The clock gene *period* in the housefly, *Musca domestica*: a molecular analysis

Thesis submitted for the degree of
Doctor of Philosophy
at the University of Leicester

by

Alberto Piccin
Department of Genetics, University of Leicester

February 1998

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

ACKNOWLEDGEMENTS

I wish to express my gratitude to the people who, directly or not, contributed to the work expounded in this thesis.

Firstly, I am greatly indebted to my supervisor, Bambos Kyriacou, for his advice, scientific and not, and for happily putting up with my changes of mood associated with "writing up". Special thanks to Ezio Rosato, whose help and continuous encouragement cannot be overestimated. I wish to thank David Chalmers, who supervised my work at the early time in lab 136, when my "Gilson thumb" was not up to scratch, and who instilled enthusiasm through continuous injections of optimism. Mike Hennessy has been an excellent guide in my passage from white wine to bitter, and a valuable teacher in many technical aspects of my work, while Martin Couchman helped me in the gargantuan task of acquainting with computers. Raymond Dalgleish rescued me from the most fearsome problem ever encountered by writing up students, the "unrecoverable disk error".

I am grateful to Martin Williamson, who provided the *Musca* genomic library, to Helen Dobbs, Alessia Bonini and Giulietta Di Benedetto who performed the *D. melanogaster* transformations, to Kathleen Siwicki and Joceline Noveral for their help with the immunochemistry and to Ralph Stanewsky for providing the Per antibody and helpful tips on RNase protection.

Thanks also to Rudi Costa for sending me to the English exile, and to Susanna Campesan for the Tuesday morning summary of Barry Norman's review and for allowing me to maintain my mother tongue to an intelligible standard. And thanks to Min Shi for gastronomical support and for running the lab smoothly, and to all the people in lab 136 for their help and support.

Bin Zhao has been an exceptional guide for my first steps towards a new life in a foreign land, and Naci Sevkal a supplier of good music and good humour.

Finally, I am indebted to Paula Wan for having made the time of writing up, certainly not shorter but more light and pleasant.

ABBREVIATIONS

ATP	Adenosine 5'-triphosphate
A_x	Absorbance at a wavelength of x nm
bp	Base pair
BSA	Bovine serum albumin
°C	Degrees Celsius
CT	Circadian time
Ci, μCi	Curie, micro-
cm	Centimetres
DEPC	Diethyl pyrocarbonate
DNase	Deoxyribonuclease
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
g	Acceleration due to gravity
g, mg, μg	Grams, milli-, micro-
HEPES	N-(2-hydroxyethyl)piperazine-N'-2-ethanesulphonic acid
IPTG	Isopropyl-β-D-thiogalactopyranoside
kb	Kilobase pairs
l, ml, μl	Liters, milli-, micro-
M, mM, μM	Molar, milli-, micro-
min	Minutes
nm	Nanometres
OD	Optical density
o/n	Overnight
PAGE	Polyacrylamide gel electrophoresis
pfu	Plaque forming units
PMSF	Phenylmethyl sulphonyil fluoride
RACE	Rapid amplification of cDNA ends
RNA	Ribonucleic acid
RNase	Ribonuclease
RPA	Ribonuclease protection assay
SDS	Sodium dodecyl sulphate
TCA	Tri-Chloro-Acetic Acid
TEMED	N,N,N',N'-tetramethyl-ethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
UTR	Untranslated region
ZT	Zeitgeber time

ABSTRACT

The *period* gene (*per*) of *Drosophila melanogaster*, which lies at the core of the fly's circadian clock, controls a number of biological rhythms, including the circadian periodicity of locomotor activity and adult emergence from the pupal case. Levels of *per* gene products cycle with a 24 hour period, as required for a clock molecule. Furthermore, a temporal delay between the peak expression of *per* mRNA and protein suggests the existence of a negative feedback loop through which Per protein regulates the synthesis of its own mRNA.

In an attempt to determine the extent to which clock molecules and mechanisms are conserved among dipterans, I cloned the *per* homologue from the housefly, *Musca domestica*. The *Musca per* gene encodes a protein of 1072 amino acids in which areas of high similarity with *Drosophila per* are interspersed by non-conserved stretches. Areas of high conservation include the amino terminus, the PAS domain and the region surrounding the per^S mutation site of *D. melanogaster*. Successively I assessed *Musca per* functionality in the *D. melanogaster* circadian machinery. Behavioural analysis of transgenic per^0 fruit flies expressing the *Musca per* homologue, demonstrates that the housefly *per* is able to replace endogenous *per* functions in the host's clock system.

Expression of *per* products was investigated in both housefly and transgenic fruitfly. Housefly *per* transcript levels display daily changes in abundance, similar to those observed in wild-type *D. melanogaster*. Circadian oscillations were also found in *per* levels in *D. melanogaster* transformants, albeit with a lower amplitude of those described for wild type. Expression of Per protein in both housefly and *D. melanogaster* transformants does not show the daily fluctuation characteristic of wild type *D. melanogaster* Per.

This comparative study suggests that the current model based on the negative feedback loop may be inadequate to explain the molecular mechanism underlying the circadian clock.

CONTENTS

'I rise at eleven, I dine about two,
I get drunk before seven, and the next thing I do,
I send for my whore, when for fear of a clap,
I spend in her hand, and I spew in her lap...
I storm, and I roar, and I fall in a rage,
And missing my whore, I bugger my page.
Then crop-sick all morning I rail at my men
And in bed lie yawning till eleven again.'

('Regime de Vivere' by John Wilmot, Earl of Rochester, 1647-1680, in one of the earliest literary accounts on circadian rhythmicity).

CHAPTER 1

Introduction

Introduction

GENES AND BEHAVIOUR

The study of the heritability of behavioural traits is one of the more mature branches of the relatively young discipline of genetics. Born in the right place (the centre of the British empire) at the right time (the height of the colonial expansion and a time of social turmoil), behavioural genetics was soon abused and exploited to give rise to social darwinism and biological determinism.

The scientific and technological achievements from the Renaissance to the present, gave to the 19th century western society, an unbreachable faith in Science. Very soon, all nature's secrets would have been unveiled and man, like Laplace's demon, would have been able to predict the Universe's own destiny. But more earthly matters could be settled with help from science. Nineteenth century society, like any other, was certainly preoccupied with the perpetuation of the social *status quo*. And, like any other society, it turned to religion to justify it. The roman poet Virgil chants about the divine "pedigree" of Augustus, descended from Eneas, son of Venus, who fled from Troy after the defeat and, after numerous dangers and peregrinations, under his divine mother's guidance settles down in Latium where he is told of the great fate reserved for his progeny. People were only too happy to pay taxes to an emperor with divine ancestry, having only *circenses* in return. Similarly, science, the new religion of the 19th century, declares that the order of the new, reformed society in which every class privilege is abolished, is predetermined by people's innate attitudes. Variation of traits is present in every population. Black cows breed black cows and white cows breed white cows, likewise entrepreneurs nurture entrepreneurs and dumb peasants give birth to dumb peasants. The heritability of these attitudes is solely responsible for the fact that inter-class exchanges are not so commonly observed. But the pseudo-scientific biological determinism went even further, offering hypocritical alibis for the dismissal of any form of social assistance as deeply damaging, because it allows the support and reproduction

of individuals of no use to the society. The variation in innate capabilities is not only confined to individuals but is also observed between races and nations. Consider a passage from a letter of H. F. Osborne, president of the American Museum of Natural History, to the New York Times in 1924:

> "The northern races invaded the countries to the south, not only as conquerors but as contributors of strong moral and intellectual elements to a more or less decadent civilisation. Through the Nordic tide which flowed into Italy came the ancestors of Raphael, Leonardo, Galileo, Titiano; also, according to Gunther, of Giotto, Botticelli, Petrarca and Tasso. Columbus, from his portraits and from busts, whether authentic or not, was clearly of Nordic ancestry."

In this extract, which is more in the line of bar-room logic than of a scientific claim, this eminent scholar sweeps away in an instant 5000 years of history of the Mediterranean basin. Obviously, the Museum of Natural History must have been located far away from the History Museum.

The study of the rational quantification of the transmissibility of human behaviours and cognitive capabilities was initiated in the second half of the nineteen century by F. Galton. His work is an example of how the current ideology and beliefs of the society can bias the conclusion of scientific reports. In his book "Hereditary genius" (1869) he states that the skills and abilities of different classes of persons, ranging from judges to wrestlers, are highly hereditary. From here to a classification of races according to their abilities intended as expression of their biological potential does not take long:

> "The average standard of the Negro race is two grades below the Anglo-Saxon", the Australian native is at least one grade below the African, the Lowland Scot and the English North-Country man is "decidedly a fraction of grade superior to the ordinary English" and the race that inhabited Attica between 530 and 430 BC were "very nearly two grades higher than our own, that is about as much as our race is above the African Negro".

The ideas expounded by Galton became very popular among the bourgeoisie of Europe and North America, and gave birth to a school of thought called Eugenics, according to which the amelioration of the human species was possible by a careful selection of marriages and birth control. The theme of biological determinism was not only confined to scientific literature: in the group of novels "Rougon-Marquart" for example, Zola tells the story of a family descended from two lovers of a woman, one a laborious farmer, and the other an idle waster. The branch of the family originating from the peasant have the some industrious attitude of the ancestor, while from the other man descend a line of criminals and social misfits.

The first world war cooled down the topic in Europe, where the biological inferiority of the millions of working class men sent to be slaughtered was suddenly no longer under question, but eugenics continued to be popular in the United States, which were facing waves of unwanted immigration and the problems caused by the integration of the various ethnic minorities. Eugenics became an official creed in Germany after the rise of the Nazi party, with its myth of the Aryan race and its brutal attempt to create a pure German breed by eliminating all "contaminations". After World War II understandably, eugenics became a taboo, tightly associated with National-Socialism and the Holocaust, and a similar fate was reserved to the study of the heritability of behaviour and cognitive abilities. Only in more recent years has the study of genetically determined behaviour been reinitiated, mainly with non-human species (not only because of the intrinsic difficulties of scientific investigation in our species, but also because of the hostility of various human rights groups towards this kind of research).

Behaviours are, by definition, quantitative characters. Even in simple reflexes, in which the response is "all or nothing", individual variability in the level of the stimulus required to trigger the response exists. It was therefore widely accepted that behaviours must be polygenic traits (Wilcock, 1969), in which a certain number of genes contribute to a given behavioural response. The isolation of individual genes affecting behavioural traits would therefore be particularly difficult. The identification of the *per* mutants in *Drosophila* for example, demonstrated that single genes can have a profound effect on quantitative aspects of behaviour (Konopka and Benzer, 1971).

According to Wilcock (1969), in order to be defined as "behavioural", a gene must be psychologically informative, that is, it must affect the ability of elaborating a

neurological response to a set of given stimuli, and not merely modify the behaviour indirectly as a consequence of "peripheral abnormalities", for example anatomical defects on sensory organs. As Wilcock put it: "no deep psychological meaning can be educed from the fact that a color-blind man drove through traffic lights set at red". This view is undoubtedly that of a psychologist. Evolutionary biologists and geneticists have a different opinion. A given locus is regarded as behavioural when it influences some aspects of the normal ethology of the organism. According to this view, the molecular function of the gene is unimportant for its classification as "behavioural". What matters is that the man passes through the traffic light set on red, with possible serious consequences for his fitness.

Progress in molecular biology though, has pointed out that this classification is somehow inadequate. Consider the locus *Shaker* (*Sh*) in *Drosophila* which encodes a potassium channel subunit (Iverson *et al.*, 1988; Timpe *et al.*, 1988). The originally isolated *Sh* mutant shakes when subjected to CO_2 anaesthesia (Kaplan and Trout, 1969) and therefore does not have particular ethological significance, since anaesthesia is not a natural factor for wild type flies, and this phenotype does not have any selective impact on natural populations. A different *Sh* point mutant though, exhibits deficits in experience-dependent behaviours (Cowan and Siegel, 1986) and in gustatory responses (Balakrishnan and Rodrigues, 1991). Therefore it is probably more pertinent to talk about "behavioural mutants" rather than "behavioural genes". Some genes cannot be considered "behavioural" altogether since only one particular mutation at the locus gives a behavioural phenotype while all the others can have different effects. This is related to the pleiotropy of gene function. When a particular gene product serves more than one function, we can reasonably expect that some mutations will affect one of these functions but not the others. The fact that a locus affects not only behaviour, but also other characters will not diminish the importance of the information we can gain by studying it in a behavioural perspective. To quote Hall (1994) "we have to ask the fly: 'would you evolve a given class of transcription factor, or a cyclic-nucleotide phosphodiesterase, or a receptor tyrosine kinase, and use that entity just *once*-during only a restricted period of neural development, or in only one portion of the mature and functioning nervous system?' If we are willing to listen, the fly is telling us: 'of course not.' " (page 25).

A large number of genes affecting behaviour have been identified in *Drosophila*. A rather incomplete list of examples includes the gene *foraging* (*for*, de Belle *et al.*, 1989a: 1989b), which determines the feeding habits of the larval stage of the fly, *sevenless*, the mutant of which lacks the phototaxis response to UV light (Tomlinson and Ready, 1986), *bendless* (Thomas and Wyman, 1982) and *passover* (Thomas and Wyman, 1984), both of which affect the jump reflex, an escape response triggered by switching off the light, *cacophony* (Kulkarni and Hall, 1987), *dissonance* (Kulkarni *et al.*, 1989) and *fruitless* (Gailey and Hall, 1989), which influence the fly's mating behaviour, the genes *dunce*, *rutabaga* and *amnesiac* (Quinn and Greenspan, 1984), three loci associated with learning and memory formation, and finally *period* (Konopka and Benzer, 1971) and *timeless* (Sehgal *et al.*, 1994: Vosshal *et al.*, 1994), two genes which are involved in the regulation of the behavioural rhythmicity.

There is little doubt that the behavioural genetics "flagship" is the *per* gene. It was the first behavioural gene cloned by mutation (Bargiello *et al.*, 1984; Reddy *et al.*, 1984), although not the first behavioural gene isolated, which was the *egg-laying hormone* gene in Aplysia (Mahon *et al.*, 1985). However *per* provides us with the best example of the many approaches which can be applied to the study of a behavioural gene, ranging from the biochemical to the evolutionary.

BIOLOGICAL CLOCKS

The habitat of all surface-dwelling organisms is a cyclically changing environment. The most obvious of these cyclic changes are brought about by the rotation of the Earth on its own axis and around the sun. Organisms must therefore adapt to the environmental variations by modifying their physiology and behaviour. Since this requires a gradual alteration of the metabolic rate, the ability to anticipate environmental changes is a very desirable characteristic for any living being which has to face this regular challenge. In order to achieve this, organisms are "equipped" with a time measuring device, an endogenous clock which allows them to time their physiology to the cyclical variations of the environment. As a result, a variety of physiological and behavioural rhythms have evolved, such as circannual (yearly, eg. the annual migration of birds,

reviewed by Gwinner, 1986), circalunar (monthly), circadian (daily) and circatidal (twice daily). The most evident and the most extensively studied among these rhythms, is the circadian cycle.

In order to be classified as clock-controlled or endogenous, a rhythmic phenotype must be maintained in constant environmental conditions. This characteristic ensures that the cyclic variable is indeed regulated by an endogenous pacemaker rather than simply being a response to the fluctuating environment. This is the first property required for an endogenous clock, namely that its oscillations persist in absence of external stimuli.

The second property is temperature compensation: the period of oscillation of a clock must be constant within the range of temperature in which the organism lives. This is not trivial. The clock's oscillations can ultimately, be reduced to chemical reactions and the effect of temperature on chemical reactions is well known. For example, consider that the rate of chemical reactions normally doubles with every 10° C rise in temperature (Q_{10} = 2). Thus, if no temperature compensation mechanisms were deployed, a clock with a 24 h cycle at 20° would tick much faster at 25°, perhaps giving a period of about 16 hours. The Q_{10} of circadian clock oscillations is generally between 0.9 and 1.2 (Takahashi, 1993).

Finally, the ability to respond to environmental cues (zeitgebers; literally, time-givers) the most common of which are light and temperature, is also a cardinal property of a biological clock. The endogenous oscillator can therefore be reset by the cyclic variations in the external environment.

A typical clock consists of three main components: an input pathway, which conveys information from the environment and can synchronise the oscillator to the external zeitgebers, a molecular oscillator, the internal source of time keeping, and an output pathway, which links the oscillator to cycling aspects of physiology and behaviour and triggers the rhythmic responses.

Input pathway: the input pathway ensures that the difference between the period of the circadian clock and the length of the environmental cycle will be reset after every oscillation. The ability of the input component to modify the phase of the oscillator can be represented by the phase response curve (PRC). To calculate the PRC, the organism under study is typically entrained in a 12 h light-12 h dark cycle (LD 12:12) so

the phase of its clock is synchronised to the phase of the LD (zeitgeber) cycle. The time is conventionally measured as zeitgeber time (ZT), where ZT=0 corresponds to the onset of the entraining stimulus (light-on in our example) and ZT=12 to its offset. After being transferred to free running conditions (constant darkness or DD), the organism is then given pulses of light at different times of the subjective day or night, and the effect of these stimuli on the phase of the clock are analysed. Light pulses during the subjective day do not influence the clock while light stimuli at various times of the subjective night do so to different extents (Saunders, 1982). Light given in the early hours of the subjective night reset the circadian clock to "light-off", so resulting in a phase delay, while a pulse administered in the late subjective night will "trick" the clock into "believing" it is ZT=0, and will result in a phase advance (Saunders, 1982).

Light is certainly the most effective of all the external stimuli able to mediate phase resetting of the circadian clock, and most work on the clock resetting described in literature makes use of this form of stimulus. However temperature has also been shown to play a part in the entrainment of the circadian clock (Chiba *et al.*, 1993; Gooch *et al.*, 1994) and possibly other sensory perceptions can have similar effects. It is generally accepted that phase resetting in mammals involves cellular structures within the suprachiasmatic nucleus (SCN) that are only responsive to light (Morin, 1994; Takahashi, 1995) but it seems that conditioned stimuli coupled to light through associative learning, can elicit the cellular response in the SCN (expression of the "immediate-early" transcription factor FOS) and the phase resetting normally induced by the light itself (Amir and Stewart, 1996). The information associated with the entraining stimulus is then passed, via the input pathway, to the oscillator.

Molecular oscillator: the molecular oscillator is the core of the biological clock. The precise mechanism that allows an organism to measure time is not yet known, but several theoretical models have been presented which try to explain the biochemical basis of an endogenous oscillator. All these models have in common the fact that the cycling level of a certain "state variable" (the variable which describes the state of the system) is unequivocally related to the circadian time. In the different models, different state variables and mechanisms for their endogenous, self-sustained oscillation are hypothesised. The early models proposed, can be grouped into three main categories:

metabolic feedback loops, consisting of oscillations in the energy metabolism and/or other biosynthetic pathways originated by cross-coupling a network of various biochemical reactions and by the appropriate value of allosteric constants and enzyme turnover rates (Winfree, 1967; Hess and Boiteux, 1971; Cummings, 1975; Selkov, 1975); transcriptional models, in which long, polycistronic replicons (chronons) are sequentially transcribed (Erhet and Trucco, 1967). The transcription rate would be slowed down by the requirement for the newly transcribed RNA to diffuse into the cytoplasm and being translated before its product can activate the transcription of the next cistron. The chronon would control a multiplicity of cellular mechanisms by means of different cistrons transcribed at various times of the day. The product of the last cistron of the chronon would then restart the whole process; membrane models, where the rhythms originated by means of a mechanism in which a certain molecule (possibly an inorganic ion) is actively transported into an organelle until it reaches a critical concentration which triggers structural changes in the membrane, like lipid or protein composition or functional changes in the associated transport channels. Passive diffusion would then re-establish the even concentration of the molecule on both sides of the membrane and would, in turn, bring the membrane back to its original structure (Njus *et al.*, 1974; Sweeney, 1974).

These models reflect the particular scientific spirit of the time they have been proposed: the advances in biochemistry and the discovery of the prokaryotic operon in the early '60s certainly prompted the first two groups of models while the third coincided with the advances in membrane biochemistry and physiology. More recently, the advent of molecular biology and the characterisation of the first clock mutants from *D. melanogaster* and *Neurospora crassa*, led the way to a new clock model, which is in fact a hybrid between the metabolic feedback loop and the transcriptional models: the transcription negative feedback loop (Hardin *et al.*, 1990).

According to this model, a "clock gene" is transcribed and the RNA accumulates in the cytoplasm where it is translated into protein. When a threshold concentration is reached, the protein undergoes post-translational modifications and is then translocated into the nucleus where it inhibits the transcription of its own gene. This causes the level of its mRNA to drop, and in turn the level of the protein follows the same destiny. The fall in protein concentration removes the transcriptional inhibition and a new cycle of

transcription can start. A delay mechanism, such as post-translational modification of the protein, is necessary in order for the system to became a self sustained oscillator rather than reaching a state of equilibrium. The "clock protein", apart from controlling the synthesis of its own mRNA, is also active in the transcriptional regulation of clock controlled genes (*ccg*), whose expression determines the circadian phenotypes.

Output pathways: they consist of the signalling cascades which convey the oscillatory informations from the pacemaker to all the overt rhythms. Since in the negative feedback loop model, the core of the pacemaker is represented by a transcriptional regulator, the output pathways invariably begin with *ccg* transcription.

It is relatively easy to separate the input pathway from the other two components of the clock, by simply segregating the organism under study in constant environmental conditions, but it is much more difficult to discriminate between the pacemaker and its output. The fact that a particular molecular rhythm is maintained in a DD regimen is not a valid criterion because the rhythm could simply be an expression of the circadian oscillator rather than its core. Similarly, mutations which obliterate a rhythm could affect the oscillator or just a component of the output pathway. Given these difficulties, some criteria have been established that must be satisfied by a pacemaker component. Aronson *et al.* (1994) summarise these rules in five points: (i) mutations in one of the pacemaker components affect the properties of the clock; (ii) its concentration (or activity) must oscillate in a self-sustained manner with the proper periodicity; (iii) induced variations in its concentration (or activity) will, via negative feedback, change its ratio of synthesis or degradation; (iv) its oscillatory phase is re-synchronised by shifts in the light-dark entrainment regimen and, conversely, changes in its concentration (or activity) will reset the overt rhythms; (v) block of its oscillation will result in cessation of the rhythmicity. These criteria constitute a valid prerequisite for defining a particular candidate molecule as a clock component.

GENETICS OF RHYTHMICITY

Genetics is a powerful tool for the analysis of many biological phenomena. Its strength consists in that with the genetic approach, we can begin to work knowing virtually nothing about a particular character. All that is required is a measurable phenotype and a good dose of patience for the screening of mutants. It is therefore not surprising that the genetic analysis of biological rhythms has provided a significant contribution in the understanding of the mechanisms at the core of the biological clock.

Initially the methodology used was typical quantitative genetics: one of the first applications of this approach was the selection of "early" versus "late" emerging strains of *D. pseudoobscura* (Pittendrigh, 1967). Pittendrigh obtained 2 strains after 50 generations of selection which eclosed about 4 hours apart. The change though, was due to a modification in the phase of the output rather than to an effect on the period of the oscillator. Subsequently, the mutagenesis approach was adopted instead, aimed at the identification of single genes. A list of the clock-affecting genes so far identified is given in table 1.1. Not surprisingly, the organisms for which most information is gathered, are the two classic genetic workhorses, *Neurospora crassa* and *Drosophila melanogaster*. A circadian phenotype in *N. crassa* was first described by Pittendrigh *et al.* (1959). It consists in a rhythmic change in the morphology of the culture. In the late night phase, aerial hyphae raise from the growth surface and originate asexual spores (conidia), while at other time of the day the culture remain as an undifferentiated layer of mycelia. The period of this rhythm is about 21.6 h at 25° C. The first *Neurospora* circadian mutants were isolated by Feldman and Hoyle in 1973 in a screening of mutagenised cultures. They all affected the periodicity of growth of the fungus and mapped to the same locus, called *frequency* (*frq*).

Two years earlier, Konopka and Benzer isolated three circadian mutants in *D. melanogaster* (Konopka and Benzer, 1971). As in *Neurospora*, the mutants mapped to the same locus, which was given the name *period* (*per*). Other mutations were identified by other phenotipic criteria, and only later shown to affect clock properties. This is for example the case of *disconnected* (*disco*) whose mutations affect the normal development of the fly's visual system (Steller *et al.*, 1987), resulting in an essentially blind mutant. *disco* flies appear arrhythmic under constant darkness (Dushay *et al.*, 1989)

Organism	affected rhythmic phenotype, wild type period length (h)	Gene	Allele	Period length (h)	Dominance	Other clock properties affected
Mesocricetus auratus Golden hamster	Locomotor activity, 24	*tau*	τ	τ/+ = 22 τ/τ = 20	Semidominant	-
Mus musculus	Locomotor activity, 23.3	*Circadian locomotor output kaput*	*clock*	*clock*/+ = 24.4 *clock*/ *clock* = 27.3	Semidominant	-
Chlamydomonas reinhardtii	Phototaxis, stickiness to glass, 24	*w-c*$_1$	-	21	-	-
		90-	-	21	-	-
		s-	-	18	-	-
		period-1	*per-1*	26-28	Dominant	-
		period-2	*per-2*	26-28	Recessive	-
		period-3	*per-3*	26-28	-	-
		period-4	*per-4*	26-28	Semidominant	-
Neurospora crassa	Conidation, 21.5	*arginine-13*	*arg-13*	19	Recessive	-
		chain elongation	*cel*	variable	Recessive	Temperature compensation
		chrono	*chr*	23.5	Semidominant	Temperature compensation
		clock affecting-1	*cla-1*	27	Semidominant	-
		frequency	*frq*^{1}	16	Semidominant	-
			frq^2 = *frq*^{4} =*frq*^{6}	19	Semidominant	-
			frq^3	24	Semidominant	Temperature compensation
			frq^7 = *frq*^{8}	29	Semidominant	Temperature compensation
			frq^9	variable	Recessive	Temperature and nutritional compensation
			frq^{10}	variable	Recessive	Temperature and nutritional compensation
		period-1	*prd-1*	26	Recessive	Temperature compensation
		period-2	*prd-2*	25.5	Recessive	-
		period-3	*prd-3*	25	Recessive	Temperature compensation
		period-4	*prd-4*	18	Semidominant	Temperature compensation
Saccharomyces cerevisiae	budding	*GTS1*	-	-	-	-

Organism	affected rhythmic phenotype, wild type period length (h)	Gene	Allele	Period length (h)	Dominance	Other clock properties affected
Drosophila melanogaster	Eclosion, locomotor activity, 24	*Andante*	*And*	24-26	Semidominant	-
		Disconnected	$disco^{1}$=$disco^{2}$ =$disco^{3}$	arrhythmic	Recessive	Entrainment
			$disco^{1656}$	arrhythmic	Recessive	Entrainment
		period	per^{01}=per^{02} =per^{03}	arrhythmic	Recessive for rhythmicity, semidominant for period	Evening eclosion peak in LD
			per^{04}	arrhythmic	Recessive for rhythmicity, semidominant for period	Evening eclosion peak in LD
			per^{S}	18-20	Semidominant with per^{+}	Very early evening eclosion peak in LD
			per^{L1}=per^{L2}	28-30	Dominant to per^{0}, semidominant with per^{+}	Temperature compensation
			per^{Clk}	22.5	Semidominant	Early evening eclosion peak in LD
		timeless	*tim*	arrhythmic	-	-
		phase angle-2	*psi-2*	-	-	Early phase of eclosion
		phase angle-3	*psi-3*	-	-	Early phase of eclosion
		soiree	*sre*	arrhythmic	-	Dark active in LD cycle
Drosophila pseudoobscura	Eclosion, locomotor activity, 24	(not named)-A	cl^{7}=cl^{9}	arrhythmic		
		(not named)-B	cl^{8}=cl^{10}	arrhythmic		
Lucilia cuprina Blowfly	Eclosion, activity	(not named)		arrhythmic		

Table 1.1. List of mutations known to affect clock related rhenotypes (modified from Dunlap, 1993). Mutations at the level of the energetic metabolism have been omitted.

although some mutants retain a residual, weak rhythmicity (Dowse *et al.*, 1989). They also display an unusual behaviour in light-dark cycles (Hardin *et al.*, 1992b). It is noteworthy that other genetically blind flies, such as *norp-A* mutants, show an almost normal rhythmic behaviour (Dushay *et al.*, 1989), meaning that the effects of the *disco* mutation on rhythmicity are not due to the impossibility of entrainment.

The search for new clock mutants continues, and has provided further material, such as *timeless* (Sehgal *et al.*, 1994; Vosshal *et al.*, 1994), which, as we will see, has contributed significantly to a deeper understanding of the molecular mechanisms of circadian timing.

Two clock mutations have also been identified in mammals. The golden hamster *tau* mutation shorten the circadian rhythm of wheel-running activity (Ralph and Menaker, 1988), while the mouse *Clock* mutation affects the rhythm in the opposite direction. In homozygous *Clock* mutants the initial weak rhythmicity eventually develops into arrhythmicity after two weeks of free-running conditions (Vitaterna *et al.*, 1994).

Clock mechanism in *Neurospora crassa*

The gene *frequency* (*frq*) represents one of the best characterised example of a clock gene. The study of *frq* has grown in parallel to that of *per*, with findings from each of the two fields often contributing to shed light on the other.

The filamentous fungus *Neurospora crassa* displays a clear circadian rhythm in its developmental growth cycle. In particular environmental conditions (growth on solid surface under low external CO_2 concentration, [Sargent and Kaltenborn, 1972]), *Neurospora* grows as undifferentiated mycelium for most of the organism's circadian cycle, but at some times of the day it elaborates aerial hyphae which will produce asexual conidia. This developmental cycle has been shown to recur every 21.6 h, with conidiation occurring at the end of the subjective night. In an attempt to identify genes involved in the regulation of the *Neurospora* circadian clock, Feldman and Hoyle (1973) used this phenotype for their mutagenesis screening and isolated several mutations which affected the conidiation rhythms. Coincidentally all mutants mapped at the *frq* locus.

The *frq* locus gives origin to four transcripts, two of which are devoid of clock functions. The two "clock" transcripts measure about 4 and 4.5 kb, contain long 5' untranslated regions and both give rise to two forms of Frq protein, a 989 aa long (Frq^{1-989}) and a 890 aa long ($Frq^{100-989}$) form, differing for choice of starting codon. Initiation of translation in the shorter form occurs at AUG^{100}, with respect to the longer form (Garceau *et al.*, 1997). The Frq^{1-989} protein mediates rhythmicity at the higher end of the temperature range of clock permissivity (which goes from 15° to 34°C, Francis and Sargent, 1979), while $Frq^{100-989}$ appears more functional at the lower end (Liu *et al.*, 1997). Both protein forms contain a nuclear localisation domain, a DNA-binding region and an acidic region (Lewis and Feldman, 1993; Aronson *et al.*, 1991), suggesting a possible role of Frq in transcriptional modulation. Frq also contains a Thr-Gly/Ser-Gly repeat region, thereby showing a weak homology with the *D. melanogaster* clock protein Per.

The levels of *frq* transcript and protein oscillate in a circadian manner, with a periodicity reflecting the rhythm of conidiation of the strain analysed (Aronson *et al.*, 1994; Garceau *et al.*, 1997). In particular the peak and trough of *frq* accumulation are reached at CT0-5 and CT14-18, respectively, while the peak of Frq follows that of its transcript by 4-6 h (CT 4-8) and its minimum is recorded at CT20. In the arrhythmic mutant strain frq^9 (in which a one bp deletion results in a truncated, non-functional protein) the level of the *frq* transcript is erratic and does not show a circadian oscillation. Furthermore, *frq* levels in this strain are about two to three times higher than peak levels in frq^+. During its accumulation, Frq protein is progressively phosphorylated. This could be related to a mechanism by which the phosphorylation state of Frq can direct the protein's own degradation (Garceau *et al.*, 1997).

This cycle of *frq*/Frq expression displays the typical characteristics of a feedback loop, promoting *frq* from the role of simple accessory clock molecule to that of true clock component. A frq^+ strain transformed with a *frq* open reading frame under the control of a quinic acid (QA)-inducible promoter, shows wild type periodicity in the absence of the inducer, but in presence of QA, which results in constitutive expression of *frq*, periodicity of banding is replaced by constant conidiation (Aronson *et al.*, 1994). No rescue of rhythmicity was obtained in a frq^9 strain transformed with the same QA-induced construct, after constitutive expression of *frq* with various concentrations of

QA. In short, cycling expression of *frq* rather than its mere presence, is necessary for displaying the rhythmic phenotype. Additionally, expression of *frq* from the inducible construct, was shown to suppress expression of the endogenous *frq* transcript (Aronson *et al.*, 1994), threby closing the feedback loop. It follows that induced *frq* expression, by bringing *frq* level to a minimum, can reset the phase of the clock to CT14-16, as Aronson *et al.* (1994) unequivocally demonstrated.

One of the salient characteristics of a circadian clock is its entrainability by environmental factors (zeitgebers), the most powerful of them being light. Light cues perceived through a photoreceptor must be conveyed to the clock via a signal transduction pathway, and here act to change the level of a state variable. In *Neurospora*, the effect of light on the pacemaker activates *frq* transcription (Crosthwaite *et al.*, 1995). A light pulse in the early evening will cause a delay in the circadian clock by raising *frq* levels at a time when they are supposed to fade, with the consequence that some extra time is required for Frq levels to fall below the threshold of *frq* repression. Similarly, a pulse in the late subjective evening will increase *frq* levels thereby anticipating the transcription rise due to Frq degradation, and resulting in a phase advance. It is not clear how light exerts its effect on *frq*, but a significative step in this direction has been made by dissecting clock functions in the mutant strains *wc-1* and *wc-2*. In these blind mutants, which lack all photoresponses typical of *Neurospora* (Harding and Turner, 1981), neither *frq* nor Frq are detected in DD at any time and race tube analysis fails to detect conidiation rhythmicity (Crosthwaite *et al.*, 1997). Exposure to light does not elicit *frq* expression in the *wc-1* mutant, but *frq* induction is achieved in the *wc-2* strain. The amount of *frq* transcript though, drops rapidly and its transient appearance is not able to start the feedback loop; this points to a role for *wc-2* in enhancing *frq* expression (Crosthwaite *et al.*, 1997). On the other hand, the *wc-1* mutant displays the ability to maintain endogenous *frq* rhythms, at least for a few cycles, once a first burst of *frq* expression has been induced from a QA-inducible *frq* construct (Crosthwaite *et al.*, 1997). Since a functional *wc-1* is required for sustained rhythmicity of the oscillator, its functions are not limited to phototransduction to the oscillator; *wc-1* though, cannot be considered a true element of the feedback loop. Both WC-1 and WC-2 contain a PAS domain, which is thought to mediate protein-protein interactions (Huang *et al.*, 1993), and represents a region of homology with the *Drosophila* Per protein and with the

mouse *Clock* gene (King *et al.*, 1997). PAS domains have also been found in several plant phytochromes and bacterial photoreceptors. As pointed out by Crosthwaite *et al.* (1997), this remarkable homology could represent an evolutionary link between clock and photoreceptor proteins, indicating that circadian clocks evolved from the molecular machinery that was used to adjust the cell metabolism in response to light.

Mammalian clock molecules

The molecular dissection of mammalian clocks lags noticeably behind that of the more experimental-friendly organisms such as *Drosophila* or *Neurospora*. Mammalian studies have traditionally been aimed at understanding the physiology of the circadian system. Nevertheless some noteworthy data exist, and the recent cloning of a mouse candidate pacemaker gene (King *et al.*, 1997) will undoubtedly mark a starting point for the molecular characterisation of the mammalian clock.

In 1988, Ralph and Menaker isolated a spontaneous golden hamster mutant which displayed a 22 h locomotor activity rhythm in DD. The mutation proved to be autosomal and semidominant, with the periodicity in homozygous individuals reduced to 20 h. The golden hamster *tau* mutant has contributed to the study of the neuroanatomy of the mammalian circadian system, but given the paucity of genetic markers in hamster, no molecular characterisation of *tau* has been possible. Rather than attempt to map and isolate the *tau* gene, it was easier to start afresh with a mutagenesis screening in mouse, an organism with adequately developed genetics. A mutant was isolated in which the wild-type 23.7 h circadian rhythm was increased to 24.8 h in heterozygotes and to 27.3 h, followed by the loss of locomotor activity periodicity after about two weeks in a DD regimen, in homozygotes (Vitaterna *et al.*, 1994). The locus was given the name of *circadian locomotor output cycles kaput*, thankfully shortened into *Clock*. The autosomal *Clock* gene has recently been cloned (King *et al.*, 1997). It spans about 100 kb of genomic sequence and contains 24 exons which give an open reading frame of about 2.5 kb. The *Clock* mutation consists of an A→T transversion at the 5' splice donor site of intron 19, resulting in the loss of the whole exon 19 from the mutant Clock protein. Interesting features of the Clock protein include a glutamine rich stretch, a basic

Helix-Loop-Helix (bHLH) domain and a PAS region, and indicate that Clock might function as a transcriptional regulator. The presence of PAS domains in clock associated molecules (*per*, *Clock*, *wc-1* and *wc-2*) suggests that these dimerisation motifs represent an evolutionary conserved feature of clock proteins.

THE *period* GENE

Genetics of *per*

The *period* (*per*) gene was isolated in 1971 in an EMS-mutagenesis designed to screen for mutants in the rhythmic eclosion of the imago from the pupal case (Konopka and Benzer, 1971). The emergence of the adult is not constant throughout the whole day, but occurs preferentially in the morning (Roberts, 1956). The rhythm is maintained in constant environmental conditions: the eclosion is "gated" to the early hours of the subjective day and the 24 h periodicity between successive peaks is temperature independent. Three mutant lines were isolated in which the rhythmicity was altered. Their period was either shortened to 19 h, speeded up to 29 h or completely abolished. Further analysis demonstrated that the same effect was discernible on locomotor activity. Locomotor activity in *D. melanogaster* is a true clock-regulated phenotype (Roberts, 1956) with a period of 24 h. The period of locomotor activity of the three mutant lines had the same value than the period of eclosion, 19 h, 29 h and aperiodic, respectively. Genetic mapping of the mutations, showed that they were indeed three alleles of the same X-linked locus, placed between *zeste* and *white*. The locus was named *period* and the three alleles per^S (S = short), per^L (L = long) and per^0, respectively.

per was later demonstrated to influence other time-related phenotypes. The male of *D. melanogaster* show a complex courtship behaviour during which he extends one wing and vibrates it, so generating a lovesong. An analysis of this lovesong (Cowling and Burnet, 1981) demonstrated that some acoustic components of the lovesong are species-specific. Furthermore, one of these components, the length of the interpulse interval (IPI, the interval occurring between successive pulses of hums) varies in a rhythmic fashion

displaying a period of about 1 minute in *D. melanogaster* (Kyriacou and Hall, 1980). In per^S, per^L and per^0 this rhythmicity is shortened, lengthened or abolished, respectively (Kyriacou *et al.*, 1980; 1982; 1986; 1989). Also the 10-day embryo-to-adult development time is affected by *per* mutations: in constant light, per^0 and per^S flies develop faster than per^+, which is on the other hand faster than per^L (Kyriacou *et al.*, 1990).

Several more *per* alleles have been discovered, like the arrhythmic $per^{0.1}$ (Hamblen-Coyle *et al.*, 1989), the long period mutant per^{L2} (Konopka, 1987) and the short period per^{Clk} (Dushay *et al.*, 1990; 1992). A number of *per* mutants were also generated by *in vitro* mutagenesis (Baylies *et al.*, 1992; Rutila *et al.*, 1992) and, interestingly, most of the new variants showed short rhythmicity.

Like most X-linked genes, *per* is dosage compensated. per^+/Y males have the same free-running period as per^+/per^+ females (Smith and Konopka, 1982). The period of females carrying three copies of the *per* gene is decreased by about half an hour while in two dose males (equivalent to four copies in a female) the period is shortened by one hour. The reduction in period length reaches a maximum at about 1.5 h less than wild-type period (Smith and Konopka, 1982), after which, extra copies of *per* no longer alter the periodicity. This observation suggested that the 5 h reduction of the original per^S mutation was due to increased activity of the mutant protein (Cote and Brody, 1977; Baylies *et al.*, 1987) rather than to overexpression. The new *in vitro* generated short period mutants (Baylies *et al.*, 1992; Rutila *et al.*, 1992) though, revealed that short period alleles are relatively easy to obtain and should therefore be viewed as loss-of-function mutations.

Molecular analysis of *per*

The period gene was cloned independently in two laboratories in 1984. At the time, *per* was known to map in the region 3B-3C of the X chromosome (Young and Judd, 1978), between the loci *zeste* (*z*) and *white* (*w*). Bargiello and Young (1984), cloned this genomic fragment by chromosome walking (see fig. 1.1), using as starting point the distal boundary of a deletion in the region 3B2-3C6 (*Df(1)62d18*, Judd *et al.*,

1972) which fails to complement the *per*^{01} mutation (Young and Judd, 1978). The second group (Reddy *et al.*, 1984) cloned the DNA in the *zeste-white* region by initial microexcision of polytenic chromosomes. In both cases, subclones of this region allowed the mapping of breakpoints of various relevant chromosomal aberrations which either did or did not affect behavioural rhythmicity, and *per* functions could thus be restricted to a small genomic location.

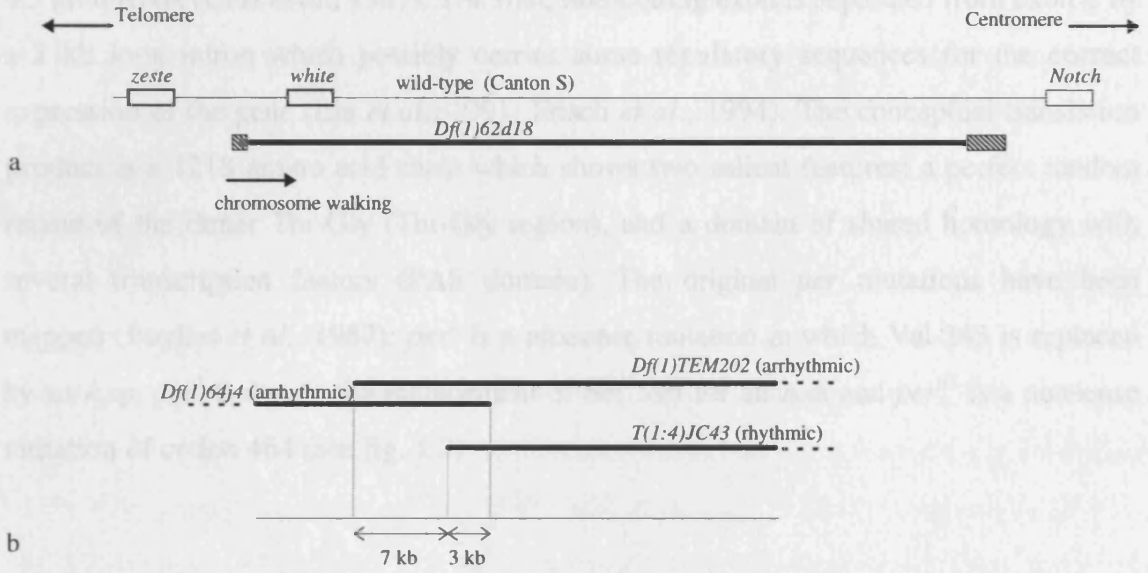


Figure 1.1 Diagram illustrating the strategy of Bargiello *et al.* (1984) for the cloning of the *per* gene. a) A DNA clone (box with diagonals) hybridising with a *Notch* probe was retrieved from the arrhythmic deficiency strain *Df(1)62d18*. Using the portion of this clone homologous to region 3B1-2 (distal to *Notch*) as a probe, a chromosome walking with a Canton S genomic library allowed to isolate ≈ 90 kb of DNA from the 3B interval. This DNA was used to molecularly characterise several deletions previously known to affect *per* functions.
b) Comparison of different, partially overlapping deletions, circumscribed *per* activity to a 7 kb fragment.
Solid bars represent deleted DNA fragments. Drawings are not scaled.

