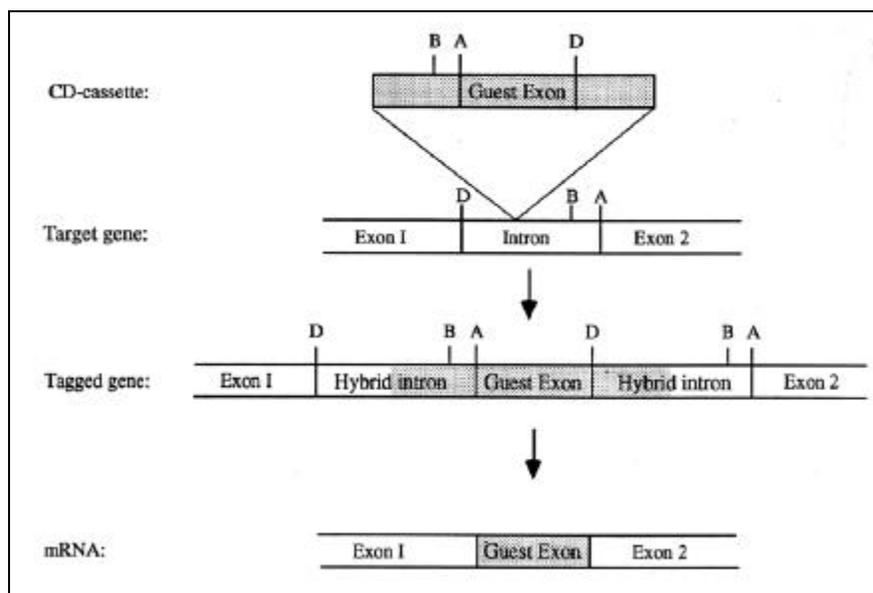


Figure 1. CD-tagging: the basic concept. The letters B, A and D represent branch, acceptor and donor sites, respectively, for RNA splicing.

Research Reports

Mannheim, Mannheim, Germany) was used at 2.5 µg/mL, and anti-RSP3 rabbit antibody (29) was used at a dilution of 1:400.

Immunoblotting

Chlamydomonas axonemes were prepared from 300 mL liquid cultures grown in HSA medium (15), solubilized in sodium dodecyl sulfate (SDS) sample buffer, and separated by SDS gel electrophoresis on 10% polyacrylamide gels. Transfected *Drosophila* SL2 cells were processed for electrophoretic separation of proteins on 10% SDS-polyacrylamide gels as described by Gavis and Hogness (12). Proteins were electrophoretically transferred to nitrocellulose, probed with anti-RSP3 antibody (1:3000) or with 12CA5 (2.5 µg/mL) and visualized with alkaline phosphatase-labeled secondary antibody (Bio-Rad, Hercules, CA, USA).

Immunostaining of *Drosophila* Embryos

Collection, fixation, devitellinization and immunohistochemical staining of *Drosophila* embryos using HRP-conjugated anti-mouse secondary antibody were performed as described previously (1). Primary antibodies were either monoclonal 12CA5, against the HA epitope in the CD tag, or monoclonal 5C.2B (19), against an epitope C-terminal to the *UBX* homeodomain. Stained embryos were examined and photographed under Differential Interference Contrast optics using a Nikon Microphot-FXA microscope (Bayshore, NY, USA).

RESULTS

CD-Tagging the *pf14* Gene of *Chlamydomonas reinhardtii*

Figure 2 shows the sequence and structure of the CD-1 cassette. The guest exon in the CD-1 cassette is 54 nucleotides (nt) in length and encodes 18 amino acids. To test the CD-tagging approach in the unicellular biflagellate green alga *Chlamydomonas reinhardtii*, CD-1 DNA was inserted in vitro into *pf14*, a three-intron gene that encodes the flagellar radial-spoke protein RSP3 (6,29). Five plasmids carrying tagged

