In vivo protein trapping in Drosophila

Xavier Morin

Date received (in revised form): 19th May 2003

Abstract

The deciphering of complete genome sequences has opened a post-genomic proteomics era. While the sequence of many proteins is now known, attention will increasingly focus on the complex interaction networks that regulate their activity, and the analysis of protein distribution in the cell will be crucial to elucidating their function. A new generation of gene trapping devices, protein trap transposons, offers a way of analysing in vivo the dynamics of sub-cellular distribution of a large number of proteins. Many transgenic lines have already been established and are available. Screens focusing on particular cell compartments can be devised.

INTRODUCTION

Much information on gene expression, regulation and function in Drosophila has come from enhancer and gene trapping, which consists of the random sprinkling over the genome of reporter genes carried by transposable elements and the subsequent analysis of reporter expression.1,2 Loss-of-function phenotypes are often induced by the inherent mutagenicity of enhancer trap transposons, and help to uncover the function. While knowledge of the expression pattern of a given gene can help to formulate hypotheses on its function, a great number of genes are ubiquitously expressed, rendering expression data of little value. The activity and distribution of their protein products, however, can be spatially and temporally modulated within complex networks of interaction with other molecules. As cell biology and physiology become increasingly integrated with developmental biology studies, it becomes obvious that knowledge of the sub-cellular distribution of proteins, and of their dynamics, is essential. A new generation of transposing elements, called protein trap transposons (PTTs), has recently emerged designed to tag randomly a great number of proteins with the green fluorescent protein (GFP), and to follow their distribution at the cellular level, in real time, in the living organism.3

THE PROTEIN TRAP: TAGGING MOLECULES EXPRESSED FROM THEIR ENDOGENOUS LOCUS

While the idea of the protein trap is very simple, and has already been used in cultured mammalian cells,4 it has only recently been applied to Drosophila.3 Rather than trapping the regulatory elements of a gene, which was the principle of enhancer traps, here it is the coding sequences which are primarily targeted. This involves the GFP being used in the form of an additional exon, devoid of start and stop codons, flanked with splicing sequences. There are three versions of the PTT, each of them presenting one of the three possible reading frames at splice junctions. When a PTT hits an intron, it becomes part of the trapped gene (Figure 1). The next phase involves the production of a chimaeric protein, whose distribution is dictated by the localisation domains of the trapped protein. Its expression is controlled by the regulatory sequences of the gene.
A TOOL FOR GENE DISCOVERY AND ANALYSIS

So far, gathering information on protein distribution has relied on the laborious process of generating specific antibodies or artificially over-expressing tagged molecules. This implies a previous knowledge of the protein sequence, and is therefore a targeted, non-random approach. The PTT offers a tool to proceed in the reverse direction: one can focus on a specific cell compartment and screen for new proteins solely on the basis of their localisation. Selected loci can easily be identified by inverse polymerase chain reaction (PCR) due to the presence of a transposon in the locus. About one-third of the GFP-positive insertions are into previously known genes, and another third into predicted CG loci. This, therefore, provides invaluable data on the distribution and localisation of those uncharacterised genes, particularly when their sequence does not contain any conserved domains which would give some indication as to which cell compartment they are targeted to (see Figure 2). Surprisingly, more than a third of the GFP-positive PTT insertions are in regions of the genome where no gene has been predicted by computer algorithms; the PTT may then be used as a gene discovery tool to reveal a number of genes with a non-conventional structure, whose existence is difficult to predict. While mutagenicity of the PTTs has not been studied systematically, lethality seems to be elevated within GFP trap lines. The addition of a new protein motif within existing proteins is likely to modify their three-dimensional conformation and interactions with other proteins. While this may not necessarily alter their

Figure 1: The principle of the protein trap. An artificial exon, coding for a start and stop codon-less GFP, and flanked by splicing sites, is used to search for coding sequences in the genome. Upon insertion in an intron, the transgene is integrated and transcribed within the pre-messenger RNA, spliced and translated to produce a chimaera, which contains the N-terminal and C-terminal parts of the trapped protein separated by the fluorescent tag. In most cases the chimaera retains the sub-cellular localisation domains of the trapped protein.
distribution significantly, as observed for
known molecules, it is likely to induce
phenotypic alterations.

GENERATING MARKERS
FOR CELL
COMPARTMENTS
Insertions which label virtually every
compartment of the cell have been
identified, from nucleoli to synaptic
compartments, via centrosomes and the
mitotic spindle (Figure 2).\(^3\) The
localisation of known molecules tagged
with GFP has confirmed previous
descriptions; for example,
Histone2av::GFP is associated with
chromatin and Lamin::GFP with the
nuclear lamina.\(^3\) Since GFP signals can be
followed in living cells, protein trap lines
can provide excellent markers for cell
compartments during developmental
events (Figure 2). Bobinnec and
colleagues have used a PDI::GFP
chimaera (protein disulphide isomerase)
to describe the dynamics of the
endoplasmic reticulum during early
embryonic development.\(^5\) These GFP
markers will be very useful in studying the
fate of specific structures in different
mutant backgrounds.