The DNA fragment isolated by Bargiello and Young (1984) was then used to probe polyA RNA isolated from a Canton S strain and shown to hybridise with a 4.5 kb transcript. The same DNA fragment was able to rescue circadian rhythmicity in *per*^{01} flies (Bargiello *et al.*, 1984), thereby demonstrating that it carried the appropriate *per* sequence information. Reddy *et al.* (1984) restricted *per* activity to a slightly bigger fragment of genomic DNA which was then used to probe Northern blots of a wild type

Drosophila strain. Four transcripts were identified which appear in wild type but not in a *per*$^{-}$ deficiency. Reddy *et al.* (1984) failed to detect cycling in the level of the 4.5 kb transcript, and concluded that the adjacent 0.9 kb mRNA, which showed reduced abundance at night compared with daytime expression, was implicated in the fly's rhythmic functions. Subsequent germline transformation experiments confirmed that the 0.9 kb transcript was not essential for clock functions and that the 4.5 kb transcript was indeed the *per* transcript (Zehring *et al.*, 1984; Hamblen *et al.*, 1986).

In *D. melanogaster* the primary transcript contains eight exons and gives rise to a 4.5 kb mRNA (Citri *et al.*, 1987). The first, non-coding exon is separated from exon 2 by a 2 kb long intron which possibly carries some regulatory sequences for the correct expression of the gene (Liu *et al.*, 1991; Frisch *et al.*, 1994). The conceptual translation product is a 1218 amino acid chain which shows two salient features: a perfect tandem repeat of the dimer Thr-Gly (Thr-Gly region), and a domain of shared homology with several transcription factors (PAS domain). The original *per* mutations have been mapped (Baylies *et al.*, 1987): per^{L} is a missense mutation in which Val 243 is replaced by an Asp, per^{S} is due to the replacement of Ser 596 for an Asn and per^{01} is a nonsense mutation of codon 464 (see fig. 1.2).

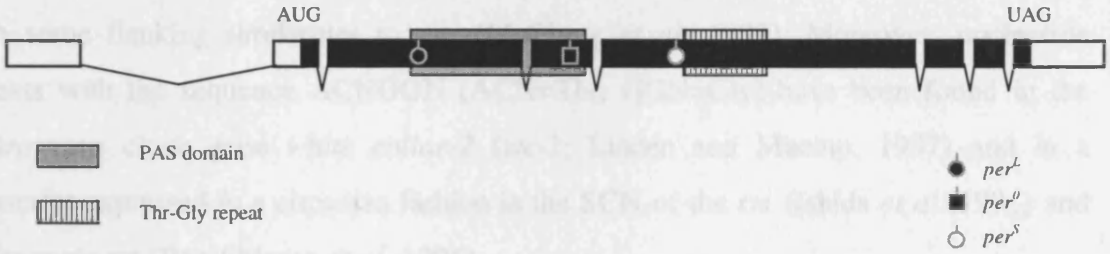


Figure 1.2. Molecular organization of the *per* gene. Filled and empty boxes represent coding and non-coding exons, respectively.

The Thr-Gly repeat region consists of about 20 Thr-Gly pairs and is present in all of the *Drosophila* species analysed, although the length of the motif can differ greatly between species (Peixoto *et al.*, 1993). Furthermore, the Thr-Gly region is polymorphic in length in *D. melanogaster* (Costa *et al.* 1991, 1992) and *D. simulans* (Rosato *et al.*

1994). The striking simplicity of the repeat region is not the only reason for the interest it stimulated. The repeat is similar to the Ser-Gly repeats found in mammalian proteoglycans (Bourdon *et al.*, 1985; Ruoslahti 1988) and several biochemical findings initially seemed to confirm that the PER protein was indeed modified by glycosylation (Reddy *et al.*, 1986; Bargiello *et al.*, 1987). Since proteoglycans are known to mediate cell-cell interaction, it was thought that a glycosylated form of PER mediated intercellular communications in the fly's CNS, and by this means exerted its influence on the fly's rhythmicity. The role of PER in cell-cell coupling was apparently demonstrated by Bargiello *et al.* (1987) who reported differential intercellular diffusion of a physiological dye in the different *per* mutants. This theory was short-lived as independent experiments failed to confirm PER's role in cell-cell communications (Siwicki *et al.* 1992a, Flint *et al.* 1993) and furthermore, recent immunological studies showed that PER is not subjected to extensive glycosylation (Edery *et al.*, 1994).

The repeat is not indispensable for basic *per* function since a gene devoid of it, is still able to rescue rhythmicity in a per^{01} fly (Yu *et al.*, 1987b), but this region is nevertheless implicated in the control over the ultradian cycling of the IPI in the male lovesong (Yu *et al.*, 1987b), and in the temperature compensation of the circadian locomotor phenotype (Ewer *et al.* 1990; Sawyer *et al.*, 1997; Peixoto *et al.*, 1998). Remarkably the *Neurospora* clock gene *frequency* (*frq*) has also a short Thr-Gly repeat with some flanking similarities to *per* (McClung *et al.* 1989). Moreover, nucleotide repeats with the sequence ACNGGN (ACN=Thr, GGN=Gly) have been found in the *Neurospora* clock gene *white collar-2* (*wc-2*; Linden and Macino, 1997) and in a transcript expressed in a circadian fashion in the SCN of the rat (Ishida *et al.* 1991) and of the mole rat (Ben-Shlomo *et al.* 1996).

The Thr-Gly region has an abnormally high mutation rate. At least eight different length variants are found in natural populations of *D. melanogaster* (Costa *et al.*, 1991;1992), five in *D. simulans* (Rosato *et al* 1994), two in *D. pseudoobscura* (Costa *et al.*, 1991) and every *Drosophila* species has a different degenerate repetitive Thr-Gly sequence (Peixoto *et al.*, 1992; 1993; Nielsen *et al.*, 1994). The mechanism by which this region mutates at such high frequency seems to include a series of duplication-deletion events (Costa *et al.*, 1991; Peixoto *et al.*, 1992; Rosato *et al.*, 1994) and indeed all the Thr-Gly regions sequenced from *Drosophila* species seem to originate from reiterated

expansions of either a pentapeptide encoding motif or a Thr-Gly encoding motif (Nielsen *et al.*, 1994). Moreover a model has been proposed which explains the origin of all the different *melanogaster* length variants by means of seven duplication or deletion steps (Costa *et al.*, 1991; Rosato *et al.*, 1996). The mutability of this region which appears to be important for the thermal compensation and pattern of circadian behaviour (Ewer *et al.*, 1990; Sawyer *et al.*, 1997; Peixoto *et al.*, 1998) and which also affects the period of the lovesong cycle (Yu *et al.* 1987b; Wheeler *et al.*, 1991) makes the Thr-Gly repeat a region possibly important for the ecology and evolution of *Drosophila.* Coyne (1992) has promoted *per* to the ranks of a "speciation" gene.

The second region of special interest has been called PAS after Per, Arnt and AhR, and Sim, the proteins in which the domain was first recognised, and consists of a ≈270 amino acid long domain which contains two degenerate 51 amino acid repeats. The gene *single-minded* (*sim*) encodes a bHLH transcription factor required for the normal development of the mild line cells of the *Drosophila* CNS (Crews *et al.*, 1988; Nambu *et al.*, 1991), while the human *aryl hydrocarbon receptor nuclear translocator* (*arnt*) encodes for the nuclear translocator of AhR (Hoffman *et al.*, 1991), a protein which mediates a series of cytotoxic responses following the exposure to several environmental pollutants (such as polycyclic and aromatic compounds, heterocyclic amines and polychlorinated hydrocarbons). It is believed that AhR is a ligand-activated transcription factor (Burbach *et al.*, 1992) which, after binding to its ligand, is translocated into the nucleus by Arnt. Thus, Arnt and the activated AhR, form a heterodimer complex via the PAS domain. Complex formation is necessary for DNA binding activity (Burbach *et al.*, 1992). A bHLH region is also found upstream of the PAS domains of both Arnt and AhR (Hoffman *et al.*, 1991; Burbach *et al.*, 1992).

Possibly, different responses to various kind of signals are mediated by Arnt. Formation of heterodimers with different partners could activate different sets of genes. Indeed it seems that Arnt has the capability of forming heterodimers with a wide range of PAS proteins *in vitro*, while the other PAS proteins appear to be more selective in their partner specificity (Swanson *et al.*, 1995). The partner choice influences the DNA target sequence recognised by the heterodimer (Swanson *et al.*, 1995). Basically, Arnt appears to possess strong activating properties which can be recruited to a specific promoter via protein-protein interactions with another factor (Lindebro *et al.*, 1995). It is interesting

to note that a recently cloned murine *sim* homologue encodes a protein which binds to Arnt with higher affinity than AhR and whose tissue-specific expression matches that of Arnt (Ema *et al.*, 1996), and that Arnt seems to mediate some cytological responses to hypoxia in mouse cells in partnership with the Hypoxia Inducible Factor α (HIF-α, Li *et al.*, 1996). The possibility exists that these two presumptive Arnt partners are indeed the same protein since a *Drosophila* homologue of the human HIF-α, the bHLH-PAS protein Similar, has been shown to share some sequence homology to *sim* (Nambu *et al.*, 1996).

Unlike the other PAS proteins, Per does not contain a bHLH region nor does it carry any other known DNA binding domain but it is believed that Per can form heterodimers with other PAS proteins through its PAS domain and in this way act as a transcriptional regulator. The PAS region of Per (PerPAS) is indeed able to mediate interactions with peptide molecules such as PerPAS, ArntPAS and SimPAS *in vitro* and in mammalian cells (Huang *et al.*, 1993; Lindebro *et al.*, 1995).

The temporal expression pattern of Per can provide additional information regarding its molecular functions. The expression of both RNA and protein cycles with a 24 h period (see fig 1.3). The mRNA level peaks at about ZT 14-15 (two-three hours after lights off), then dampens rapidly and a minimum is reached towards ZT 4 (Hardin *et al.*, 1990). The phase of the protein fluctuation is shifted about 6 h later with respect to the mRNA. The peak of protein expression occurs in the late night at ZT 22 while the trough is recorded at ZT 8 (Siwicki *et al.*, 1988; Zerr *et al.*, 1990; Edery *et al.*, 1994). Per also undergoes post-transcriptional modifications. As detected on SDS-PAGE, Per is gradually phosphorylated between the time of its initial appearance and disappearance (Edery *et al.*, 1994).

In agreement with rules established for biological clocks, the molecular cycling continues in free-running conditions. RNA expression in per^{01} mutants does not show any circadian fluctuation regardless of the environmental conditions. per^{01} RNA is constitutively expressed with an abundance intermediate between the peak and trough levels in wild-type flies (Hardin *et al.*, 1990), while in per^{S} mutants, the phase of the RNA and protein cycles is shifted several hours earlier in LD (Hardin *et al.*, 1990; Edery *et al.*, 1994). The similar effect of *per* mutations on molecular and behavioural cycles suggests that Per may regulate its own transcription via a negative feedback loop.

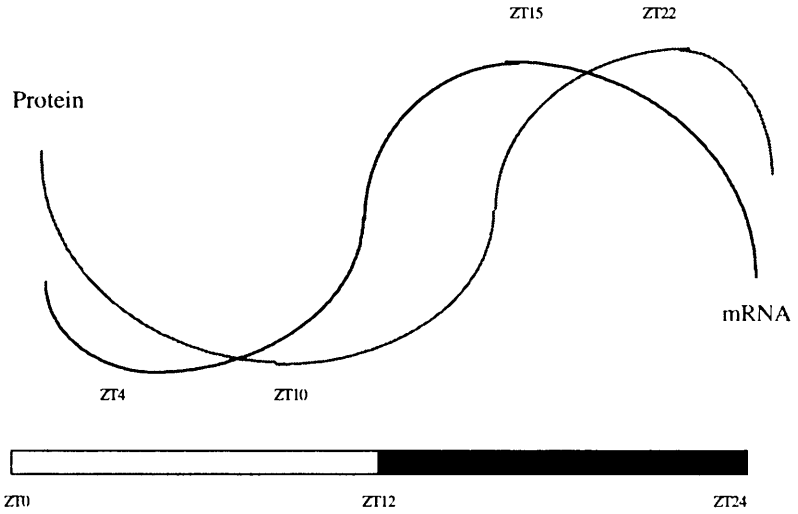


Figure 1.3. Temporally regulated expression of *per* mRNA and protein.

Consistent with this view, the cycling of *per* RNA has been shown to be regulated at the transcriptional level, since 1.3 kb of upstream sequence are sufficient to direct cycling of a reporter gene in wild-type flies (Hardin *et al.*, 1992a). Moreover, constitutive expression of Per in the photoreceptors from a *ninaE* promoter results in the repression of endogenous *per* RNA expression (Zeng *et al.*, 1994). At the subcellular level, Per is exclusively localised in the cytoplasm during the accumulation phase, between ZT 10 and 17, is then found in both cytoplasmic and nuclear compartments at ZT 18, and is predominantly nuclear for the rest if its circadian cycle (Curtin *et al.*, 1995). The per^S mutation apparently does not influence the time of nuclear entry in LD conditions, while in per^L flies, Per is translocated at ZT 23 (Curtin *et al.*, 1995).

Taken together, these facts suggest that Per is a transcriptional repressor. When Per concentration is low, a heterodimer with activator properties promotes the rhythmic transcription of its own gene, but when Per concentration raises, Per is translocated into the nuclear compartment where it can compete for the dimerisation domain of the activator thereby inhibiting transcription (see fig. 1.4). Interestingly, Per has been showed to bind with high affinity to AhR and Arnt, disrupting their ability to form DNA-binding dimers, and its ectopic expression in transfected mammalian cells blocks Ahr-Arnt mediated response to dioxin (Lindebro *et al.*, 1995).

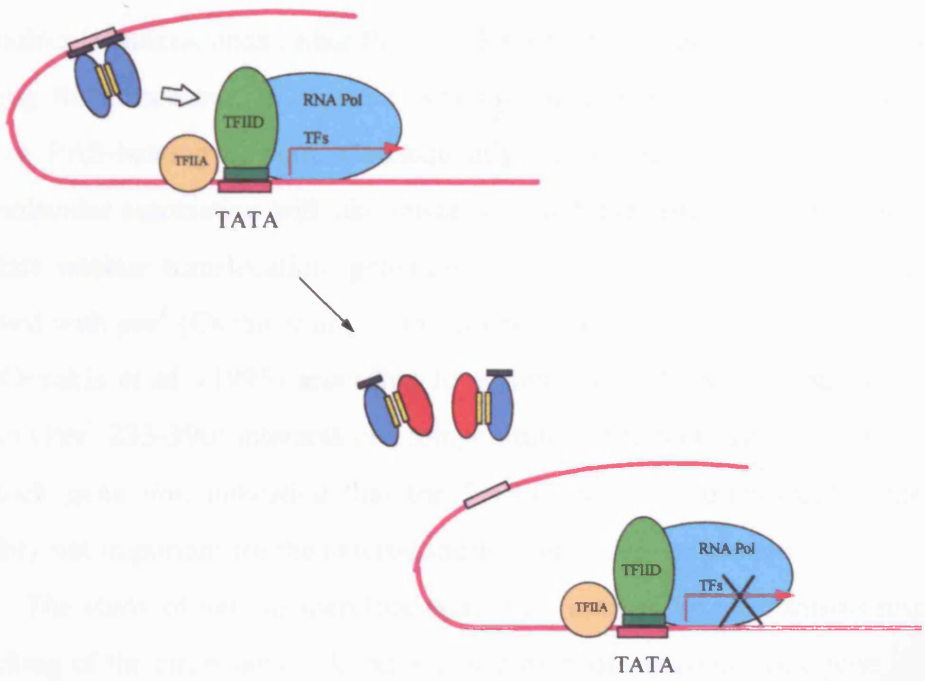


Figure 1.4 Model illustrating the role of Per on transcription. During the subjective day, when Per level is low, transcription occurs normally from a Per responsive promoter. At later ZT, nuclear Per level increases and Per (in red in the picture) can bind to the transcriptional activator (in blue) preventing the formation of a DNA binding dimer and blocking transcription.

PAS-mediated interactions can also account for the delay in nuclear entry showed by the Per^L protein and for another characteristic of the circadian rhythmicity of per^L mutants: temperature sensitivity. The period length of per^L flies increase from 27 to 33 h over a temperature range from 15° to 30° C (Konopka *et al.*, 1989; Ewer *et al.*, 1990). Homodimerisation of the PAS region of the Per protein has been observed *in vitro* (Huang *et al.*, 1993). Moreover the PAS domain interacts with a second fragment of Per located in C-terminal with respect to PAS itself, at amino acid position 524-685 (Huang *et al.*, 1995), called the C domain, indicating that intramolecular interactions within the Per protein are also taking place. The temperature coefficient of these two kinds of interaction is equivalent, so that increasing the temperature has the effect of incrementing the formation of PAS-PAS and PAS-C domain dimers to the same extent, leaving their ratio unvaried (Huang *et al.*, 1995). Replacement of valine 243 with an aspartic acid in the PAS fragment (as in the per^L mutation) leads to a different result. In per^L the temperature coefficient of the PAS-C domain interaction favour the formation of

temperature coefficient of the PAS-C domain interaction favour the formation of intramolecular interactions rather than the PAS-PAS intermolecular complexes, *de facto* reducing the concentration of "free" Per protein available for dimer formation with a putative PAS-bearing protein. Consequently, according to Huang *et al.* (1995), the intermolecular association will take longer to reach the critical concentration in order to stimulate nuclear translocation, generating the delay in the phase of the cycle that is observed with per^{L} (Curtin *et al.*, 1995). Some critiques to this conclusion though, come from Gekakis *et al.* (1995) according to whom a Per^{L}-PAS fragment devoid of the C domain (Per^{L} 233-390) interacts in a temperature dependent fashion with the product of the clock gene *tim*, indicating that the PAS-C domain intramolecular interaction are probably not important for the heterodimerisation of the Per protein.

The study of *per* has therefore started to unravel the mechanisms responsible for the ticking of the circadian clock, but the discovery of a second clock gene in *Drosophila* (Sehgal *et al.*, 1994; Vosshall *et al.*, 1994) has given a definitive boost to our understanding of the molecular basis of the oscillator.

timeless; a second *Drosophila* clock gene

The gene *timeless* (*tim*) has been identified following P-element mutagenesis and screening for recessive mutations altering the phase of eclosion. In tim^{0} flies, as in per^{01} mutants, rhythmicity in eclosion is lost (Sehgal *et al.*, 1994). In a DD cycle, flies can emerge from the pupal case at any given time of the subjective day or night. Also the period of locomotor activity is affected; tim^{0} flies show arrhythmic behaviour in free-running conditions (Sehgal *et al.*, 1994). At the molecular level, the tim^{0} mutation blocks *per* RNA and protein cycling (Sehgal *et al.*, 1994; Price *et al.*, 1995), and the nuclear translocation of Per (Vosshall *et al.*, 1994).

Unexpectedly, the isolation of *tim*, carried out by positional cloning (Myers *et al.*, 1995) and in a yeast two hybrid screening for proteins able to interact with the PAS domain of Per (Gekakis *et al.*, 1995), revealed no sequence homology between PAS and the Tim protein. Nevertheless, intermolecular interactions *in vitro* and in a yeast two-hybrid system between the PAS domain of Per and fragment 505-906 of the Tim protein

have been reported (Gekakis *et al.*, 1995). The conceptual Tim protein is 1389 amino acid long and contains a putative nuclear localisation signal (NLS) and an acidic region, typical of the activation domain of several transcription factors. No significant sequence homologies have been found in database searches.

A more accurate mapping of the regions involved in the heterodimerisation of these two proteins has been performed by Saez and Young (1996). Two sites of intermolecular interactions have been described: the first PAS repeat binds *in vitro* to a portion of Tim containing the NLS, and a Per fragment localised between the second PAS repeat and the per^S site interacts with a second Tim portion.

Temporal expression of *tim* RNA and protein closely resemble those for *per*. Cycling of *tim* RNA is practically superimposed to that of *per* in LD and DD in both per^+ and per^S (Sehgal *et al.*, 1995), indicating a role of Per in the expression of *tim*. No cycling of *per* is observed in tim^0 flies (Sehgal *et al.*, 1994; 1995) and, conversely, *tim* RNA oscillations are abolished in per^{01} flies, in LD and free-running conditions (Sehgal *et al.*, 1995). Tim oscillations differ slightly from those of Per, since Tim decrease is rapid and precedes the more gradual decline of Per (Zeng *et al.*, 1996). Temporal phosphorylation occurs also in the Tim protein, although its apparent increase in size on SDS-PAGE is less dramatic than in Per (Zeng *et al.*, 1996). Tim cycling is observed in a per^{01} background in LD but not in DD (Zeng *et al.*, 1996). This cycle is driven by light, which causes the rapid degradation of Tim at ZT0 (Zeng *et al.*, 1996; Myers *et al.*, 1996). The removal of the photic stimulus permits the accumulation of Tim after ZT12. On the other hand, Per levels are low throughout the 24 h in tim^0 flies, irrespective of the environmental conditions (Vosshall *et al.*, 1994; Price *et al.*, 1995).

It appears clear that *per* and *tim* are indeed part of the same autoregulatory loop. During the accumulation phase, Tim and Per are located in the cytoplasm. No free-Per or free-Tim are detected but, while all Per is found in a complex with Tim, part of Tim dimerises with another factor, possibly a second Tim molecule (Zeng *et al.*, 1996). Per and Tim are subsequently phosphorylated and then translocated into the nucleus, where phosphorylation still continues (Edery *et al.*, 1994). In the nucleus, the Per-Tim complex is stable for several hours, until the sudden degradation of Tim by light releases the monomeric form of Per (Zeng *et al.*, 1996; Myers *et al.*, 1996; Lee *et al.*, 1996). Further phosphorylation of the free Per could result in a signal for its own degradation (Zeng *et*

al., 1996; Lee *et al.*, 1996). How the Per-Tim complex regulates transcription is not known, since neither protein contains DNA binding domains.

Both proteins must be expressed in order for nuclear translocation to occur. In per^{01} flies Tim accumulates in the cytoplasm but, since no Per-Tim complex is formed, fails to translocate to the nucleus (Vosshall *et al.*, 1994). In tim^{0} flies, no cytoplasmic accumulation of Per is detected (Zeng *et al.*, 1996), suggesting that Tim protects Per from early degradation by binding to it during the accumulation phase. A cytoplasmic localisation domain (CLD) has been individuated on both Per and Tim by deletion experiments (Saez and Young, 1996). In Per this domain correspond to the second dimerisation-promoting fragment while in Tim the CLD has been mapped to a 160 amino acid long C-terminal fragment. The monomeric form of the proteins is maintained in the cytoplasm but Per-Tim heterodimers, in which the CLDs are masked, are free to migrate into the nucleus.

This model of Per-Tim action, explains how phase advances or delays can be caused by light pulses. Pulses during the subjective day do not affect the phase of the clock since no high levels of Tim are yet available. Stimuli in the early hours of the night delay the clock, by degrading the cytoplasmic Tim protein and therefore, indirectly, by promoting Per degradation. Since RNA levels of both *per* and *tim* are still high, the degraded Per and Tim are readily replaced. Late night light pulses trigger the premature nuclear degradation of Tim, so advancing Per phosphorylation and degradation. New synthesis of Per or Tim cannot take place because no *per* or *tim* RNA is available at this stage of the circadian cycle. One of the fundamental properties of a biological clock (its entrainability by environmental factors), can therefore be easily explained in terms of transcript-protein levels.

Molecular evolution of the *per* locus

Circadian periods have been described for all living beings. Twenty four-hour biological clocks have been shown to tick in prokaryotes (Kondo *et al.*, 1993; Huang and Grobbelaar, 1995), unicellular eukaryotes (Pittendrigh *et al.*, 1959), plants (McClung and Kay, 1994) and metazoan (Roberts, 1956). All circadian clocks have similar

properties: they interact with the external environment in similar ways, and are temperature compensated. Given their ubiquity, circadian clocks probably evolved very early in the biosphere's history and so the molecular mechanisms responsible for their cycling would supposedly be well conserved throughout the various kingdoms. Identification of the key clock molecules should in this case be a relatively easy task (by a simple homology screening) even in distantly related organisms. In contrast, *per* homologues have proved difficult to isolate. Not long after its cloning, the *per* gene was sequenced in *D. melanogaster* (Jackson *et al.*, 1986; Citri *et al.*, 1987), in *D. pseudoobscura* and *D. virilis* (Colot *et al.*, 1988) and in *D. yakuba* (Thackeray and Kyriacou, 1990). Given the universality of the clock, it was assumed that very similar mechanisms, and highly conserved genes, provided the molecular basis for the pacemaker for a great variety of organisms. Thus the hunt was on for the *per* mammalian homologues, and given additional hope by reports that sequences homologous to *per* were found to cycle in the rodent SCN (Ishida *et al.*, 1991; Ben-Shlomo *et al.*, 1996). However mammalian *per* has proven to be elusive until very recently, when a candidate *per* homologue has been cloned in mouse and human (Sun *et al.*, 1997; Tei *et al.*, 1997). Its search has been initially hampered by failure not only in mammals, but also in closer relatives of *Drosophila*. The "drosophilidae barrier" was broken with the cloning of *per* from the housefly *Musca domestica* (this thesis) followed by the lepidopteran *Antheraea pernyi* (Reppert *et al.*, 1994). *per* fragments have been sequenced from several other species of dipterans (Peixoto *et al.*, 1992, 1993; Nielsen *et al.*, 1994; Rosato *et al.*, 1994), from the lepidopteran *Manduca sexta* (K.K. Siwicki, personal communication), from the cockroach *Periplaneta americana* (Reppert *et al.*, 1994) and from the honeybee *Apis mellifera* (G. Robinson, pers. comm.).

Why has so much effort been invested into the identification of *per* homologues? The initial task was to confirm the supposed universality of the clock. Findings that anti-Per antibodies labelled circadian pacemaker neurons in the gastropods *Aplysia* and *Bulla* (Siwicki *et al.*, 1989) and in the SCN of the rat (Siwicki *et al.*, 1992b) seemed to confirm this idea, but inital attempts to clone the *per* homologues have been unsuccessful. It appears now clear that *per* has evolved rapidly even within insects, so *per* homologues in different classes of organisms have proved rather difficult to identify. Perhaps clock molecules undergo different selective pressures in different classes of organisms. For

example we know that *per* displays some pleiotropic effects by influencing, besides circadian rhythmicity, the cycle of the lovesong in *Drosophila* (Kyriacou *et al.*, 1980; 1986).

When the sequences from *D. virilis* and *D. pseudoobscura* (Colot *et al.*, 1988), *D. yakuba* (Thackeray and Kyriacou, 1990) and *A. pernyi* (Reppert *et al.*, 1994) are compared to *D. melanogaster* (Citri *et al.*, 1987), the gene appears to be a patchwork of conserved sequences interspersed with regions of poor homology. One of the best conserved regions is the area designated as c2 (see fig. 3.1), which includes the PAS domain. The N-terminal half of the protein is generally better conserved than the C-terminus. This is dramatically evident in *A. pernyi*, where the C-terminal half is largely missing.

This observation does not necessarily mean that the second half of the dipteran protein is devoid of functions. It could rather be the case that this region is not involved in the molecular mechanisms of basic clock function, but is important in defining the species-specific aspects of the behaviour, the "fine tuning" of clock functions. This is indeed true, at least in *Drosophila*, for the Thr-Gly domain which, together with its 5' flanking region, affects the temperature compensation of the clock (Ewer *et al.*, 1990; Sawier *et al.*, 1997; Peixoto *et al.*, 1998) and the cycling of the male lovesong (Yu *et al.*, 1987b; Wheeler *et al.*, 1991). Furthermore, the repetitive nature of the Thr-Gly encoding DNA, makes this region extremely mutable (Dover, 1987) and explains its rapid evolution within the dipterans (Nielsen *et al.*, 1994).

The initial aim of my project was to identify the *per* gene from the housefly *M. domestica*. *Musca* is assuming increasing importance as a comparative model for evolutionary studies involving *Drosophila* homologues (Sommer and Tautz, 1991; Lukowitz *et al.*, 1994). *Musca* is also used for studies of circadian behaviour (Helfrich *et al.*, 1985; Smietanko and Engelmann 1989a; 1989b) and clock neurobiology (Pyza and Meinterzhagen, 1995). The time of divergence between *Musca* and *Drosophila* (100 Myr; Hennig, 1981) potentially permits a fair degree of divergence at the molecular level, but probably not enough to make the gene impossible to isolate by homology screening.

After isolating and characterising the *Musca per* clone, I tested it in interspecific transformations for its ability to restore circadian rhythmicity of locomotor activity in arrhythmic per^{01} *D. melanogaster* mutants. I examined the possibility that *Musca per*

confers species-specificity of behaviour. *D. melanogaster* transformants were also obtained, which carry chimaeric *D. melanogaster/Musca* genes; the locomotor activity of these transgenic flies was analysed in order to investigate the model of intramolecular coevolution proposed by Peixoto *et al.* (1993; 1998) discussed in chapter 5.

Temporal and anatomical expression of the products of the *per* gene has been described for *D. melanogaster* in fine detail (Siwicki *et al.*, 1988; Hardin *et al.*, 1990; Edery *et al.*, 1994), and has lead to the development of the negative feedback loop model (Hardin *et al.*, 1990). The expression of the *per* gene products in *Musca* has therefore been assayed in both *Musca* and *D. melanogaster* transformants, with unexpected results.

CHAPTER 2

Materials & Methods

MOLECULAR TECHNIQUES

Bacterial Strains

The LE392 (*supE44 supF58 hsdR514 galK2 galT22 metB1 trpR55 lacY1*) strain of *E. coli* was used in the library screening, while the strains DH5α (*supE44 ΔlacU169 (ϕ80 lacZΔM15) hsdR17 recA1endA1gyrA96 thi-1 relA1*) and XL1-Blue (*supE44 hsdR17 recA1endA1gyrA46 thi relA1 lac* F'[*proAB*$^{+}$ *lacI*^{q} *lacZΔM15 Tn10 (tet*^{r}*)*]) (Stratagene) were used for most of the cloning work.

Plasmids

The vector pUC 19 was used for most of the cloning. pBluescript II KS+ (Stratagene) was used for the synthesis of the RNA probes. The transformation vector pW8 (Klemenz *et al.*, 1987) was the vector of choice for the P element-mediated transformation of *D. melanogaster*. The vector pGEX-4T3 (Pharmacia) was used for *in vitro* expression of a fragment of the *Musca* Per protein.

pJMH3, kindly provided by M. Hennessy, containing a 9 kb fragment of the *per* gene of *D. melanogaster* was used in the construction of *D. melanogaster/M. domestica* chimaeric *per* genes.

RP49-EH was given by Paul Hardin and contains a fragment of the *RP49* gene of *D. melanogaster* cloned into pBluescript KS-. This plasmid was used in the RNase protection assay.

The genomic library used for the cloning of the *Musca domestica per* gene and provided by M. Williamson (IACR, Rothamsted), was constructed using an EMBL-3 vector.

Basic techniques

Basic molecular techniques were as in Sambrook *et al.* (1989). Restriction digests were performed according to the manufacturer's instructions and ligation usually carried out at 16° C for 3h to overnight. Typically, insert DNA was in a 3-fold molar excess with respect to the vector, with the amount of vector usually ranging between 5 to 50 ng.

Agarose electrophoresis was usually carried out in a 0.6 - 1.5% gel. Ethidium bromide was added to the molten agarose to a final concentration of 0.5 µg/ml. Gels were run in TBE (89 mM Tris-borate (pH 8.3), 2 mM EDTA). The required amount of sample DNA was mixed with 1/5 volume of loading buffer (62.5 mM Tris-Cl (pH 6.8), 25% glycerol, 0.25% bromophenol blue) before loading. Gels were run at constant voltage, the value depending on the particular conditions of the run but never exceeding 10 V/cm.

Oligonucleotide primerswere synthesised by the Protein & Nucleic Acid Chemistry Laboratory (PNACL) of the university of Leicester on a 40 nM scale.

Genomic DNA extraction

Fly DNA was extracted as in Gloor and Engels (1990). Single *Musca* heads were ground in 100 µl of extraction buffer (10mM Tris-Cl pH 8.2, 1 mM EDTA, 25 mM NaCl, 200 µg/ml proteinase K), incubated at 37° C for 30 min. and then heated to 100° C for 10 min. 2 µl of this DNA preparation were used for each 20 µl PCR.

Recovery of DNA from agarose gel

DNA was recovered from the agarose gel by electroelution. Once the desired fragment's separation was achieved in the electrophoretic run, an agarose slice containing the band of interest was cut from the gel. The agarose slice was then transferred into a dialysis tube (previously boiled for 10 min. in 5 mM EDTA and stored in 100% ethanol at 4° C) together with an equal volume of TBE. The dialysis tube was clamped at both ends and put into an electrophoresis tank. A difference of potential of

100 V was then applied for 30 min., after which the polarity of the current was reversed for a further 5 min. The TBE was recovered from the dialysis bag and subjected to phenol/chloroform extraction and ethanol precipitation.

Determination of Nucleic Acid concentration

The concentration of sample DNA was assayed using an Ultrospec Plus (LKB) spectrophotometer, assuming that 50 μg/ml of double stranded DNA or 40 μg/ml of RNA give an Absorbance of 1 unit at a wavelength of 260 nm (A_{260} = 1) in a 1 cm path length cuvette.

Bacterial transformation

E. coli cells (XL1-Blue strain) were transformed by electroporation. Cells were grown in Luria-Bertani (LB) broth (1% bacto-tryptone, 0.5% bacto-yeast extract, 1% NaCl) plus 12.5 μg/ml tetracycline until they displayed an A_{600} = 0.5 and pelleted at 4000 x g 15 min. at 4° C. The cells were washed twice in decreasing volumes of H_2O (1 volume and 0.5 volume of the original culture) and once in 0.02 volume of 10% (w/v) glycerol. Finally they were resuspended in 1/500 volume of 10% glycerol and frozen in a dry ice-ethanol bath in 40 μl aliquots or used immediately.

The recombinant plasmid was mixed with the competent cells into an electroporation cuvette and placed in the electroporator (Gene Pulser, BioRad) set on 25 μF, 1.5 kV. A pulse of current was delivered to the cuvette and the cells were immediately transferred to a new tube containing 1 ml of LB medium, and shaken for 1h at 37° C. The cells were then plated on petri dishes containing LB agar (as LB broth, with the addition of 1.5% bacto-agar) supplemented with 50 μg/ml of both X-Gal and IPTG plus the appropriate antibiotic.

Radio-labelling of DNA

Two methods were used for the labelling of DNA, depending on whether the DNA fragment was to be used as a probe (high specific activity required) or as a marker.

Random priming (Feinberg *et al.*, 1983)

20 ng of DNA template were denatured by boiling for 5 min. and chilled on ice for a further 5 min. To the denatured probe were then added 5 µl of 10x labelling buffer (0.5 M Tris-Cl (pH 8.0), 50 mM Mg Cl_2, 20 mM DTT, 2 mM HEPES (pH 6.6)), 5 µl of 50 A_{260} units/ml hexadeoxynucleotides (Pharmacia), 1 µl of dNTP (20 mM each of dATP, dGTP and dTTP), 2 µl of [α-^{32}P]-dCTP (3000 Ci/mmol, 10 µCi/µl, Amersham), 5 units of DNA polymerase I (Klenow fragment) and H_2O to a total volume of 50 µl. The reaction was incubated at 37° C for 1 h and terminated with 5 µl of 0.5 M EDTA. Unincorporated nucleotides were removed by ethanol precipitation and the probe resuspended in H_2O and denatured before use.

Fill-in labelling (Sambrook *et al.*, 1989)

200 ng of DNA (usually pBR322 cut with *Msp*I restriction endonuclease) were added to 2 µl of 10x Klenow fill-in buffer (0.5 M Tris-Cl (pH 7.6), 0.1 M $MgCl_2$), 1 µl 5 mM of each dATP, dGTP, dTTP, 4 µl of [α-^{32}P]-dCTP (3000 Ci/mmol, 10 µCi/µl, Amersham) and 2.5 units of DNA polymerase I (Klenow fragment). The volume of the reaction was then adjusted to 20 µl with H_2O. The reaction was allowed to proceed for 1/2 h at 30° C and stopped by heating the tube for 10 min. at 70° C. Ethanol precipitation was then performed in order to remove the unincorporated [α-^{32}P]-dCTP.

Library screening

E. coli (LE392 strain) cells were grown overnight at 37° C in LB containing 0.2% (w/v) maltose. The cells were then pelleted by centrifugation (1000 g for 10 min.) and resuspended in 10 mM $MgSO_4$ at such a dilution resulting in an A_{600} = 0.5.

200 µl aliquots of the resuspended cells were then mixed with 50 µl of λ buffer containing 3000 pfu of an EMBL 3 genomic library and incubated for 15 min. at 37° C. At the end of the incubation 3 ml of molten H top agar (1% bacto-tryptone, 0.8% NaCl, 0.7% bacto-agar) at 42° C were added to the infected cells, mixed and overlaid onto H agar (as H top agar, but with 1.5% bacto-agar) plates prewarmed at 37° C. The petri dishes were left to set at room temperature and then incubated overnight at 37° C.

After overnight incubation the plates were cooled at 4° C for 1 h. Hybond N filters were then laid over the agar and left for 5 min. Meanwhile, three sheets of 3 MM paper were prepared, soaked with denaturing solution (1.5 M NaCl, 0.5 M NaOH), with 1 M Tris-Cl (pH 7.5) and with 1 M Tris-Cl (pH 7.5), 1.5 M NaCl, respectively. The filters were then peeled off the agar and placed, DNA side up, on the denaturing solution-soaked paper for 10 min., then on the second sheet for 1 min. and on the third for a further 10 min. Finally the filters were briefly rinsed in 2x SSC, allowed to dry and fixed by exposing for 45 s. on an UV transilluminator.

Up to 12 Hybond membranes were placed into hybridisation boxes and 25 ml of hybridisation solution were added to each box. The filters were prehybridised for 2 h in hybridisation solution (0.1% Ficoll 400, 0.1% poly vinyl pyrrolidone, 0.5% BSA fraction V, 0.5% SDS, 0.2% non-fat dried milk, 6 x SSC) at 65° C inside a hybridisation oven (Hybaid), then freshly denatured, ^{32}P-labelled probe was added and the filters incubated shaking overnight at 65° C.

The membranes were then washed in 6x SSC, 0.1% SDS for 20 min at 65° C, after which a series of increasingly stringent washes was performed. The series would typically consist of 4x SSC, 0.1% SDS; 2x SSC, 0.1% SDS; 1x SSC, 0.1% SDS; 0.5x SSC, 0.1% SDS washes, each for 20 min. at 65° C. Between each wash the filters were monitored with a Geiger counter and the washes were stopped when the radiation associated with the filters reached background level. The filters were then dried and covered with Saran wrap, placed in an autoradiography cassette with intensifying screens, together with a sheet of Fuji RX 100 X-ray film and exposed at -80° C for 2 to 4 days.

Once a positive clone was detected, the position of the plaque giving the signal was identified and the plaque recovered using the wide end of a blue tip. The plug of agar removed was put in a microfuge tube containing 0.5 ml of λ buffer and 50 µl of chloroform and eluted for several hours. A second round of screening was then

performed, plating several different dilutions of the potential positive. The filters were then hybridised with the same probe and another clone isolated from the dish with the lowest number of plaques using the thin end of a sterilised Pasteur pipette in order to pick a single positive. A third screening confirmed that no other phages are contaminating the positive clone.

λ DNA isolation

1 ml of λ buffer was poured onto a high plaque-density dish originating from the third round of screening, and the dish gently shaken for 30 min. at room temperature. The buffer was then collected into a microfuge with an added 100 µl of chloroform. This phage culture was used to infect 15 aliquots of 300 µl of *E. coli* LE392 (resuspended in 10 mM $MgSO_4$ at an OD_{600} = 0.5). An appropriate quantity of phage was used to give confluent plaques. The infected bacteria were plated out on plates prepared with LB containing agarose rather than bacto-agar (1.5% and 0.7% for base and top, respectively) and incubated overnight at 37° C. 3 ml of λ buffer were then poured onto the plates and the dishes placed at 4° C with gentle shaking. After 2 h the top layer of agar from each Petri was scraped off and collected in 30 ml Corex tubes. An equal volume of chloroform was added to each tube and the agarose dispersed using a pipette. The tubes were centrifuged at 12000 g for 10 min. at 4° C, the supernatant transferred to fresh tubes and spun again with the same parameters. The supernatant was then loaded into polyallomer tubes. The tubes were balanced with λ buffer and spun in a SW41 swing-out rotor (Beckman) at 40000 rpm for 55 min. at 4° C. The supernatant was discarded and each pellet resuspended in a tube volume of λ buffer. The tubes were spun at 12000 g for 10 min. at 4° C, after which the supernatant was collected into fresh tubes and centrifuged again at 40000 rpm for 55 min. at 4° C. The supernatant was discarded and the tubes left upside down to remove the excess moisture from the pellet. Each of the pellets were resuspended in 0.5 ml of λ buffer and 2 µg/ml of DNase I added. The phage suspension was incubated for 15 min. at 37° C then 1/10 vol 0.5 M EDTA (pH 8.0), 1/10 vol 10% SDS and 100 µg/ml proteinase K were added. The λ preparation was further incubated for 10 min. at 65° C and 50 min. at 37° C.

A phenol/chloroform extraction and ethanol precipitation were then performed on the sample and the λ DNA pellet resuspended in 100 µl of STE (10 mM Tris-Cl (pH 8.0), 1 mM EDTA, 0.1 M NaCl). 20 µg of RNase were then added and the DNA was incubated for 15 min. at 37° C. After the incubation, 5 µl of 20 x STE, 10 µl of 10% SDS and 1 µl proteinase K (20 mg/ml) were added and the incubation at 37° C extended for a further 15 min. Finally, a new phenol/chloroform extraction and ethanol precipitation were carried out and the DNA pellet resuspended in 50 µl of TE (10 mM Tris-Cl (pH 8.0), 1 mM EDTA).