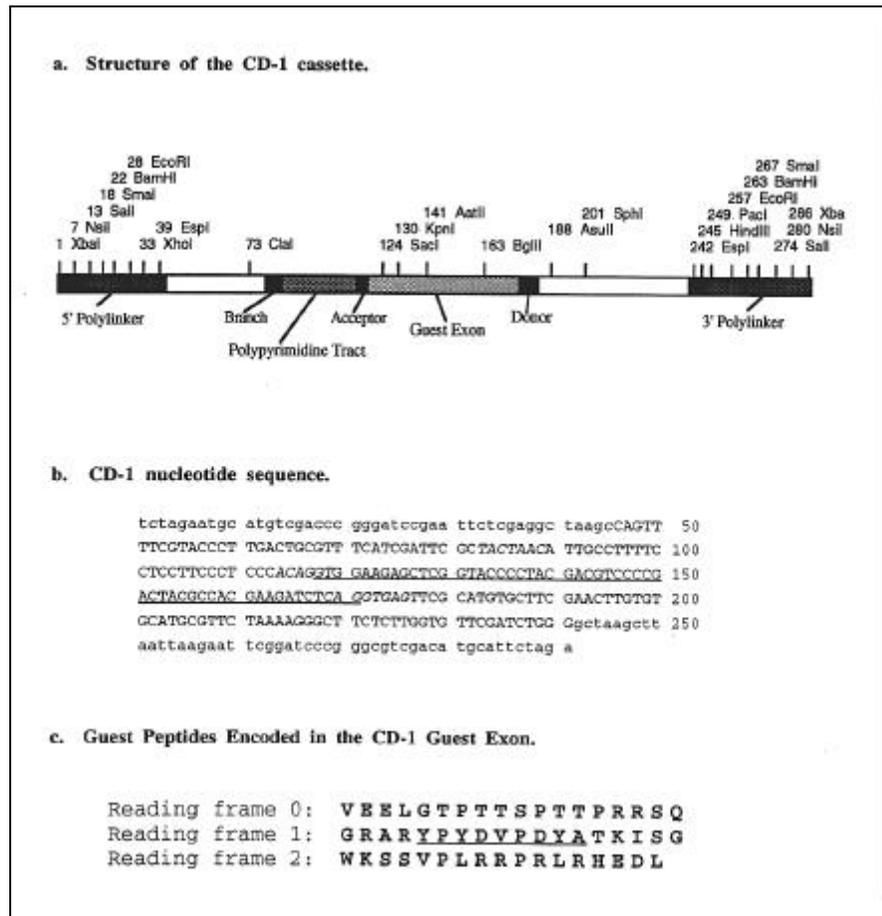


Figure 2. CD-1 cassette. (a) Structure of the CD-1 cassette. (b) Nucleotide sequence of the CD-1 cassette. Nucleotides 1–45 and 242–291 (lower-case letters) are polylinker sequences. Nucleotides 46–116 are derived from the 3' end of intron 2 of the $\alpha 1$ tubulin gene of *Chlamydomonas reinhardtii* (26). The sequence has been modified to include consensus splice branch and acceptor sites, shown in italics. Nucleotides 118–172 (underlined) is the guest exon. It contains no stop codon in any reading frame. The consensus splice donor site is italicized. Nucleotides 173–241 are derived from intron 2 of the $\alpha 2$ tubulin gene of *Chlamydomonas reinhardtii* (26). (c) The amino acids encoded in the CD-1 guest exon. The underlined sequence is the HA epitope recognized by monoclonal antibody 12CA5. The sequence encoding the HA epitope was designed so that it would not contain a stop codon in any reading frame.

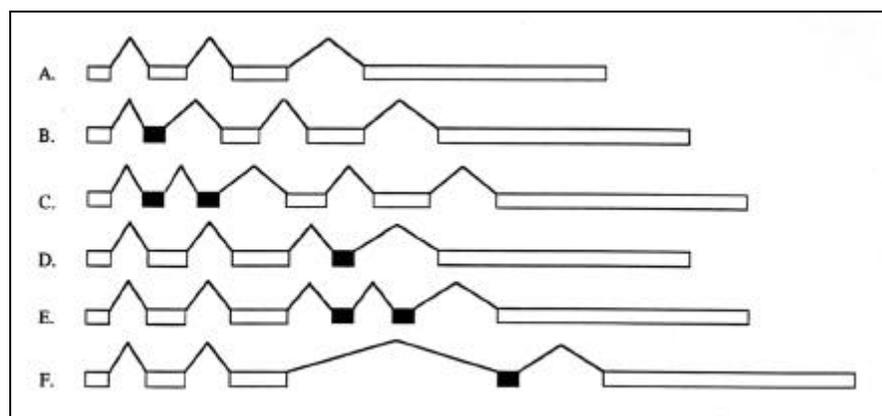


Figure 3. Expected splicing patterns for the five *pf14* constructs. The regions shown extend from the start codon to the termination codon. Open boxes represent the four *pf14* exons. Filled boxes represent the guest exons. (A) Untagged gene. (B) Single tag in intron 1. (C) Double tag in intron 1. (D) Single tag in intron 3. (E) Double tag in intron 3. (F) Single tag with *npt1* insert in intron 3.

Research Reports

pf14 genes were constructed, two with single CD-1 insertions in intron 1 or 3, two with tandem insertions in intron 1 or 3 and one with a single CD-1 insertion in intron 3 within which a complete *E. coli npt1* (kanamycin resistance) gene was inserted upstream of the guest exon. The structures and predicted splicing patterns for these five constructs are shown in Figure 3. In the reading frame determined by intron 1, a Class 0 intron (25) that interrupts the coding sequence between codons, the CD-1 cassette is predicted to introduce the guest peptide VEELGTPTTSPIT-PRRSQ into the gene product; and in the reading frame determined by intron 3, a Class 1 intron that interrupts the coding sequence between the first and second nucleotides of a GGT codon, it is predicted to introduce the guest peptide GRARYPYDVPDYATKISG. The latter sequence contains within it the 9-amino acid HA epitope YPYDVPDYA recognized by monoclonal antibody 12CA5 (10).