OBTENTION OF PROTEIN
TRAP EVENTS
Obtention of protein trap events is by
design a rare event: P-element
transposons rarely hit introns,\(^6\) and those
that do will produce a GFP chimaera only
when the insertion is in the correct
reading frame and orientation. In
addition, the approximately 20% of
intronless genes contained in the
Drosophila
genome will never be hit.
These limitations therefore impose
selection for protein trap events in a
screen being based on direct detection of
the GFP signal, either under the dissecting
microscope equipped with
epifluorescence or with an automated
embryo sorter. Huge numbers of animals
can be screened in a relatively short
period of time. New genes are discovered
solely on the basis of their expression.
This in turn relies on two constraints: the
trapped gene must be expressed at the
stage of screening (usually embryonic or
first instar, which offers a good signal
background ratio), and at a level sufficient
to allow detection, which also means that
a significant number of trapped genes will
not be detected. On the positive side, this
type of screen is easy to handle. It does

---

Figure 2: Examples of distribution of
protein trap fusions. A) Time-lapse analysis
of the distribution of a DEK::GFP fusion over
an 8-minute period of time during division 12
in the syncytial blastoderm. DEK is a histone-
interacting protein which labels the
chromatin, and is here observed from late
interphase, through mitosis and into early
interphase of the next cycle. B) Axons are
labelled by the GFP::CG17238 fusion in this
ventral view of a late (st16) embryo.
CG17238 does not show any homology to
known proteins. C) The same fusion as in B)
labels elongating microtubules in the mitotic
spindle during late prophase (left) and
metaphase (right) cells of the syncytial
blastoderm. D) A GFP::CG8895 fusion labels
the endoplasmic reticulum, observed here in
late metaphase/early anaphase in the
syncytial blastoderm. CG8895 belongs to the
reticulon family. E) GFP::Tropomysin2
fusion, seen in larval body wall muscles,
shows a characteristic sarcomeric
distribution.
not rely on preselection of lethal insertions, lines are established only after a positive event is detected and there are no false-positives, since the selection is direct.

Using P-element-based transposons, it was possible to detect insertions at a frequency of up to one in 1,600 screened larvae.\(^3\) With a transposition frequency of 15 per cent in the mutator line, and considering that only one in six intronic insertions will be in the correct orientation and reading frame, this means that approximately one in 40 new insertions is in an intron of a gene expressed at a detectable level in larval stages. Compared with an intron content of 1/6th in the fly genome,\(^7\) this is consistent with a strong bias against P-element insertions in introns.

**A BRIGHT FUTURE?**

Setting these limitations aside, the protein trap approach offers a unique way to characterise the distribution and dynamics of a large number of proteins. Hundreds of different lines have now been generated and are being characterised in several laboratories. Saturation of the genome will not be achieved with P-element based PTTs, but several hundred independent loci should ultimately be hit, providing an unprecedented wealth of information at the protein level.

Improved efficiency may also come from the use of other families of transposable elements, such as piggyBac, Hermes or Minos, which can, in addition, be used in other non-drosophilid and even non-insect species. Upcoming data suggest that piggyBac-based vectors are actually biased towards hits in introns and are not as prone as P-elements to hit hotspots.\(^8\) Several groups are currently developing new versions of protein trap vectors based on piggyBac (RS Mann, M Buszczak, personal communications) and on other colour variants of the fluorescent protein. Future generations of automated embryo sorters may also provide a means to detect lower levels of fluorescence.

With increasing numbers of lines available to the fly community, it is becoming possible to screen large existing collections for molecules located in specific cell compartments (eg synaptic junctions), or at specific times in the life of a cell (eg centrosomes or mitotic spindles).

**FURTHER INFORMATION AND LINKS**

Mutator lines, as well as insertions in a number of characterised genes and predicted CGs, are now referenced in FlyBase (http://flybase.bio.indiana.edu) and are available from the Bloomington Stock Center (http://flystocks.bio.indiana.edu/protein-trap-GFP.htm). Several laboratories are generating their own collections. The contents of two collections can be browsed online: http://flytrap.med.yale.edu contains a list of lines generated in Cooley’s (Yale, USA) and in Chia’s (IMCB, Singapore and King’s College London, UK) laboratories. A selection of lines generated in the laboratory of Debec (Villefranche-sur-Mer, France) can be viewed at http://biodev.obs-vlfr.fr/gavdos/protrap.htm. Time-lapse movies of a number of lines are presented.

**References**


5. Bobinnec, Y., Marcaillou, C., Morin, X. and