Plasmid DNA preparation

Small scale

Plasmid minipreps of high purity were obtained according to Feliciello and Chinali (1993). 1.5 ml of an overnight bacterial culture were microfuged for 5 min. at low speed. The bacterial pellet was resuspended in 1 ml of ice-cold STE and centrifuged for 5 min. at low speed. The supernatant was removed and the cells resuspended in 200 µl of DNA solution I (10 mM Tris-Cl (pH 8.0), 1 mM EDTA, 50 mM glucose) and left to chill on ice. After 5 min. 400 µl of freshly made DNA solution II (0.2 N NaOH, 1% SDS) were added to the tube and the content mixed by inverting the tube several time. The samples were kept on ice for exactly 5 min. after which 600 µl of ice-cold DNA solution III (4 M potassium acetate, 2 M acetic acid) were added. The tubes were shaken to mix the contents and stored on ice for an additional 5 min. The samples were then microfuged at maximum speed for 5 min. at 4° C and 1 ml of supernatant transferred to a new tube. 0.5 ml of isopropanol were added to each sample, the tubes microfuged for 10 min. and the supernatant discarded. The pellet was resuspended in 125 µl of TE containing 20 µg/ml of Rnase A and incubated for 30 min. at 37° C. After the incubation, 150 µl of 88% isopropanol, 0.2 M potassium acetate were added, the tubes vortexed and left for 10 min. at room temp. The tubes were centrifuged for 5 min. and the supernatant carefully removed. The pellet was then resuspended in 10 µl of TE.

Large scale

For bulk preparations, essentially the same procedure described in the previous paragraph (plasmid DNA preparation: small scale) was used, appropriately scaled up. A 200 ml overnight culture was spun at 6000 g for 20 min. at 4° C. After cell washing in 200 ml of STE, the pellet was resuspended in 4 ml of DNA solution I in a 30 ml Oak Ridge tube. The lysis step was carried out with 8 ml of DNA solution II for 5 min. and 12 ml of DNA solution III were then added. The lysate was incubated on ice for a further 10 min. and spun at 12000 g for 20 min. at 4° C to separate the bacterial debris. After isopropanol precipitation (12000 g for 20 min.), the pellet was resuspended in 4 ml of TE, and mixed with an equal volume of cold 4 M ammonium acetate. The tube was left on ice for 10 min. and centrifuged at 12000 g for 5 min. at 0° C. The supernatant was transferred to a fresh tube and the DNA precipitated with 0.5 volumes of isopropanol (12000 g for 20 min.). The resulting pellet was then processed to remove RNA contamination, precipitated with 88% isopropanol, 0.2 M potassium acetate and finally resuspended in TE.

The preparations of the plasmids used to transform *D. melanogaster* embryos were obtained using the Qiagen tip-500 (Qiagen Inc.) following manufacturer's instructions.

Sequencing of DNA

Double stranded DNA was sequenced with the dideoxy chain-termination method (Sanger *et al.*, 1977) using the Sequenase version 2.0 kit (United States Biochemicals). The sequencing reactions were electrophoresed on a 8.3 M urea, 6% polyacrylamide (19:1 acrylamide:bisacrylamide) gel, using a Sequi-Gen (BioRad) nucleic acid sequencing cell, for 3-8 h.

PCR amplification

PCR was carried out according to Jeffreys *et al.* (1988). Each reaction contained 2.7 µl of 7.4 x PCR mix (0.5 M Tris-Cl pH 8.8, 0.12 M ammonium sulphate, 50 mM $MgCl_2$, 74 mM 2-mercaptoethanol, 11 mM each of the four deoxynucleotide triphosphate, 1.25 µg/µl DNase free BSA), 1 µl of each 10 µM primer and 0.3 µl of Taq polymerase (5 U/µl, Applied Biosystems) plus template DNA and water to a final volume of 20 µl.

The correct cycling of the PCRs was empirically determined for each particular reaction. As a rule of thumb, 1 min. extension time (at 72° C) was allowed for each kb in length of the fragment to be amplified. Reactions were performed in a Perkin-Elmer Cetus 4800 or in a Biometra Trio Thermal Cycler. When higher replication fidelity was needed, Taq polymerase was substituted with Vent polymerase (New England Biolabs). A typical 20 µl reaction would contain 2 µl of the manufacturer's supplied buffer, 1 µl of each 25 mM dNTP, 1 µl of each 10 µM primer, template DNA and 1 µl of Vent pol. (2U/µl).

If the PCR product was to be used for cloning purposes, it was purified from the other components of the reaction by agarose gel purification.

Total RNA extraction

Total RNA was isolated with the guanidinium-CsCl method (Sambrook *et al.*, 1989). All glassware was baked in an oven at 220° C for 3 h prior to use and plastics used were taken from sealed packs. Solutions were treated with DEPC or made with DEPC treated water (0.1% DEPC was added to the solutions, stirred overnight, and autoclaved for 1 h in liquid cycle).

Flies were collected, frozen in liquid N_2 and vortexed to separate the body segments. Roughly, 300 heads were collected and ground in liquid N_2 using a porcelain mortar and pestle. When the N_2 was completely evaporated, 20 ml of homogenisation buffer (4M Guanydine isothiocyanate, 50 mM Tris-Cl (pH 7.6), 10 mM EDTA, 2% (w/v) lauryl sarkosynate, 1% (v/v) β mercaptoethanol) were added and allowed to thaw

at room temp. The head extract was then further homogenised using a Sorvall homogeniser.

After the homogenisation, the extract was centrifuged at 10000 g for 10 min. at 20° C and 1 gram of CsCl was added to every 2.5 ml of the resulting supernatant. This solution was then layered in ultracentrifuge tubes over 9 ml of 5.2 M CsCl. The tubes were then filled with H_2O and spun at 28000 rpm for 16 h at 20° C in a TST 2838 rotor (Kontron). Carefully, the supernatant was removed and the RNA pellet resuspended in 2 ml of extraction buffer and transferred to a 15 ml polypropylene tube. The RNA solution was extracted with an equal volume of chloroform-butanol (4:1) and the aqueous phase transferred to a new tube. The organic phase was re-extracted with 2 ml of extraction buffer and the aqueous phase was added to the previously extracted one. The RNA was finally precipitated with 1/10 volume of 3 M Na acetate (pH 5.2) and 2.2 volumes of ethanol, and washed with 70% ethanol.

A quicker procedure, a modification of the method described by Chomczynski and Sacchi (1987), was used for multi-sample RNA preparation for the RNase protection assay. About 30 fly heads per each sample were isolated in the same fashion described above. The heads were homogenised using a glass homogeniser in 0.8 ml of RNazol B (Cinna/Biotecx) following the manufacturer's instructions.

Isolation of mRNA

mRNA was purified from total RNA by oligo(dT) cellulose chromatography (Aviv and Leder, 1972). The whole procedure was carried out in microfuge tubes. Addition of solutions to the slurry was accompanied by gentle 10 min. mixing. Between each subsequent addition the tubes were centrifuged for 2 min. at room temperature and low speed (1500 g). 0.1 g of oligo(dT) cellulose were activated by adding 1 ml of binding buffer. After centrifugation the supernatant was discarded and 1 ml of 0.1 M NaOH, 5 mM EDTA was added to the resin. The oligo(dT) cellulose was then washed a few times with H_2O until the pH of the supernatant became < 8 and finally with 1 ml of binding buffer. Meanwhile an equal volume of 2x binding buffer was added to the RNA sample and the RNA denatured by heating for 10 min. at 65° C and immediately followed by chilling in ice. The RNA was added to the resin and mixed. After centrifugation, the

supernatant was collected and the denaturation and binding step repeated once more. The slurry of oligo(dT) cellulose was then washed several times with binding buffer until the A_{260} of the supernatant stabilised at about 0.01 (typically 5 washes). A final wash was performed using 1 ml of wash buffer. The mRNA was eluted by adding to the resin 0.5 ml of elution buffer, incubating for 10 min. at 65° C and spinning. The RNA was then ethanol precipitated.

Reverse transcription

Reverse transcription reactions were set up by mixing together 2µl 10 x RT buffer, 0.8 µl of 25 mM each of the four dNTP, 1 µl of 10 µM primer, 1 µl (100 U) Mo-MLV reverse transcriptase, 0.5 µl (20 U) RNasin (RNase inhibitor, Promega), 1µg of total RNA or up to 50 ng of mRNA. The reaction was incubated for 15 min. at room temperature, 65 min. at 37° C and 10 min. at 95° C. The cDNA was then amplified via PCR in order to obtain quantities suitable for cloning.

Radio-labelling of RNA

The DNA fragment to be transcribed was subcloned into pBluescript II KS+ (Stratagene). Prior to the transcription reaction the plasmid was linearized at the insert's 5' end with a restriction endonuclease producing either 5' overhangs or blunt ends. Following the restriction, the DNA was treated with proteinase K (30 min. at 37° C), phenol/chloroform extracted, ethanol precipitated and resuspended in RNase-free water at 1 µg/µl.

In a microfuge tube were added, in the listed order, 4 µl of 5x transcription buffer (0.4 M Hepes-KOH (pH 7.5), 160 mM $MgCl_2$, 10 mM spermidine, 200 mM DTT), 2 µl of 100 mM DTT, 0.5 µl (20 units) of RNasin, 1 µl each of 10 mM ATP, CTP and GTP, 1 µl of template DNA, 5 µl of [α-^{32}P]-UTP (800 Ci/mmol, 20 µCi/µl, Amersham), 1 µl of T3 or T7 RNA polymerase (20 U) and H_2O to a final volume of 20 µl. The reaction was carried out for 2 hours at 37° C after which 2 U of RNase-free DNase I were added. The mixture was further incubated for 15 min. at 37° C, then an equal volume of

RNA gel loading buffer was added. The RNA probe was denatured by heating 5 min. at 90° C and run on a 5% acrylamide gel (Mini-PROTEAN II, BioRad) at 100 V until the bromophenol blue band reached the bottom edge (about 1 hour). The band corresponding to the full-length probe was detected by exposing a sheet of Fuji RX 100 X-ray film to the gel for 10 s. to 1 min., cut out of the gel, and eluted in acrylamide gel elution buffer (0.5 M ammonium acetate,1 mM EDTA, 0.2% SDS) overnight at 37° C. After elution, the probe was phenol-chloroform extracted and 1 µl used to determine the activity. The probe was then ethanol precipitated and resuspended to the appropriate concentration.

RNase protection assay

The total RNA was resuspended in 100 µl of RQ1 DNase mix (5 mM DTT, 0.05 mg/ml BSA, 40 mM Tris pH 8.0, 10 mM NaCl, 6 mM $MgCl_2$, 0.05 u/µl RQ1 DNase (Promega), 0.4 u/µl RNasin) and incubated 1 h at 37° C, phenol-chloroform extracted and precipitated. For each time point, eight µg of total RNA were placed in eight µl of H_2O, and 1 µl of *per* probe (1 x 10^6 cpm) plus 1 µl of *RP49* probe (1 x 10^5 cpm) were added, together with 20 µl of hybridization buffer (80% deionized formamide, 100 mM Na citrate (pH 6.4), 300 mM Na acetate pH 6.4, 1 mM EDTA) and incubated overnight at 42° C.

Once the hybridisation was completed, 200 µl of RNase digestion buffer (Ambion) containing 10 U of RNase ONE (Promega) were added, and the samples incubated for 50 min. at 32° C. The digestion was terminated by adding 300 µl of RNase inactivation solution (Ambion), 1 µl of glycogen (20 mg/ml), and 200 µl of ethanol. The tubes were placed on dry ice for 15 min and spun to precipitate the undigested probe. Pellets were resuspended in 8 µl of RNA gel loading buffer, heated for 5 min. at 90° C and run on a 8.3 M urea, 5% polyacrylamide gel. Fill-in labelled pBR322/*Msp*I DNA was used as a molecular weight marker. The gel was then fixed in 10% methanol, 10% acetic acid for 30 min., dried and placed into an autoradiography cassette with a sheet of film. The cassette was stored at -80° C for 1-3 days before developing the film.

Protein extraction

Flies were collected and frozen in liquid N_2 and the heads separated from the rest of the body. About 15 heads were used for each sample. The heads were placed in a microfuge tube and homogenised in 200 µl of protein extraction buffer (20 mM HEPES (pH 7.5), 0.1 M KCl, 10 mM EDTA. 1 mM Mg Cl_2, 5% glycerol, 1 mM DTT, 0.1% Triton X-100, 10 µg/ml aprotinin, 5 µg/ml leupeptin, 5 µg/ml pepstatin A, 0.5 mM PMSF) using a disposable pestel (Kontes). The tubes were centrifuged for 5 min and the supernatant transferred to a new tube and centrifuged a second time.

Polyacrylamide gel electrophoresis of proteins

Denaturing PAGE was carried out on a 16 x 18 cm slab gel. The resolving gel was 5.7% acrylamide (37.5:1 acrylamide:bisacrylamide) in 0.375 M Tris-Cl (pH 8.8), 0.1% SDS. Ammonium persulphate and TEMED were added to 0.07% and 0.1% respectively. The stacking gel consisted of 4.5% acrylamide in 0.125 M Tris-Cl (pH 6.8), 0.1% SDS, 0.07% ammonium persulphate and 0.1% TEMED. The gel was run in Tris-glycine buffer (25 mM Tris, 0.25 M glycine) with added 0.1% SDS.

Protein concentration of the samples was determined using the Bradford quantitative assay (BioRad) and an Ultrospec Plus (LKB) spectrophotometer. Equal amounts of proteins (about 500 µg of total proteins) were used for each time point. Samples were boiled for 10 min in the presence of loading buffer, and loaded on the gel. The electrophoresis was carried out overnight at 15 mA constant current. Prestained high molecular weight range marker (BioRad) was used.

Electroblotting of proteins

Proteins were blotted on nitrocellulose filters according to Towbin *et al.* (1979). After the electrophoresis, the gel was soaked for 10 min in transfer buffer (20 mM Tris, 0.15 M glycine, 20% methanol) before being placed onto 3 pieces of Whatman 3 MM

paper. A nitrocellulose filter was then put on the gel and covered with the other 3 pieces of 3 MM paper. The transfer was carried out for 3 h under a 0.4 A constant current.

Immunoblotting

The membrane with the immobilised proteins was incubated for 1 h at 37° C in blocking solution (2% (w/v) non-fat dried milk, 0.05% (w/v) Tween 20 in TBS. TBS is: 10 mM Tris-Cl (pH 7.5), 150 mM NaCl) and subsequently for 2 h at room temperature in blocking solution containing the appropriate dilution of the primary antibody. The filter was then subjected to three 5 min. washes in TBST (TBS containing 0.05% Tween 20), incubated for a further hour at room temperature with the secondary antibody (horseradish peroxidase conjugated, Sigma) diluted to the manufacturer's recommendation in blocking solution, and washed for 5 min. (3x) in TBST. The filter was then submerged in 0.1 M Tris-Cl, pH 8.5, 6.25 μM Luminol, 6.38 μM p-Coumaric acid, 2.7 mM H_2O_2 for 1 min and the chemiluminescent signal detected by exposing the filter to an autoradiography film.

CONSTRUCTION OF THE CHIMAERIC GENES

The "raw material" for the production of the chimaeric constructs was represented by clone 5 (see fig. 3.3), which contains the complete coding sequence of the *Musca per* gene plus about 2 kb of upstream untranslated sequence and 1 kb of downstream flanking sequence, and by pJMH3 (fig. 2.1), which consists of a 9 kb *Bam*HI-*Bam*HI *D. melanogaster per* fragment cloned into the transformation vector pW8 (Klemenz *et al.*, 1987). This 9 kb fragment contains *per* sequences from -3.5 kb to +5.5 kb (ca. 350 bp downstream of the perfect Thr-Gly repeat) with respect to the transcription start. The 3.5 kb of 5' regulatory sequence + UTR (exon1 and the 5' half of exon2) contained in this fragment is the same upstream sequence which drives correct expression of *per* transcript in pAP12.8, the plasmid originally used by Hamblen *et al.* (1986) to identify the *per* locus, and in the pAP12.8-derived 13.2 construct.

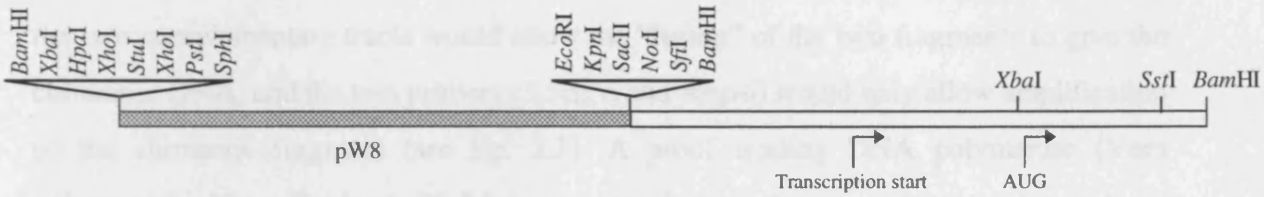


Figure 2.1. Linear map of pJMH3.

Construction of the *D. melanogaster/M. domestica* chimaeric *per* gene *mm1*

This chimaeric gene was intended to address the question whether or not the *M. domestica* PER is functional in a *D. melanogaster* genetic background. The *Musca per* gene was therefore fused to the 5' regulatory sequence and UTR of the *D. melanogaster* gene at the starting codon. The 6 kb fragment of *D. melanogaster* carries sequences necessary for the normal expression of *per* (Hardin *et al.*, 1992a) while the *M. domestica* fragment contains the whole coding sequence, with the exception of the intron corresponding to intron 2 of *D. melanogaster*, and 3' untranslated sequences up to 1 kb downstream to the poly-adenylation signal (see fig. 2.2). The large intron 2 was eliminated from the chimaeric gene in order to reduce the size of the construct and so to increase the efficiency of P element-mediated transformation.

Step I: construction of the *Xba*I-*Bgl*I chimaeric fragment

In order to obtain the chimaeric junction, a PCR based strategy was adopted (see fig. 2.3). First, a 422 bp *D. melanogaster* DNA fragment, containing *per* sequence from -422 to -1 with respect to the translation start, was amplified using pJMH3 as a template and the primers `55dro` (5'-CGAAGCAACATTCGGAATTTG-3') and `53dro` (5'-ATTCACCTTCCATGGTGCTTAGGTTCTCCAGCTTG-3'). `53dro` carries a tail (underlined) complementary to part of the *Musca* sequence. At the same time a 160 bp *Musca* cDNA fragment (from 0 to +160 with respect to the starting methionine) was amplified using `35mus` (5'-AACCTAAGCACCATGGAAGGTGAATCTACGGAAT-3') in conjunction with `amp6` (5'-GCGGGATCCGATGGTTTGCCGCCATAACC-3'). The underlined region of `35mus` does not bind to the *M. domestica* sequence but represents

a region of complementarity to the *D. melanogaster* sequence. Comparable amounts of the amplified fragments were then pooled together and again subjected to PCR so that the two complementary tracts would allow the "fusion" of the two fragments to give the chimaeric DNA, and the two primers (*55dro* and *amp6*) would only allow amplification of the chimaeric fragment (see fig. 2.3). A proof reading DNA polymerase (Vent polymerase, New England Biolabs) was used in order to minimise the risk of mutagenesis. The chimaeric fragment was then cut with the enzymes *Xba*I and *Bgl*I to give the resulting 0.5 kb fragment (see fig. 2.2).

Step II: joining the various fragments

The chimaeric fragments had then to be joined to a 6 kb *Bam*HI-*Xba*I *D. melanogaster* fragment and to a 5 kb *Bgl*I-*Sal*I *M. domestica* fragment (see fig. 2.2). The latter was cut out from the *Hind*III-*Sal*I subclone of the gene, while the former one was obtained by digesting pJMH3 (fig. 2.1) with *Xba*I and *Xho*I; the resulting 14 kb fragment contains the whole pW8 sequence plus the 6 kb fragment. The *Sal*I site in the 3' end of the *Musca* gene is compatible with the *Xho*I site of the pW8 polylinker. The three fragments were gel purified and then joined together in a single step by a three way ligation. The PCR generated chimaeric tract of the gene was sequenced to ensure that no mutations were incorporated into the construct.

Construction of the *D. melanogaster/M. domestica* chimaeric *per* gene *mm2*

To test whether the hypothesis of the molecular coevolution of the Thr-Gly repeat and its flanking region also applies to species relatively distant from Drosophilids, two further chimaeric constructs were made, one of which (pMM2, fig. 2.4) contains 6.4 kb of *D. melanogaster* upstream region plus *melanogaster* coding sequence up to the beginning of the Thr-Gly region, fused to *Musca* sequences which include the Thr-Gly region through to the end of the clone, 1 kb downstream to the poly-adenylation signal.

Step I: construction of the *Sst*I-*Sty*I chimaeric fragment

The first step in the construction of the plasmid involved the generation of a chimaeric fragment by PCR (see previous paragraph and fig. 2.3). The primers used to

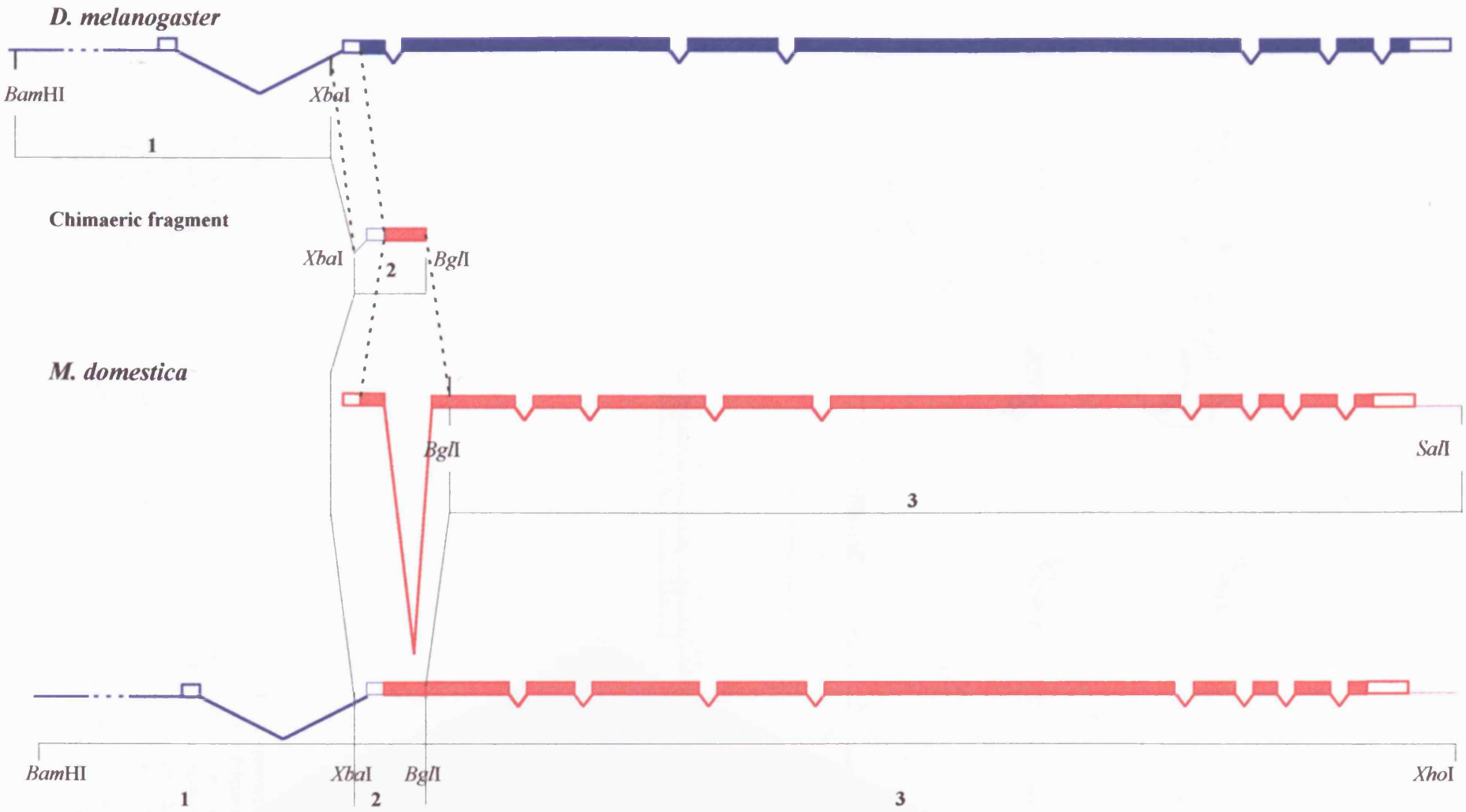


Figure 2.2. Diagram of the construction of per^{mm1}. The three fragments used for the construction of the gene and the relevant restriction sites are shown, together with the *D. melanogaster* (blue) and *M. domestica* (red) genes. Translated and non-translated exons are represented by filled and empty boxes, respectively. *D. melanogaster* 5' non-transcribed sequences and intron I are not scaled.

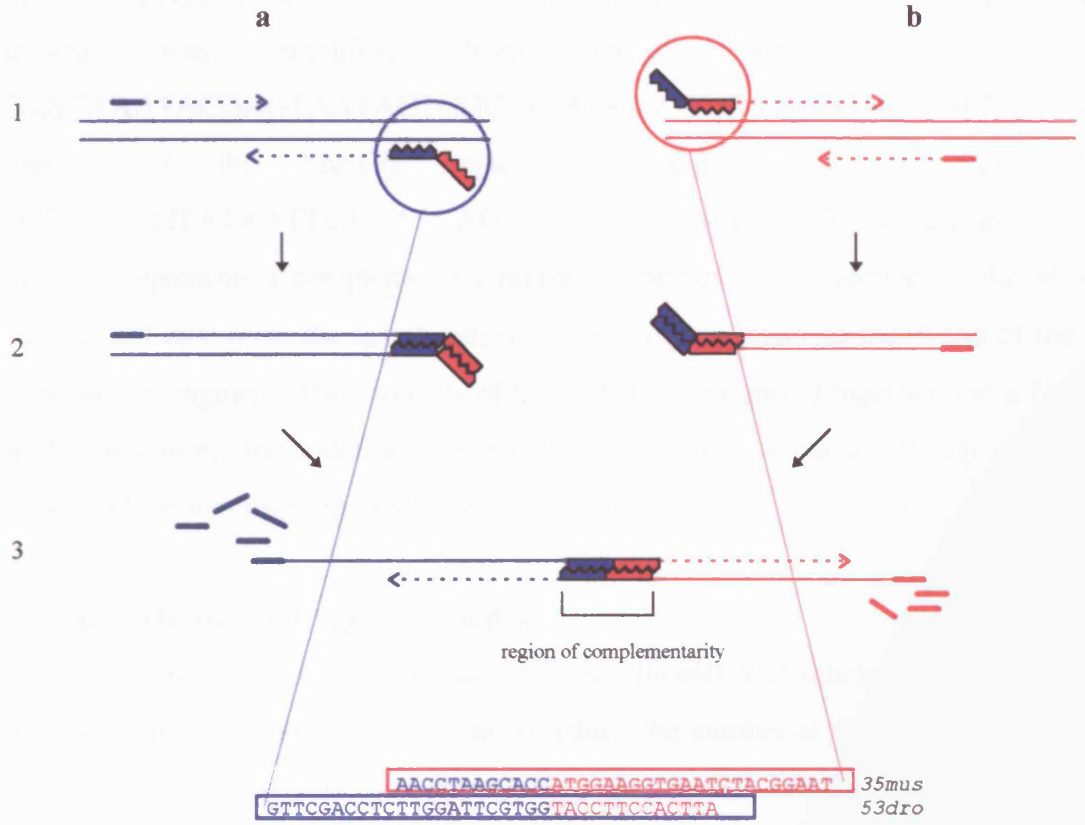


Figure 2.3. Diagram of the 2-step construction of a chimaeric fragment by PCR. 1) The fragments of DNA to be joined together are amplified by a first PCR (typically 30 cycles). The 3' primer of fragment **a** and the 5' primer of fragment **b** carry a tail complementary to the 5' end of frag. **b** and to the 3' end of frag. **a** respectively. 2) The PCR products are then purified and mixed together at a 1:1 molar ratio. 3) In a second PCR, which uses both the fragments as a template, there will be at first, extension of the strands paired together through the short region of complementarity created by the primer's tails, and then amplification of the chimaeric fragment. The second PCR is normally carried out for 10 to 15 cycles to minimise misincorporation of nucleotides and therefore a considerable amount of template is employed (50-100 ng). The relationship between primers *53dro* and *35mus* is shown at the bottom of the picture.

amplify the 250 bp *D. melanogaster* fragment from pJMH3 were *tga* (5'-ACCAGCTAAACTATAACGAGA-3' binding at position 4871-4891 of the sequence deposited in the SWISS-PROT database, accession n° P07663) and *chidm53* (5'-GCCAGTATTACTCGTATTTGTCACACTGCTCATGT-3'). The 400 bp *Musca* fragment was amplified from clone5 with *chimu35* (5'-GTGACAAATACGAGTAATACTGGCACGGGTACTTC-3', pos. 4054-4077, at the beginning of the Thr-Gly region) in conjunction with *fla3* (5'-CACTGCGAGTATAATTCAGACCAG-3', pos. 4451-4474). The underlined tail of *chidm53* represents a complementary region to the first 12 nucleotides in the *Musca* sequence and *vice verse* the tail of *chimu35* is complementary to the 3' end of the *D. melanogaster* fragment. The products of these PCRs were mixed together and a further amplification using the external primers *tga* and *fla3* gave a 650 bp chimaeric fragment which was then digested with *Sst*I and *Sty*I.

Step II: joining fragment 3 and 4

The 3' end of the construct was cut from a *Bam*HI-*Sal*I subclone of clone 5. The reason why this subclone was used, was to reduce the number of *Sty*I sites from 7 to 2, which would greatly simplify the whole cloning. The *Bam*HI-*Sal*I subclone was prepared by digesting clone 5 (see fig. 3.3 and 2.5) with *Bam*HI, and then gel purifying and re-circularizating the 5.6 kb fragment containing the desired fragment fused to pUC19 at the *Sal*I site. The poly-linker of this plasmid was cut with *Sst*I and *Bam*HI. When the products of the partial digestion are gel fractionated, the 5.3 kb band will not be entirely resolved from the 5.6 kb uncut fragment so the final ligation of step II could yield in large proportion the unwanted *Bam*HI-*Sal*I subclone. By double cutting the poly-linker of this plasmid, its self-circularisation will be prevented (see fig. 2.5). The linearised fragment was then subjected to partial digestion with *Sty*I. The product of this later digestion was electrophoresed and the 5.3 kb band corresponding to the *Sty*I-*Sal*I fragment plus the attached pUC19 (fig. 2.5) was recovered from the gel. This fragment was then ligated to the *Sst*I-*Sty*I chimaeric fragment to give pUC 3+4. The chimaeric fragment was sequenced to check that no misincorporation of nucleotides were caused by PCR.

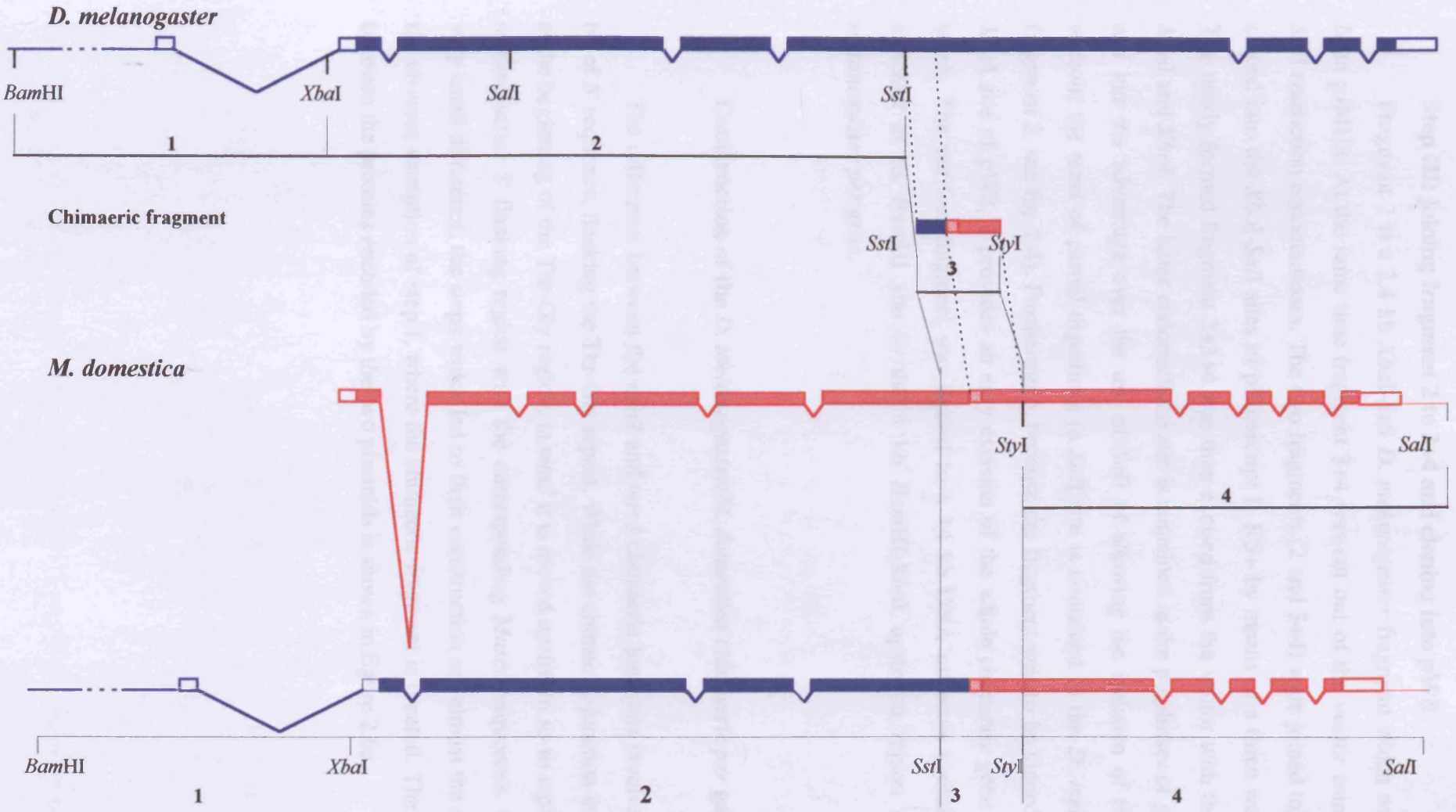


Figure 2.4. Diagram of the construction of the chimaeric gene per^{mm2}. The four fragment used are displayed. The Thr-Gly repeat of *M. domestica* is represented by a box filled with diagonal lines and translated and non-translated exons are represented by filled and empty boxes, respectively. *D. melanogaster* 5' non-transcribed sequences and intron I are not scaled.

Step III: joining fragment 2 to 3+4 and cloning into pW8

Fragment 2 is a 2.4 kb *Xba*I-*Sst*I *D. melanogaster* fragment which was isolated from pJMH3. At the same time fragment 3+4 was cut out of the vector using *Sst*I and *Sal*I restriction endonucleases. The two fragments (2 and 3+4) were joined together and cloned into the *Xba*I-*Sal*I sites of pBluescript II KS+ by means of a three way ligation. The newly formed fragment 2+3+4 was then excised from the vector with the enzymes *Xba*I and *Xho*I. The latter endonuclease site is contained in the polylinker of pBluescript and has the advantage over the use of *Sal*I of allowing the isolation of frag 2+3+4 without the need of partial digestions (a *Sal*I site is contained in the *D. melanogaster* fragment 2, see fig. 2.4). Furthermore, because the fragment was to be cloned using the *Xho*I site of pW8, it provides an easy excision of the whole chimaeric gene for future work. The purified fragment was ligated to a 14 kb DNA molecule containing pW8 attached at its *Bam*HI site to the 6 kb *Bam*HI-*Xba*I upstream region of the *D. melanogaster per* gene.

Construction of the *D. melanogaster/M. domestica* chimaeric *per* gene *mm3*

The difference between the *mm2* and *mm3* chimaeric junctions involves just 160 bp of 5' sequence, flanking the Thr-Gly repeat. While the chimaeric junction in *mm2* falls at the beginning of the Thr-Gly region, in *mm3* it is moved upstream so to replace the *D. melanogaster* 5' flanking region with the corresponding *Musca* sequences. Given this very small difference, the steps which led to their construction are almost the same, with the obvious exception of step I, where the chimaeric fragment is created. The difference between the proteins encoded by the two plasmids is shown in figure 2.6a.

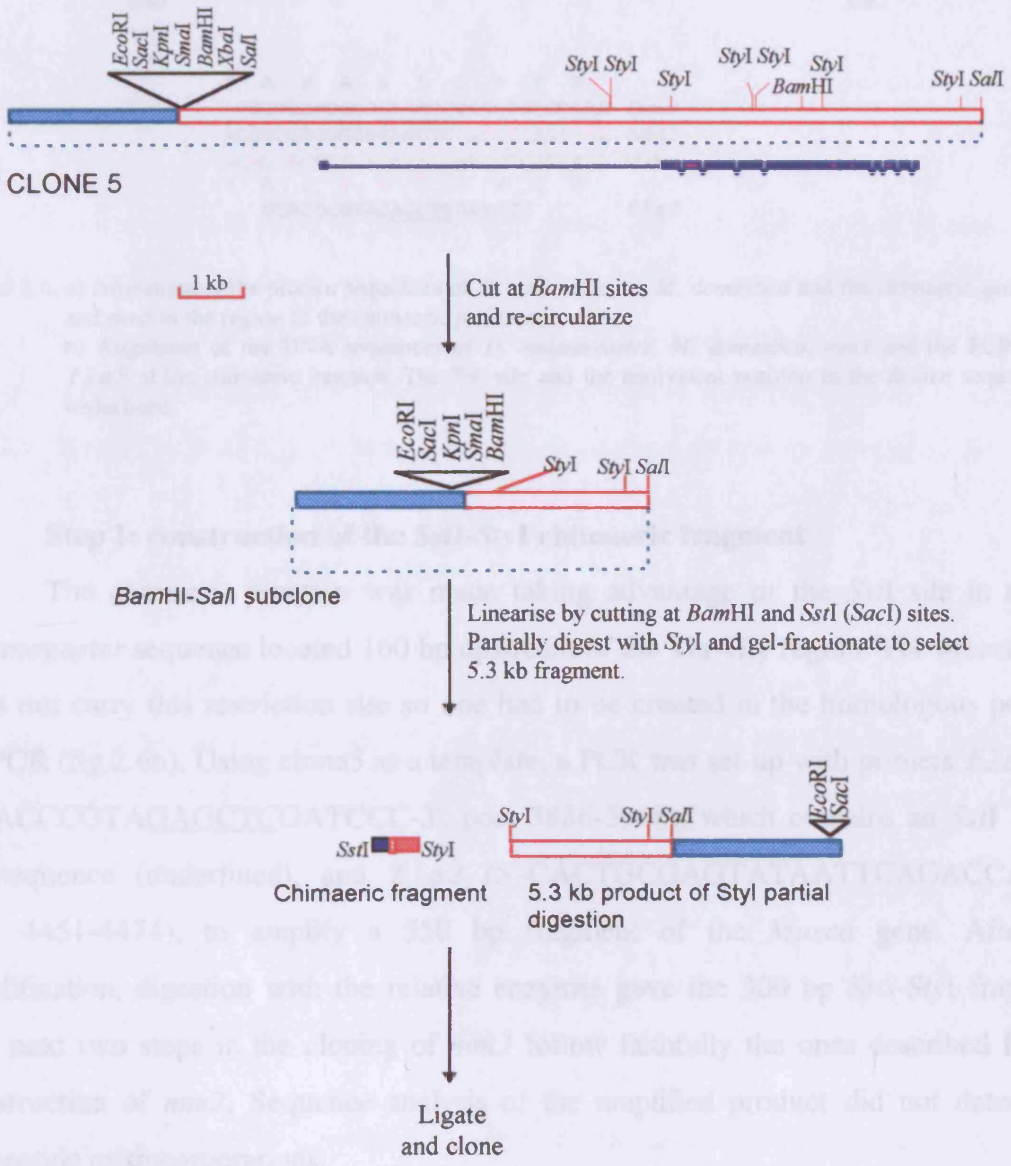


Figure 2.5. Scheme showing how fragment 3 (chimaeric fragment) and 4 (3' end of the *Musca* gene) were joined together (for details see text).

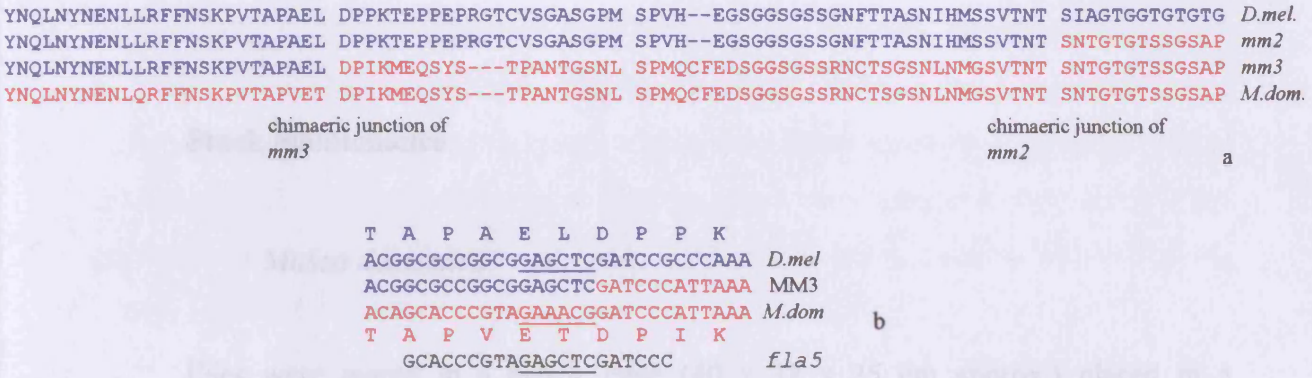


Figure 2.6. a) Alignment of the protein sequences of *D. melanogaster*, *M. domestica* and the chimaeric genes *mm2* and *mm3* in the region of the chimaeric junctions.
b) Alignment of the DNA sequences of *D. melanogaster*, *M. domestica*, *mm3* and the PCR primer *fla5* at the chimaeric junction. The *Sst*I site and the equivalent position in the *Musca* sequence are underlined.

Step I: construction of the *Sst*I-*Sty*I chimaeric fragment

The chimaeric junction was made taking advantage or the *Sst*I site in the *D. melanogaster* sequence located 160 bp upstream of the Thr-Gly region. The *Musca* gene does not carry this restriction site so one had to be created in the homologous position by PCR (fig 2.6b). Using clone5 as a template, a PCR was set up with primers *fla5* (5'-GCACCCGTAGAGCTCGATCCC-3', pos. 3886-3906), which contains an *Sst*I site in its sequence (underlined), and *fla3* (5'-CACTGCGAGTATAATTCAGACCAG-3', pos. 4451-4474), to amplify a 550 bp fragment of the *Musca* gene. After the amplification, digestion with the relative enzymes gave the 300 bp *Sst*I-*Sty*I fragment. The next two steps in the cloning of *mm3* follow faithfully the ones described for the construction of *mm2*. Sequence analysis of the amplified product did not detect any nucleotide misincorporations.

Drosophila and *Musca*

Stock maintenance

Musca domestica

Flies were reared in a plastic cage (40 x 32 x 25 cm approx.) placed in a controlled temperature room set on 25° C and in a 12 h light-dark cycle (12:12 LD). Adult flies were fed on sucrose and dried milk and provided with fresh water. Eggs were laid in larval medium (prepared by mixing 50 g of bran, 1 teaspoon of dried yeast, 80 ml of milk, 30 ml H_2O and 1.5 ml of 20% nipagine) where they developed into adults.