Chlamydomonas strains carrying each of the five constructs were obtained by transformation (8). Such

transformants carry ectopic copies of the transforming DNA in their genomes, and so each carries both the transgene and the *pf14* gene that was resident in the recipient. Transformants were obtained in two genetic backgrounds: one with a wild-type *pf14* gene and one with a *pf14* ochre mutation that renders the flagella immotile. All transformants, including the five in the mutant background, had fully motile flagella, indicating that each of

the tagged genes encodes a functional product. As demonstrated in Figure 4, Western blot analysis of the transformants' flagella, probed with polyclonal antibody raised against gel-purified RSP3, showed that the singly tagged genes encode RSP3 species about 2 kDa larger than the untagged protein (89 kDa vs. 87 kDa) and that the doubly tagged genes encode RSP3 species about 4 kD larger than the untagged protein (91 kDa vs. 87 kDa). These

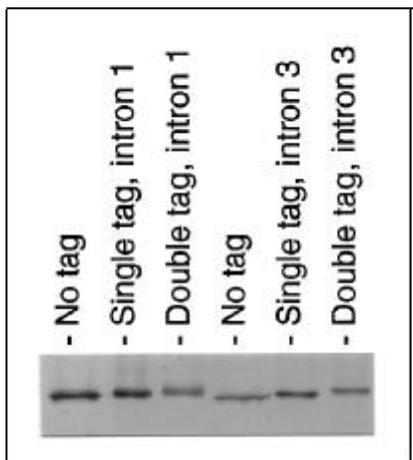


Figure 4. Immunoblots of flagella from singly and doubly tagged transformants of the *pf14* ochre mutant. Transformants were: Lane 1, untagged pKE2-RSP3; lane 2, intron 1 single-tag construct pRS011-1; lane 3, intron 1 double-tag construct pRS011-2; lane 4, untagged pKE2-RSP3; lane 5, intron 3 single-tag construct pRS03-1; lane 6, intron 3 double-tag construct pRS031-2. Based on the mobility of the molecular weight standards β -galactosidase (116 kDa), fructose-6-phosphate kinase (85 kDa) and glutamate dehydrogenase (56 kDa), the apparent molecular weight of the untagged RSP3 was approximately 87 kDa; this agrees well with the previously published value (29).

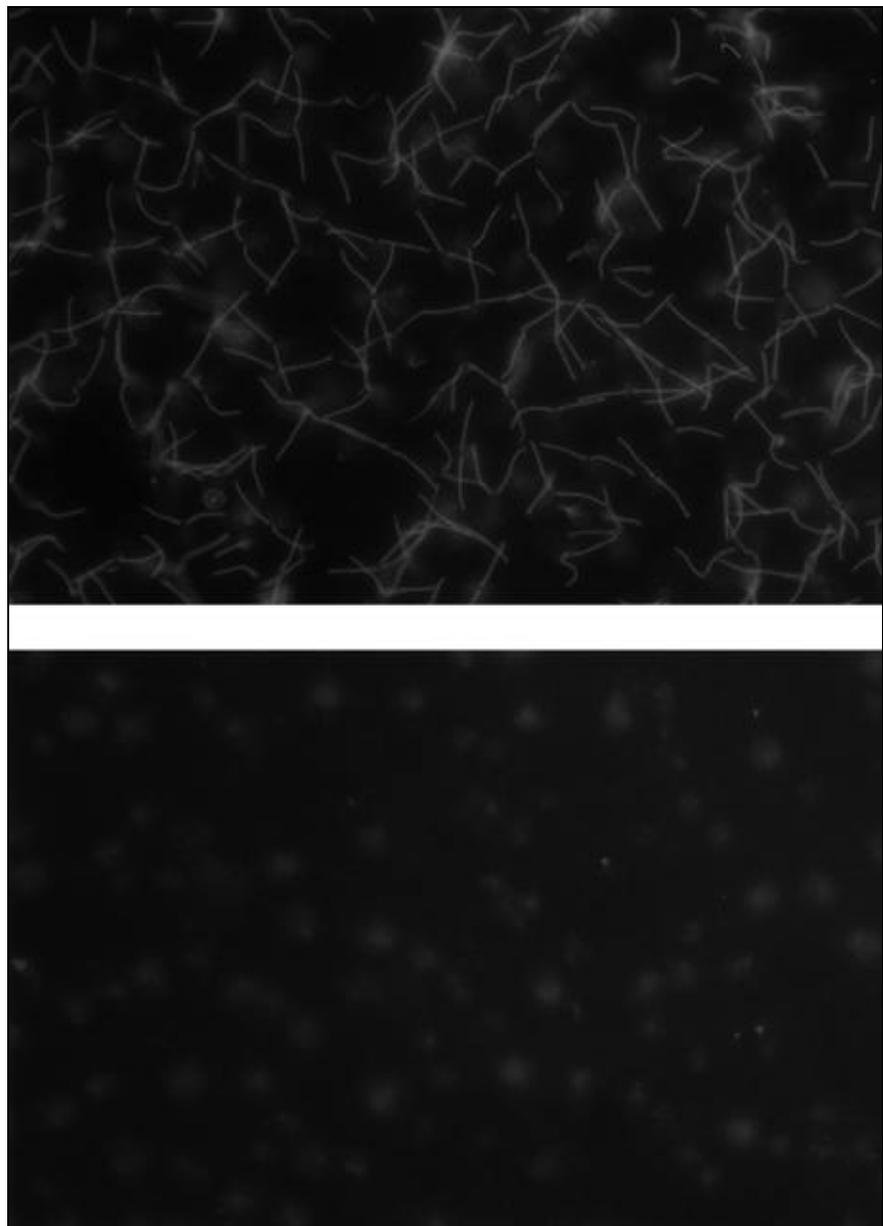


Figure 5. Indirect immunofluorescence micrographs of *Chlamydomonas* cells stained with anti-HA antibody 12CA5. Upper panel: *pf14*- transformed with intron 3 double-tag construct pRS031-2. Lower panel: *pf14*- transformed with untagged *pf14* plasmid pKE2-RSP3.

results are exactly as expected if the singly tagged proteins contain an 18-amino acid guest peptide and the doubly tagged proteins contain two 18-amino acid guest peptides.

Indirect immunofluorescence analysis using antibody to RSP3 revealed specific axonemal staining in all of the CD-tagged strains. Furthermore, monoclonal antibody 12CA5 (specific to the HA-epitope) stained the axonemes of the intron 3 tagged cells but none of the others (Figure 5). Not unexpectedly, the doubly tagged proteins gave brighter signals than the singly tagged proteins; and the transformed *pf14+* strains, in which the tagged proteins have to compete with untagged protein for assembly into flagella, showed less intense staining than the corresponding transformed mutant strains.

CD-Tagging the *Ubx* Gene of *Drosophila Melanogaster*

The *Ubx* gene of *Drosophila melanogaster* contains 3 introns and 4 exons. It encodes a primary transcript of about 77 kb that is alternately spliced in a tissue- and stage-specific manner to yield six different protein isoforms that

function as transcription factors (17, 20). A P-element containing a *Ubx* minigene with shortened introns measuring 0.5, 1.3 and 2.9 kb (27) was CD-tagged at the *Xba*I site in intron 3 to produce an element whose structure is shown in Figure 6. Since Intron 3 is a constitutively spliced Class 1 intron that interrupts the coding sequence between the first and second nucleotides of a GGT codon (17,20), the CD-1 cassette was expected to introduce the

guest peptide GRARYPYDVPDYATK-*ISG*, containing the HA-epitope, into the *UBX* protein. pUMG101-CD1 DNA was transfected into Schneider cells along with plasmid pAKG1, which provides expression of the *S. cerevisiae GAL4* protein to activate transcription of the *Ubx* minigene. After allowing 24 hours for expression of the transfected DNA, the cells were harvested and their proteins were separated by SDS gel electrophoresis and

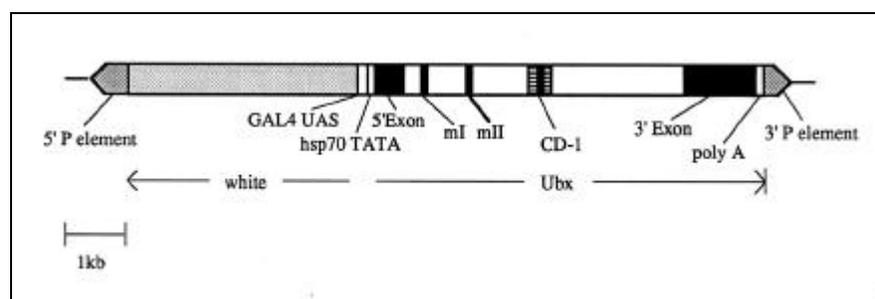


Figure 6. Structure of the CD-tagged *Ubx* minigene. The structure of the P-element-mediated germline transformation construct pUMG101-CD1 is shown as inserted into the *Drosophila* genome. The 5' and 3' P-element ends are indicated by dark stippling. The *white+* marker gene is indicated by light stippling. The *Ubx* exons (5' exon, mI, mII and 3' exon) are indicated as black boxes. The CD tag cassette CD-1 is indicated as a horizontally striped box with the guest exon in black. *GAL4* UAS: upstream activating sequence responsive to *GAL4* protein. *hsp70* TATA: minimal promoter derived from the *Drosophila hsp70* gene. *polyA*: *Ubx* polyadenylation site. The arrows at the bottom of the figure represent the primary transcripts of the *white* and *Ubx* genes.