Drosophila melanogaster

Drosophila stocks were kept in either 1/3 pint milk bottles or glass vials (10 x 2.2 cm) and fed on sugar medium (6.5% sucrose, 11.5% baker's yeast, 1% bacto-agar, 0.2% nipagin). The flies were kept in either an 18° or 25° C controlled temperature room in 12:12 LD.

P element-mediated transformation

Transformation of *Drosophila* embryos was carried out according to Spradling (1986). The strain used was *w*; +/+; *Sb e Δ2-3/TM6*, which contains a stable P element (Δ2-3) on the third chromosome as a source of transposase (Robertson *et al.*, 1988). The transformation vector of choice was pW8 (Klemenz *et al.*, 1987) which allows red eye selection of the transformed flies.

Eggs were laid on egg laying medium (1.25% sucrose, 1.75% bacto-agar, 0.2% nipagin, 50% pineapple-grapefruit juice; red or green food colourant was added to allow an easy detection of the eggs on the surface of the medium) and manually dechorionated with the aid of a dissection needle. After dechorionation, the eggs were allowed to desiccate for 10 min and covered with Voltalef oil (grade10 S). The eggs were then

injected with plasmid DNA resuspended in injection buffer (5 mM KCl, 0.1 mM Na_2HPO_4 (pH 6.8)) at a concentration of 200 ng/µl. The injection apparatus consisted of a PV830 pneumatic picopump (World Precision Instruments) connected to a microneedle operated by a Narishige micromanipulator which was mounted on an Olympus CK2 inverted microscope. The needles were prepared by pulling glass microcapillary tubes (GC100TF-15, Clark Electromedical Instruments) with a Flaming Brown micropipette puller.

After being injected, the eggs were placed on a petri dish containing sugar medium and incubated at 18° C for several days. Each larva was then transferred to a glass vial and kept at 18° C until it became adult, then was crossed to a *w* strain. The progeny from this cross was analysed for the presence of red eyes which would indicate a transformed individual. Any G_1 red eyed progeny were used to found the transformed lines. Analysis of the G_2 progeny allowed the insertion to be chromosomally mapped, as both chromosomes III in G_0 carry morphological markers, one of which will be passed onto G_1. Linkage between these autosomal markers and the red eye phenotype indicated a chromosome III insertion and sex-linkage of the red eye phenotype indicated X or Y insertion. Lack of linkage of the red eye marker with the sex or 3rd chromosome was interpreted as linkage with chromosome II given the small size of chromosome IV (Ashburner, 1989).

The construct pMM1 was injected by H. Dobbs (Genetics department, University of Leicester), flies transformed with pMM2 and pMM3 were obtained by A. Bonini and G. Di Benedetto in the Biology department, Universita' di Padova.

Locomotor activity monitoring

The locomotor experiments were carried out with the use of an activity event recorder produced by Biodata Ltd (Manchester, UK), consisting of various units, each containing three series of infra-red photocells. Single flies were loaded into 80 x 3 mm cylindrical glass tubes in which one end was filled with approximately 20 mm of sugar medium and sealed with tape to prevent dessiccation, while the other end was closed with some tissue paper. Each glass tube was clamped between the two diodes of the photocell, which would detect every time the fly interrupts the infra-red beam. Infra-red

light was chosen for the reason that flies are insensitive to these wavelengths (Bertholf, 1932; Kyriacou *et al.*, 1979). The number of events recorded for each fly every 30 min time-window were stored in a 80386 personal computer.

Flies used in the experiment were reared at 25° C in LD 12:12 with lights on/off at 09:00/21:00 h. At day 0, flies were loaded into the tubes and placed into incubators set on the selected temperature and with the same light/dark cycle. At 21:00 h of the following day, the lights of the incubators were turned off for the duration of the whole experiment, and 18 hours later (15:00 h of day 2) the seven days of data collection began.

Data were subsequently edited. All the flies which died before the 192nd bin (fourth day) were excluded from the analysis. For those channels in which data collection ended after the fourth day, the last, empty bins were deleted from the data file. Also excluded were the flies catalogued as hyper- or hypoactive, that is, flies exhibiting an abnormally high events-count in many bins, or flies with a very low number of counts in most of the bins. The occasional bins with abnormally high number of events were replaced with the value of the average of the two adjacent bins. As a rule, any bin with a value higher than 100 was replaced in this fashion unless several of the neighbouring bins also displayed a relatively high number. The occasional appearance of these high values seems to be related to problems with the software but their rare frequency did not cause serious problems for the data analysis. Similar problems have been reported by Hamblen *et al.* (1986).

The periodicity of the locomotor activity of each fly was calculated with two methods: autocorrelation and Fourier analysis. Autocorrelation was performed with the SPSS/PC+ version 2 package. The string of data is compared with a copy of itself beginning at bin n° 2 and the correlation coefficient calculated. Then the starting point of the string copy is shifted one bin forward and the new coefficient computed. The whole procedure is reiterated n/2 times (where n is the total number of bins) and the correlation coefficient for each starting bin is then used to draw the autocorrelation plot. The highest peak in the autocorrelation plot, within the range 20-80 bins was taken as the period of activity only if its value is higher than the confidence limit.

In the Fourier analysis (performed with the CLEAN algorithm of Roberts *et al.* (1987) which was run on on a Silicon Graphics platform), the string of data is transformed into a mathematical function which assumes the form $F(x) = f_1(x) + f_2(x) +$

... + $f_n(x)$ where f(x) are harmonic functions differing for phase, period and amplitude. A frequency curve was then plotted in which the X-axis represents the period and the Y-axis the amplitude of these harmonics. Peaks in the curve correspond to the major spectral components in the data. The confidence limit was calculated by randomly shuffling the bins and repeating the spectral analysis on the new string of data. This procedure was repeated 100 times. For any given point of the X-axis (period), the 100 amplitude values were ordered and the 95^{th} and 99^{th} in ascending order plotted. These approximated to two straight lines across the period domain and represented the 95 and 99% confidence limits, respectively. The period of the highest peak was considered to represent the period of rhythmicity of the fly. If no peak was found above the 95% confidence limit the fly was classified as arrhythmic. A similar procedure has been used to generate confidence limits for song cycle (Konopka *et al.*, 1996). An example of locomotor analysis output (which refers to the DD locomotor activity of a housefly) is presented in fig. 2.7.

To analyse the pattern of locomotor activity in DD, the numbers of locomotor activity events recorded in each time bin of any individual fly is superimposed for the successive circadian cycles so the entire 7 days activity monitoring is "contracted" into a time frame representing the circadian period of that fly, and can be plotted as a chart where the abscissa represents the fly's circadian time and the ordinate represents the fly's activity. In order to compare these superimposed profiles of activity (wraps) between flies with different periods, to the highest peak of activity is assigned a value of 1 and the whole activity is normalized to this value. The length of the wrap is subsequently expanded or squeezed to 48 bins. This is achieved by generating 48 equidistant points on the abscissa and assigning to each of them an ordinate value calculated in the following fashion. b_n and b_{n+1} are two adjacent bins in the activity-profile plot and b_s is one of the 48 bins that will convey the squeezed or expanded activity information and which falls between b_n and b_{n+1}. The ordinate value of b_s [$Y(b_s)$] is obtained by plotting a straight line on the X-Y plane, between the points $Y(b_n)$; b_n and $Y(b_{n+1})$; b_{n+1}, and finding the corresponding ordinate value on this line for the abscissa b_s (see fig. 2.8).

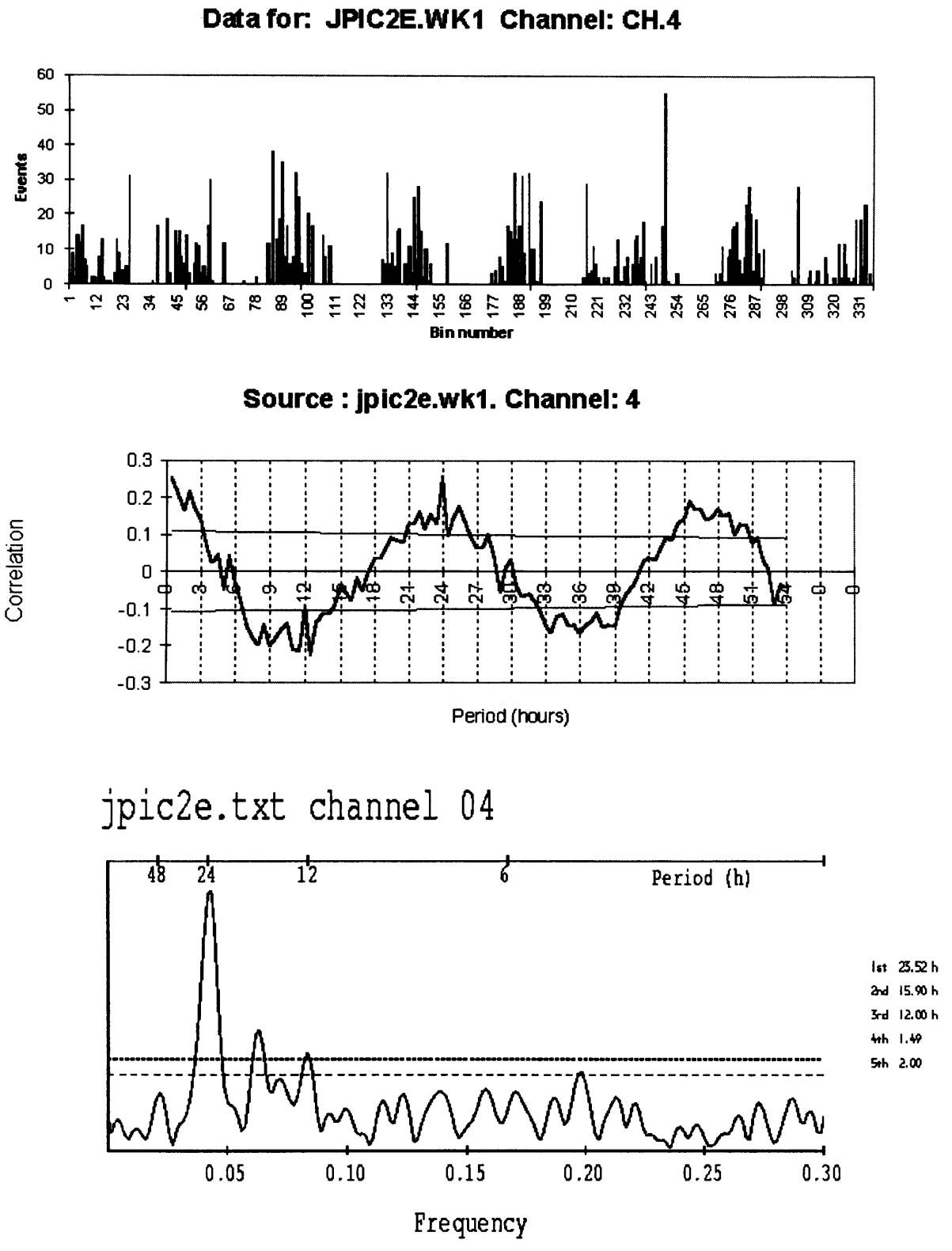


Figure 2.7. **a:** Histogram of raw activity data plotted over 7 days in 30 minute bins. **b:** Autocorrelation plot for each time lag up to 54 h in 30 minute bins; horizontal lines represent the 95% level of confidence. **c:** Spectrogram of the data. The period of each rhythmic component is plotted against its amplitude; horizontal lines represent the 95 (---) and 99% (···) confidence level. All charts refer to the same data.

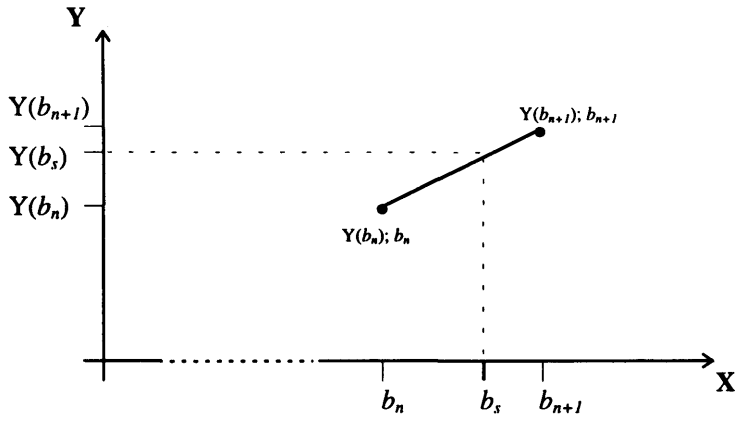


Figure 2.8. Plot showing how the activity values of the squeezed or expanded DD data were calculated. The value of the measured activity for each bin corresponds to a point on the plane X-Y. The abscissa, representing one circadian period, is divided into 48 intervals and the corresponding activity value is intermediate between the values of the adjacent original bins (b_n and b_{n+1} in the example).

CHAPTER 3

Identifying the *Musca domestica period* gene

Identifying the *Musca domestica period* gene

INTRODUCTION

In order to identify the *M. domestica period* gene, PCR primers were initially designed with reference to the available *per* sequences, which were from *D. melanogaster* (Citri *et al.*, 1987; Jackson *et al.*, 1986), *D. pseudoobscura* and *D. virilis* (Colot *et al.*, 1988) and *D. yakuba* (Thackeray *et al.*, 1990). The coding region of the gene spans between 4.0 kb (*melanogaster*) and 4.9 kb (*virilis*) including the introns. The encoded polypeptide is a 1218 amino acid chain in *D. melanogaster*, 1208 aa in *yakuba*, and 1241 aa and 1087 aa in *pseudoobscura* and *virilis*, respectively. Comparison of the coding sequence of the four orthologues shows that they can each be divided into 6 highly conserved regions (c1-c6) separated by 5 non-conserved areas (nc1-nc5) (Colot *et al.* 1988; see fig. 3.1).

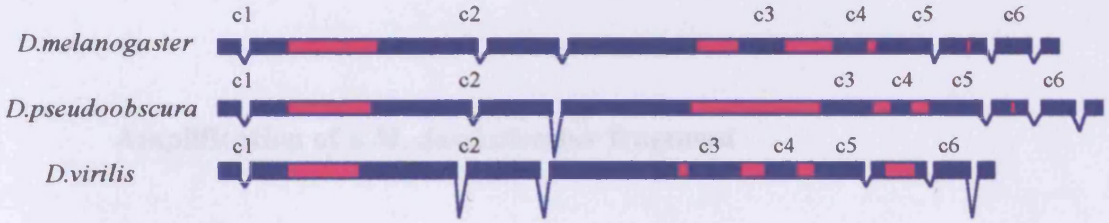


Fig. 3.1 Diagram of the conserved and non-conserved regions of the *per* gene of *D. melanogaster, D. pseudoobscura* and *D. virilis* according to Colot *et al.* (1988). Conserved regions are indicated in blue and non-conserved regions in magenta. The position of the introns is indicated. The map starts at the first codon and terminates at the end of the coding sequence.
c1-6 = conserved region 1-6.

A high degree of conservation is found in c1 (the N-terminal block of the coding sequence), in c2 (the longest of these blocks, containing the sites affected by the per^{L1} and per^{S} mutations and the PAS region), c3 and in c6 (localised at the C-terminus). c4 and c5 show a lower amount of conservation even though an alignment

by eye is still possible. The non conserved blocks differ between the species for both sequence and length so alignment is difficult. However they still present some common features such as the polarity of the encoded amino acid sequence (Colot *et al.*, 1988; Thackeray *et al.*, 1990) or short stretches of homology found in pair-wise comparisons. The non-conserved regions make up 30% of the length of the entire protein.

All four *per* coding sequences contain 6 introns. The intron-exon boundaries are perfectly conserved for the introns falling within regions of high homology but intron lengths vary irrespective of whether they are found in relatively conserved or unconserved areas. In *D. melanogaster* introns from 2 to 7 are, respectively, 61, 64 62, 70, 64 and 58 nucleotides long, in *D. yakuba* they measure 73, 77, 63, 56, 57 and 61 bp, those of *D. pseudoobscura* 62, 64, 532, 124, 62 and 197 and in *D. virilis* 70, 461, 475, 72, 66 and 484 nt long. In *D. melanogaster* a 2501 nt long intron is located in the 5' untranslated sequence; a large intron is found at an equivalent position in *D. yakuba* (M. Couchman, personal communication) while no data are available for the length of the 5' UTR of *D. pseudoobscura* or *D. virilis*.

METHODS

Amplification of a *M. domestica per* fragment

A fragment from the *M. domestica per* gene was amplified by PCR. *M. domestica* genomic DNA was extracted as described in chapter 2 and used in a polymerase chain reaction in conjunction with degenerate oligonucleotide primers, *deg5'* (5'- CCCGAATTCATGGARACNYTNATGGAYGA-3') and *deg3'* (5'-CCCGAATTCRTCRTARTARTCRTGRTG-3'). The two degenerate primers, carrying an *Eco*RI cleavable extension (*Eco*RI sites underlined), were designed in an area which encompasses the per^S site within c2 and displays high homology between the available sequences, spanning from positions 1678 to 1809 of the *D. melanogaster* coding sequence (see fig 3.2).

```
            M  E  P  L  M  N  E  V  A  R  Q  D  L  K  L  E  L  P  N  E  N  E  L
Musca       ATGGAACCCCTGATGAATGAAGTGGCCCGTCAAGATCTCAAATTAGAGCTACCAAATGAAAATGAACTA
mel. (1678) ATGGAAACGCTCATGGACGAGGTGTCCCGGGCGGATCTCAAGCTGGAGCTGCCGCACGAGAACGAGTTG
            ****** * ** *** * ** *** ****    ********  * ***** **  * ** ** **  *
                  T       D       S    A                      H

            T  V  S  E  R  D  S  V  M  L  G  E  I  S  P  H  H  D  Y  Y  D
Musca       ACCGTTTCCGAACGTGACTCGGTTATGTTGGGTGAAATTTCACCACATCACGACTACTACGAC
mel.        ACCGTCTCGGAGCGGGACAGCGTGATGCTCGGCGAGATTTCGCCGCACCACGACTACTATGAC (1809)
            ***** ** ** ** ***   ** *** * ** ** ***** ** ** ********** ***
```

Fig. 3.2 Sequence of the fragment amplified from *M domestica* genomic DNA aligned to the homologous fragment of *D. melanogaster* (numbered). The amino acid translation is given in bold type; only amino acid replacements are shown for the *melanogaster* sequence. Asterisks denote nucleotide matches. The regions where primer *deg5'* and *deg3'* bind, are highlighted by a line on the top and bottom of the sequence, respectively.

The amplified 132 bp *Musca* PCR fragment was cloned into pUC19 and sequenced, and was subsequently used as a probe for the *M. domestica* genomic library screening. In order to do this, a bulk preparation of the fragment-containing plasmid was performed, and the gel purified fragment was labelled by hexanucleotide random priming (see chapt. 2).

Library screening

The genomic library was plated onto *E. coli* strain LE392 and screened at high stringency conditions: one positive clone was isolated, from which DNA was obtained using the λ DNA preparation protocol described in chapter 2.

Characterisation and sequencing of the clone

The λ DNA was cut with *Sal*I (the same enzyme used for the construction of the library) to give the two phage arms (19.9 kb and 8.8 kb) plus a 4 kb and a 12 kb

fragment. Southern blotting showed that the probe hybridised to the larger of the two fragments. Both DNA fragments were subcloned into pUC19. A restriction map of the 12 kb clone (clone 5) is shown in fig 3.3 overleaf.

In order to sequence the *M. domestica* gene, primers were designed based on the sequence of the probe and used to extend the region of known sequence in both 5' and 3' directions. New primers were then designed close to the edge of the sequences and the process repeated until a putative polyadenylation signal was reached at the 3' end, and the intron corresponding to the *Drosophila* intron 2 was reached in the 5' direction (see fig. 3.8). About 1.2 kb of this intron were sequenced without finding any evidence for a candidate exon 2 (by comparison with the *Drosophila* exon 2 or by a splice junction search) so a different approach was used; the 5' end of the gene was subcloned using a modification of the RACE method originally described by Frohman *et al.* (1988). *Musca* total RNA was reverse transcribed using primer *amp6* (5'-GCGGGATCCGATGGTTTGCCGCCATAACC-3') which binds in c1 at the beginning of exon 3 in pos. 1991-2010 of the genomic sequence (fig.3.8), and the resulting cDNA fragment was poly T tailed (fig. 3.4). The 0.5 kb 5' end of the cDNA was then amplified using primers *amp6* and *amp5* (5'-GCGGGATCC$(A)_{20}$-3'), subcloned into the *Bam*HI site of pUC19 and sequenced (the *Bam*HI site is underlined in the primer sequences). After the determination of the 5' end of the sequence I could verify that the original genomic clone contained exon 2 and by restriction mapping measure the length of intron 2; the genomic clone was cut at a *Hin*dIII site near the 3' end of intron 2 (pos. -68 from the intron2:exon3 splice junction) and the 7.1 kb *Hin*dIII-*Sal*I fragment containing the 5' end of the gene was subcloned in pUC19. A restriction map of this fragment is shown in fig. 3.5.

The 2.2 kb *Sal*I-*Pst*I segment was also subcloned and partially sequenced; in this way it was established that the *Pst*I site of this fragment corresponds to the *Pst*I site at nucleotide pos. +66 with reference to the starting codon.

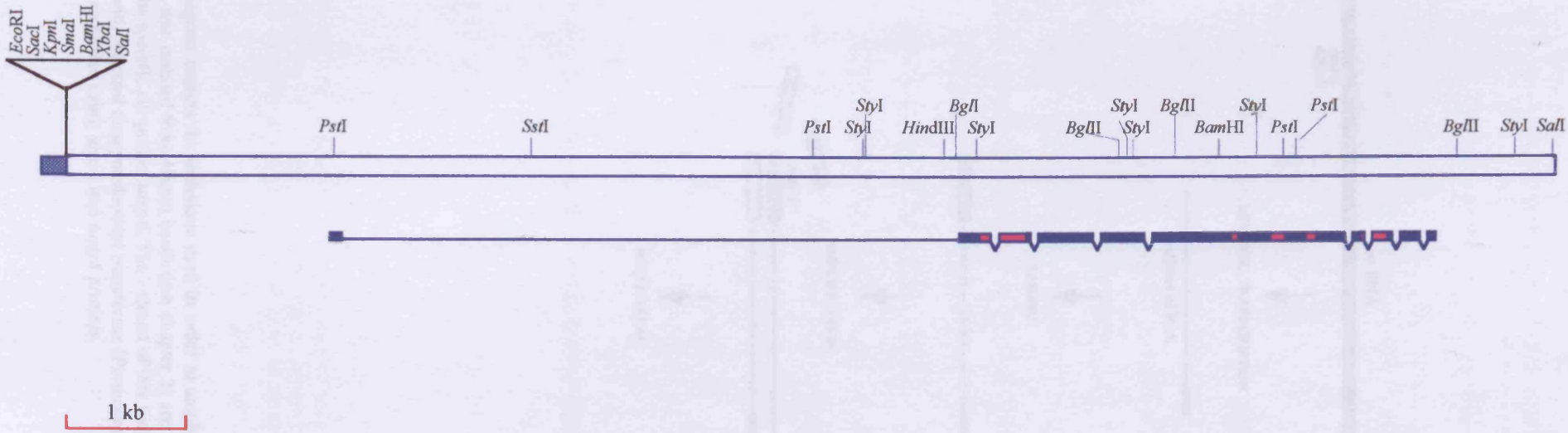


Fig. 3.3. Restriction map of the 12 kb clone containing the *Musca domestica per* gene. The clone derived as a 16 kb fragment from an EMBL3 genomic library, was cut with the restriction endonuclease *Sal*I and subcloned into pUC19.

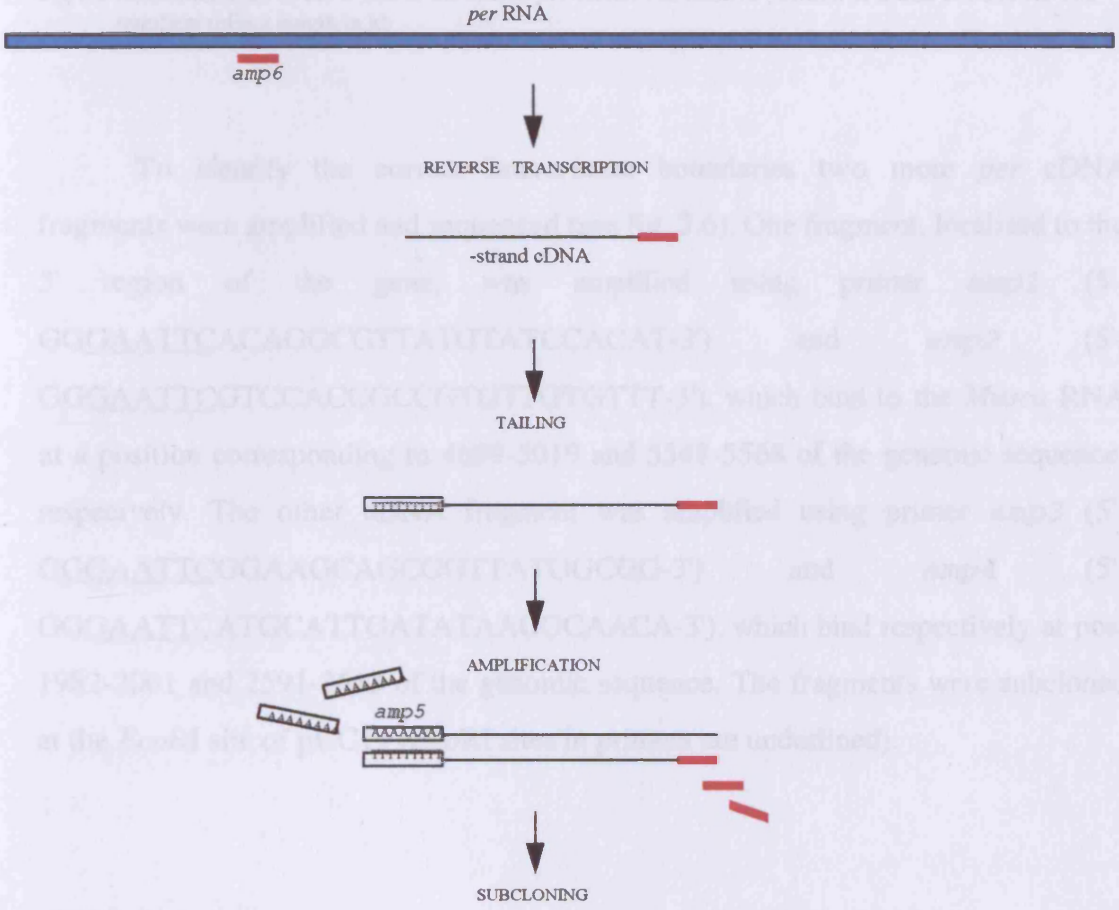


Fig. 3.4 The diagram depicts the technique used in order to amplify the 5' end of the *Musca per* transcript. mRNA was isolated from *Musca* heads (see chapter 2) and a reverse transcription reaction was set up using the specific *per* primer *amp6*. The - strand of this fragment of *Musca per* cDNA was then poly-T tailed with terminal deoxynucleotidyl transferase (Promega) following manufacturer's instructions, and finally amplified using *amp5* and *amp6* primers.

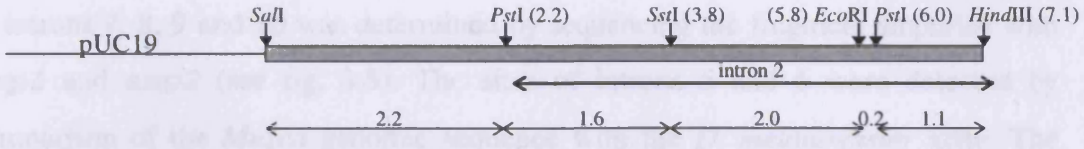


Fig 3.5 Restriction map of the 5' end of the *Musca per* clone. The relative position of intron 2 is shown. The numbers reflect length in kb

To identify the correct intron-exon boundaries two more *per* cDNA fragments were amplified and sequenced (see fig. 3.6). One fragment, localised to the 3' region of the gene, was amplified using primer *amp1* (5'-GGGAATTCACAGGCGTTATGTATCCACAT-3') and *amp2* (5'-GGGAATTCGTCCACCGCCGTGTTGTGTTT-3'), which bind to the *Musca* RNA at a position corresponding to 4699-5019 and 5548-5568 of the genomic sequence, respecively. The other cDNA fragment was amplified using primer *amp3* (5'-GGGAATTCGGAAGCAGCGGTTATGGCGG-3') and *amp4* (5'-GGGAATTCATGCATTGATATAACGCAACA-3'), which bind respectively at pos. 1982-2001 and 2591-2613 of the genomic sequence. The fragments were subcloned at the *Eco*RI site of pUC19 (*Eco*RI sites in primers are underlined).

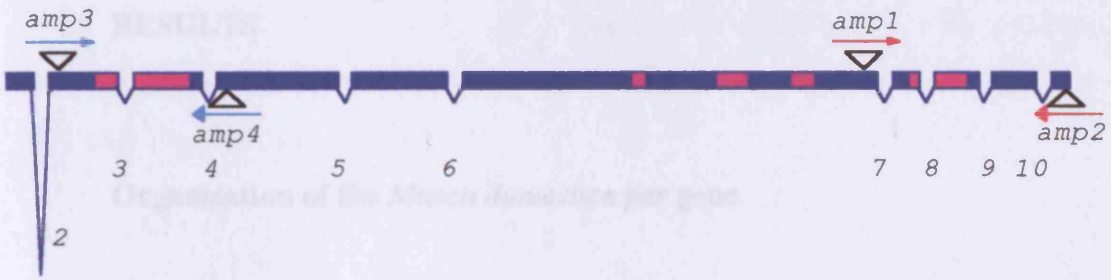


Fig 3.6 Scheme of the *Musca domestica per* gene. Conserved and non-conserved regions are indicated with the same colour code used in fig. 3.1 The position of primers used to amplify cDNA fragments in order to identify intron/exon boundaries is shown. Introns are numbered.

Comparison of the sequence of the mRNA fragment amplified by *amp3* and *amp4*, permitted the localisation of two introns, introns 3 and 4, while the position

of introns 7, 8, 9 and 10 was determined by sequencing the fragment amplified with *amp1* and *amp2* (see fig. 3.5). The sites of introns 5 and 6 were detected by comparison of the *Musca* genomic sequence with the *D. melanogaster* gene. The position of intron 1 was identified by comparing the cDNA sequence of the 5' end of the gene with the corresponding genomic sequence; only a fragment of this intron has been sequenced and its size has not been determined.

Computer analysis

DNA and protein data analysis were performed using various programs of the GCG package for molecular biology (version 7; University of Wisconsin Genetics Computer Group, Madison, WI, USA; Devereux *et al.*, 1984). Multiple sequence alignment was performed with the program ClustalW (Higgins *et al.* 1988) and corrected by eye. The phylogenetic analysis was computed with the PHYLIP (Phylogeny Inference Package, version 3.57c) package provided by J. Felsenstein (University of Washington, Seattle, Washington USA). PEST sequences analysis was performed with the PEST-FIND program (Rogers *et al.*, 1986).

RESULTS

Organisation of the *Musca domestica per* gene

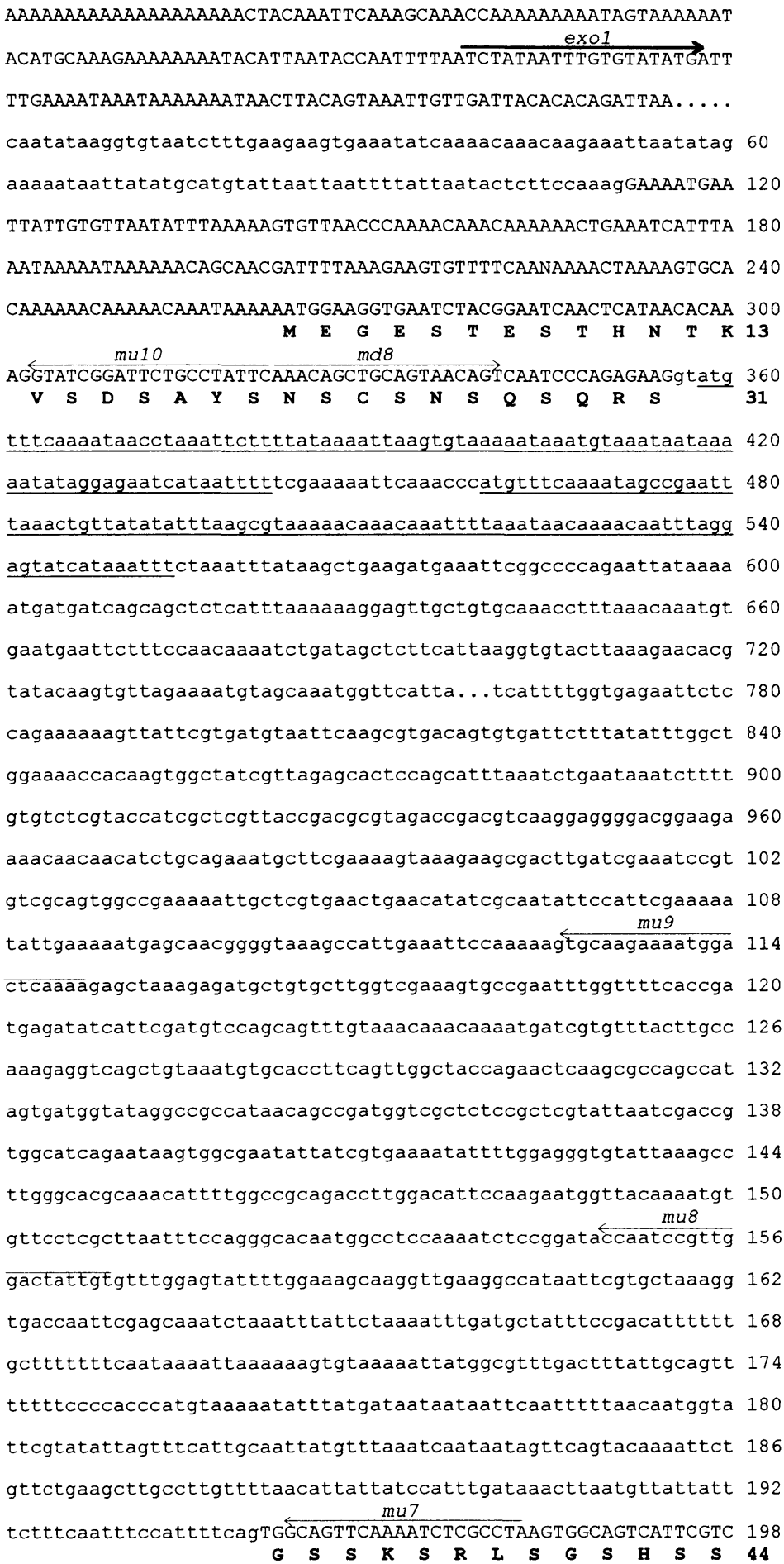
While the sequencing of the coding region of the *Musca per* gene has been completed by the time of writing, a 3.5 kb portion of the intron corresponding to the *melanogaster* intron 2 was left unsequenced. Fig. 3.8 displays the sequence of the *per* gene of *M. domestica*. The sequence information available for the 5' UTR was mostly derived from cDNA so it is not possible to determine the genomic length of this region. However evidence for an intron was found upon comparison of the 5' UTR with some genomic sequences. The presence of additional introns in the 5'

untranslated region of the gene therefore cannot be excluded, so the numeration of exons and introns must be regarded as provisional.

per in *Musca domestica* measures 9 kb from the starting codon to the putative polyadenylation signal. This dramatic difference in size with respect to its *Drosophila* homologues is somewhat deceptive; the encoded protein, with its 1072 amino acid length is shorter than that of *D. virilis*. The size of the gene is increased by a modification in its intron-exon structure, three more introns are present in the *Musca* transcript and the arbitrarily called intron 2 has "expanded" to 5 kb. Apart from the large intron 2, the other introns falling within the coding sequence are small, ranging in size between 50 and 68 bp. Attempts to measure the length of intron 1 with PCR have been unsuccessful (the primers used were *exo1* (5'-TCTATAATTTGTGTATATGA-3') and *mu10* (5'-GAATAGGCAGAATCCGATAC-3'), see fig 3.8). Intron 2, the first intron of the primary transcript to fall within the coding sequence, resides in a region of high conservation, where only synonymous substitutions are observed in an area of about 160 bp around the point of insertion; the intron is located at exactly the same position in all *per* genes and the intron/exon boundaries are well conserved. Yet, in *Musca* this intron presents a vast increase in size when compared with its positional homologues. Intron 2 measures 5 kb in *Musca* while in *Drosophila* its length varies between 60 (*melanogaster*) and 70 bp (*virilis*). The sequence available from *Musca* intron 2 (1.6 kb) shows traces of at least one duplication event involving approximatively 90 bp (fig. 3.7), indicating how this intron may have expanded to its final size from a shorter ancestor.

```
358 ATGTTTCAAAATAACCTAAATT...CTTTTATAAAATTAAGTGTAAAAATAAA.....TGTAAATAATAAA..AATATAGGAGAATCATAATTTT 442
    |||||||||||||| || || ||   || ||||| | ||||| ||||||| |||     | ||||||| |||  ||| |||||| ||||||| |||
460 ATGTTTCAAAATAGCCGAATTTAAACTGTTATATATTTAAGCGTAAAAACAAACAAATTTTAAATAACAAAACAATTTAGGAGTATCATAAATTT 554
```

Fig 3.7 Alignment of the repeats found within the second intron of the *per* gene in *Musca domestica*. Numbers correspond to the position of the sequence within the intron. Gaps (represented by dots) were introduced in the sequence in order to improve the alignment.


exo1
mu10
md8
mu9
mu8
mu7

amp6
amp3
1981 GGGAAGCAGCGGTTATGGCGGCAAACCATCAACACAAACCAGCAGGTTAAAACTTAAATC 2040
45 G S S G Y G G K P S T Q T S R L K L K S 64

2041 CCCATTTCCATATTATTCTGAAACTAATTGTAACTTTTTCTCCATTAATACACAAAGTGA 2100
65 P F P Y Y S E T N C N F F S I N T Q S E 84

2101 AATTCACATTACCAAAAGGGGTAAGGACAAGGGCAGGAAAAAGAAGAAACAAAAGTGTTC 2160
85 I H I T K R G K D K G R K K K K Q K C S 104

2161 CTCAACCGTGACAAATGCCTTGGATTCACAAGAAGATGTGGCCAAAACAAATGAAACAAA 2220
105 S T V T N A L D S Q E D V A K T N E T K 124

2221 GGCCGATGAAGAAATTGGTCTCAATGATATTACAATGAgtaggtggtaaacgaaatagat 2280
125 A D E E I G L N D I T M T 137

mu6
2281 catgtttgatatttaaaatatacctcttttatcatccatagCTCAAGGTAATCGACAAAA 2340
138 Q G N R Q K 143

2341 GGAGAATAAAGAGAATTTGGAAAAGGAAAATAATGAACCcAAAATAGAGAAGTCCCTTCA 2400
144 E N K E N L E K E N N E P K I E K S L Q 163

2401 GTCGCCCATACCATCACCATTGTCGGCAACCACTAATCAAGGTGTAAAATCTGAGAAAAC 2460
164 S P I P S P L S A T T N Q G V K S E K T 183

2461 TTGCGAATCGGCTCCTGGTAAATTGGAATCTTCGGGAAAAACTGAAAAACTTAAAGAGgt 2520
184 C E S A P G K L E S S G K T E K L K E 202

2521 aacttaatgagagaattaaaaaaacaaagaaatcgcctaatgttttgacattattttcca 2580

amp4 mu5
2581 gGACAGTTTCTGTTGCGTTATATCAATGCATGATGGCATTGTTTTATATACCACACCCAG 2640
203 D S F C C V I S M H D G I V L Y T T P S 222

2641 TATTACAGATGTACTTGGTTTCCCCAGAGATATGTGGCTGGGACGTTCATTTATTGATTT 2700
223 I T D V L G F P R D M W L G R S F I D F 242

2701 TGTACATCCCAAAGATCGCGCCACATTTGCCAGCCAGATAACTACGGGCATTCCCATTGC 2760
243 V H P K D R A T F A S Q I T T G I P I A 262

2761 CGAGTCGAGGAGCAGTATACCCAAAGATGCCCGCAGCTCGTGCTGTGTAATGCTAAGGCG 2820
263 E S R S S I P K D A R S S C C V M L R R 282

mu4
2821 TTATCGTGGCCTCAAATCGGGGGGTTATGGTGTTATAGGACGCTCGGTGAATTACGAGCC 2880
283 Y R G L K S G G Y G V I G R S V N Y E P 302