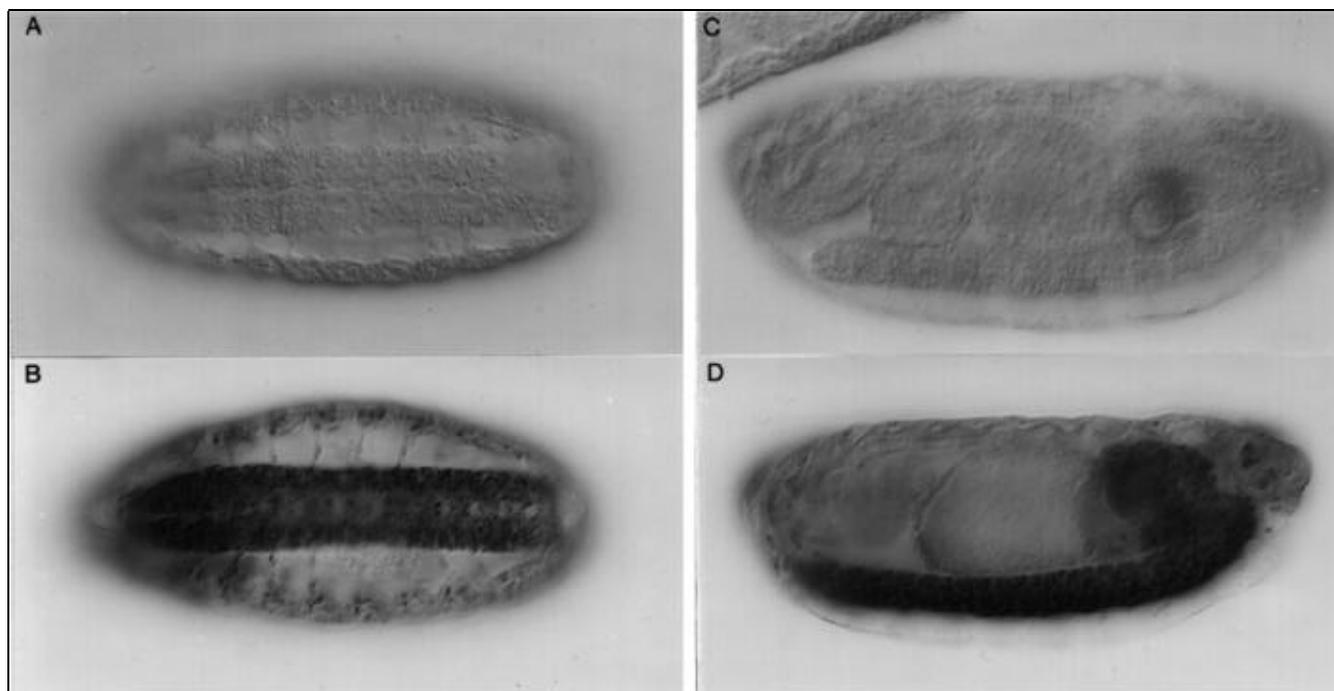


Figure 7. Immunostaining of embryos expressing the CD-tagged *Ubx* minigene. Expression of the CD tag, detected by immunohistochemical staining with monoclonal 12CA5, is visible as red-brown stain in cell nuclei of the nervous system in Panels B and D. (A, C) Untagged control. (B, D) CD-tagged germ line transformants. (A) and (B) show ventral views, with anterior to the left. (C) and (D) show side views, with anterior to the right.

Research Reports

blotted to nitrocellulose. When the blots were probed with antibody 12CA5, an immunoreactive protein band of the expected size was observed (data not shown).

Several *Drosophila* lines with germ line insertions of pUMG101-CD1 were obtained as described under Materials and Methods. One of these, designated CDT 1-11, was crossed to c1003, a line that contains an insertion of the pGawB enhancer trap vector (2) that previously had been shown to express *GAL4* protein in the embryonic nervous system (27). Because *GAL4* regulates the expression of the CD-tagged *Ubx* minigene in CDT 1-11, it was expected that progeny with both the *Ubx* minigene and the pGawB enhancer trap would express the *Ubx* minigene and therefore accumulate epitope-tagged *UBX* protein in the embryonic nervous system. This expectation was confirmed by immunostaining embryos derived from

the cross with the antibody 12CA5. The stained embryos showed exactly the predicted stage and tissue-specific pattern of immunostaining (Figure 7). The CD tag detected with antibody 12CA5 was expressed in the nervous system in the same pattern as ectopic *UBX* protein that was expressed from the minigene and detected with antibody 5C.2B; and within the stained cells it was localized to the nuclei, as expected and confirmed for the *UBX* protein (Figure 7 and data not shown).