2881 GTTTCGTTTGGGTCTGACCTTTCGTGAAGCTCCCGAAGAGGCaAGATCAGATAATTCTTT 2940
303 F R L G L T F R E A P E E A R S D N S L 322

2941 ACCCAGTGGTACAAACATGCTATTGGTAATATGTGCCACACCCATAAAAAGTGCTTACAA 3000
323 P S G T N M L L V I C A T P I K S A Y K 342

3001 AGgtatctcaaaagaaccgaaactttccacagttttctctaattttcctatttttcagTA 3060
343 V 344

3061 TGCGATGAATTGCTATCCCGAAAAACTCCTAAATTTGCAATACGTCACACAAAGACTGGT 3120
345 C D E L L S R K T P K F A I R H T K T G 364

3121 ATTATTTCAACAGTAGATAGTGGAGCTGTTTCAGCATTGGGTTATTTGCCCCAGGACCTT 3180
365 I I S T V D S G A V S A L G Y L P Q D L 384

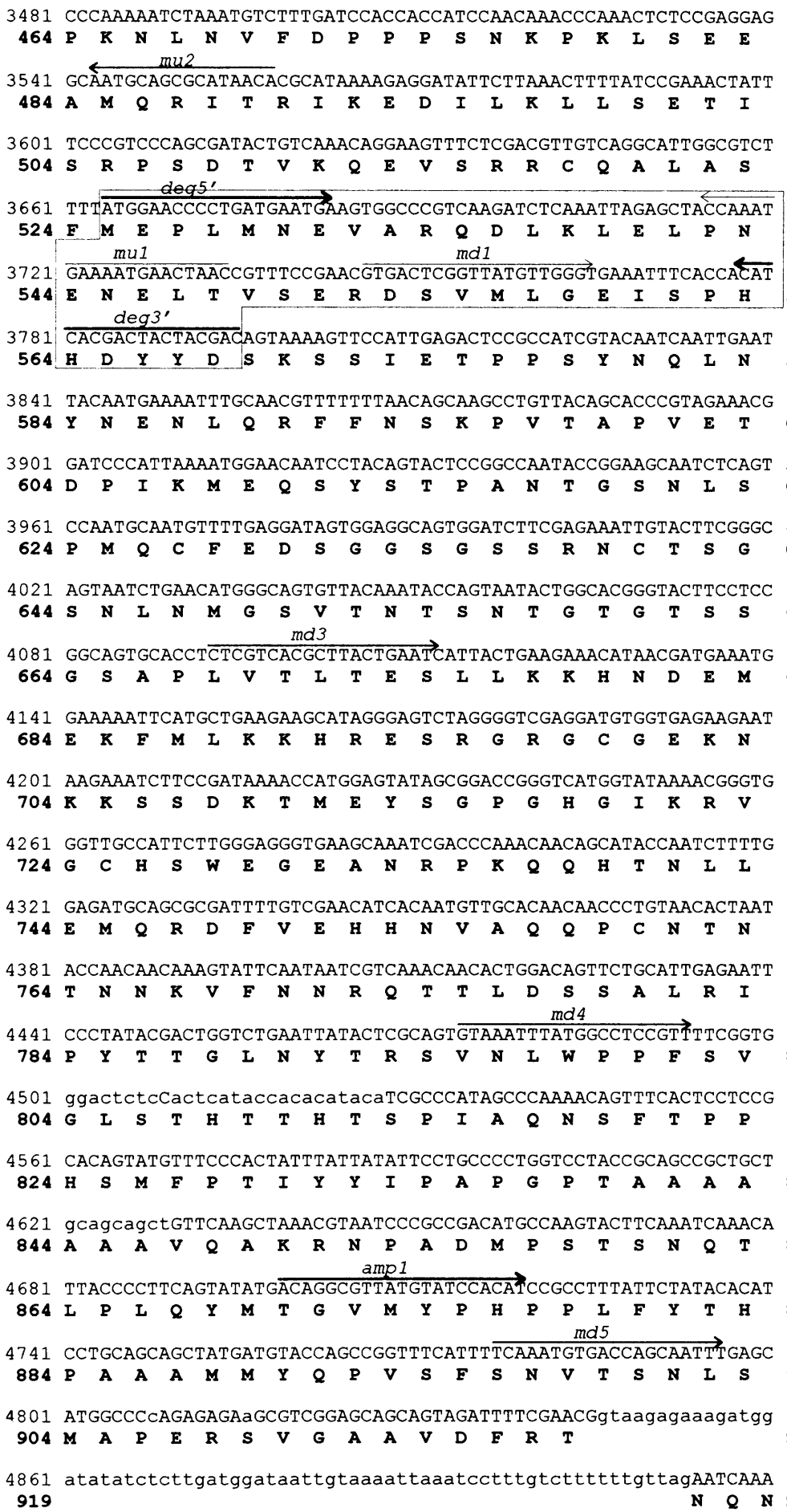
3181 ATTGGTCGctccatattggacttttatCACCATGAAGATCTCACTGTCTTtAAAGGAAATC 3240
385 I G R S I L D F Y H H E D L T V L K E I 404

mu3
3241 TACGAGACTGTTATGAAGAAAGGCCAAACGGCTGGTGCTTCATTTTGTAGCAAaCCCTAT 3300
405 Y E T V M K K G Q T A G A S F C S K P Y 424

3301 AGATTCCTTGTGCAAAATGGCTGCTATGTTCTCTTGGATACAGAGTGGACCAGCTTCGTA 3360
425 R F L V Q N G C Y V L L D T E W T S F V 444

3361 AATCCATGGTCAAGAAAATTGGAATTTGTCATTGGACACCATCGCGTCTTTCAGGgtaag 3420
445 N P W S R K L E F V I G H H R V F Q G 463

3421 ttaaaacactattatgcctaggagctttcctaaagcatatattcattgcttatttcagGT 3480


mu2
deg5'
mu1
md1
deg3'
md3
md4
amp1
md5

```
4921 ATTTAATGGTGGCCCCAAATAGTTCTAAACCTCAACAACAACAGCCACATCAGCAGCAAC 4980
 922   L  M  V  A  P  N  S  S  K  P  Q  Q  Q  Q  P  H  Q  Q  Q  Q  941

4981 AACAGGGAGCTTTTCATTCAATAACACCCGTGCAATTGCAGCGACCATCGTCCCAGGCCA 5040
 942   Q  G  A  F  H  S  I  T  P  V  Q  L  Q  R  P  S  S  Q  A  T  961

5041 CTTCGGTAAAAGCCGAGCCAGGATCAaATATGgtaggtttagattagaagagatattgtt 5100
 962   S  V  K  A  E  P  G  S  N  M                                971

5101 atggggttatttacactgttcatattattttgttttcgaatagGCTCCATCGGATTCTTC 5160
 972                                              A  P  S  D  S  S  977

                md6
5161 TAAAAAGGGAATTGCTGGTTCTCCCATTGTCTCTGTCATGGGTGATTATGTTTCTGATCA 5220
 978  K  K  G  I  A  G  S  P  I  V  S  V  M  G  D  Y  V  S  D  Q  997

5221 GCACAACGAAAACACTCTTAAGgtaattttgaaacttacacatttaaaaaatatgtttaa 5280
 998  H  N  E  N  T  L  K                                        1004

5281 ggatcgtatctaatttttttacagCCCAACACCGATAGTAATGGTAATAGTGATGATATGG 5340
1005                        P  N  T  D  S  N  G  N  S  D  D  M  D 1017

5341 ACGGTTCCAGCTTTTCATCGTTTTATTCTTCATTTATTAAAACAACTGACGGCTCAGATA 5400
1018   G  S  S  F  S  S  F  Y  S  S  F  I  K  T  T  D  G  S  D  S 1037

5401 GTCCCCAGGAGAACGACAACAGCAAAGAGGCACGAAAATTTAAGgtaagttttacttaaa 5460
1038   P  Q  E  N  D  N  S  K  E  A  R  K  F  K                  1051

5461 aacaccagtaattcccaaaattgatctattctactttgaaaaagGTACAGACTGAAGAGA 5520
1052                                                 V  Q  T  E  E  K 1057

                md7                          amp2
5521 AAATAATGGAAGATGCCGAGGAGGATGAAACACAACACGGCGGTGGACTGTAAATTATAA 5580
1058   I  M  E  D  A  E  E  D  E  T  Q  H  G  G  G  L  *           1073

5581 GATATTCGTGATATGGATATTTTTTTTCTCGTCACAAGCATTTTGGATGGTCGCTCAAGT 5640

5641 ACAAATCAGCATAGATGGAATGCAGTAGGTTCCCTAAGCACTTCGTACTATATTTCTATA 5700

5701 TTTCATAAACGTTTTTCTTTTTATTTCATTTTTACTTGCAGTTTTCAAAAAAACACTTTG 5760

5761 TGGACACATTTAGTTATAGCAGTTTTCTACTATATACTATTTTGTTAACAGTTGGCTATT 5820

5821 TTGTAATACAATAGATATGTAATTAAGAATGAAAGGATACCTTATTTGTATTTAAGAGAA 5880

5881 AAGAACTAAAGATAAAATAAAGAATATTTTAATTTTACTTACAAATCATATTGTACATCA 5940

5941 CTTCTTTCGTTTCTTGTTGGTTTTGCTCGGCGATTTTTCAACACCCTTTTCTGTTTGGAA 6000

6001 TTGTTGCAATTCGGAGCAGAATAAGACCTAGAAAATGAATTAAAAAAGCCTTAAATATAA 6060

6061 AAAGAATTTATTTTGAAAATGTATGTCACAAACCGCCTGCGTCCATCCAGCATAGGGTCC 6120

6121 ATAAATATTTCGAAAGTTCTCCGCAATTTCTTCGTAGATCT 6162
```

Fig. 3.8 DNA and amino acid sequence of the *Musca domestica per* gene. Amino acids are indicated in bold type with the one-letter code below the first nucleotide of each codon. Introns are represented in lower case. Note that part of the sequence of intron 1 and 2 is missing. Dots have been inserted to indicate this fact. The numeration of the DNA sequence starts in intron 1 and must be regarded as provisional. The two repeats found in intron 2 are underlined. The black arrows on top of the sequence represent primers used for the sequencing, the blue arrows delineate primers used to amplify fragments of the gene. The region amplified in the degenerate PCR has been boxed in blue. Also boxed (in red) is the putative polyadenylation signal.

All the remaining introns of the *Musca per* gene are short, resembling those in *D. melanogaster*, including introns 3, 4 and 8 which do not have a *Drosophila* counterpart. In region c2 are introns 4, 5 and 6. The latter two, unlike intron 4, have a *Drosophila* positional homologue occurring in the same codon and in the same phase. Introns 7, 9 and 10 of *Musca* have a *Drosophila* equivalent even though the divergence between the *Musca* and *Drosophila per* genes does not allow an alignment of the neighbouring exon sequence, while intron 8, which is found in nc5, does not have a *Drosophila* counterpart. Alignment of the intron sequences does not reveal any significant sequence similarity. All the introns in the *Musca* gene are rich in A+T in contrast with *Drosophila*. In fact in *D. melanogaster* none of the introns are A+T rich, while in *D. pseudoobscura* and *virilis* a bias is noticed in the nucleotide composition of the larger introns; intron 4 and 7 in *pseudoobscura* and intron 3, 4 and 7 in *virilis* have a high A+T content. The overall A+T content of the whole of the available intron sequences is therefore 52% and 53% for *melanogaster* and *yakuba*, respectively, 62% in *pseudoobscura*, 60% in *virilis* and 66% in *Musca*.

A rather different situation is found for the coding sequence. Here all the *Drosophila* species show a relatively high C+G content (62%, 64%, 60% and 56%, respectively in *melanogaster, yakuba, pseudoobscura* and *virilis*), while *Musca* displays a 44% C+G content. This observation reflects a different codon usage in *per* between the different groups: among Drosophilids C- and G- ending codons are strongly preferred while in *Musca* codons are favoured which terminate in either A or T. Published codon usage tables for *D. melanogaster* (Sharp *et al.* 1992) display similar *per*-like values for highly expressed genes. Given the low number of genes sequenced to date in *M. domestica*, a codon usage table is not available, but comparison of codon usage in different genes shows that the preference for A- and T- terminating codons is not a general rule for this species. In particular, *armadillo*, *nanos* and a gene coding for the alpha1 subunit of a calcium channel have an high ratio of A- and T- ending codons while in *bicoid*, *hunchback*, *Ultrabithorax* and the gene coding for a member of the cytochrome P-450 family, the codon usage seems to be more uniform. In the gene coding for the glutathione transferase 1a, C- and G- ending codons are predominant (see table 3.1); it must be noted, though, that this

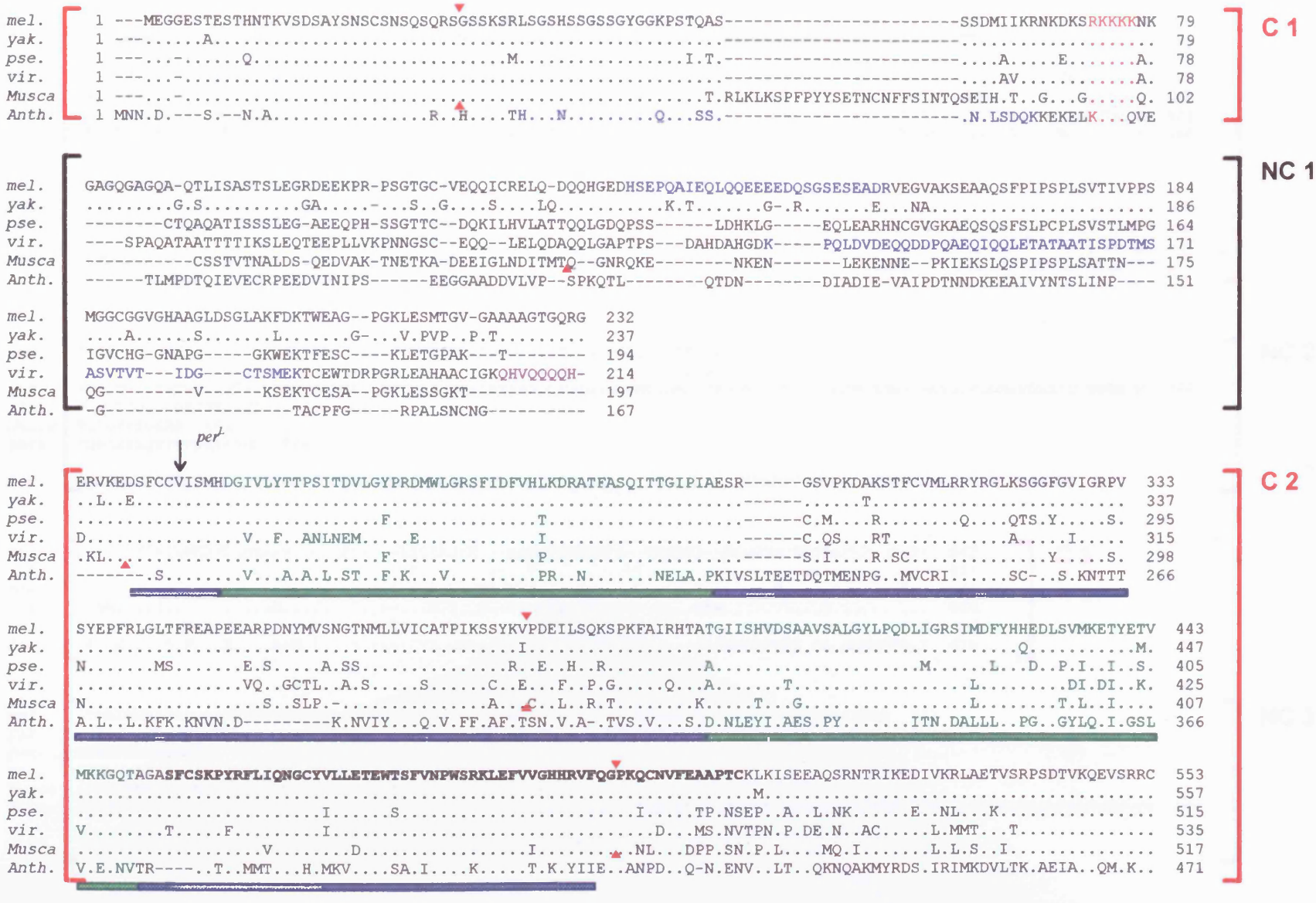
gene has a rather short coding sequence. The significance of these observations is unclear.

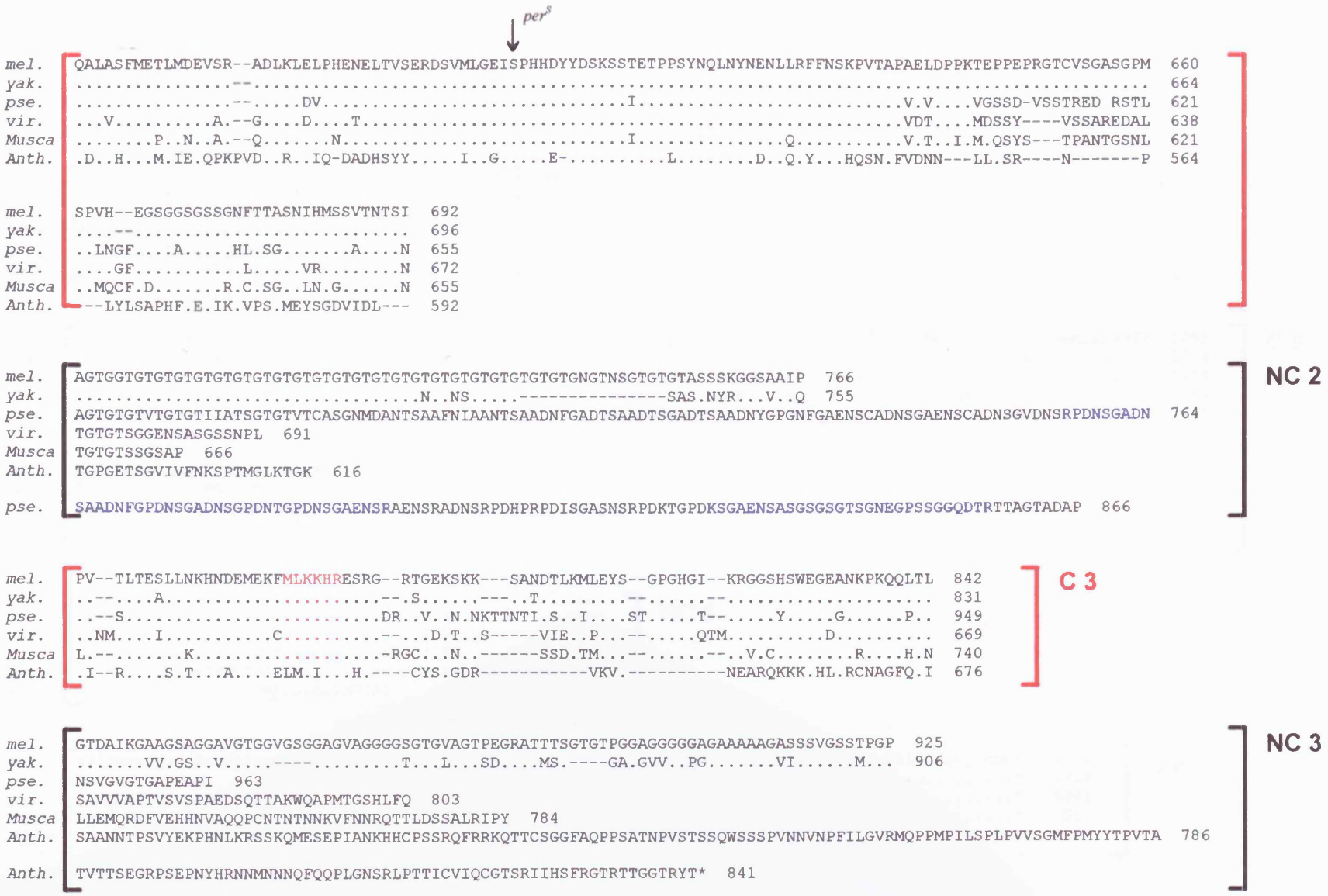
	A- + T- ending	C- + G- ending	Total codons
period	61.9%	38.1%	1072
armadillo	57.7%	42.3%	814
nanos	67.1%	32.9%	416
α1 subunit calcium channel	70.2%	29.8%	1688
*bicoid**	54.9%	45.1%	317
*hunchback**	42.5%	57.5%	527
*Ultrabithorax**	52.4%	47.6%	248
cytochrome P 450	45.7%	54.3%	536
glutathione transferase1a	31.6%	68.4%	209

Table 3.1. Percentage of A- + T- and C- + G- ending codons in various genes of *M. domestica*. The sequence of the genes indicated by an (*) is incomplete.

In *D. melanogaster* evidence supporting the existence of alternative splicing pathways has been given by Citri *et al.* (1987). Three different transcripts are described by these authors; the most abundant being type-A, the translation of which gives the protein shown in fig. 3.9a. In the type-B transcript an intron is excised from exon 5, resulting in a 96 amino acid deletion of the type-A protein. In type-C, introns 5, 6 and 7 are not spliced and translation proceeds through "intron 5" and arrives out-of-frame into exon 6, the result of which is that a 107 amino acid fragment at the C-terminus in type-C Per protein replaces the 149 amino acid C-terminus of type-A. Other investigators failed to confirm these findings (reviewed in Hardin and Siwicki, 1995) and there is considerable doubt in the existence of type-B and type-C transcripts, in particular Hall (1995) suggests that they were an artefact. Therefore I did not attempt an analysis of the *Musca* transcripts, limiting this investigation to some considerations based on the gene sequence.

As reported by Citri *et al.* (1987), in type-B a splicing event occurs which involves a non-consensus splicing site (Padgett *et al.* 1986) at the 3' junction. A simple visual inspection of the *Musca* sequence revealed no consensus sequences for splice sites within exon 5, although a non-canonical splice site cannot be excluded.


C 1
NC 1
C 2
per


per^S
mel. QALASFMETLMDEVSR--ADLKLELPHENELTVSERDSVMLGEISPHHDYYDSKSSTETPPSYNQLNYNENLLRFFNSKPVTAPAELDPPKTEPPEPRGTCVSGASGPM 660
yak. ....--..... 664
pse. ..............--.....DV.....I.....V.V....VGSSD-VSSTRED RSTL 621
vir. ...V.....A.--G....D....T.....VDT....MDSSY----VSSAREDAL 638
Musca .....P..N..A.--Q.....N.....I.....Q.....V.T..I.M.QSYS---TPANTGSNL 621
Anth. .D..H...M.IE.QPKPVD..R..IQ-DADHSYY....I..G.....E-.....L.....D..Q.Y...HQSN.FVDNN---LL.SR----N-------P 564
mel. SPVH--EGSGGSGSSGNFTTASNIHMSSVTNTSI 692
yak. ....--..... 696
pse. ..LNGF....A.....HL.SG.....A....N 655
vir. ....GF.....L.....VR.....N 672
Musca ..MQCF.D.....R.C.SG..LN.G.....N 655
Anth. --LYLSAPHF.E.IK.VPS.MEYSGDVIDL--- 592
mel. AGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGNGTNSGTGTGTASSSKGGSAAIP 766
yak. .....N..NS.....----------SAS.NYR...V..Q 755
pse. AGTGTGTVTGTGTIIATSGTGTVTCASGNMDANTSAAFNIAANTSAADNFGADTSAADTSGADTSAADNYGPGNFGAENSCADNSGAENSCADNSGVDNSRPDNSGADN 764
vir. TGTGTSGGENSASGSSNPL 691
Musca TGTGTSSGSAP 666
Anth. TGPGETSGVIVFNKSPTMGLKTGK 616
pse. SAADNFGPDNSGADNSGPDNTGPDNSGAENSRAENSRADNSRPDHPRPDISGASNSRPDKTGPDKSGAENSASGSGSGTSGNEGPSSGGQDTRTTAGTADAP 866
NC 2
mel. PV--TLTESLLNKHNDEMEKFMLKKHRESRG--RTGEKSKK---SANDTLKMLEYS--GPGHGI--KRGGSHSWEGEANKPKQQLTL 842
yak. ..--....A.....--.S.....---..T.....--.....--..... 831
pse. ..--S.....DR..V..N.NKTTNTI.S..I....ST.....T--.....Y.....G.....P.. 949
vir. ..NM....I.....C.....--...D.T..S-----VIE..P...--.....QTM.....D..... 669
Musca L.--.....K.....-RGC...N..------SSD.TM...--.....--..V.C.....R....H.N 740
Anth. .I--R....S.T...A....ELM.I...H.----CYS.GDR----------VKV.----------NEARQKKK.HL.RCNAGFQ.I 676
C 3
mel. GTDAIKGAAGSAGGAVGTGGVGSGGAGVAGGGGSGTGVAGTPEGRATTTSGTGTPGGAGGGGGAGAAAAAGASSSVGSSTPGP 925
yak. .....VV.GS..V.....---.....T...L...SD....MS.----GA.GVV..PG.....VI.....M... 906
pse. NSVGVGTGAPEAPI 963
vir. SAVVVAPTVSVSPAEDSQTTAKWQAPMTGSHLFQ 803
Musca LLEMQRDFVEHHNVAQQPCNTNTNNKVFNNRQTTLDSSALRIPY 784
Anth. SAANNTPSVYEKPHNLKRSSKQMESEPIANKHHCPSSRQFRRKQTTCSGGFAQPPSATNPVSTSSQWSSSPVNNVNPFILGVRMQPPMPILSPLPVVSGMFPMYYTPVTA 786
Anth. TVTTSEGRPSEPNYHRNNMNNNQFQQPLGNSRLPTTICVIQCGTSRIIHSFRGTRTTGGTRYT* 841
NC 3

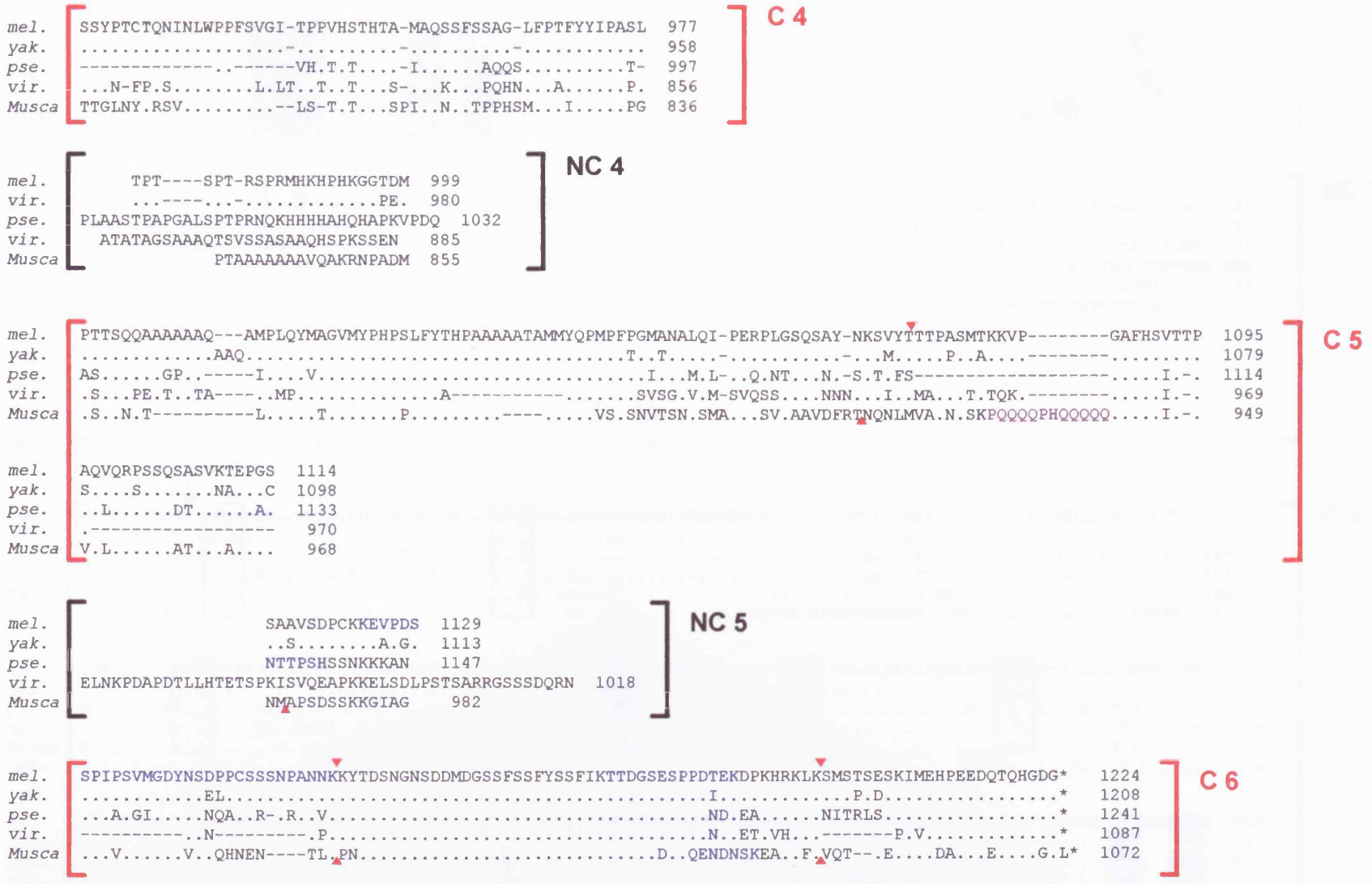
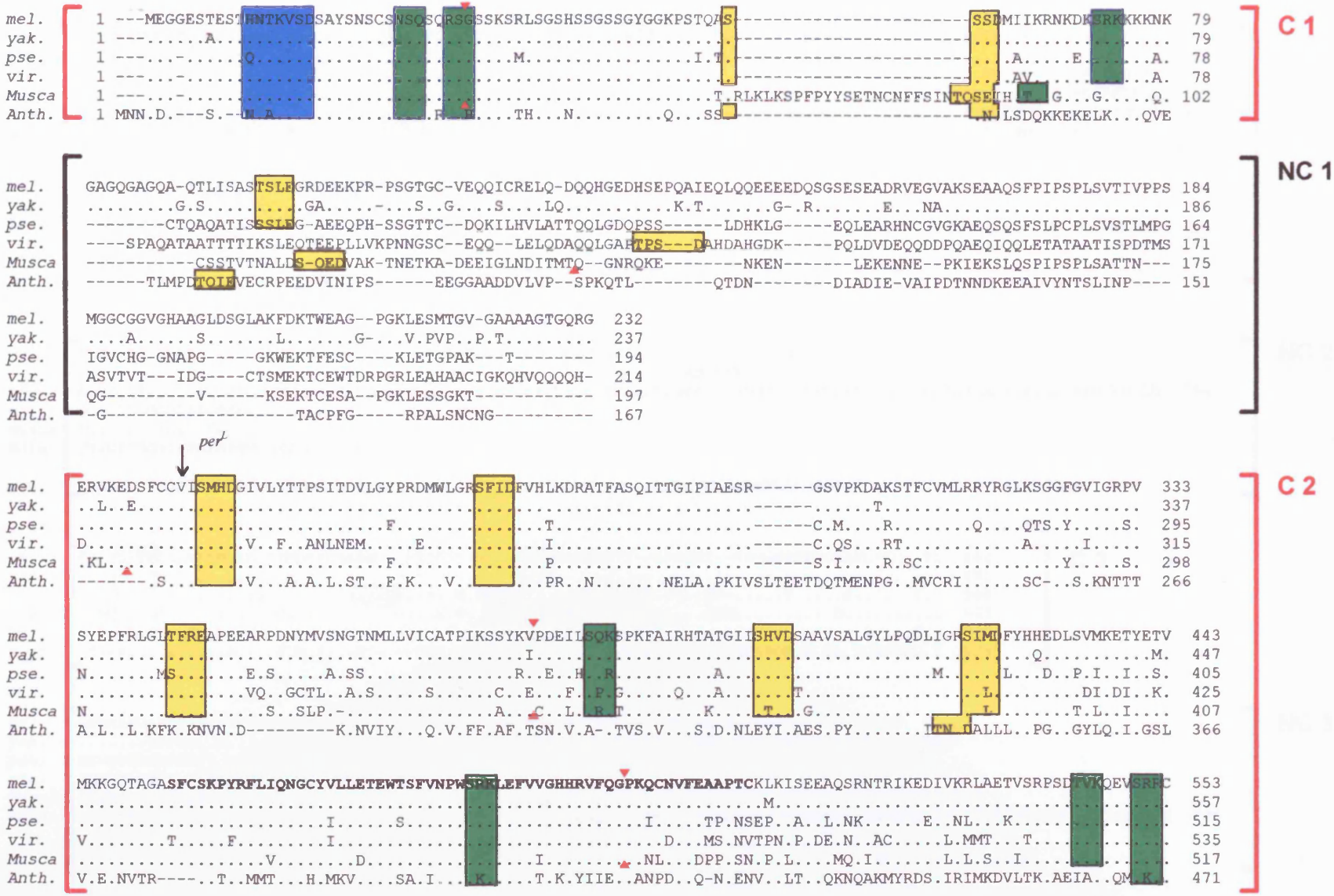


Fig. 3.9a. Alignment of the Per proteins from *D. melanogaster*, *D. yakuba*, *D. pseudoobscura*, *D. virilis*, *M. domestica* and *A. pernyi*. The alignment was performed with the ClustalW program (Higgins *et al.*, 1988) and corrected by eye. An alignment is also attempted in nc1, nc4 and nc5. The end of the sequences is indicated by asterisks. Dots represent equal amino acids and dashes represent deletions. The position of introns is indicated by red arrows. The sequence written in red is the nuclear translocation signal, in magenta the poly Q stretch and in blue the PEST regions. The PAS region is underlined and the 2 degenerate repeats are highlighted in green. The cytoplasmic localisation domain is shown in bold type. C=conserved regions, NC=non-conserved regions


C 1
NC 1
C 2
per


per^S
mel. QALASFMETLMDEVSR--ADLKLELPHENEITVSERDSVMLGEISPHHDYYDSKSSTETPPSYNQLNYNENLLRFFNSKPVTAPAELDPPKTEPPEPRGTCVSGASGPM 660
yak. 664
pse. 621
vir. 638
Musca 621
Anth. 564
mel. SPVH--EGSGGSGSSGNFTTASNIHMSSVTNTSI 692
yak. 696
pse. 655
vir. 672
Musca 655
Anth. 592
mel. AGTGGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGNGTNSGTGTGTASSSKGGSAAIP 766
yak. 755
pse. AGTGTGTVTGTGTIIATSGTGTVTCASGNMDANTSAAFNIAANTSAADNFGADTSAADTSGADTSAADNYGPGNFGAENSCADNSGAENSCADNSGVDNSRPDNSGADN 764
vir. TGTGTSGGENSASGSSNPL 691
Musca TGTGTSSGSAP 666
Anth. TGPGETSGVIVFNKSPTMGLKTGK 616
pse. SAADNFGPDNSGADNSGPDNTGPDNSGAENSRAENSRADNSRPDHPRPDISGASNSRPDKTGPDKSGAENSASGSGSGTSGNEGPSSGGQDTRTTAGTADAP 866
NC 2
mel. PV--TLTESLLNKHNDEMEKFMLKKHRESRG--RTGEKSKK---SANITLKMLEYS--GPGHGI--KRGGSHSWEGEANKPKQQLTL 842
yak. 831
pse. 949
vir. 669
Musca 740
Anth. 676
C 3
mel. GTDAIKGAAGSAGGAVGTGGVGSGGAGVAGGGGSGTGVAGTPEGRATTTSGTGTPGGAGGGGGAGAAAAAGASSSVGSSTPGP 925
yak. 906
pse. NSVGVGTGAPEAPI 963
vir. SAVVVAPTVSVSPAEDSQTTAKWQAPMTGSHLFQ 803
Musca LLEMQRDFVEHHNVAQQPCNTNTNNKVFNNRQTTLDSSALRIPY 784
Anth. SAANNTPSVYEKPHNLKRSSKQMESEPIANKHHCPSSRQFRRKQTTCSGGFAQPPSATNPVSTSSQWSSSPVNNVNPFILGVRMQPPMPILSPLPVVSGMFPMYYTPVTA 786
Anth. TVTTSEGRPSEPNYHRNNMNNNQFQQPLGNSRLPTTICVIQCGTSRIIHSFRGTRTTGGTRYT* 841
NC 3

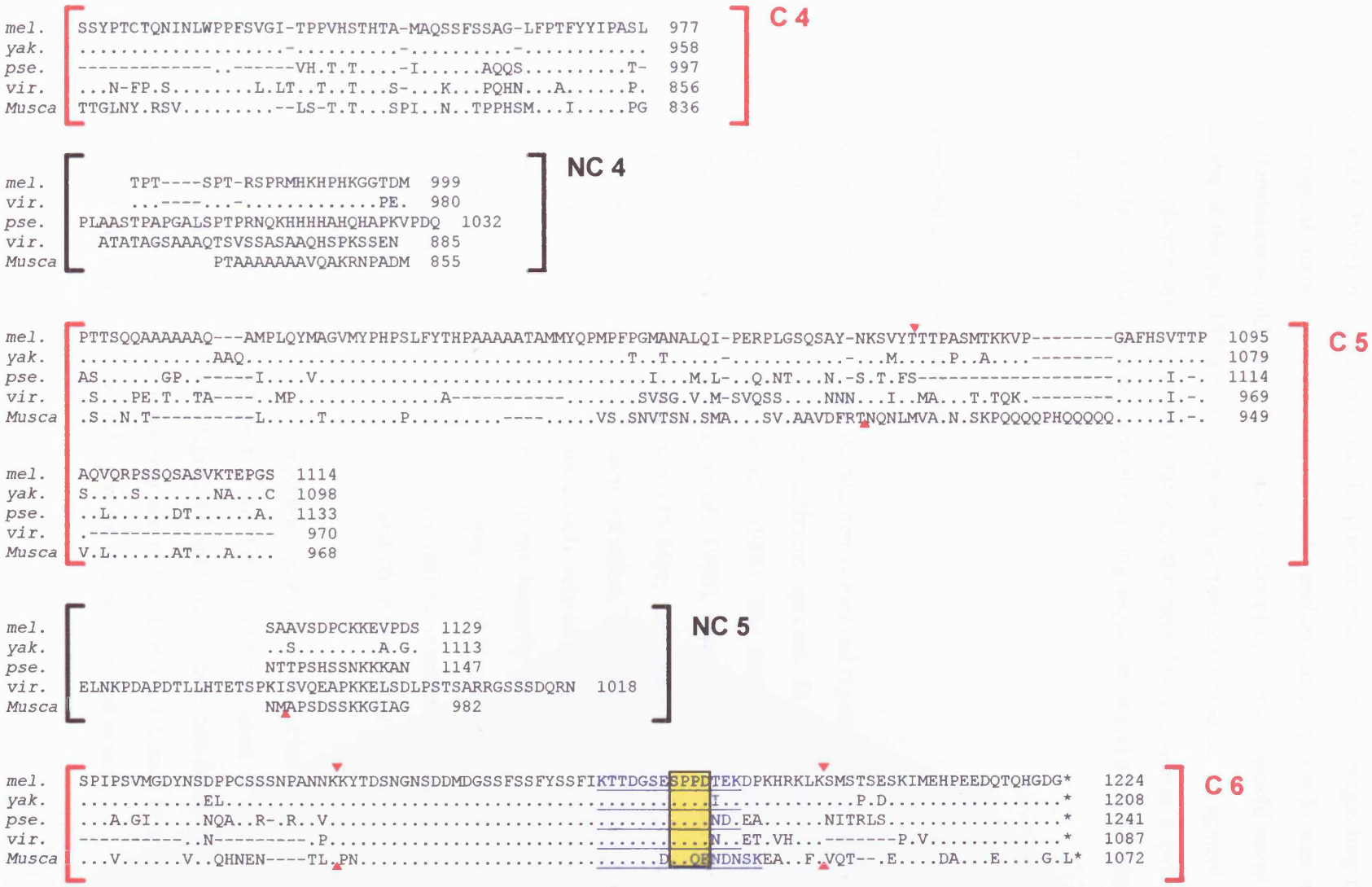


Fig. 3.9b. Putative phosphorylation sites of PER. Shaded in blue is the substrate of tyrosine protein kinases, in green the protein kinase C phosphorylation site and in yellow the casein kinase II site. Only the sites conserve at least among the dipteran proteins are shown. The N-terminal PEST sequence is highlighted and underlined in blue. The position of the introns is marked by red arrows on the *D. melanogaster* and *M. domestica* sequence. C=conserved regions, NC=non-conserved regions.

However it must be said that the non-canonical splice site described by Citri *et al.*, is the only known example of 3' splice junction deviating from the "AG rule" in *Drosophila* (other exceptions have been found in human (Rodriguez *et al.*, 1994; Watanabe *et al.*, 1996) and *Euglena gracilis* (Henze *et al.*, 1995)). Regarding the type-C transcript, if intron 7 ("homologous" to the *melanogaster* intron 5) was not spliced out, translation would stop at the fourth codon of exon 6 and would result in the substitution of the last 155 amino acids with a mere 26 amino acid fragment. If any alternative splicing involves reading through the neighbouring intron 8 (which does not have a *Drosophila* counterpart), the coding sequence would stop 4 codons within this "intron".

Conserved coding sequences

The division of *per* into conserved and non-conserved regions was introduced upon comparison of the gene in three different species, *D. melanogaster*, *D. pseudoobscura* and *D. virilis* (Colot *et al.*, 1988). The fourth *Drosophila* gene, cloned later from *D. yakuba* (Thackeray *et al.*, 1990), given the short evolutionary distance of this species from *melanogaster* (15 Myr, Lachaise *et al.* 1988; 6 Myr, Russo *et al.*, 1995), does not show much variation, even in the so called non-conserved regions, when compared to its closely related homologue. This overall pattern of variation is completely preserved in the housefly gene (fig. 3.9a); the 6 conserved blocks are clearly apparent upon comparison of *Musca per* with any of the *Drosophila* homologues. As in *Drosophila*, c1 and c2 constitute most of the N-terminal part of the protein, while c3, c4, c5 and c6 are localised in the C-terminal half and are generally less well conserved.

With the aid of the *Musca* sequence, *per* has been cloned from the lepidopteran *Antheraea pernyi* (Reppert *et al.* 1994). The *per* gene of this insect shows, as expected, a much lower degree of similarity when compared with its dipteran relatives; the regions of homology are reduced to c1, c2 and, to a lower extent, c3 while the C-terminal half of the protein is much reduced in size.

In *Musca* the similarity of c1, c2 and c3 to the *Drosophila* proteins is very high, between 80 and 92%, slightly lower in c6 (75-82%) and considerably lower in c4 and c5 (57-71%) as scored by the GCG program "bestfit". In the N-terminal block c1 there is a 24 amino acid insertion (aa 59-83) in *Musca* (see fig. 3.9a). Probably an insertion occurred in the lineage which led to *Musca* after the separation of Schizophora into Acalyptratae (*Drosophila*) and Calyptratae (*Musca*), rather than a deletion in the group Acalyptratae, since no trace of an insertion in that position is found in *Antheraea pernyi*. Apart for this, *Musca* c1 is very similar to the *Drosophila* homologous region: 8 amino acids are replaced between *Musca* and *melanogaster*, a number comparable with the 7 substitutions in the *pseudoobscura/melanogaster* pairing. In addition, an amino acid deletion at position 4 is shared by *Musca*, *D. pseudoobscura* and *D. virilis* with respect to *melanogaster* and *yakuba*, indicating that the glycine at this position of *D. melanogaster* (in the fragment MEGGES) may have originated by slippage after the separation of *D. pseudoobscura* from the *melanogaster* subgroup.

c2 is the longest conserved block, representing almost half the length of the entire *Musca* Per protein. The sites to which the per^L, per^S and per^0 mutations have been mapped (Baylies *et al.* 1987, Yu *et al.* 1987 a) are included within this region and are perfectly conserved in all Per proteins. In this area an amino acid deletion is observed in *Musca* compared to *Drosophila*; it lies in a short stretch where several substitutions are clustered (pos. 351-363 in *D. melanogaster*), between the two degenerate 51 residue PAS repeats (Crews *et al.*, 1988). *Musca* c5 contains a cluster of glutamines at pos. 931-941, a feature associated with transcriptional activators (Courey *et al.* 1988, Emili *et al.* 1994); a poly-Q stretch is also found in nc1 of *D. virilis* and this region is glutamine-rich in all the *Drosophila* proteins.

Protein database searches performed with the *Musca* conserved regions, fail to detect any significative similarity in polypeptides other than those with SIM, AHR and ARNT in c2.

Non-conserved coding sequences

The non-conserved regions amount to about 20-30% of the entire length of the Per proteins. The most N-terminal of these non conserved blocks, nc1, consists of 100-150 amino acid situated in the best conserved half of the protein. The remaining non-conserved regions are localised in the C-terminal half. nc2, with its long Thr-Gly perfect repeat in *D. melanogaster* and *yakuba*, the long pentapeptide repeat in *D. pseudoobscura* and the short Thr-Gly repeat in *Musca* and *D. virilis*, is the most outstanding non-conserved region both in the nature of its sequence and the extent of the variability observed. This region will be described in greater detail in a different section. Analysis of this region by Thackeray *et al.* (1990) based on the algorithm of Chou and Fasman (1974), was subsequently supported by structural studies and molecular dynamics simulations of Thr-Gly polymers (Ishida *et al.*, 1994b; Castiglione-Morelli *et al.* 1995) and indicate that nc2 exists in a β-turn conformation and may act as a flexible spacer between two globular domains of the protein.