DISCUSSION

The results presented here demonstrate that the CD-tagging method can be used to epitope tag the products of cloned genes. When both tagged and untagged versions of the gene were present in the cell, as was the case for the *Ubx* construct and for the *pfl4⁺ Chlamydomonas* constructs, the tagged pro-

tein was observed in the appropriate cellular location. We anticipate that codominance of this kind will be a frequent property of CD-tagged genes.

If CD-tagging is to be a generally useful method, it must work with a wide variety of genes; furthermore, the tagged protein must frequently retain biological function. Although we cannot make empirical generalizations based on the limited data presented here, a number of strong arguments can be made in favor of the expectation that CD-tagged proteins will, in many cases, retain biological function. There exists a very large body of experimental data on conventional epitope tagging in which an epitope-encoding sequence is inserted in-frame into an expressed cDNA, typically at or near the 5' or 3' end of the coding sequence. In recent years numerous proteins have been tagged in this way and shown to possess normal or nearly normal biological

function (for example, References 10, 13, 16, 21, 22). In addition, there are many examples in the literature of protein fusions that have been shown to retain activity for one or both members. In a recently published study, for example, a 3' *lacZ* tag was randomly introduced into yeast genomic DNA, and over a thousand tagged clones were examined to identify those that showed specific localization of the β -galactosidase fusion protein within the cell (4). Approximately 10% of the clones showed such localization, even though most of the fusion proteins carried substantial C-terminal deletions and all carried a C-terminal addition of approximately 1000 amino acids.

Cassette Design Considerations

In designing the CD-1 cassette, one of our goals was to maximize the likelihood that splicing would occur as diagrammed in Figure 1. Thus, we gave the guest exon a size that would not present special challenges to the splicing machinery, and we surrounded it with sequences derived from introns in two highly expressed *Chlamydomonas* genes (the genes encoding $\alpha 1$ and $\alpha 2$ tubulin). The natural sequences were altered at a few sites to create consensus branch, acceptor and donor sites and to increase the U+C content of the polypyrimidine tract. Because the CD-1 cassette contains consensus splice sites, it ought to function effectively whenever the target intron also contains consensus sites. If the target intron departs significantly from consensus, as do many alternatively spliced introns, the outcome is less predictable. Since most introns conform to consensus, however, we expect that most of them will in fact be suitable targets for CD-tagging. Based on the similarity between *Chlamydomonas* splice sites and those of other eucaryotes, we expected that the cassette would function in *Drosophila*, and this expectation was confirmed. For the same reasons, we anticipated that it would function in mammalian cells as well; preliminary experiments (data not shown) indicate that this expectation was also correct.

Because all but the guest exon is spliced out of the transcribed CD cassette, there is great flexibility in terms

of what sequences can be placed in the cassette arms; indeed, the non-exon portions of the cassette ought to accept almost any DNA sequences that do not contain splice sites, including entire functional genes that can be used for selection or as reporters. In demonstration of this feature, we placed a 1252-bp insert into the left arm of CD-1 to produce an expanded cassette that contains an entire functional *npt1* gene—i.e., we placed a new gene in one of the two hybrid introns created by the CD-cassette insert. When transformed into *Chlamydomonas* this expanded cassette functioned just as well as CD-1 itself. We have also shown that when two cassettes are present in tandem, two guest exons are inserted into the gene, leading to a protein containing tandem guest peptides. There is no reason to believe that three adjacent CD-cassettes would not lead to a triple peptide tag, etc. While our intent in designing the CD-1 cassette was to minimally disturb native protein structure and function, it should be clear to the reader that the CD-tagging method can also be used to intentionally engineer new and predetermined functions into proteins.

If an intron in an uncharacterized gene is tagged using the CD-1 cassette described here, the probability of being able to obtain a protein that could be recognized by the 12CA5 antibody is only one in three, because only one intron in three will be Class 1. In order to increase this probability to unity, we have generated derivatives of CD-1 which will tag Class 0 or Class 2 introns in such a way that the tagged protein receives the HA epitope. We have also generated polyclonal antisera to the peptides (Figure 2c) encoded in the other two reading frames of the CD-1 guest exon. Using these new sera we are able to detect tagged proteins no matter what class of intron receives the CD-1 cassette.

Gene Discovery

We have shown that CD-tagging provides a new means to tag the products of cloned genes. But the technology's greatest utility, we believe, will be in the realm of gene discovery. Unlike gene discovery methods in current use, CD-tagging can be used to discover

genes through a primary screen at the true level of function for most genes—the protein level. With this approach, the very process by which a gene is discovered provides specific empirical information about its biological function. This is a very significant advantage over current methods that rely on sequence homologies, rather than experimentation, to infer the possible functions of new genes. We envisage tagging the genome at numerous random sites, either directly (using retroviral or transposon vectors containing the CD-cassette, for example) or indirectly (introducing the CD-cassette into a phage or cosmid library using an appropriately designed bacterial transposon and transforming the tagged library into eucaryotic cells, for example), and then screening individual tagged clones or organisms for the presence of tagged protein. Many primary screens—such as radioimmunoassays, flow cytometry or ELISAs—are possible. If the primary screen is light microscopy of immunostained cells or organisms, as in this paper, then for each new gene the cellular and/or subcellular location of its product would be learned in conjunction with its discovery. Having identified tagged cells or organisms of interest, one could then take advantage of the unique nucleotide tags in the transcript and gene for cloning and sequencing (using reverse transcription polymerase chain reaction [RT-PCR], for example), and, likewise, one could take advantage of the guest epitope to affinity-purify and analyze the gene product.

Of course, not every cell with a CD-cassette in its genome will express an epitope-tagged protein—the cassette must be in an intron, it must be in the correct orientation and the gene must be expressed at a sufficient level in the cells that are screened. Nonetheless, we expect that cells or organisms expressing the CD-tag will not be rare. We have, in fact, commenced a CD-tagging gene discovery endeavor in *Drosophila melanogaster* using a P-element construct to deliver the CD-cassette to numerous locations in the genome. In a significant fraction of the transposants examined to date, we have observed new patterns of embryonic immunostaining (data not shown).

Research Reports

Finally, we emphasize that CD-tagging is particularly well-suited for gene discovery in intron-laden genomes such as that of the human; indeed, the more intron sequence a genome has, the more effective the approach should be. In addition, and in contrast to cDNA-dependent methods for gene discovery, CD-tagging favors the discovery of large genes encoding large proteins, not only because their total intron content is high but also because they have many introns per gene and thus provide numerous distinct sites in the protein for the guest peptide to go.

ACKNOWLEDGMENTS

We thank Bruce Taillon and Peter Berget for numerous valuable discussions. Plasmid pKE2-RS3 and the anti-RSP3 serum were kindly provided by Drs. Dennis Diener and Joel Rosenbaum. The derivatives of CD-1 that tag Class 0 and Class 2 introns so as to insert the HA epitope were constructed by Dr. Paul Gold. Support was provided in part by Grant MCB 9119168 from the National Science Foundation and Grants GM 46903 and HD 28664 from the National Institutes of Health. CT was supported by an NSERC of Canada Postdoctoral Fellowship. The CD-tagging technology is the subject of US and international patent applications.