No prominent features are found in nc3, nc4 and nc5; these regions are quite short (except for *melanogaster* and *yakuba* nc3, which is about 90 residue long), but this is the only common characteristic. No similarity in sequence, electric charge or hydrophobic moment was found upon their comparison using the PEPPLOT program of the GCG package (data not shown).

Putative functional motifs in the Per protein

In the proximity of the C-terminal end of conserved region 1 there is a short amino acid sequence which conforms to rules established for nuclear translocation signals (NTS, Kalderon *et al.* 1984). This stretch of lysines is found in the same area of all the Per proteins. Indeed this signal is sufficient in promoting nuclear localisation of Per in *D. melanogaster*, but only if a fragment of the protein (aa 453-511) is removed (Vosshall *et al.*, 1994; Saez and Young, 1996); the wild type protein can only enter the nucleus upon interaction with the Per partner, Timeless. A

second putative nuclear translocation signal is localised at the beginning of c3 (see fig. 3.9a) in the dipteran proteins, but no trace of it is found in *Antheraea*.

c2 contains the PAS region. Limiting our examination to the dipteran proteins, in the first PAS repeat (as shown in fig. 3.9a) only replacements involving similar residues are observed, while in the second repeat, in *D. virilis* an aspartic acid replaces a hydrophobic valine (pos. 435 in the *D. melanogastrer* protein). However the balance of charge is kept by the nearby substitution of a polar amino acid (threonine in *D. melanogaster*) for a lysine at position 442. In the region between the repeats there are some more radical replacements: *D. pseudoobscura* acquires 3 net negative charges by losing an arginine (317) and a lysine (322), both replaced by glutamine, and obtaining a glutamic acid instead of an alanine (351), *D. virilis* loses 3 positively charged amino acids (Lys 322 and Arg 352 and 393, replaced, respecively, by Ala, Gln, Gln) and acquires an extra Glu (replacing Val 377) and *Musca* acquires 2 more positive charges (one Arg and one Lys replacing Gln 384 and Ala 396, respectively). All substitutions and positions refer to the *D. melanogaster* sequence. Interestingly the number of amino acid substitutions observed in the PAS region is lower for the *D. melanogaster/M. domestica* pair than for *D. melanogaster/D. pseudoobscura* or *D. melanogaster/D. virilis*, contrary to expectations on the basis of phylogenetic distance. Also the cytoplasmic localisation domain (CLD) is found in c2, more precisely within the stretch of amino acids 453-511 of the *D. melanogaster* protein sequence (Saez and Young, 1996). This domain has been found to bind to a fragment of Tim localised within amino acids 715-914 (Saez and Young, 1996). Per-Tim interactions, by masking the Per CLD, allow the protein complex to be exported to the nucleus (Saez and Young, 1996). In contrast with what is erroneously reported by Saez and Young (1996), the CLD overlaps for about 3/4 of its length with the C-terminus of the PAS domain. Several amino acid substitutions are observed in the PAS portion of the CLD, but most of the mutations are located in the remaining fragment (see fig. 3.9a). The total number of substitutions in the CLD is in agreement with the phylogenetic distance between the species.

An *opa* repeat is found in c5 of *Musca per. opa*s consist of tandem degenerate repeats of the trinucleotide CAG which are often associated with developmental genes (Wharton *et al.* 1985). They can be part of the untranslated

sequence of the mRNA or be translated in any of the three coding frames. In *Musca* this repeat is translated to give the polyQ stretch at pos. 931-941 (see fig. 3.9a). Interestingly, SIM, ARNT and AHR possess a Q-rich domain at their C-terminal end, which is thought to act as an activation domain, and these proteins also have a bHLH domain at the N-terminus, which is not found in any of the Per proteins. Despite the polyQ motif of *Musca* being localised in a conserved region, a similar repeat is found only in *virilis*, among the other *per* orthologues. In *virilis* the *opa* repeat lies at the end of nc1, at pos. 207-214 of the coding sequence (fig. 3.9a). Interestingly, nc1 is relatively rich in glutamine even in *D. melanogaster* (14 out of 90 residues), *D. yakuba* (14 out of 90) and *D. pseudoobscura* (10 out of 90), where no *opa*s are found.

Various PEST sequences are found within the Per proteins. PEST sequences are regions of a protein rich in the amino acids proline, glutamic and aspartic acid, serine and threonine, generally flanked by clusters of positively charged residues, which are thought to act as proteolytic signals (Rogers *et al.*, 1986). They are normally associated with proteins with a rapid turnover and indeed the negative feedback model of *per* function (Hardin *et al.*, 1990) requires that Per is rapidly degraded. Only one PEST sequence is found in *Musca* at position 1070-1085; similarly *Antheraea* and *D. yakuba* have one PEST sequence (located in c1 in the former species and in c6 in the latter) while one to three more are found in the other *Drosophila* proteins (see fig 3.9a and table 3.2). One of these PEST sequences is conserved among all the dipteran Per proteins. The conserved PEST signal displays the highest PEST score in those proteins containing multiple PEST (table 3.2).

A high number of putative phosphorylation sites are spread along the proteins. The consensus sites conserved (or found at equivalent positions within non-conserved regions) at least between the dipteran PERs are reported in fig. 3.9b. One putative site for caseine kinase II-phosphorylation is found in all Dipterans within the N-terminal, conserved PEST sequence. Its possible significance will be discussed later in this chapter.

	Sequence	pos.	PEST score
Antheraea pernyi	HSGSNSSGSSGYGGQPSTSSSSNDLSDQK	38-67	9.6
D. melanogaster	HSEPQAIEQLQQEEEEDQSGSESEADR	136-163	15.64
	KEVPDSSPIPSVMGDYNSDPPCSSSNPANNK	1132-1163	7.45
	KTTDGSESPPDTEK	1190-1204	25.0
D. yakuba	**KTTDGSESPPDIEK**	1210-1224	17.7
D.pseudoobscura	RPDNSGADNSAADNFGPDNSGADNSGPDNTGPDNSGAENSR	786-827	10.6
	KSGAENSASGSGSGTSGNEGPSSGGQDTR	862-891	9.4
	KTEPASNTTPSH	1173-1184	11.0
	KTTDGSESPPDNDK	1244-1258	21.0
D. virilis	KPQLDVDEQQDDPQAEQIQQLETATAATISPDTMSASVTVTIDGCTSMEK	142-193	5.1
	KTTDGSESPPDNEK	1119-1127	22.7
Musca domestica	**KTTDGSDSPQENDNSK**	1070-1085	16.3

Table 3.2. Sequence of the PEST regions found in the Per proteins. The PEST score was calculated by the program PEST-FIND (Rogers *et al.*, 1986). Only sequences with a PEST score ≥ 5 are reported by the program. In bold type is the PEST sequence shared by all the dipteran Per proteins and, underlined, the phosphorylation site for the casein kinase II.

Molecular phylogeny of the *per* genes

Molecular phylogeny makes use of molecular data to establish evolutionary relationships between various groups of organisms. Mutation is a phenomenon generated by a number of stochastic processes. In neutrally evolving DNA[1] such as pseudogenes, random nucleotide changes accumulate with constant frequency, therefore the number of mutations between two sequences belonging to different species is proportional to the time elapsed since the two species had a common ancestor or, in other words, is a measure of the genetic distance between two species. In principle it is possible to draw the phylogeny of different species simply by counting the number of nucleotide differences in pair-wise comparisons of a given DNA sequence, applying various corrections to take into account multiple hits, back

[1] The term "neutrally" is applied here in its strictest sense and indicates the complete lack of selective constraint. Later in the text, "neutrality" will assume a looser meaning; in coding sequences selective constraint will usually mean that evolution proceeds more slowly, but nevertheless coding sequence differences between species may still be proportional to the time of divergence. In fact many proteins show this phenomenon which led originally to the development of the "molecular clock" hypothesis (Zuckerkandl and Pauling, 1965). In this contest, "neutrality" indicates constancy in the rate of evolution.

substitutions, number of transitions/transversions etc, and comparing the values for the different pairs. A different situation exists in DNA under selection; mutations arise, which can be either eliminated, maintained at low frequency or fixed. Because of selective pressure, changes generally accumulate at a much slower pace than in neutrally evolving DNA, and the frequency of accumulation of changes can vary between or within lineages depending on the extent of the selective constraint.

The two types of algorithms most commonly employed in the determination of a phylogeny with molecular data, are either distance methods or character-state methods. The former involve a two-step procedure: first a distance matrix is constructed in which the genetic distance is calculated for each sequence pair, in the second step the distance matrix is "translated" into the sequence tree. In the latter methods, the state of a particular character (such as the nucleotide at a particular site, or the presence or absence of a restriction site in a DNA sequence) is taken into account for all the sequences examined and the tree inferred is the one in which the smallest number of changes are necessary to explain the state of the characters. The phylogenetic tree that represents the evolutionary relationships of a group of species is called the "species tree", while the tree inferred by analysis of a gene is called the "gene tree". Species trees and gene trees do not necessarily coincide, particularly if the gene is under selection. To infer the real species phylogeny with molecular data, more than one gene is needed.

The phylogeny of the species mentioned here is well known by means of more traditional taxonomic approaches even though there is not a general agreement about the timing of the branching points. *Antheraea pernyi* belongs to the order *Lepidoptera* which was already well differentiated at the end of the Triassic era (Boudreaux, 1978), 200 million years ago. The group Calyptratae (to which *Musca domestica* belongs) diverged from the group Acalyptratae (which includes the drosophilidae) 100 Myr ago (Hennig, 1981), the time of divergence of *D. melanogaster* and *D. virilis* is estimated in about 40 Myr (Schlotterer *et al.*, 1994), the *obscura* group (to which *D. pseudoobscura* belongs) separated from the *melanogaster* group between 25 Myr (Russo *et al.*, 1994) and 30 Myr ago (Schlotterer *et al.*, 1994) and the phylogenetic distance between *D. melanogaster* and *D. yakuba* amounts to between 6 (Russo *et al.*, 1994) and 15 (Lachaise *et al.*

1988) Myr; a species tree based on these divergence times is shown in fig. 3.10. Although there are uncertainties about the exact time of divergence, there are no ambiguities in the branching order of these species that require a molecular solution.

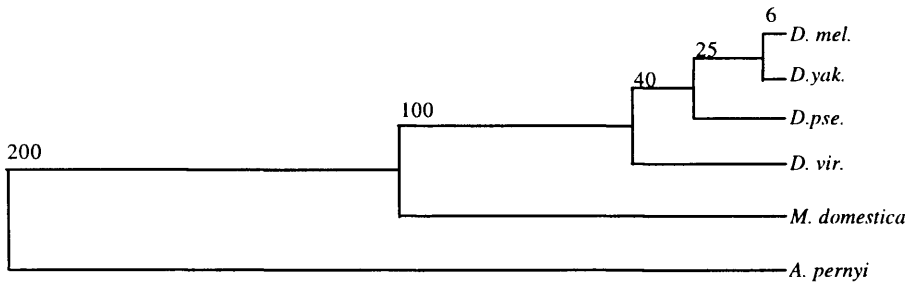


Figure 3.10. Species tree showing the time of divergence of the species analysed. The numbers on each branching represent the approximate time of divergence.

Therefore I have used the phylogenetic approach to examine whether there is any significant difference between the species tree and the Per protein tree which could be taken as an indicator of non-neutrality in the evolution of the Per sequences. For this reason the analysis was carried out at the amino acid level, as this is the level at which selection is expected to act. All the neutral changes (degenerate positions within codons) will simply be ignored in the evaluation of the protein distance.

The tree building principles are the same for proteins as for nucleic acid sequences: the distance values of each pair-wise sequence comparison (the protein distance is based on the number of amino acid replacements) are used to create a distance matrix which is subsequently interpreted to give the sequence tree. Various methods are available for calculating phylogenetic distance: Kimura (1983) assumes that any amino acid can change to any other with the same probability and only corrects the total number of replacements for multiple changes; more complicated methods may take into account that some amino acid replacements occur at higher frequency than others (PAM distance matrix, Dayhoff, 1979) or the chemical similarity of the replaced amino acid and the nucleotide changes necessary to have a particular amino acid replacement (Felsenstein, 1988). Dayhoff's and Felsenstein's

approach may seem equivalent but the first is based on empirical data while the second is based on a theoretical assumption; thus the two methods might produce different phylogenies. The distance matrix is then "translated" according to the UPGMA method (Sneath and Sokal, 1973) to draw the phylogenetic tree. This method makes the assumption that the rate of evolution is constant for all the lineages in the tree; by comparing the UPGMA tree with the known species tree we can test whether the sequences evolve neutrally.

Only regions c1, c2 and c3 are clearly alignable for all six species considered and consequently we limited our analysis to these blocks. Multiple alignments of the sequences were performed using the program ClustalW and the resulting aligments (fig. 3.11 for c1+c3 and fig. 3.13 for PAS) used for the phylogenetic analysis. Conserved regions 1 and 3 were joined together in order to have a longer, more informative sequence.

The tree inferred by analysis of regions c1+c3 with the PAM algorithm (fig. 3.12a), shows an altered phylogenetic relationship of the sequences under examination compared to the species tree (depicted in fig. 3.10); in particular *D. virilis* appears here more closely related to *D. melanogaster* than *D. pseudoobscura*. A similar phylogeny is obtained when the UPGMA method is applied to the Kimura's distance matrix (fig. 3.12b), while the distance calculated with Felsenstein's algorithm (fig 3.12c) gives a tree comparable to the species tree. This difference between expected and observed phylogeny could be either due to evolutionary constraint in particular lineages, or to the size of the sample data. Indeed, the bootstrap values are relatively low at all branches except for the one dividing *D. melanogaster* and *yakuba*, indicating that the significance of the trees inferred on the basis of c1-3 is rather low.

The second fragment of Per used in this analysis was the PAS region. This 250 amino acid region localised in c2, represents a functional domain of Per implicated in protein-protein interactions (Huang *et al.* 1993) and initially it was noticed that it is characterised by a smaller number of amino acid replacements in the *melanogaster/Musca* pairing (29 aa changes + 1 aa deletion) than between *melanogaster* and *pseudoobscura* (33 replacements). The crude number of amino acid changes is not a very precise parameter for describing the divergence of two

CLUSTAL W(1.5) multiple sequence alignment

```
A.pernyi     1 MNNMDG---SE--NNAKVSDSAYSNSCSNSQSRRSHSSKSTHSGSNSSGSSGYGGQPSTSSS----------------------SNDLSDQKKEKELKKKKQVE  77
D.mel.       1 ---MEGGESTESTHNTKVSDSAYSNSCSNSQSQRSGSSKSRLSGSHSSGSSGYGGKPSTQAS----------------------SSDMIIKRNKDKSRKKKKNK  79
D.yakuba     1 ---MEGGESAESTHNTKVSDSAYSNSCSNSQSQRSGSSKSRLSGSHSSGSSGYGGKPSTQAS----------------------SSDMIIKRNKDKSRKKKKNK  79
D.pseud.     1 ---MEG-ESTESTQNTKVSDSAYSNSCSNSQSQRSGSSKSMLSGSHSSGSSGYGGKPSIQTS----------------------SSDMAIKRNKEKSRKKKKAK  78
D.virilis    1 ---MEG-ESTESTHNTKVSDSAYSNSCSNSQSQRSGSSKSRLSGSHSSGSSGYGGKPSTQAS----------------------SSDMAVKRNKDKSRKKKKAK  78
M.dom.       1 ---MEG-ESTESTHNTKVSDSAYSNSCSNSQSQRSGSSKSRLSGSHSSGSSGYGGKPSTQTSRLKLKSPFPYYSETNCNFFSINTQSEIHITKRGKDKGRKKKKQK 102
                   *.*   .*   *.****************.** ****  ***.*********.**  .*                      *         .. .***.


A.pernyi   617 PIR--LTESSLTKHNAEMEKELMKIHREHRC--YS--KGDR---------VKVSNE----ARQKKK------EHLARCNAGFQTI 676
D.mel.     767 PV--TLTESLLNKHNDEMEKFMLKKHRESRG--RTGEKSKK---SANDTLKMLEYS--GPGHGIKRGGSHSWEGEANKPKQQLTL 842
D.yakuba   756 PV--TLTEALLNKHNDEMEKFMLKKHRESRG--RSGEKSKK---SATDTLKMLEYS--GPGHGIKRGGSHSWEGEANKPKQQLTL 831
D.pseud.   867 PV--SLTESLLNKHNDEMEKFMLKKHRESRGDRRTVEKNKNKTTNTIDSLKILEYSSTGPGHGTKRGGSYSWEGEGNKPKQQPTL 949
D.virilis  692 PVNMTLTEILLNKHNDEMEKCMLKKHRESRG--RTGDKTKK----SVI-EKMPEYSGPGHGQTMKRGGSHSWEGDANKPKQQLTL 769
M.dom.     667 LV--TLTESLLKKHNDEMEKFMLKKHRESRG-RGCGEKNKK----SSD--KTMEYS--GPGHGIKRVGCHSWEGEANRPKQQHTN 740
                 .   ***  * *** **** ..* *** *    .  *                      .  *.     *          *
```

Fig. 3.11. Alignment of the protein sequence of conseved region 1 and 3 used to construct the phylogeny shown in fig. 3.10. Identical residues are indicated by asterisks, similar residues by dots. Dashes represent deletions. The numbers refer to the position of the sequences within the protein. c1 (upper block of sequences) was joined together to c2 (lower block) in order to have a more informative sequence.

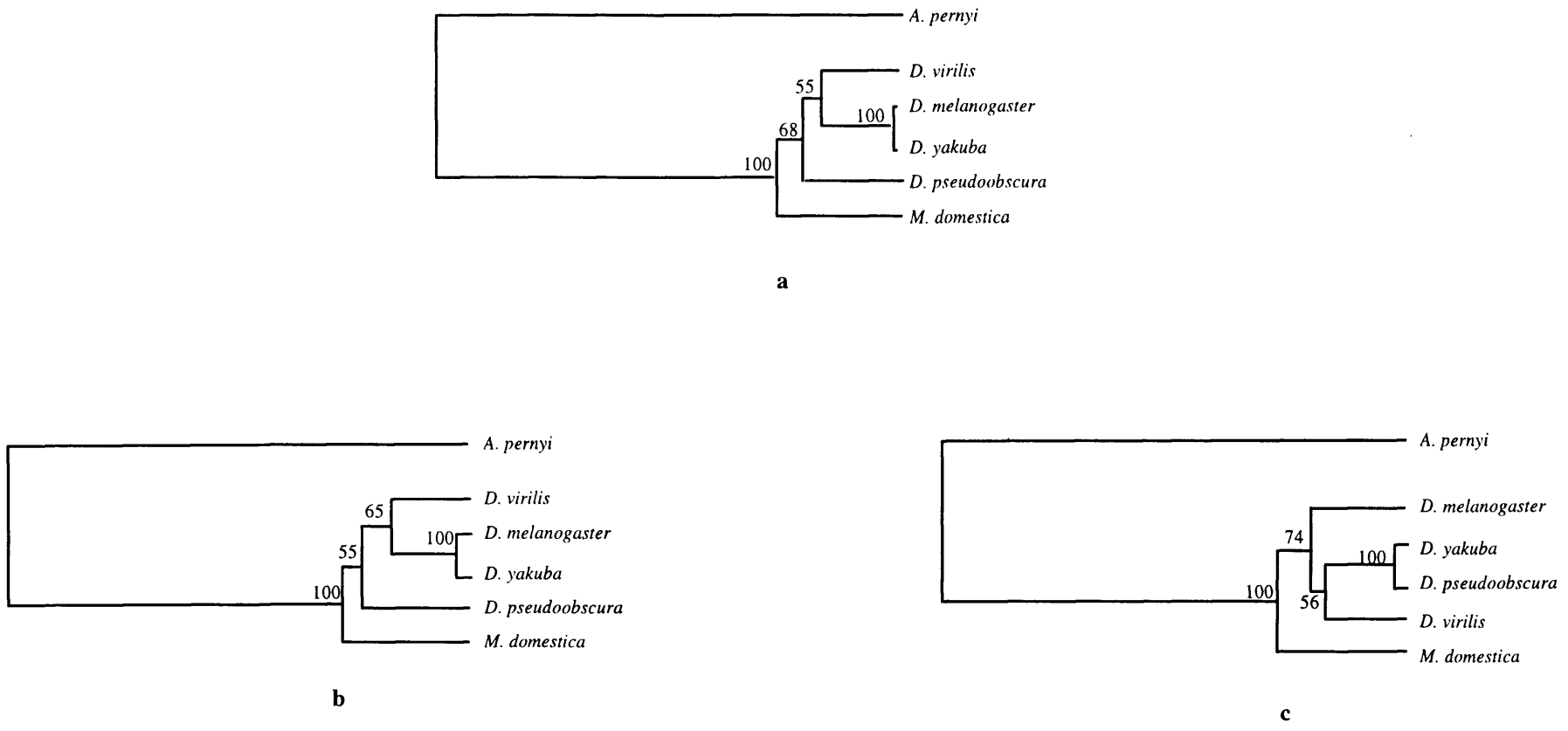


Fig. 3.12. a: Phylogeny tree based on the amino acid sequence of c1+c3 region. The tree was obtained using the UPGMA method (Sneath *et al.*, 1973) and the PAM distance matrix (Dayhoff, 1979). b and c: protein trees of c1+c3. tree b was derived by applying the UPGMA method and Kimura's distance (1983) and tree c with the UPGMA method and Felsestein's distance (1988). The values reported on each node represent the statistical significance of the branching, assessed with the bootstrap procedure.

sequences. It is better to consider the overall chemical similarity of the residues involved in these replacements. The similarity was computed by the program "bestfit" of the GCG package which performs pair-wise alignments and assigns similarity scores on the basis of the chemical similarity of the amino acids occupying the aligned position (mismatches are scored according to the distance between the amino acids as measured by Dayhoff (1979) and normalised by Gribskov *et al.*, (1986)).

The similarity score given by the program is normally not very informative because it depends only on the nature of the sequence in the aligned fragment and not on the proportion of alignable sequence. However in the case of the PAS region this does not influence the score because the total length of the sequences is perfectly alignable for every pairing (with the exception of the pairings between *A. pernyi* and the other species; *Antheraea* is anyway the outgroup of the phylogeny and even if the similarity scores are increased by non-perfect alignment, it is scored by far as the most dissimilar sequence). The similarity scores are reported in table 3.3; *melanogaster* PAS is slightly more similar to *Musca*'s than the PAS region of *pseudoobscura*. This is a rather interesting observation given that *D. melanogaster* and *pseudoobscura* diverged 30 Myr ago (Schlotterer *et al.*, 1994), *D. melanogaster* and *M. domestica* 100 Myr ago (Hennig, 1981) and *D. melanogaster* and *Antherea pernyi* at least 200 Myr ago (Boudreaux, 1978) and suggests that this region is not evolving neutrally. On purely neutral criteria we might have expected a *M. domestica/D. melanogaster* similarity between 65 and 90%. These observations and their possible functional and evolutionary implications will be further discussed in chapter 5.

	Anth.	*D. mel.*	*D. yak.*	*D. pse.*	*D. vir.*	*Musca*
Anth.		67.9	67.5	67.5	65.8	68.3
D. mel.			99.2	93.0	91.1	93.8
D. yak.				92.6	90.7	93.0
D. pse.					87.3	91.5
D. vir.						86.9
Musca						

Table 3.3. Percentage of amino acid similarity between pair-wise comparisons of the PAS regions of various species. The similarities were calculated with the program "bestfit".

The molecular tree inferred on the basis of the PAM distance matrix places *Musca* closer to *D. melanogaster* and *yakuba* compared to *D. pseudoobscura* and *virilis* (fig 3.14a). A similar tree was also obtained when a distance matrix is calculated according to Kimura (fig. 3.14b). When Felsenstein's method to calculate the distance matrix is applied, the derived tree shows that *D. pseudoobscura* is a closer relative to *melanogaster* than *Musca* (fig. 3.14c) but the latter species still appears closer to *D. melanogaster* than *virilis*. The bootstrap values for some of the branching points in the phylogeny of the PAS peptide are below the 95% probability level, however this limit is probably too severe for trees based on short sequences (Pelandakis *et al.*, 1991). As a control, the phylogeny of PAS was also calculated using the DNA sequence; the distance was calculated according to Kimura's (1980) two-parameters model and the tree with the UPGMA method (fig. 3.14d). As expected the branching points in the DNA tree agree with the phylogeny of the species.

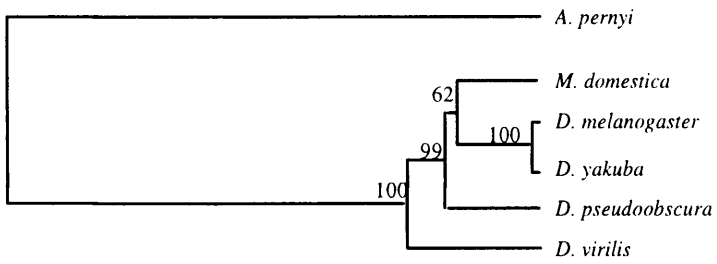
The Thr-Gly repeat

The Thr-Gly region of *M. domestica* is quite different from the homologous region of *Drosophila*. It contains one single copy of a hexapeptide unit and two Thr-Gly pairs at pos. 656-659 (see fig. 3.9). For this reason, Nielsen *et al.* (1994) investigated whether the Thr-Gly region had undergone extensive changes in length in other dipterans or whether such a dramatic expansion of the repeat is only

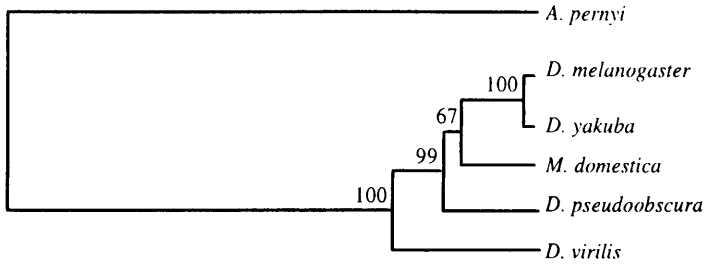
CLUSTAL W(1.5) multiple sequence alignment

```
A.pernyi      168 --FSCVISMHDGVVLYATASLTSTLGFPKDMWVGRSFIDFVHPRDRNTFASQITNELAIPKIVSLTEETDQTMENPGSTMVCRIRRYRGLSCG-  258
D.mel.        238 DSFCCVISMHDGIVLYTTPSITDVLGYPRDMWLGRSFIDFVHLKDRATFASQITTGIPIAESR------GSVPKDAKSTFCVMLRRYRGLKSGG  325
D.yakuba      242 ESFCCVISMHDGIVLYTTPSITDVLGYPRDMWLGRSFIDFVHLKDRATFASQITTGIPIAESR------GSVPKDTKSTFCVMLRRYRGLKSGG  329
D.pseud.      200 DSFCCVISMHDGIVLYTTPSITDVLGFPRDMWLGRSFIDFVHTKDRATFASQITTGIPIAESR------CSMPKDARSTFCVMLRQYRGLQTSG  287
D.virilis     220 DSFCCVISMHDGVVLFTTANLNEMLGYPREMWLGRSFIDFVHIKDRATFASQITTGIPIAESR------CSQSKDARTTFCVMLRRYRGLASGG  307
M.dom.        203 DSFCCVISMHDGIVLYTTPSITDVLGFPRDMWLGRSFIDFVHPKDRATFASQITTGIPIAESR------SSIPKDARSSCCVMLRRYRGLKSGG  290
                    *.********.**..*  ..  **.*..**.********* .** *******. . *                ..    .*.**** .

A.pernyi      259 FSVKNTTTAYLPFLLKFKFKNVNEDK-------G--NVIYLVIQAVPFFSAFKTSNEVLA-KTVSFVIRHSADGNLEYIDAESVPYLGYLPQDI  342
D.mel.        326 FGVIGRPVSYEPFRLGLTFREAPEEARPDNYMVSNGTNMLLVICATPIKSSYKVPDEILSQKSPKFAIRHTATGIISHVDSAAVSALGYLPQDL  419
D.yakuba      330 FGVIGRPVSYEPFRLGLTFREAPEEARPDNYMVSNGTNMLLVICATPIKSSYKIPDEILSQKSPKFAIRHTATGIISHVDSAAVSALGYLPQDL  423
D.pseud.      288 YGVIGRSVNYEPFRLGMTFREAPEEERSDNYMVANSSNMLLVICATPIKSSYRVPEEIHSQRSPKFAIRHTAAGIISHVDSAAVSALGYLPQDL  381
D.virilis     308 FGIIGRPVSYEPFRLGLTFREAPEEVQSDGCTLSNATSMLLVISATPIKSCYKEPDEFLSPKGPKFAIQHTAAGIISHVDTAAVSALGYLPQDL  401
M.dom.        291 YGVIGRSVNYEPFRLGLTFREAPEEARSDNSLPS-GTNMLLVICATPIKSAYKVCDELLSRKTPKFAIRHTKTGIISTVDSGAVSALGYLPQDL  383
                  . . .    * ** *   *.   *.             . *** * *  *...   *  . .   * *.*.  * .  .*. .*  *******.

A. pernyi     343 TNRDALLLYHPGDLGYLQEIYGSLVKEG----NVTRSKTYRMMTQNGHYMKVETEWSAFINPWSKKLEFVTGKHYII  416
D.mel.        420 IGRSIMDFYHHEDLSVMKETYETVMKKGQTAGASFCSKPYRFLIQNGCYVLLETEWTSFVNPWSRKLEFVVGHHRVF  496
D.yakuba      424 IGRSIMDFYHQEDLSVMKETYEMVMKKGQTAGASFCSKPYRFLIQNGCYVLLETEWTSFVNPWSRKLEFVVGHHRVF  500
D.pseud.      382 MGRSIMDLYHHDDLPVIKEIYESVMKKGQTAGASFCSKPYRFLIQNGCYILLETEWSSFVNPWSRKLEFVVGHHRVF  459
D.virilis     402 IGRSILDFYHHEDLSDIKDIYEKVVKKGQTAGATFCSKPFRFLIQNGCYILLETEWTSFVNPWSRKLEFVVGHHRVF  479
M.dom.        384 IGRSILDFYHHEDLTVLKEIYETVMKKGQTAGASFCSKPYRFLVQNGCYVLLDTEWTSFVNPWSRKLEFVIGHHRVF  461
                   .*  .  **  **  ... *  ..* *        ** .* . *** *. ..***..*.****.***** *.* .
```

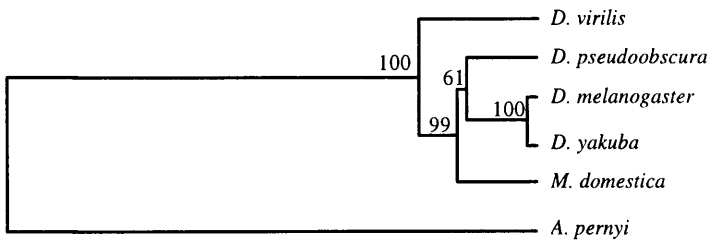
Fig. 3.13. Alignment of the protein sequence of the PAS region used to construct the phylogeny shown in fig. 3.13. The two degenerate repeats are printed in green. Identical residues are indicated by asterisks, similar residues by dots. Dashes represent deletions. The position of the sequences within the protein is indicated.



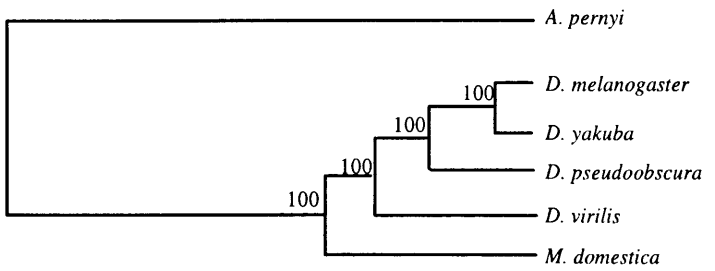
a



b



c



d

Fig. 3.14. a, b and c Phylogeny based on the amino acid sequence of the PAS region. Tree a was obtained applying the UPGMA method (Sneath *et al.*, 1973) to the PAM distance matrix (Dayhoff, 1979), b was derived by using the UPGMA method and Kimura's distance (Kimura, 1983) and c with the UPGMA method and Felsestein's distance (Felsenstein, 1988).
d: phylogeny of the DNA sequence of the PAS region; the tree was constructed with the UPGMA method (Sneath *et al.*, 1973) and Kimura's two parameters distance (1980). The numbers and represent the confidence limits referred to the branching on their right. The length of the branches is proportional to the value of distance calculated by the programs.

confined to the *Drosophila* genus. Nielsen *et al.* (1994) used a PCR approach based on the *Musca* sequence to amplify the Thr-Gly repeat plus flanking sequences both in the 5' and 3' direction, corresponding to amino acid 567 to 676 of the *Musca* protein. This region was divided into 5 areas, conserved area 1, 2 and 3 and variable area 1 and 2 (see fig. 2 and 3 in the appendix), not to be confused with the conserved and non-conserved regions in which the *per* gene is divided by Colot *et al.* (1988). The evolution of this region has been described in considerable detail in Nielsen *et al.* (1994). The interested reader can find a copy of the paper in the appendix.

Two Thr-Gly pairs are encoded at the 3' boundary of cons2 in most of the species examined, and since the first amino acid of var2 is always a threonine (with the exception of *D. melanogaster*), the Thr-Gly region contains at least two Thr-Gly couples in all flies observed. Region var2 displays the greatest length variability, ranging from nine amino acids in all the non drosophilids, to the 209 amino acids of the *pseudoobscura* protein (see fig. 5 in appendix). A distinction can thus be made between the *Drosophila* genus and all the other dipterans, the former accounting for all the length variability observed, the latter presenting a much shorter and more stable var2. In the outgroup *Beris vallata* (soldier fly) var2 begins with the hexapeptide TSSGSF, which is found in slightly mutated form in all the other non drosophilids. From this sequence a series of point mutations, deletions and duplications can give rise to the two basic pentapeptides, TSXXX or NSXXX, the latter being derived from the former, which are found in various disguise in the Drosophilids.

DISCUSSION

Conservation and divergence

Evidence for the successful cloning of the *Musca domestica per* homologue has been presented in this chapter. Stretches of similarity are found in all the conserved regions first identified by Colot *et al.* (1988) in Drosophilids *per* genes.

The structure of the *Musca* gene is different from its *Drosophila* orthologue; even though the full length of the primary transcript is not known, it must measure at least 3 kb more than that of *D. melanogaster* given the increase in the size of intron 2. This increase in size probably occurred relatively recently in the *Musca* lineage; traces of a duplication are evident at pos 358-554 of the sequence (fig. 3.8). We can roughly estimate the time of occurrence of this duplication: adopting a mutation rate of about 10^{-9} per generation per site, as in *D. melanogaster* (Kreitman, 1991), and assuming 5 generations per year, it requires a total of 15 Myr to generate 13 point mutations in a 2x84 bp fragment. Not only the size, but also the number of introns is increased in the *Musca* gene. In agreement with this observation, the genome of *Musca domestica* is about 5 times bigger than that of *D. melanogaster* (840 Mb and 165 Mb, respectively; John and Miklos, 1988) and it would be tempting to say that the increase in genomic complexity must correlate with an increase in both number and size of introns. Unfortunately, most of the available *Musca* gene sequences come from cDNA libraries, so there is not sufficient genomic data to test this hypothesis. In *D. melanogster*, *per* is sex-linked, being located at the 3B1-2 region of the X chromosome (Konopka, 1972; Young *et al.*, 1978) between *yellow* (*y*) and *white* (*w*). Also in *D. pseudoobscura per* has been mapped to the X chromosome (Petersen *et al.*, 1988) while the chromosome location of *per* in *D. virilis* is simply not known. The genomic mapping of the *Musca per* gene was unsuccessfully attempted by *in situ* hybridisation on polytene chromosomes, but some considerations suggest that it could be autosomal. The X chromosome of *M. domestica* is entirely heterochromatic[2] (Milani, 1967) and should therefore contain very few, if any, genes. Indications that *per* might be located on chromosome III of *Musca* come from Milani (1967) and Malacrida *et al.* (1985), who described the correspondence between various linkage groups of *D. melanogaster* and *M. domestica.* In particular, the *Musca* genes corresponding to *Drosophila*'s *y* and *w* (respectively *brown body* (*bwb*) and *w*) are found on the right arm of chromosome III.

The A+T content is high in intron sequences from *Musca*, *D. pseudoobscura* and *D. virilis* and lower in *D. melanogaster* and *D. yakuba*. The A+T content is also

[2] In most european strains the Y chromosome is the sole sex-determinant (Milani *et al.*, 1967) but several strains have been collected in which the sex-determining factors are autosomal (Franco *et al.*, 1982).

relatively high in the coding sequence of the *Musca* gene, in contrast to *Drosophila*. This reflects a bias in *Drosophila per* for the usage of codons which either terminate in C or G. Indeed it has been reported (Sharp *et al.*, 1992) that some *melanogaster* genes have a high proportion of C- and G- terminating codons, while others have a more uniform codon usage. This normally correlates with the level of expression of the genes, so that C- and G- ending codons are more frequent in highly expressed genes, and for this reason Sharp *et al.* (1992) refer to them as "optimal codons". The few *Musca* genes analysed cannot provide much light for elucidating the codon usage in this organism but it seems that a situation similar to *Drosophila* occurs here, with some genes adopting a random distribution of codons and others displaying a bias in codon usage towards a higher frequency of optimal codons which in *Musca* appear to be A- and T- ending.

At the protein level, very high homology is found in both c1 and c2, underscoring the importance of PAS in the biochemical function of Per, and suggesting that an equally important role may be played by c1. Two sites are found within the PAS domain which, when mutated, decrease the dimerization efficency; the per^L site and a cluster of amino acids at position 413-419 (in the second PAS repeat) of the *D. melanogaster* protein (Huang *et al.*, 1993). The effect on protein-protein interactions of this amino acid cluster was assayed because the residues contained in the fragment are highly conserved in the PAS regions of AHR, ARNT, SIM and Per. Both of these areas of PAS are highly conserved in *Musca*. Another area in c2 has been identified as a "short period domain" in which mutations consistently shorten the circadian period (Baylies *et al.*, 1992; Rutila *et al.*, 1992). It extends from 3 amino acids upstream to 16 downstream of the per^S site. The high degree of homology suggests that in *Musca* this area retains its functional importance.

In contrast to the *Drosophila* proteins, which contain multiple PEST regions (with the exception of *D. yakuba*), *Musca* Per only contains one. Moreover, the PEST-score of this region is considerably lower than its *Drosophila* counterpart (again excluding *D. yakuba*). How this could affect the rate of Per degradation in *Musca* is not clear. While, with one exception, all eukaryotic rapidly degraded proteins of known sequence contain PEST regions (Rechsteiner, 1988), various

PEST-containing proteins do not have a short half life (Rogers *et al.*, 1986; Rechsteiner, 1988). Moreover, the correlation between half-life of the protein and PEST-score is weak (Rogers *et al.*, 1986, Rechsteiner, 1988). Therefore some of the PEST regions in the *Drosophila* proteins could be irrelevant for degradation, and the fact that the *D. yakuba per* gene rescues the circadian phenotype in *D. melanogaster* (Couchman, personal communication), suggests that only the N-terminal PEST regions (among the ones reported in table 3.2) might be important.

Phosphorylation has been shown to play an important part in Per function (Edery *et al.*, 1994) and it may also contribute to Per degradation. In the yeast fructose-1,6-biphosphatase for example, phosphorylation transforms a weak PEST region into a strong proteolytic signal (Rechsteiner, 1988). Similarly, the degradation of Per might occur mainly at the level of the phosphorylated form, whose appearance triggers the negative feedback on *per* transcription. In this case conditional PEST region(s), different from those listed in table 3.2, could be activated by phosphorylation. It must be noted in this regard that many consensus phosphorylation sites are found in the different Per proteins, and a good fraction of them is conserved at least among the dipteran proteins (fig. 3.9b). This does not promote them as being functionally active, since they should statistically occur in a random amino acid sequence once every 100 residues, but, interestingly, one of these sites falls in the PEST sequence shared by all dipteran PERs. Future identification of the relevant phosphorylation sites will significantly contribute to understanding the proteolytic pathway of Per which is a critical factor for the negative feedback model.

The *opa* repeat found in c5 of *Musca* does not have a homologous region in *Drosophila*. Instead, one *opa* is found in *virilis* in a different position of the gene (nc1) and in the other *Drosophila* species nc1 is relatively Q-rich. No particularly high frequency of Glu is recorded here in *Musca*. Both poly-Q stretches and Q rich regions have been recognised as capable of activating transcription (Courey *et al.*, 1988) but it seems unlikely that this is the role of the Q rich regions in the Per proteins, given the lack of conservation of position. It is interesting to note that clusters of glutamine are often found in proteins associated with the development or the normal functioning of the nervous system and that several human neurological

disorders are associated with proteins containing abnormal expansions of this amino acid (Karlin *et al.*, 1996).

Phylogeny of Per

With the analysis of the molecular phylogeny of Per I tested whether different regions of the protein were evolving neutrally. Sequence comparisons show a conservation of the N terminal half of the protein, indicating that basic Per functions may be performed by sequences in this area. Within this half is located the PAS domain, which by homology with other proteins and by *in vitro* studies is believed to promote interactions with the Per partner Timeless (Huang *et al.*, 1993; Zeng *et al.*, 1996; Myers *et al.*, 1996; Lee *et al.*, 1996). This means that PAS cannot evolve independently, but it may do so together with the dimerization domain of Tim. Most of the amino acid substitutions originating in the PAS region may be deleterious for the protein-protein interaction unless they are "counterbalanced" by complementary changes in the partner's dimerization domain. Thus we might reasonably expect a coevolution between the Per and Tim dimerization domains. This could be reflected in heterospecific transformation experiments where a more similar PAS domain to the host (*D. melanogaster*) would generate enhanced rescue of arrhythmicity. Corresponding experiments could also test this coevolution at a biochemical level by examining heterospecific PAS-Tim binding affinity in the yeast two hybrid system.

In the phylogeny of the PAS region presented in this chapter, *Musca* clusters more closely to *D. melanogaster* and *D. yakuba* than the phylogenetically closer *D. pseubdoobscura* and *D. virilis*. This is supported by the values of amino acid similarity for the PAS region (table 3.3). The method of choice for calculating the distance matrix was Dayhoff's PAM method. By comparing the changes in two different sequences, it assumes that some amino acid replacements are more likely to occur than others. The likelihood of any possible change is given by the frequency with which that change is observed in a database of some thousands of proteins from different kingdoms. The method of Kimura has a different starting assumption, namely that the protein undergoes neutral evolution, so every amino acid

replacement is allowed with the same probability. Even with this assumption the tree inferred with the Kimura matrix is equivalent to the one derived with the PAM matrix, strengthening the validity of the phylogeny obtained. Probably, most of the amino acid positions in the PAS region (the ones involved in the protein-protein interaction) are under constraint, but for the few ones that can mutate, and indeed do so, there could be virtual neutrality (intended here as absence of any constraint). Supporting this view is the observation that most of the substitutions are found in the region between the two degenerate repeats, including the single amino acid deletion in *Musca* and various replacements involving charged residues.

The other technique adopted for the compilation of a distance matrix, Felsenstein's method, seems inappropriate in our case; by rating the observed changes on the basis of the genetic code it places more emphasis on the time factor required to generate the changes, rather than on the constraints imposed on the chemical similarity of the residues. This is therefore a more powerful method for inferring "real" phylogeny but is less appropriate for detecting selective constraints on a protein. It is not surprising that the tree resulting from the application of this method does not coincide with the others since, compared to *D. melanogaster* PAS, it postulates a higher number of nucleotide changes to derive the more similar *Musca* PAS, than for the less similar (at the amino acid level, see table 3.2) *D. pseudoobscura* PAS. However even this correction for the genetic code is not enough to compensate for *D. virilis* PAS which branches earlier than *Musca*. When a PAS phylogeny was constructed using the DNA data instead of the protein sequence, the resulting tree is equivalent to the species tree; in this case all the nucleotide changes are analysed, whether synonymous or not, and the sequence is inevitably much more informative about the relative time of divergence of the different species. The DNA tree demonstrates that there is less constraint on this region at the DNA level, and selection occurs at the level of the protein. The protein tree of the PAS region inferred using the "maximum parsimony" method (a character-state method) of the PHYLIP package is equivalent to the species tree (data not shown). It must be noted though, that the informative characters considered for this region are very limited; in the case of the four species whose phylogeny is here under question (*D.*

melanogaster, *pseudoobscura*, *virilis* and *Musca*) only eight amino acids are informative so the statistical significance of this tree is very low.

The phylogeny inferred using c1+c3 is also inconsistent with the species tree. The difference between expected and observed phylogeny could be due to evolutionary constraint in particular lineages, but not much can be said in this regard since the biochemical role of these regions is unknown. Alternatively it could be due to the size of the sample data as the bootstrap values are quite low at all branches except for the one dividing *D. melanogaster* and *yakuba*. When analysing molecular phylogenies, it has become a current practice to calculate the bootstrap values for the trees inferred in order to estimate its statistical significance. Bootstrapping (Felsenstein, 1985) consists in making new data sets by resampling the existing one introducing random deletions and duplications. This method though, lacks rigorous justifications. An important critique (Brown, 1994), is that the method relies on the assumption that the resampled units are independent, but this is clearly not true in the case of coding sequences. Therefore the bootstrap values reported here must be taken as purely indicative.

Evolution of the Thr-Gly repeat

The Thr-Gly repeat of Drosophilids has undergone extensive evolution in both amino acid sequence and length (Colot *et al.*, 1988; Costa *et al.*, 1991; Peixoto *et al.*, 1992, 1993). The sequences obtained from outside the *Drosophila* genus show that at least two Thr-Gly pairs are encoded at the interface between areas cons1 and var2 in all species examined. In *Musca*, and indeed in all other dipterans outside the *Drosophila* genus, no further expansions of simple motifs occurred, and the structure of the Thr-Gly region is homogenous in all species examined. The Thr-Gly region of these flies always contains two threonine-glycine pairs followed by one copy of a hexapeptide motif. It seems therefore, that long stretches of Thr-Gly (or indeed of other amino acids) are not required for the basic functions of the *per* gene.

Theoretical models of repetitive DNA evolution (Gray and Jeffreys, 1991; Harding *et al.*, 1992) initially require two tandem repeats in order to allow for mispairing and subsequent generation of high copy number repeats by unequal exchanges. After a certain period of expansion followed by a consolidation time, in which large numbers of repeats are maintained, there is a contraction to a small copy number and eventually a collapse to a single, stable copy. For example, in great apes, Old and New World monkeys, the MS32 non-coding hypervariable minisatellite appears stable and monomorphic, having only three to four repeat units. In humans however, MS32 shows extreme variation in copy number (Gray *et al.*, 1991). Computer simulations of minisatellite evolution using estimates of mutation rate and assuming a direct relationship between allele length, revealed that only one in 250 lineages would attain a phase of large numbers (>200) of repeat units (Gray *et al.*, 1991). Similar results were also obtained by Harding *et al.* (1992) using a different mutational model. Most lineages maintained only one or two copies of the repeat, which eventually collapsed to a single copy, which can no longer expand. However these models were developed for neutrally evolving, non-coding DNA. For coding DNA we would expect that some constraints must be involved to limit expansion. These models suggest that perhaps the Drosophilidae represent a rare lineage in which the expansion of the ancestral Thr-Gly repeat unit or units has occurred, even though the general rule is that this region of *per* will be short and non variable in most lineages.

There are indications that the Thr-Gly repeat is a flexible β-turn spacer which separates two globular domains of Per (Ishida *et al.*, 1994b; Castiglione-Morelli *et al.*, 1995). In this case it is conceivable that, in order to maintain the proper spatial relationship between the two globular domains, the repeat region and its immediate surroundings (even though the latter are not structurally part of the β-turn) must be regarded as a functional and therefore evolutionary unit. Changes at the level of the repeat region will mostly involve the length of the domain, and will probably require compensatory changes within the flanking sequences of the Thr-Gly repeat. This hypothesis has been advanced by Peixoto *et al.* (1993) and Nielsen *et al.*, (1994) after sequence analysis from different *Drosophila* species, and is supported by functional data from interspecific transformations (Peixoto *et al.*, 1998).

It did not escape attention that the *Musca per* gene could be used to test this hypothesis of intramolecular coevolution at the Thr-Gly domain level by constructing various *D. melanogaster/Musca* chimaeric *per* genes. In addition, transforming the intact *Musca per* coding sequence into *D. melanogaster* hosts offered the possibility of investigating the PAS-Tim coevolution at the functional level. Finally, the *Musca per* gene could also be used in *D. melanogaster* transformants to test the suggestion of Petersen *et al.*, (1988) that *per* contains species-specific circadian behavioural instructions. Details of the chimaeric genes construction are given in chapter 2, and the results obtained with the locomotor analysis of the transformant chimaeric flies are reported in chapter 5. However before this, chapter 4 briefly analyses the characteristics of the circadian locomotor activity of *M. domestica*.

CHAPTER 4

Locomotor activity rhythms in the housefly, *Musca domestica*

Locomotor activity rhythms in the housefly *Musca domestica*

INTRODUCTION

A common behavioural character analysed in the study of circadian rhythms, is locomotor activity. This parameter has been studied in a great variety of organisms, one of the earliest examples being the house cricket, *Acheta domestica* (Lutz, 1932). Locomotor activity under light-dark cycles (LD) or under constant darkness (free run, DD) are the two most frequently used assays to study the circadian clock. While the latter reveals the period of the endogenous oscillator which controls the activity, the former can also display species-specific details of the pattern of activity. The locomotor activity rhythms of *Musca* have been documented (Parker, 1962; Tsutsumi, 1973) and the period displayed by this organism appears to be fairly precise within a given population. However, a variant strain displaying longer than normal rhytmicity in locomotor activity experiment has been described by Aizawa and Yoshino (1986).

The analysis of circadian locomotor activity rhythms of the housefly was repeated here using the *Drosophila* locomotor activity monitor which had been amended for use with *Musca* (see chapter 2). Activity rhythms were measured on the Cooper strain of *Musca*, the same strain used for the construction of the genomic library from which the *per* clone was isolated. The onset and the offset of light does not only synchronise the circadian clock, but actively triggers a set of locomotor responses, the so-called "masking effect". For example in *Drosophila erecta*, a species belonging to the *melanogaster* subgroup, bursts of activity are recorded in light-dark cycles immediately after lights on, but no bursts were recorded at the equivalent circadian time in free running experiments (Thackeray, 1989). A similar effect has been described for the flight activity of the mosquito *Anopheles gambiae*, in which the burst of activity disappeared when the sudden dark to light transition was replaced by the gradual increase in light intensity (Jones *et al.*, 1972). For this reason the analysis of the pattern of activity was

conducted preferentially in free running conditions even though a less extensive analysis in LD conditions at 25° C has been performed. Activity monitoring in DD would result in patterns which reflect the activity profile as driven by the endogenous clock.

The results are reported in this chapter.

METHODS

Activity monitoring

Activity monitoring was performed as for *D. melanogaster* (see chapter 2) with the exception that flies were placed in glass cylinders with dimensions 200 x 12 mm. Analysis of the locomotor activity data was performed as described in chapter 2. Statistical analysis was performed with the STATISTICA package (Statsoft, Tulsa, Oklahoma).

RESULTS

Period of locomotor activity

The activity of *M. domestica* males was monitored in free running conditions at 18°, 25° and 29° C and the data collected was analysed with both the autocorrelation and spectral procedure. The rhythms of *Musca* flies are very robust and precise (see table 4.1), resulting in an extremely low percentage of arrhythmic flies, and a compact distribution of period values.

ANOVA analysis of the data collected does not show any significant difference of period between the three recording temperatures ($p = .109$), so that the temperature compensation of the clock appears to be very effective between 18° and 29° C. In the ANOVA the period recorded for one fly at 18° was omitted (see below). The average

period and SEM of the new set of data is 24.19 ± 0.14 h, according to the values calculated with the spectral analysis.

temp. (°C)	n° flies tested	n° flies arrhythmic	mean period ±SEM (spectral)	mean period ±SEM (autocorr.)
18	11	1	24.19 ± 0.13	24.16 ± 0.32
25	20	0	23.86 ± 0.09	23.89 ± 0.05
29	10	0	23.89 ± 0.22	23.52 ± 0.13

Table 4.1. Results of the locomotor activity monitoring of *Musca domestica*.

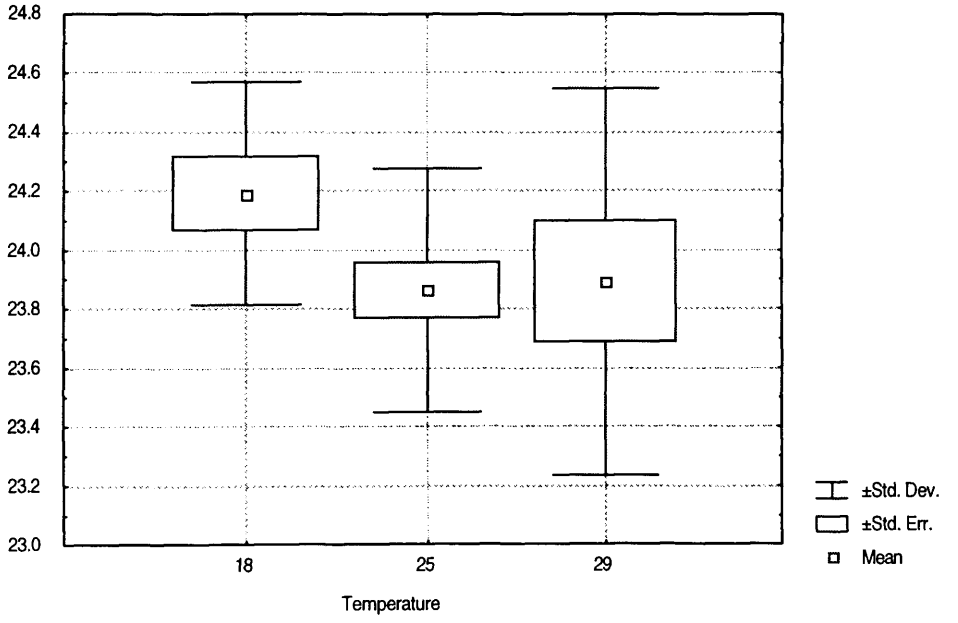


Figure 4.1. Graphical rapresentation of the results reported in table 4.1. The average period of *M. domestica* at different temperatures is depicted together with the standard deviation and the standard error of the mean.

One value was omitted from the 18° C data set because the fly, although periodic, displayed two different rhythmicities (see the actogram in fig. 4.2), one with a period of 27.56 h and the second at 23.78 h, as observed in the spectral analysis. The occurrence

of double rhythms in the housefly, although uncommon in physiological conditions, has been described in several reports (Han *et al.*, 1987; Smietanko and Engelmann, 1989a; 1989b), and has been interpreted as the uncoupling of two different oscillators, both of which concur to give the 24 h rhythms in normal condition (Smietanko and Engelmann, 1989b). Several factors have been shown to increase the percentage of flies displaying *split* or *compound* rhythms, as they have been referred to by various authors, such as continuous light conditions, Li^+ ions, and the triterpenoid azadirachtin (Smietanko and Engelmann, 1989b). It must be noted though, that in most cases the split rhythm appears as a bimodality of the locomotor activity since the two hypothetical oscillators maintain the same period and assume an altered phase, while in certain cases, the clocks can assume different periods, resulting in the simultaneous appearance of two different rhythms of activity or in the loss of rhythmicity (Smietanko and Engelmann, 1989a). The splitting of rhythmicity can be visualised in figure 4.2a.

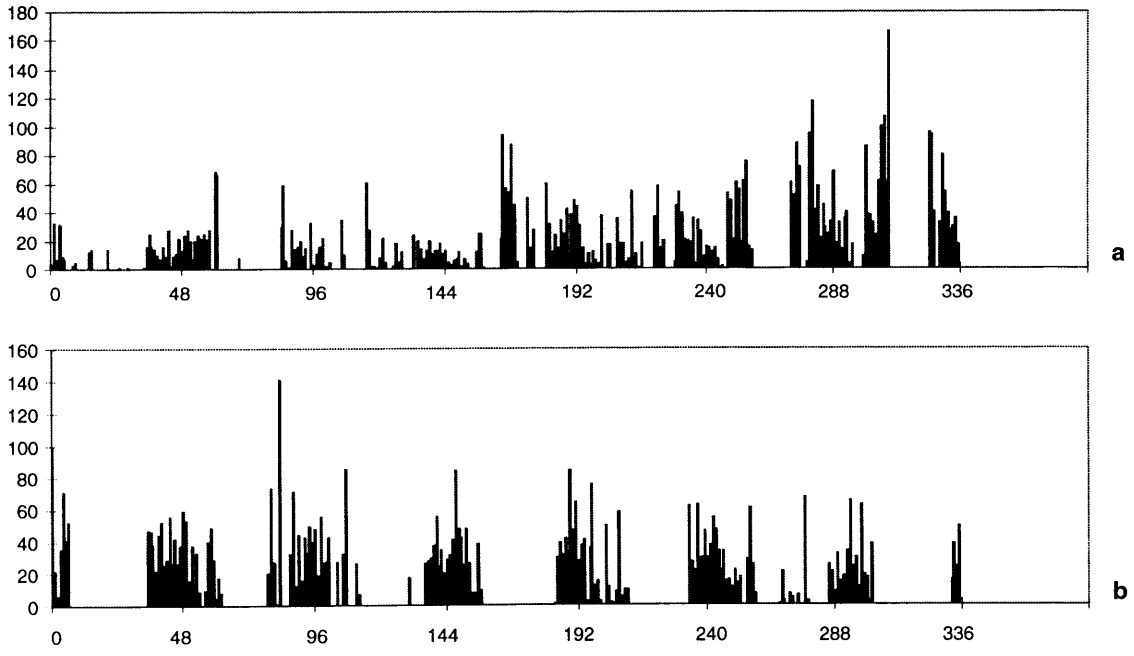


Figure 4.2. Histogram of the activity of two flies, recorded at 18°. Histogram **a** represents the activity of the fly with double rhythmicity, while histogram **b** shows a more typical locomotor activity. In the abscissa is reported the bin number (1 bin = 1/2 hour) and in the ordinate the number of counts for each bin. Bin 0 correspond to Circadian Time 6:00.

In the histograms displayed in figure 4.2 we can observe the profile of the activity of individual flies throughout the seven day experiment. Variability in the activity profile occurs not only between individual flies at any given temperature, but even between different circadian cycles of a given fly. In order to obtain the average profile of a fly, the locomotor data were superimposed for the successive circadian cycles, approximately seven. In this way, an average pattern of activity was obtained (see chapter 2 for details). The normalised profiles (in which the highest activity peak is given the arbitrary value of 1 and the other values are assigned proportionally) from different flies could also be summed together in order to give a better estimate of the average activity pattern.

The profiles at the different temperature are shown in fig. 4.3 below. Note that while the peak of activity for flies kept at 18° C occurs roughly in the middle of the circadian day (CT 6), flies at 25° and 29° display a later peak at about CT 9.

The absolute level of activity is higher for flies kept at 25° C than at the other two recording temperatures. The counts per bin of *Musca* kept at 25° C are about 100 for the maximum peak of activity, while at 18° and 29° C the peaks of activity, averaged between the various flies, recorded about 50 and 25 counts per bin respectively.

In a light-dark regimen at 25° C, the activity of the flies is restricted to the light phase of the cycle. In the histogram in figure 4.4 it is possible to appreciate how the flies activity starts abruptly with the onset of light and terminates at the time of lights off. Rather than reaching a peak at a particular time of the cycle as in constant darkness, houseflies maintain a constant activity for the whole light phase.

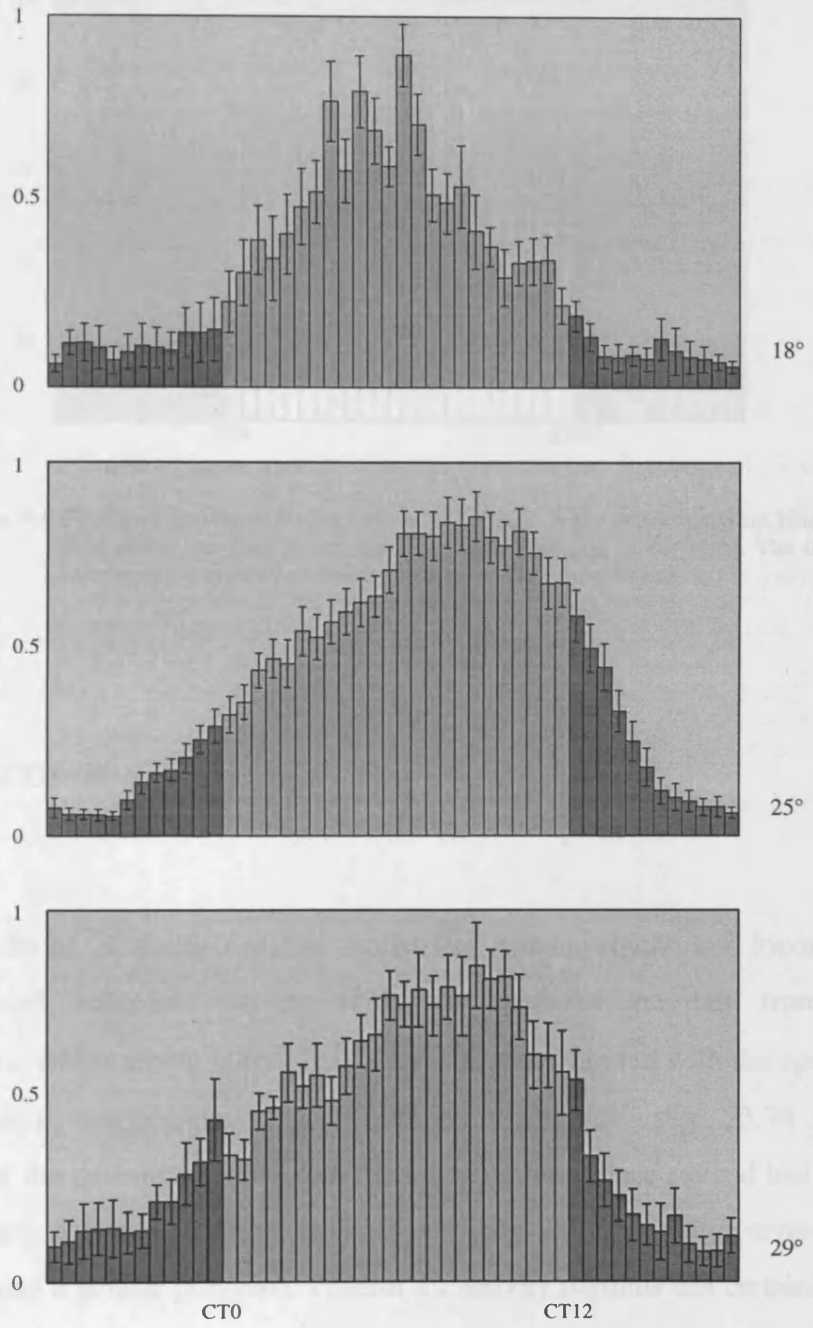


Figure 4.3. Profile of activity of *Musca* flies at various temperatures in DD. Light and dark boxes represent bins belonging to the subjective day and night, respectively. The ordinate represents the proportion of activity (see text). The patterns of activity at 18° and 29° C result from the average of 8 flies, while the profile at 25° is calculated as the average activity of 19 flies. The absolute levels of activity are higher at 25° C than at the other temperatures (see text).

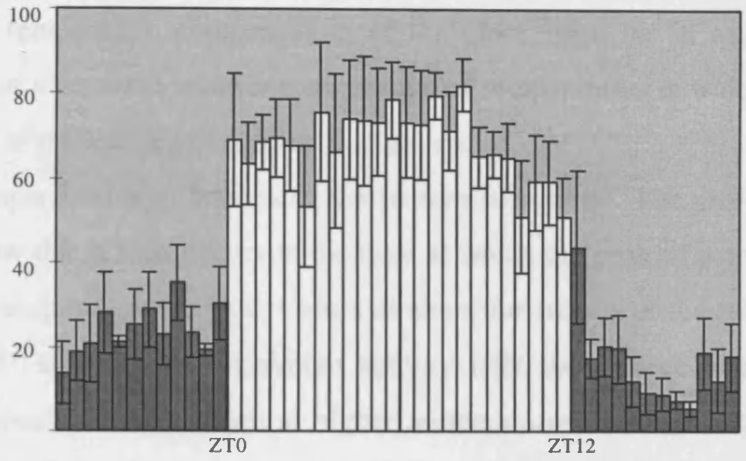


Figure 4.4. Profile of activity of *Musca* flies in LD at 25°. White boxes represent bins of the light phase and dark boxes represent bins belonging to the night. The ordinate represents the number of activity events (average from 4 flies).

DISCUSSION

The fly *M. domestica* exibits robust free running rhythms of locomotor activity, with a period, calculated as the arithmetic mean of the data from the various temperatures, which amounts to 23.92 ± 0.07 h, as computed with the spectral analysis. The less precise autocorrelation procedure gives a similar value, 23.74 ± 0.11 h. The variability of the periods of individual flies is quite low. Since several loci are known to affect the period of free running locomotor activity in *Drosophila* (reviewed in Hall *et al.*, 1987), and a similar polygenic control on activity rhythms can certainly be expected in *Musca*, this lack of variation can simply reflect a genetic homogeneity of the flies used. Indeed this strain, called Cooper, was obtained from the Wellcome Public Health laboratory in the early 1970s and is therefore highly inbred. The temperature compensation of the clock is extremely good, resulting in virtually unchanged periodicities at the three temperatures considered.

However the level of activity is affected by the temperature. At relatively cold temperatures, as expected, flies are not very active, but display a higher activity at the intermediate temperature. Finally, at 29° the level of activity is decreased to values even lower than the ones at 18°. This can give us a good idea of how refined the mechanism

controlling temperature compensation of the clock must be, in order to maintain the periodicity at a constant value across a range of temperatures in which variation in other parameters of the locomotor activity is observed.

Temperature also influences the pattern of activity. The profiles of free running activity show that a shift occurs in the time at which the peak of activity is recorded. At the lowest temperature the peak occurs at about the middle of the subjective day (CT 6) while at 25° and 29° the maximum activity falls about three hours later. This may underlie a mechanism by which at higher temperatures, the fly's activity is confined to cooler times of the day. In contrast with the activity of *D. melanogaster* (Thackeray, 1989 and next chapter), no bimodality of locomotor activity was found in *Musca* at our recording conditions. With the term "bimodality" I refer to a characteristic profile of locomotor activity in which two peaks of activity are recorded in each circadian cycle, typically one in the subjective morning and one in the evening, separated by a "resting" time during which the activity is either very low or nil. A bimodal profile is for example characteristic of the locomotor activity of *D. melanogaster* at 25° and 29° C (see next chapter).

The appearance of double peaks of activity is seen by some authors, as an indication of the multioscillatory nature of circadian rhythmicity (Helfrich, 1985; Han and Engelmann, 1987). Uncoupling of two of the oscillators would result in two bursts of activity but the length of the circadian cycle will not be affected. If the out-of-phase oscillators assume different periods, the fly's activity either exhibits two periodicities or becomes arrhythmic (Smietanko and Engelmann, 1989b). This theory must wait for the elucidation of the output pathway connecting the oscillatory information to the locomotor activity responses, but other mechanisms responsible for the appearance of double peaks of activity could be imagined without invoking extranumerary oscillators. For example, if the level of a certain state variable (such as concentration of an oscillating molecule), whose time of expression is under clock control, can trigger the behavioural response when reaching a given value, overexpression of the state variable could potentially result in a double burst of activity (see fig. 4.5).

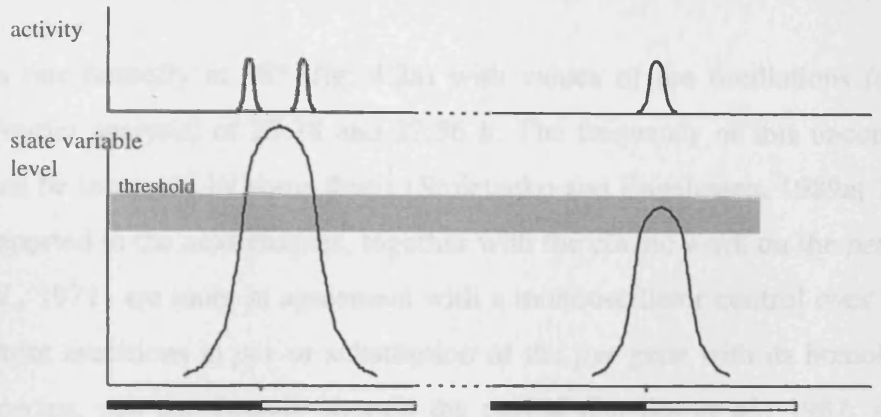


Figure 4.5. Hypothetical pathway leading to bimodality of the activity profile.

Bimodality of locomotor activity at warm temperatures, seems to be the rule rather than the exception in *D. melanogaster* (see chapter 5). This has been interpreted as adaptation of the behaviour to the arid climate typical of the original habitat of this fly, and the fact that several other *Drosophila* species from Afro-tropical climates show double peaks of activity (Thackeray, 1989), supports this view. Since bimodality probably reflects a physiological requirement of the fly (avoiding hot and dry environmental conditions), it could be desirable if this characteristic of the behaviour was due to an effect of temperature on the output pathway rather than on the desynchronisation of any underlying oscillator, this latter option being potentially disruptive. If activity monitoring was performed for several circadian cycles at 29° (conditions which maximise bimodality), followed by a second phase at 18°, persistence of bimodality at low temperature would favour of the out-of-phase oscillators hypothesis, while immediate switching to the unimodal pattern would be more in agreement with an effect on the output (even though resetting of one oscillator by the other cannot be excluded).

Smietanko and Engelmann, (1989b) also report that the oscillators can assume two different rhythms, which results in splitting of the circadian period into two components. They assume that the two apparently different phenomena, bimodality and split rhythms, are to be ascribed to the same cause, that is desynchronisation of the oscillators, and refer to both with the same term "splitting". Split rhythms, intended as the coexistance of two distinct periodicities in the same individual, were first described by Pittendrigh (1960) in the rodent *Spermophilus undulatus*. Split rhythms are observed

in one housefly at 18° (fig. 4.2a) with values of the oscillations (calculated with the Fourier analysis) of 23.78 and 27.56 h. The frequency of this uncommon phenomenon can be increased by some drugs (Smietanko and Engelmann, 1989a; 1989b). The results reported in the next chapter, together with the classic work on the *per* gene (Konopka *et al.*, 1971) are more in agreement with a monooscillator control over locomotor activity. Point mutations in *per* or substitution of the *per* gene with its homologue from another species, can consistently shorten the period (Baylies *et al.*, 1987; see also chapter 5) without giving rise to split rhythms, meaning that the molecular oscillator of which *per* is one of the components, is the only one affecting the locomotor activity.

An alternative view is that Per is part of a mechanism which controls different oscillators acting with a shorter time scale and is able to transduce into circadian periodicity these shorter oscillations, as an orchestra director can obtain a symphony from a network of different musicians playing different tunes on different instruments (Power *et al.*, 1995).

Either way, it must be remembered that oscillations in *per* levels occur in many anatomical structures of *D. melanogaster* (Siwicki *et al.*, 1988; Liu *et al.*, 1988; Ewer *et al.*, 1992; Frisch *et al.*, 1994; Giebultowicz and Hege, 1997) as well as in cell culture (Emery *et al.*, 1997), and that the lateral neurons (LN), supposedly the site of the fly's locomotor activity pacemaker (Frisch *et al.*, 1994), present a bilateral symmetry. Therefore a minimum of two potential pacemakers exist in the brain of *D. melanogaster* and this number can be even greater considering that the glial cells surrounding the LN can drive long and weak rhythms in *D. melanogaster* (Ewer *et al.*, 1992). Uncoupling between the left and right cluster of lateral neurons, or cessation of the hierarchical predominance of the LN rhythms over the rhythms driven by other anatomical structures could certainly result in the appearance of a double periodicity in the locomotor activity. A detailed study of split rhythmicity could unravel the mechanisms by which the molecular oscillation in a small subset of cells can influence the circadian rhythms of the whole organism.

After this quick view on the circadian locomotor behaviour of *Musca*, in the next chapter attention will focus on the circadian activity of *D. melanogaster* transformants carrying *D. melanogaster/Musca* chimaeric transgenes.

CHAPTER 5

Rescue of circadian rhythmicity in *D. melanogaster* by the *M. domestica per* gene

Rescue of circadian rhythmicity in *D. melanogaster* by the *M. domestica per* gene

INTRODUCTION

Chapter 3 describes the cloning and analysis of the *period* gene from *Musca domestica*. The phylogenetic analyses suggest that a portion of the gene is coevolving with another component of the clock mechanism. The dimerization domain of Per, the PAS region, has been found to physically interact with the clock component Timeless (Tim, Sehgal *et al.*, 1994; Vosshal *et al.*, 1994) both *in vitro* (Saez and Young, 1996) and *in vivo* (Zeng *et al.*, 1996; Myers *et al.*, 1996; Lee *et al.*, 1996); the interaction is essential for nuclear entry of Per (Vosshal *et al.*, 1994) and, in turn, for a correct expression of the clock output. So an altered Per-Tim interaction could affect the timing of nuclear translocation of the complex, with consequences for the circadian phenotype. Indeed the nuclear entry of Per^{L}-Tim is delayed for several hours with respect to the Per^{+}-Tim complex, accounting for the lengthening of the period of a per^{L} mutant (Curtin *et al.*, 1995). With this scenario in mind, it is easy to see that most of the interspecific amino acid changes which occur in the dimerization domain will affect interspecific protein-protein interaction with deleterious effects. However some mutations could arise which will be compensated by changes in the dimerization domain of the conspecific partner protein.

To adopt a familiar analogy, the dimerization domains must adapt to each other like a key to its lock; changes in the key must be followed by compensatory changes in the lock in order to preserve the functionality of the key-lock system. The two elements are not free to evolve independently, but are coevolving, each one influencing, and being influenced by, the other. We can expect then that the more similar the PAS domains of two species, the more similar will be the dimerization domains of the partner protein. Since *D. melanogaster* PAS is more similar to *M. domestica*'s than it is to *D.*

pseudoobscura's, then the same might be observed for the PAS-interacting domain of Tim.

One example of molecular coevolution is given by the intergenic spacers of the rDNA transcription units and the interacting RNA-polymerase I complex (Dover, 1982). The coding sequences for rRNA 18S, 5.8S and 28S are grouped in a single transcription unit. Tandem repeats of these units are clustered at one or several genomic locations, each unit separated from the other by an intergenic spacer, containing the promoter, enhancers and terminator. Length and sequence of the intergenic spacers are species-specific and, in spite of their high copy number, high homogeneity is observed within a given species (reviewed by Federoff, 1979; Dover, 1982). Various mechanisms might be responsible for maintaining the observed homogeneity (such as unequal exchange, gene conversion, RNA transfer, etc.) which are collectively referred to as "molecular drive" (Dover, 1982). The specificity of the RNA-pol. I (the pol. I complex only transcribes the rDNA of its own species, reviewed by Grummt *et al.*, 1982) probably means that a newly arisen variant will not be transcribed, but this will hardly affect the fitness of the mutant given the rDNA high copy number. Molecular drive can then spread the new variant to replace all the old units, but this will only be possible after "adaptation" of the RNA-pol. I complex to the mutant intergenic spacer.

A second example of coevolution is provided by the *per* gene. The Thr-Gly region of *per* has been described in Chapter one and three. The encoded tandem repeat shows great length variability among *Drosophila* species (Colot *et al.*, 1988; Peixoto *et al.*, 1993). The work of Peixoto *et al.* (1993) suggests that the length of this region coevolves with the adjacent segments, which are relatively conserved between species. There is a correlation between the logarithm of the length difference of the repetitive array from pairwise comparisons of *Drosophila* species, and the number of amino acids changes recorded at the flanking segments. This does not reflect the time of divergence between the two species because no correlation is found between the log. of the length difference and the percentage of third-base nucleotide changes at the flanking area (most of which are synonymous and therefore should better estimate the time of divergence than the amino acid substitutions). Thus, slippage-like mechanisms or unequal crossing over may change the length of the repetitive array driving compensatory changes in the flanking segments of the repeat or *vice versa*.

In an interesting experiment, Peixoto *et al.*, (1998) analysed per^{01} flies transformed with two chimaeric *per* genes for the rescue of circadian rhythmicity. One hybrid construct, called pMPS2, contained the *D. melanogaster* 5' UTR plus coding sequence up to the start of the Thr-Gly region, fused to the *D. pseudoobscura* repeat and *D. pseudoobscura* C-terminal sequences. In the other construct, pMPS3, the chimaeric junction was moved slightly upstream, so replacing a 55 amino acid long 5' flanking segment of *D. melanogaster* with the equivalent *D. pseudoobscura* sequence (see fig. 5.1). Despite containing a higher proportion of *D. pseudoobscura* sequence, per^{mps3} is able to fully rescue the circadian phenotype in *D. melanogaster* whereas per^{mps2} fails to do so. Thus, the structural modification caused by the exogenous repeat seems to be compensated by the correct conspecific 5' flanking region, thereby giving experimental support to the coevolution proposed between the two adjacent sequences.

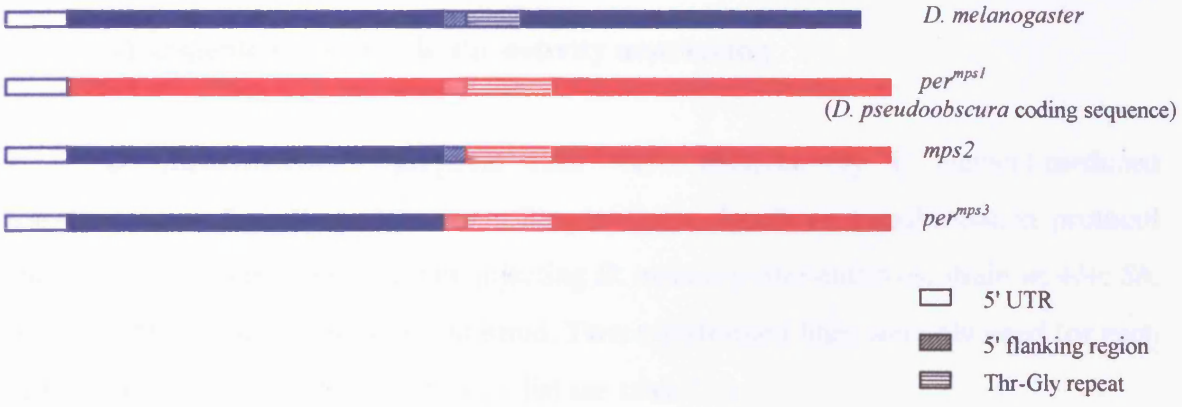


Figure 5.1. Diagram of the chimaeric genes per^{mps1}, per^{mps2} and per^{mps3}. *D. melanogaster* sequences are illustrated in blue while *D. pseudoobscura* sequences are depicted in red. The position of the Thr-Gly repeat and its 5' flanking region is indicated.

Two major differences are noticeable between the examples given. In the first case (rRNA), coevolution occurs between two different genes (intermolecular), whereas in the second, two adjacent regions of the same molecule are coevolving (intramolecular). Besides, in the case of the intergenic spacer of the rDNA, redundancy of the transcription unit means that mutations in the spacer will not affect fitness until the new variant starts to spread, while in the single copy *per* gene, the effect of changes on the fitness could be felt immediately.

In order to further test the validity of the intramolecular coevolution hypothesis in *per*, new chimaeric constructs were made using the *D. melanogaster* and *M. domestica* coding sequences. Chimaeric constructs pMM2 and pMM3 are analogous to pMPS2 and pMPS3, with the difference that the *D. pseudoobscura* sequences are substituted by *Musca* homologous regions. Plasmid pMM1 (analogous to pMPS1) was used to test the functionality of the *Musca per* coding region in *D. melanogaster*, so to indicate whether coevolution at the intermolecular level (between PAS and the partner protein's dimerization domain) is effectively occurring. Maps and details of the chimaeric genes construction are given in Chapter 2.

METHODS

Transgenic lines used in the activity monitoring

D. melanogaster transgenic flies were obtained by P element-mediated transformation (Spradling, 1986; see Chapter 2 for details on transformation protocol and selection of transgenic flies) by injecting *D. melanogaster* embryos, strain *w*; +/+; *Sb*, *e*, *Δ2-3/TM6*, with the chimaeric plasmid. Two transformed lines were obtained for each of the chimaeric genes described (for a list see table 5.1).

Four individual G_1 flies were obtained from crossing pMM1-injected fly 34 with a *w* strain. Three had the *Sb*, *e*, *Δ2-3* chromosome (34a, 34b and 34d) and one carried *TM6* (34c). These flies were then individually crossed to a *w* strain and analysis of their progeny indicated that two insertional events occurred in the germline of fly 34, one on the *Sb*, *e*, *Δ2-3* chromosome and the other on *TM6* since there was no segregation between the red eye phenotype given by the w^+ gene carried by the transformation vector pW8 and the *Sb* marker in lines 34 a, b and d; similarily, no segregation was observed in line 34c between the w^+ gene and the *Ubx* marker carried by *TM6*. The *TM6* chromosome in line 34c was subsequently balanced with *MKRS* by crossing fly 34c with a *w*; *MKRS/TM6* strain. In the other three lines, the insert was separated from the *Δ2-3* element by recombination, using *Sb* as a marker for *Δ2-3*, and finally rendered

homozygous. The loss of Δ2-3 was verified by PCR (see later). A single copy of the chimaeric construct is present in lines per^{mm1}-34a and per^{mm1}-34c, as shown by Southern analysis of total genomic DNA (see fig. 5.2). The blot also shows that the site of insertion differs in the two lines since bands of slightly different sizes were obtained by restriction with the same endonuclease (especially evident in the double digests with *Bam*HI-*Eco*RI and *Kpn*I-*Sma*I; in the first case the band from per^{mm1}-34a appears heavier than that of per^{mm1}-34c; the opposite is true for the *Kpn*I-*Sma*I digest). Since the two lines, which bear the insert on two different chromosome III, originated in the same injected egg (fly 34), we can presume that a double insertion occurred in the germ line of fly 34 (Spradling, 1986), one on each of the two chromosome III or, most likely, that a transposition event occurred in the germ line of fly 34 after the initial insertion.

Two independent transformant lines were obtained for each of the chimaeric constructs per^{mm2} and per^{mm3}. In line per^{mm2}-22 the chimaeric gene inserted into chromosome II, while in line per^{mm2}-37 the insertion occurred in chromosome III. Line per^{mm3}-19 carried the insert on chromosome III, and in line per^{mm3}-32 the insertion occurred in chromosome II (see Chapter 2 for a description of how insert position was mapped). The inserts on chromosome III were separated from the Δ2-3 element by recombination (using *Sb* as a marker). Simple crosses using *CyO/Sco* marked second chromosomes and *TM6/MKRS* or $Df(3)Ve^{il}/TM6c$ third chromosomes were used to obtain the strains listed in table 5.1. *In situ* hybridisation using a 12 kb per^{mm2} probe excised from pMM2 cutting with *Bam*HI and *Xho*I, was performed in Padova University by A. Bonini and G. Di Benedetto and confirmed that the transformant lines per^{mm2}-22, per^{mm2}-37, per^{mm3}-19 and per^{mm3}-32 originated by independent integration of a single copy of the chimaeric construct (see fig. 5.3). PCR was used to unequivocally determine the loss of Δ2-3 from the genome of the transgenic lines used in the locomotor activity analysis, in which the insertional event occurred on the *Sb*, *e*, Δ2-3 chromosome III. Given the distance between Δ2-3 (99B, Ashburner, 1989) and *Sb* (89B, Lindsley and Zimm, 1992), loss of the latter by recombination is an indication of possible loss of Δ2-3, but not a guarantee of it. Therefore, fly genomic DNA was amplified using primers *pΔ-5* (5'-AATTATGCGAACAGGAATTC-3') and *pΔ-3* (5'-GTGAGACACCTCGTCGACGT-3') which bind to the Δ2-3 element, and the analysis of

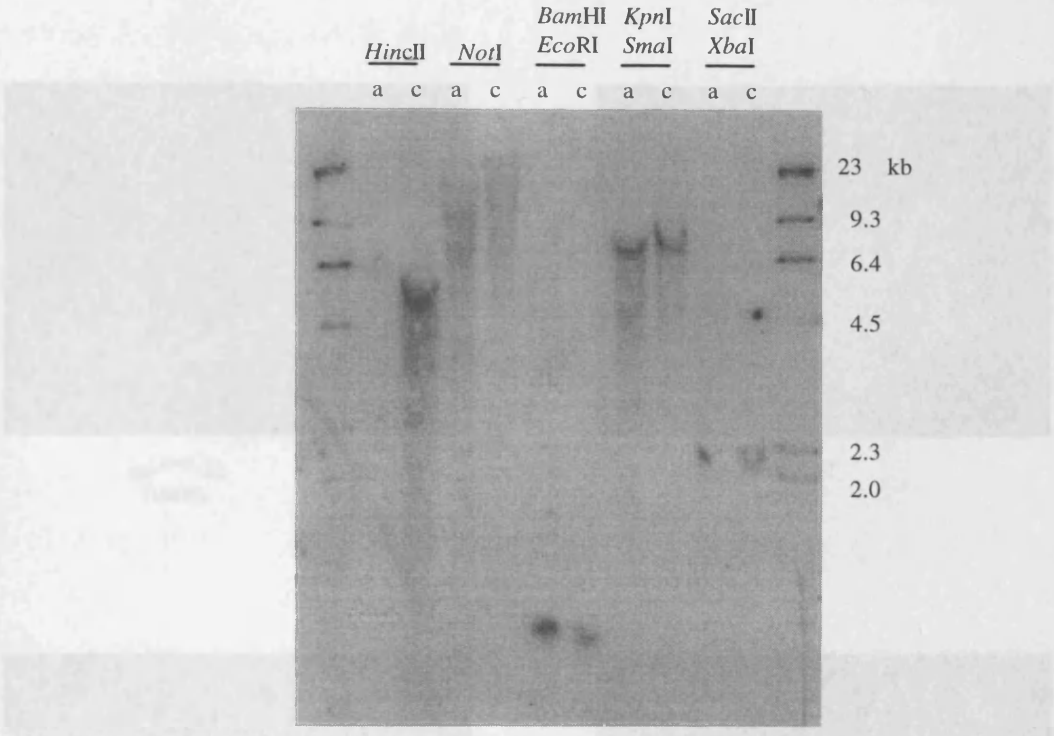
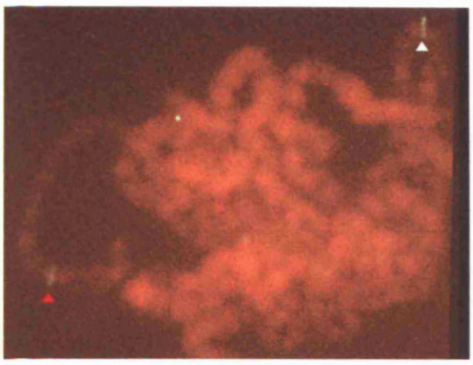


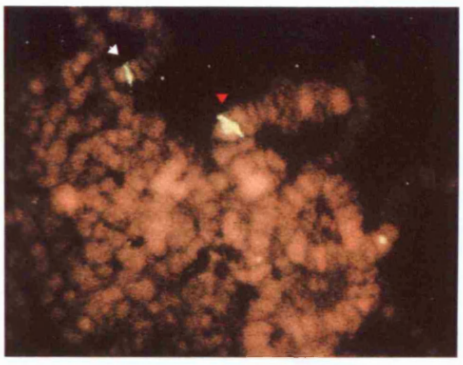
Figure 5.2. Southern blot of genomic DNA of transformant lines *per*^{mm1}-34a (a) and *per*^{mm1}-34c (c). 5 µg of DNA were digested with the endonucleases listed on the top, electrophoresed and blotted on a nylon membrane. The blot was probed with a 600 bp *Hin*dIII-*Xba*I fragment cut from the transformation vector pW8 (Klemenz *et al.*, 1987) containing the 5' P-element end.