REFERENCES

1. **Bomze, H.M. and A.J. López.** 1994. Evolutionary conservation of the structure and expression of alternatively spliced *Ultrabithorax* isoforms from *Drosophila*. *Genetics* 136:965-977.
2. **Brand, A. and N. Perrimon.** 1993. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118:401-415.
3. **Buckler, A.J., D.D. Chang, S.L. Graw, J.D. Brook, D.A. Haber, P.A. Sharp and D.E. Houseman.** 1991. Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. *Proc. Natl. Acad. Sci. USA* 88:4005-4009.
4. **Burns, N., B. Grimwade, P.B. Ross-Macdonald, E.-Y. Choi, K. Finberg, G.S. Roeder and M. Snyder.** 1994. Large-scale analysis of gene expression, protein localization and gene disruption in *Saccharomyces cerevisiae*. *Genes Dev.* 8:1087-1105.
5. **Crick, F.H.C.** 1958. On protein synthesis. *Symp. Soc. Exp. Biol.* 12:138-167.
6. **Diener, D.R., L.H. Ang and J.L. Rosenbaum.** 1993. Assembly of radial spoke proteins in *Chlamydomonas*: identification of the axoneme binding domain of radial spoke protein 3. *J. Cell Biol.* 123:183-190.
7. **Di Nocera, P.P., and I.B. Dawid.** 1983. Transient expression of genes introduced into cultured cells of *Drosophila*. *Proc. Natl. Acad. Sci. USA* 80:7095-7098.
8. **Dunahay, T.G.** 1993. Transformation of *Chlamydomonas reinhardtii* with silicon carbide whiskers. *BioTechniques* 15:452-460.
9. **Duyk, G.M., S. Kim, R.M. Myers and D.R. Cox.** 1990. Exon trapping: a genetic screen to identify candidate transcribed sequences in cloned mammalian genomic DNA. *Proc. Natl. Acad. Sci. USA* 87:8995-8999.
10. **Field, J., J.-I. Nikawa, D. Broek, B. Macdonald, L. Rodgers, I.A. Wilson, R.A. Lerner and M. Wigler.** 1988. Purification of a *RAS*-responsive adenylyl cyclase complex from *Saccharomyces cerevisiae* by use of an epitope addition method. *Mol. Cell. Biol.* 8:2159-2165.
11. **Friedrich, G. and P. Soriano.** 1991. Promoter traps in embryonic stem cells: a genetic screen to identify and mutate developmental genes in mice. *Genes Dev.* 5:1513-1523.
12. **Gavis, E.R. and D.S. Hogness.** 1991. Phosphorylation, expression and function of the *Ultrabithorax* protein family in *Drosophila melanogaster*. *Development* 112:1077-1093.
13. **Gloor, S., O. Pongs and G. Schmalzing.** 1995. A vector for the synthesis of cRNAs encoding Myc epitope-tagged proteins in *Xenopus laevis* oocytes. *Gene* 160:213-217.
14. **Gossler, A., A.L. Joyner, J. Rossant and W.C. Skarnes.** 1989. Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* 244:463-465.
15. **Harris, E.H.** 1989. The *Chlamydomonas* Sourcebook. Academic Press, San Diego, CA.
16. **Kolodziej, P. and R.A. Young.** 1991. Epitope tagging and protein surveillance. *Methods Enzymol.* 194:508-594.
17. **Kornfeld, K., R.B. Saint, P.A. Beachy, P.J. Harte, D.A. Peattie and D. Hogness.** 1989. Structure and expression of a family of *Ultrabithorax* mRNAs generated by alternative splicing and polyadenylation in *Drosophila*. *Genes Dev.* 3:243-258.
18. **Krasnow, M.A., E.E. Saffman, K. Kornfeld and D.S. Hogness.** 1989. Transcriptional activation and repression by *Ultrabithorax* protein in cultured *Drosophila* cells. *Cell* 60:597-610.
19. **López, A.J. and D.S. Hogness.** 1991. Immunohistochemical dissection of the *Ultrabithorax* homeoprotein family in *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. USA* 88:9924-9928.
20. **O'Connor, M.B., R. Binari, L.A. Perkin and W. Bender.** 1988. Alternative products from the *Ultrabithorax* domain of the bithorax complex. *EMBO J.* 7:435-445.
21. **Philipson, L.H., A. Kuznetsov, P.T. Toth, J.F. Murphy, G. Szabo, G.H. Ma and R.J. Miller.** 1995. Functional expression of an epitope-tagged G protein-coupled K⁺ channel. (GIRK1). *J. Biol. Chem.* 270:14604-14610.
22. **Rabouille, C., N. Hui, F. Hunte, R. Kieckbusch, E.G. Berger, G. Warren and T. Nilsson.** 1995. Mapping the distribution of Golgi enzymes involved in the construction of complex oligosaccharides. *J. Cell Sci.* 108:1617-1627.
23. **Rubin, G.M. and A.C. Spradling.** 1982. Genetic transformation of *Drosophila* with transposable element vectors. *Science* 218:348-353.
24. **Senapathy, P., M.B. Shapiro and N.L. Harris.** 1990. Computer analysis of protein and nucleic acid sequences. *Methods Enzymol.* 183:252-278.
25. **Sharp, P.A.** 1981. Speculations on RNA splicing. *Cell* 23:643-646.
26. **Silflow, C.D. and J. Youngblom.** 1986. *Chlamydomonas reinhardtii* tubulin gene structure. *Ann. N.Y. Acad. Sci.* 466:18-30.
27. **Subramaniam, V.** 1995. Ph.D. Thesis. Carnegie Mellon University.
28. **Taillon, B.E., S.A. Adler, J.P. Suhan and J.W. Jarvik.** 1992. Mutational analysis of centrin: an EF-hand protein associated with three distinct contractile fibers in the basal body apparatus of *Chlamydomonas*. *J. Cell Biol.* 119:1613-1624.
29. **Williams, B.D., M.A. Velleca, A.M. Curry and J.L. Rosenbaum.** 1989. Molecular cloning and sequence analysis of the *Chlamydomonas* gene coding for radial spoke protein 3: mutation *pf-14* is an ochre allele. *J. Cell Biol.* 109:235-245.
30. **Witman, G.B.** 1986. Isolation of *Chlamydomonas* flagella and flagellar axonemes. *Methods Enzymol.* 134:280-290.
31. **Wright, R.L., J.G. Spanier, S.A. Adler and J.W. Jarvik.** 1989. The nucleus-basal body connector in *Chlamydomonas*: evidence for a role in basal body segregation and against essential roles in mitosis or in determining cell polarity. *Cell Motil. Cytoskeleton* 14:516-526.

Received 19 July 1995; accepted 31 October 1995.

Address correspondence to:

Jonathan W. Jarvik
Department of Biological Sciences
Carnegie Mellon University
4400 Fifth Avenue
Pittsburgh, PA 15213-3890, USA
Internet: jarvik+@andrew.cmu.edu