Transformed line	insert on chromosome	strain
mm1-34a	3	*per*$^{+}$; *per*^{mm1}/*per*^{mm1}
mm1-34c	3	*per*$^{+}$; *per*^{mm1}*(TM6)/MKRS*
mm2-22	2	*per*$^{+}$; *per*^{mm2}/*per*^{mm2}
mm2-37	3	*per*$^{+}$; *per*^{mm2}*/MKRS*
mm3-19	3	*per*$^{+}$; *per*^{mm3}*/TM6c*
mm3-32	2	*per*$^{+}$; *per*^{mm3}/*per*^{mm3}

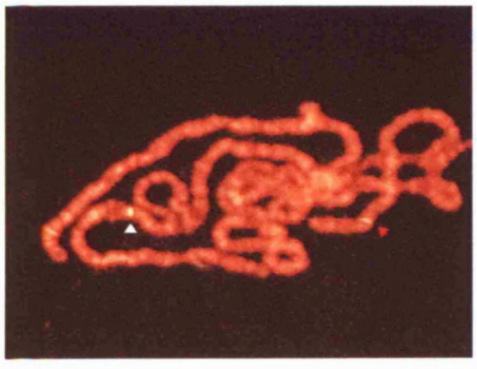
Table 5.1. List of the strains containing the different inserts used for the locomotor activity monitoring.



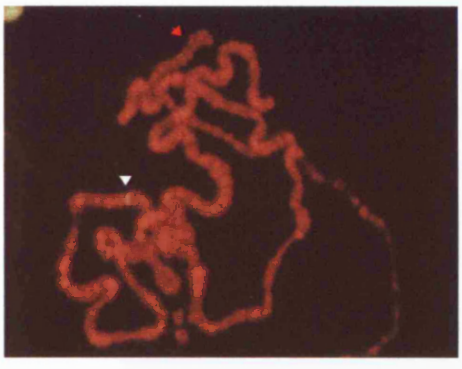
per^{mm2}-22
(1000X)



per^{mm2}-37
(1000X)



per^{mm3}-19
(400X)



per^{mm3}-32
(400X)

Transgenic line	Insert position
per^{mm2}-22	Not mapped
per^{mm2}-37	3R 99
per^{mm3}-19	3L 69F
per^{mm3}-32	2L 38C

Figure 5.3. *In situ* hybridisation of polytenic chromosomes of *D. melanogaster*. The mass of the chromosomes is stained with ethidium bromide, while the FITC signal is localised in regions of probe binding. Two bands are evidenced by the probe, one on chromosome X (endogenous *per* gene, red arrow) and one at the point of transgene insertion (white arrow).

the PCRs confirm that none of the transformed lines with the insert on chromosome III retain the $\Delta-3$ element (see fig. 5.4).

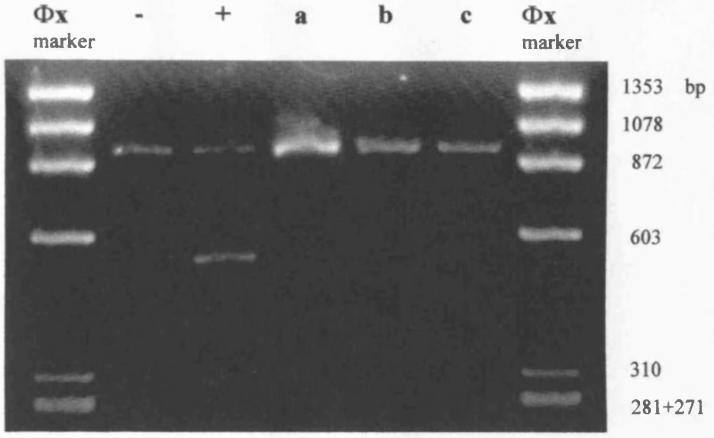


Figure 5.4. PCR on the genomic DNA of transformed lines per^{mm1}-34a (**a**), per^{mm2}-37 (**b**) and per^{mm3}-19 (**c**). Positive (**+**) and negative (**-**) controls are included. Note how the $p\Delta-5$ and $p\Delta-3$ primers amplify a non-specific band at about 0.9 kb.

As a control for the behavioural analysis, per^{01} mutants were used, carrying a 13.2 kb fragment of genomic DNA containing the *D. melanogaster* wild-type *per* sequence plus 5' and 3' adjacent sequences sufficient to drive a wild type-like rhythmicity in transformation experiments (Citri *et al.*, 1987). Transgenic 13.2 lines were obtained by J.M. Hennessy who subcloned the original 13.2 kb fragment from the transformation vector cp20.1 into pW8, and then injected the pW8-based 13.2 construct into *w*; +/+; *Sb*, *e*, *Δ2-3/TM6* flies.

Prior to locomotor activity monitoring, male flies from the different transformant lines were crossed to per^{01} females; 3-7 day old male progeny bearing the chimaeric insert in a per^{01} background were used in the activity experiments.

Computer analysis

The ANOVA analysis was performed with the STATISTICA package developed by Statsoft (Tulsa, Oklahoma). The pattern of locomotor activity was investigated with

the aid of Excel 5 (Microsoft corp. Redmond, Washington) routines written by M. Couchman.

RESULTS

Circadian analysis of the transgenic flies

The rhythmic behaviour of duplicate lines of transgenic flies carrying one autosomal copy of three different chimaeric genes (per^{mm1}, per^{mm2} and per^{mm3}) was monitored. As a control, per^{01} mutants carrying the *D. melanogaster* per^{+} 13.2 kb fragment were examined. Two lines of per^{+} 13.2 transformants (2a and 34a) were generated using the transformation vector cp20.1 (Citri *et al.*, 1987), which carries $rosy^{+}$ (ry^{+}) as a selectable marker. Two additional lines (17a and 116a) were obtained by J.M. Hennessy and have been transformed with the vector pW8, which carries the *white* minigene. Given the large quantity of per^{+} and per^{01} data collected simultanously in our laboratory by my colleagues, I did not analyse these two genotypes. Two different transgenic lines were tested for each transformed genotype in order to minimise possible position effects. Indeed variation in the activity between different transformant lines of the same construct is often observed (Hamblen *et al.*, 1986; Ewer *et al.*, 1990). The expression of the marker gene (*white*) contained in the construct was quite weak in line *mm1*-34c and *mm2*-37, indicating the possibility that in these two lines the chimaeric gene could be expressed at low levels. Activity recording for all lines were performed at three different temperatures (18, 25 and 29 °C) to check whether the chimaeric genes, besides restoring rhythmicity, were able to maintain circadian temperature compensation and to ascertain their effect on the pattern of activity.

The results obtained with the transformants are listed in table 5.2. The per^{+} construct, as expected, re-establishes wild-type circadian rhythmicity of locomotor activity in per^{01} flies. Surprisingly, the per^{mm1} construct is also able to restore high level of circadian rhythmicity in per^{01} mutants. This result is totally unexpected, given that the

genotype	line	temp. (°C)	n° flies tested	% arrhythmic	n° flies with split rhythms	mean period ±SEM (spectral)	mean period ±SEM (autocorr.)
per^{+}	2a	18	61	32.79		24.61 ± 0.11	24.16 ± 0.13
per^{+}	34a	18	51	41.18		24.21 ± 0.16	24.57 ± 0.23
per^{+}	17a	18	59	47.46		24.07 ± 0.18	23.60 ± 0.32
per^{+}	116a	18	50	18		23.51 ± 0.07	23.37 ± 0.15
per^{+}	2a	25	25	8		25.04 ± 0.12	24.66 ± 0.18
per^{+}	34a	25	47	17.02		25.02 ± 0.13	24.62 ± 0.09
per^{+}	17a	25	24	16.67		24.34 ± 0.69	24.55 ± 0.17
per^{+}	116a	25	33	9.09		25.52 ± 0.11	25.58 ± 0.10
per^{+}	2a	29	53	15.09		24.80 ± 0.09	24.94 ± 0.08
per^{+}	34a	29	32	-		24.92 ± 0.16	25.03 ± 0.10
per^{+}	17a	29	46	43.48		24.18 ± 0.19	24.33 ± 0.11
per^{+}	116a	29	33	6.06		25.42 ± 0.16	25.58 ± 0.12
per^{mm1}	34a	18	35	17.14	-	22.70 ± 0.18	22.48 ± 0.20
per^{mm1}	34c	18	43	14.0	1	21.83 ± 0.19	21.97 ± 0.20
per^{mm1}	34a	25	35	0	-	22.42 ± 0.15	22.17 ± 0.12
per^{mm1}	34c	25	36	19.44	1	22.19 ± 0.21	22.05 ± 0.19
per^{mm1}	34a	29	55	12.73	4	22.25 ± 0.15	22.00 ± 0.13
per^{mm1}	34c	29	52	23.08	1	21.89 ± 0.11	21.69 ± 0.13
per^{mm2}	22	18	24	12.50	-	24.13 ± 0.19	24.40 ± 0.35
per^{mm2}	37	18	35	17.14	-	23.82 ± 0.35	24.00 ± 0.32
per^{mm2}	22	25	52	0	2	23.96 ± 0.09	23.59 ± 0.09
per^{mm2}	37	25	23	0	2	24.15 ± 0.21	23.30 ± 0.17
per^{mm2}	22	29	34	0	5	24.60 ± 0.23	24.34 ± 0.14
per^{mm2}	37	29	29	10.34	5	24.41 ± 0.30	23.79 ± 0.26
per^{mm3}	19	18	32	6.25	-	23.59 ± 0.14	23.70 ± 0.20
per^{mm3}	32	18	40	10.0	-	23.17 ± 0.13	23.01 ± 0.17
per^{mm3}	19	25	28	0	9	24.06 ± 0.23	23.73 ± 0.18
per^{mm3}	32	25	30	3.33	3	23.54 ± 0.22	23.53 ± 0.17
per^{mm3}	19	29	102	4.90	8	23.62 ± 0.20	23.43 ± 0.15
per^{mm3}	32	29	25	4.0	10	23.73 ± 0.39	23.67 ± 0.20

Table 5.2. Summary of the activity data for all genotypes and lines analised.

per gene from the evolutionarily more closely related *D. pseudoobscura* can only poorly rescue circadian functions in *D. melanogaster* (Petersen *et al.*, 1988; Peixoto *et al.*, 1998). Flies transgenic for per^{mm1} display a shorter-than-normal period with a percentage of statistically rhythmic flies, as detected by autocorrelation and spectral analysis (see Chapter 2), always above 75%. Table 5.2 shows the results for the other genotypes and it can be seen that the autocorrelation and spectral analysis generally give very similar periods. As the spectral analysis gives a better resolution for the period (see Chapter 2 for a description of the two methods) only the values calculated according to this procedure were used in further analysis.

In order to check any significant effects of genotype, line or temperature on the period of circadian rhythmicity, ANOVAs were performed on the entire data set. As a first step ANOVAs were performed on each genotype separately. Two independent variables were considered, Line and Temperature. The results are reported in table 5.3 where it can be seen that the four per^{+} 13.2 transformant lines show a statistically significant difference for the main effects and the interaction considered (Line, $p < .0001$; Temperature, $p << .0001$; LxT interactions, $p << .0001$). The two lines of per^{01}; per^{mm1} transformants are significantly different ($p = .0005$), but are both temperature compensated (no Temperature effect shown). Transformant per^{01} mutants carrying either the per^{mm2} or per^{mm3} gene do not show any significant effect of Line or Temperature.

	Effect	df numerator	MS numerator	df denominator	MS denominator	F	p-level
per^{01}; per^{+} 13.2	Line (L)	3	6.97629	375	.986011	7.07527	**.000123**
	Temperature (T)	2	28.34091	375	.986011	28.74299	**.000000**
	LxT	6	7.76735	375	.986011	7.87755	**.000000**
per^{01}; per^{mm1}	Line	1	12.50381	212	1.000495	12.49763	**.000500**
	Temperature	2	1.26167	212	1.000495	1.26104	.285476
	LxT	2	1.96681	212	1.000495	1.96584	.142582
per^{01}; per^{mm2}	Line	1	.473711	179	1.551303	.305364	.581229
	Temperature	2	4.617551	179	1.551303	2.976564	.053497
	LxT	2	.987124	179	1.551303	.636319	.530430
per^{01}; per^{mm3}	Line	1	3.839476	238	2.474075	1.551883	.214083
	Temperature	2	2.856030	238	2.474075	1.154383	.317011
	LxT	3	1.940910	238	2.474075	.784499	.457525

Table 5.3. Summary table for ANOVAs on the period of per^{01} transformants. Individual ANOVAs were run within each genotype. Bold type, $p < .05$.

Planned comparisons revealed that per^{+} 13.2 line 116a appears to react differently to temperature from all the other per^{+} 13.2 lines ($p < .0002$), where lines 2a, 17a and 34a represent an homogeneous group. On the other hand the high variance associated to the periodicity of line 17a means that, concerning the crude line effect (collapsed over temperature), this line appears significantly different from the remaining three per^{+} 13.2 lines ($p << .0001$). The interaction of the various genotypes with temperature is a particularly interesting characteristic of the locomotor behaviour and I reasoned that the analysis of this feature could prove more revealing than that of the simple Genotype effect. Therefore I decided to group together lines 2a, 17a and 34a (see fig. 5.5 for a graphical representation of the whole behavioural data).

In a subsequent step, the four different genotypes (per^{+} 13.2, per^{mm1}, per^{mm2} and per^{mm3}) were compared against each other. Since there appear to be two different groups within the per^{+} 13.2 genotype (lines 2a, 17a and 34a versus line 116a) and within the per^{mm1} genotype (line 34a and line 34c), pooling these groups for ANOVA would add considerable "noise" to the analysis. Consequently four separate analysis were performed, in which genotypes per^{mm2} and per^{mm3} are always represented by the pooled

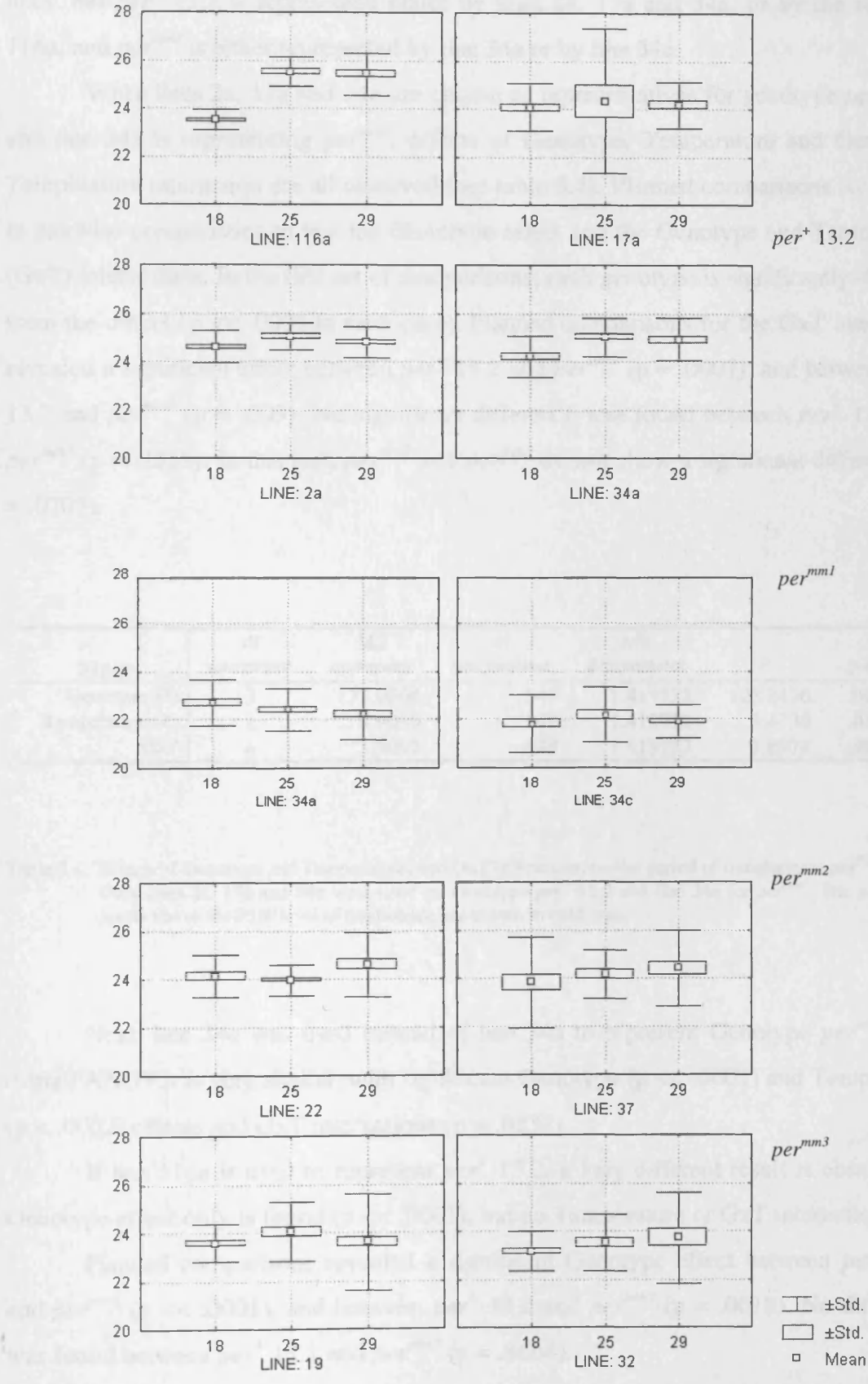


Figure 5.5. Histograms depicting the behavioural data collected for the various lines at the three different temperatures of 18°, 25° and 29° C. Temperature values are on the X axis, while the Y axis represents period length in hours.

lines, but per^{+} 13.2 is represented either by lines 2a, 17a and 34a, or by the sole line 116a, and per^{mm1} is either represented by line 34a or by line 34c.

When lines 2a, 17a and 34a are chosen as representatives for genotype per^{+} 13.2 and line 34a is representing per^{mm1}, effects of Genotype, Temperature and Genotype-Temperature interaction are all observed (see table 5.4). Planned comparisons were used in pairwise comparisons to test the Genotype effect and the Genotype and Temperature (GxT) interactions. In the first set of comparisons, each genotype is significantly different from the others ($p << .0001$ in each case). Planned comparisons for the GxT interaction revealed a significant effect between per^{+} 13.2 and per^{mm1} ($p = .0007$), and between per^{+} 13.2 and per^{mm2} ($p = .009$). No significant difference was found between per^{+} 13.2 and per^{mm3} ($p = .2355$). In this test, per^{mm2} and per^{mm3} do not show a significant difference ($p = .0705$).

Effect	df numerator	MS numerator	df denominator	MS denominator	F	p-level
Genotype (G)	3	175.0808	849	1.413732	123.8430	**.000000**
Temperature (T)	2	4.9099	849	1.413732	3.4730	**.031466**
GxT	6	5.5097	849	1.413732	3.8973	**.000757**

Table 5.4. Effects of Genotype and Temperature, and GxT interaction, on the period of transformant per^{01} mutants. Only lines 2a, 17a and 34a were used for Genotype per^{+} 13.2 and line 34a for per^{mm1}. The probability levels above the 95% level of confidence are shown in bold type.

Next, line 34c was used instead of line 34a to represent Genotype per^{mm1}. The overall ANOVA is very similar, with significant Genotype ($p << .0001$) and Temperature ($p < .0002$) effects and GxT interactions ($p = .0232$)

If line 116a is used to represent per^{+} 13.2, a very different result is obtained. A Genotype effect only is found ($p << .0001$), but no Temperature or GxT interactions.

Planned comparisons revealed a significant Genotype effect between per^{+} 13.2 and per^{mm1} ($p << .0001$), and between per^{+} 13.2 and per^{mm3} ($p = .0018$). No difference was found between per^{+} 13.2 and per^{mm2} ($p = .8664$).

Essentially the same results are obtained when line 34a is substituted with line 34c for the per^{mm1} genotype (data not shown).

Finally, the data for the various lines of each genotype were pooled together. This way the increased variance associated with the per^{+} 13.2 GxT interactions would make the test more conservative. Table 5.5 shows that significant Genotype and Temperature effects, and GxT interactions are observed in the ANOVA with the pooled data. Even with this new, more conservative analysis, planned comparisons reveal that GxT interactions are significantly different between per^{+} 13.2 and per^{mm1} (p = .0129) and between per^{+} 13.2 and per^{mm2} (p = .0136) but no such a difference is detected between per^{+} 13.2 and per^{mm3} (p = .4380).

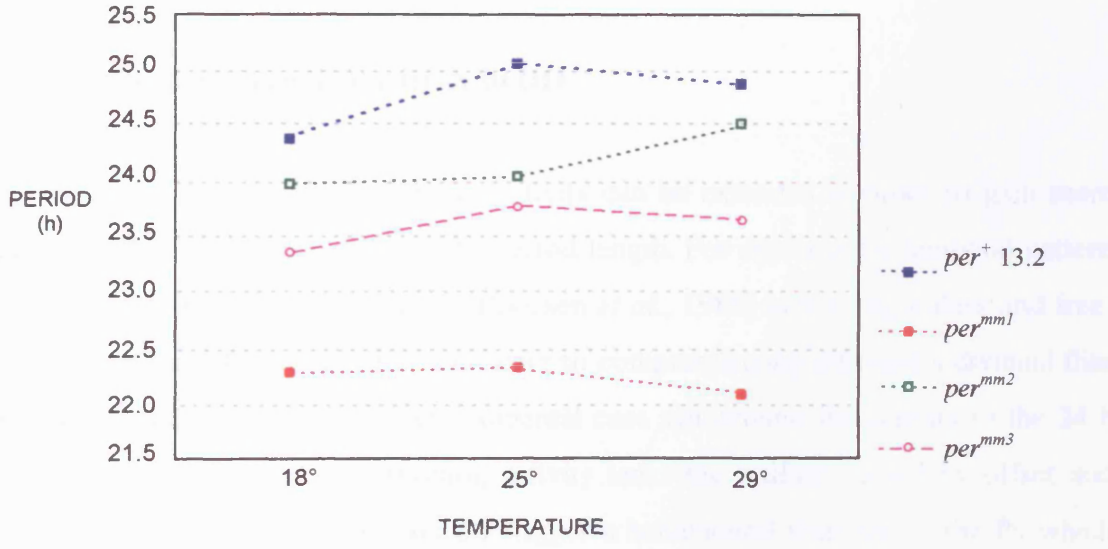


Figure 5.6. Temperature-genotype effect (collapsed over lines) on the period of transformant flies.

Effect	df numerator	MS numerator	df denominator	MS denominator	F	p-level
Genotype (G)	3	304.6116	1031	1.519293	200.4957	**.000000**
Temperature (T)	2	10.3636	1031	1.519293	6.8213	**.001140**
GxT	6	4.4662	1031	1.519293	2.9396	**.007543**

Table 5.5. Effects of Genotype and Temperature, and GxT interaction. The whole set of data collected for each genotype were pooled together. In bold type are shown the p level above the 95% level of confidence.

The difference in the behaviour of the various genotypes at the three temperatures assayed is summarized in figure 5.6. The period values reported in the chart are given by the unweighted means of the period for each of the two lines (four in per^+ 13.2). Flies transformed with the per^+ *D. melanogaster* gene, and with the per^{mm1} and per^{mm3} constructs, show an increase in their period of rhythmicity passing from 18° to 25°, and when the temperature is raised to 29° this tendency is inverted, while in per^{mm2} transformants, rising the temperature from 25° to 29° has the effect of increasing the period. Thus, the per^{mm2} flies have a temperature profile unlike that of the other transformants.

Pattern of circadian activity in DD

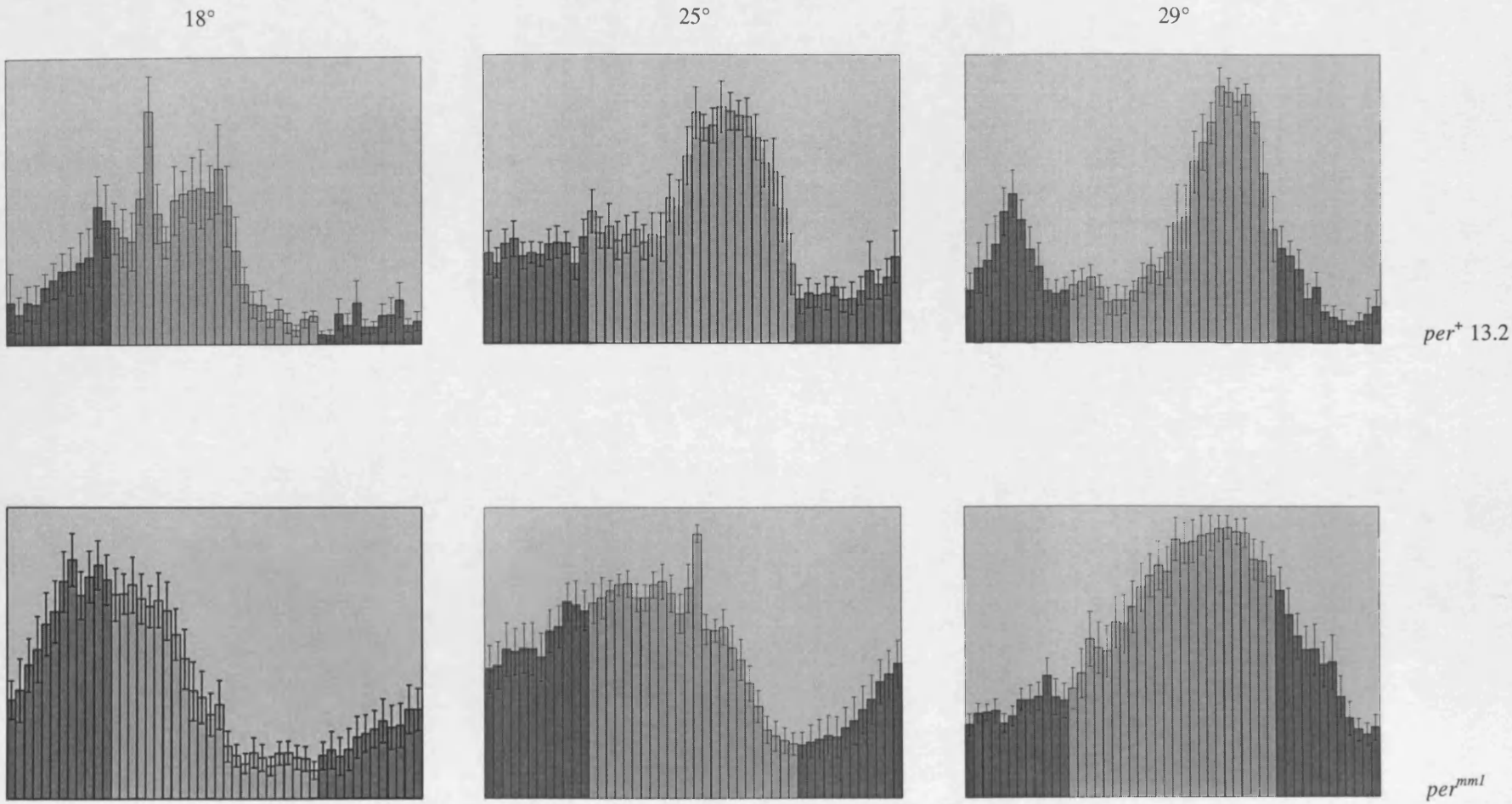
The analysis of the locomotor activity can be extended in order to gain more information than simply the value of the period length. For example the temporal pattern of the fly's activity can be studied (eg. Petersen *et al.*, 1988) in both light-dark and free-running conditions. Patterns in LD are easier to compare among different individual flies and between genotypes, since the enviromental cues synchronise the activity to the 24 h clock. On the other hand, free-running activity lacks the artifact caused by offset and onset of light. In fact, lights on and off trigger a behavioural response in the fly which can potentially mask its endogenous pattern of activity (Jones *et al.*, 1972; Thackeray, 1989).

The behaviour of transgenic flies in free-running conditions was monitored as described in Chapter 2. The number of locomotor activity events per time bin was superimposed for the several successive circadian cycles so that the entire 7 days activity monitoring is "contracted" into a window representing a subjective day, and is subsequently expanded or squeezed to 48 bins (see Chapter 2). This procedure permits a comparison of the activity patterns of flies with different periods.

It has been observed (Petersen *et al.* 1988) that the *per* gene of *D. pseudoobscura* driven by the *D. melanogaster* promoter (per^{mps1}) confers to *D. melanogaster* per^+ hosts, a *pseudoobscura*-like pattern of activity, which at 25° C is characterised by one peak of activity occurring during the subjective late evening, in

contrast to the two peaks (a morning and an early evening peak) which distinguish the circadian activity of *D. melanogaster*. The interpretation given by Petersen *et al.* (1988) is that *per* carries species-specific instructions which affect the circadian behaviour. In contrast, the *D. melanogaster/D. pseudoobscura* chimaeric construct per^{mps3} (containing a *D. melanogaster* fragment up to about 60 aminoacids before the Thr-Gly repeat, and followed by *D. pseudoobscura per* sequences, which corresponds to the *D. melanogaster/Musca* construct per^{mm3}) give to the transformants a wild type like activity profile. *D. pseudoobscura* and per^{mps1} flies have a unimodal pattern of activity in a temperature range from 18° to 29° C, while the profile of per^{+} *D. melanogaster*, and per^{+} 13.2 and per^{mps3} transformants shows a single peak at 18° which becomes bimodal at 25° (Petersen *et al.*, 1988; Hennessy, pers. comm.).

The DD activity profile of flies transformed with the *D. melanogaster/M. domestica* chimaeric constructs has been analysed. The activity of *Musca* has been described in the previous chapter. *Musca* flies have one peak of activity in DD, which occurs at CT 6 at 18° C, and at CT 9 at higher temperature (25° and 29° C, see fig. 4.3). Similarly, in per^{mm1} flies the pattern of activity consists of a single peak, which occurs about 1-2 hours before subjective dawn at 18° and is shifted towards later hours of the circadian time by increasing the temperature (see fig. 5.7). At 25°, the burst of activity appears at CT 4-5 and at 29° at CT 9-10 (see fig. 5.7). The circadian time at which the peaks occur must not be taken very strictly in the case of per^{mm1} transformants. Because the period of these flies is about 22 h (see table 5.2), from the moment the flies enter DD conditions (CT12) to the time I began recording the activity data (CT6 on the following circadian cycle, 18 h later) the circadian clock of per^{mm1} flies could potentially advance 1.5 h, with respect to the 24 h Zeitgeber time. The activity recording is delayed 18 h with respect to the DD start, in order to avoid the anomalous activity triggered by the offset of the lights which could have negative repercussions in the calculation of the periodicity. Despite this discrepancy concerning the absolute CT of the activity peaks, the shape of the daily activity of per^{mm1} flies, with its unimodal pattern at all temperatures analysed, is characteristic of *Musca*, particularly at 29° C. Any phase delay (see above) should not be a problem in per^{mm2} and per^{mm3} flies which have a 24 h period. per^{mm2} transformants activity follows a different profile (fig. 5.7). A burst of activity is recorded at CT 2-3 at 18° C. At 25° the peak shifts to CT 7-8 and a second, small peak appears in several flies


18°
25°
29°
per+ 13.2
permm1

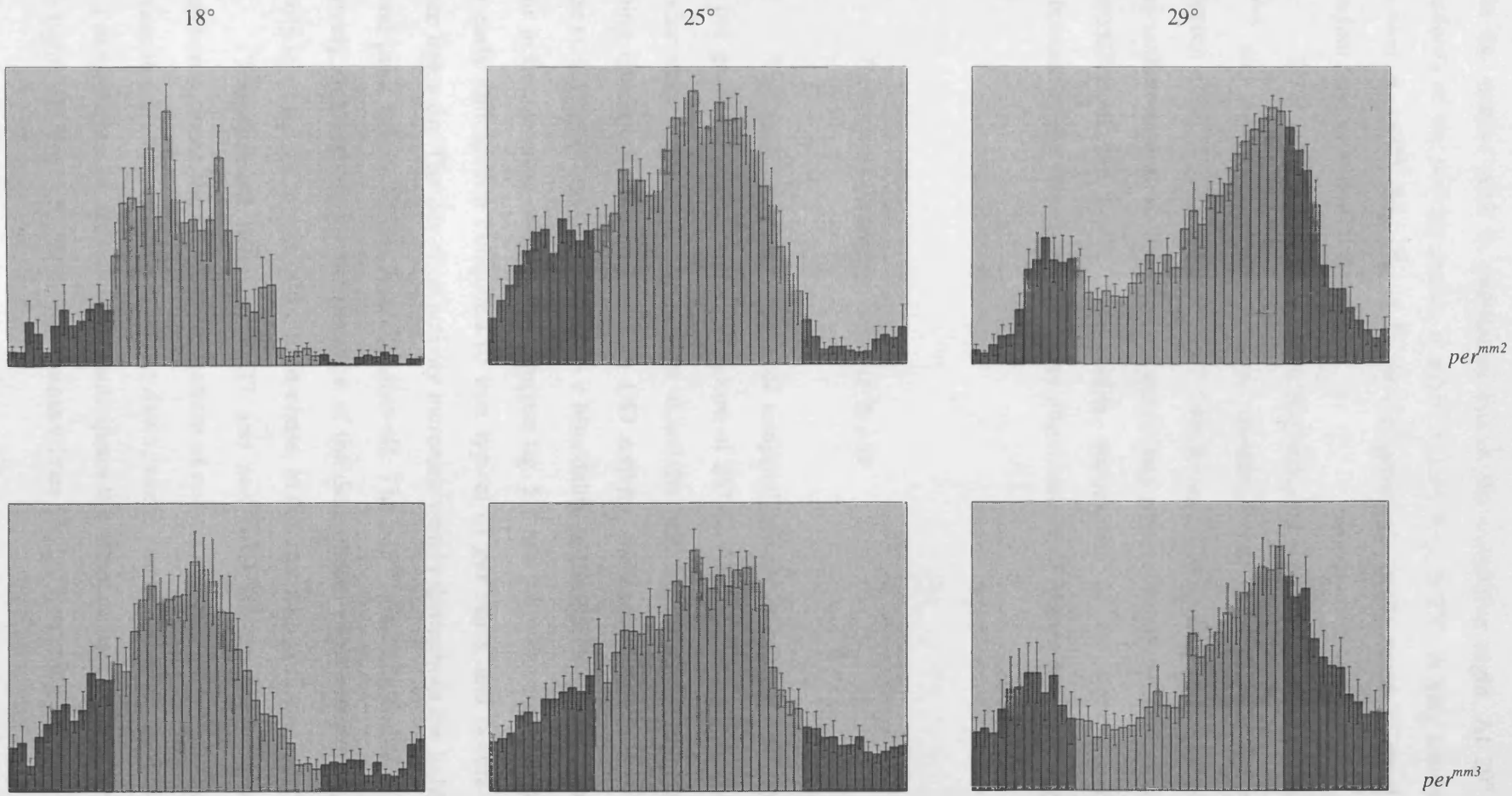


Figure 5.7. Profiles of locomotor activity of *per*$^{+}$ 13.2 and transformant flies. The X-axis delineate the circadian time (CT), with filled boxes and empty boxes representing the subjective night and day, respectively. CT0 is defined as the beginning of the subjective day. The level of activity, depicted by the Y-axis, is an average of the normalised activity level of 10 flies. Error bars are shown in the plots.

at CT 20-21. The major peak shows a further shift at 29° where it occurs at CT 11-12 while the smaller peak is maintained late in the subjective night. At 29° though, the bimodality of the activity profile is more defined than at 25°. A very similar situation is observed in per^{mm3} flies (fig. 5.7). In this genotype, peaks occur at roughly the same circadian time as in per^{mm2} flies.

So it is clearly apparent that the free-running pattern of locomotor activity of per^{mm2} and per^{mm3} transformant flies resembles that of wild type *D. melanogaster* (Petersen *et al.*, 1988) and per^{+} 13.2 transformants, even though the absolute value of the circadian time at which the peaks occur may not be exactly the same. Conversely, the unimodality of the per^{mm1} DD profile, particularly at the hottest temperature, is reminiscent of the free-running activity characteristic of *Musca* (compare figures 4.3 and 5.7).

Pattern of circadian activity in LD

The analysis of the pattern of temporal activity in LD has been limited to one line per genotype and to the temperature of 25° C. At this temperature, the per^{mm1}-34c transformants show a peculiar pattern of activity, which differs from that typical of free-running conditions. In contrast with DD activity, the entrainment cues offered by the cyclic switching of the lights result in a bimodality of the activity pattern which does not occur in free-running conditions (compare fig. 5.7 and 5.8). The morning peak shows a very early anticipation compared to that typical of per^{+} flies, and occurs about 3-4 h before lights-on. The day-time activity increases weakly throughout the light phase and a second peak is seen at the time of lights-off. The per^{mm1}-34c transformants then keep a relatively high activity for the duration of the dark-phase, while maintaining a fairly low activity level for the whole of the light-phase, in contrast with per^{+} (see fig. 5.8).

Transformant lines per^{mm2}-37 and per^{mm3}-19 do not show such a dramatic difference between the DD and LD pattern of activity. In both genotypes there is first an increase in activity a few hours before dawn, which reaches its peak just before lights-on and a second burst of activity which anticipates the offset of light and suddenly decreases after lights-off (fig. 5.8). Both transformant lines display a morning and an evening burst.

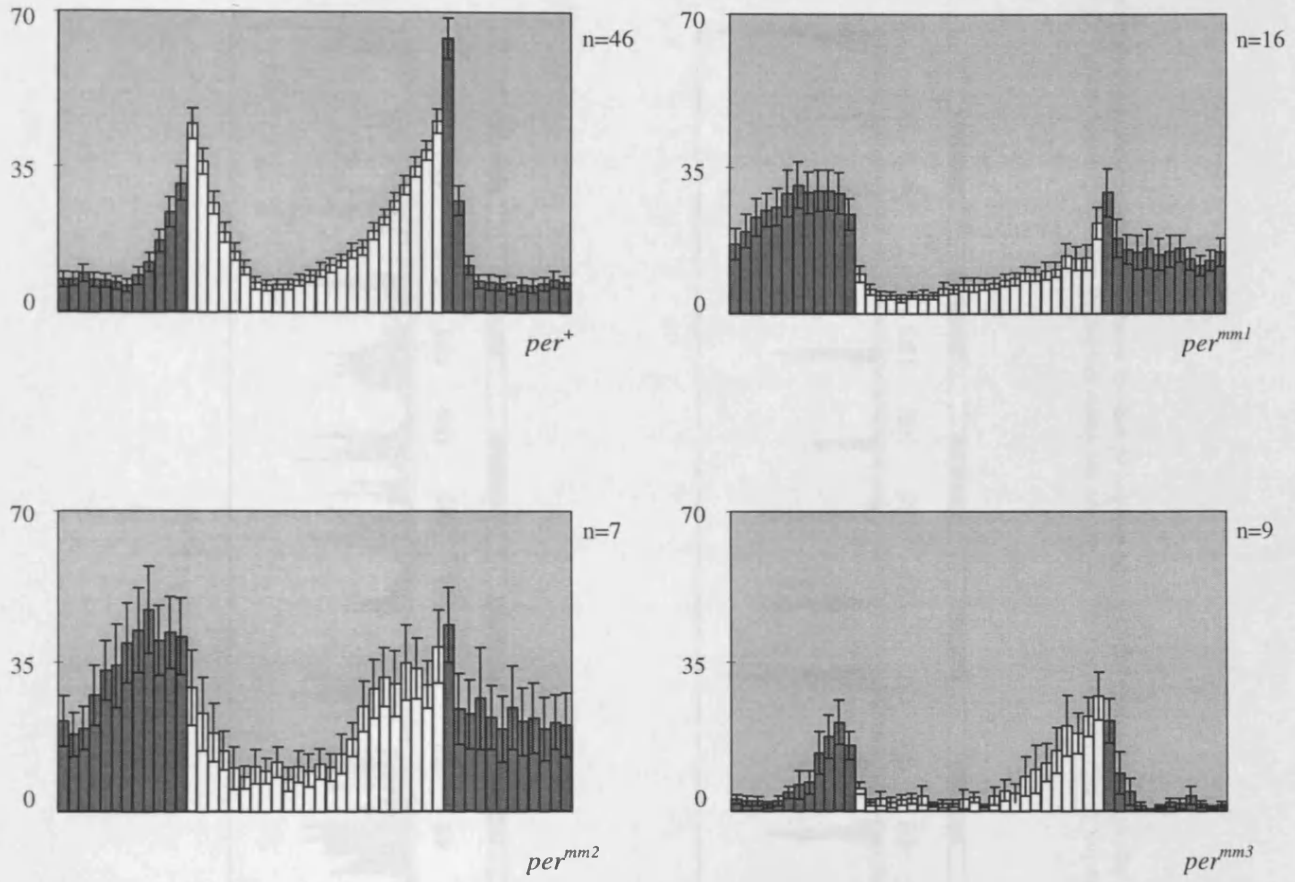


Figure 5.8. Profiles of locomotor activity of transformant flies. The X-axis delineates the zeitgeber time (ZT), with filled and empty boxes representing night and day, respectively. ZT0 is defined as the beginning of the day. The level of activity, depicted by the Y-axis, is an average of the activity of the transformants. The number of flies used for each chart is shown on the right hand corner. Error bars are shown in the plots.

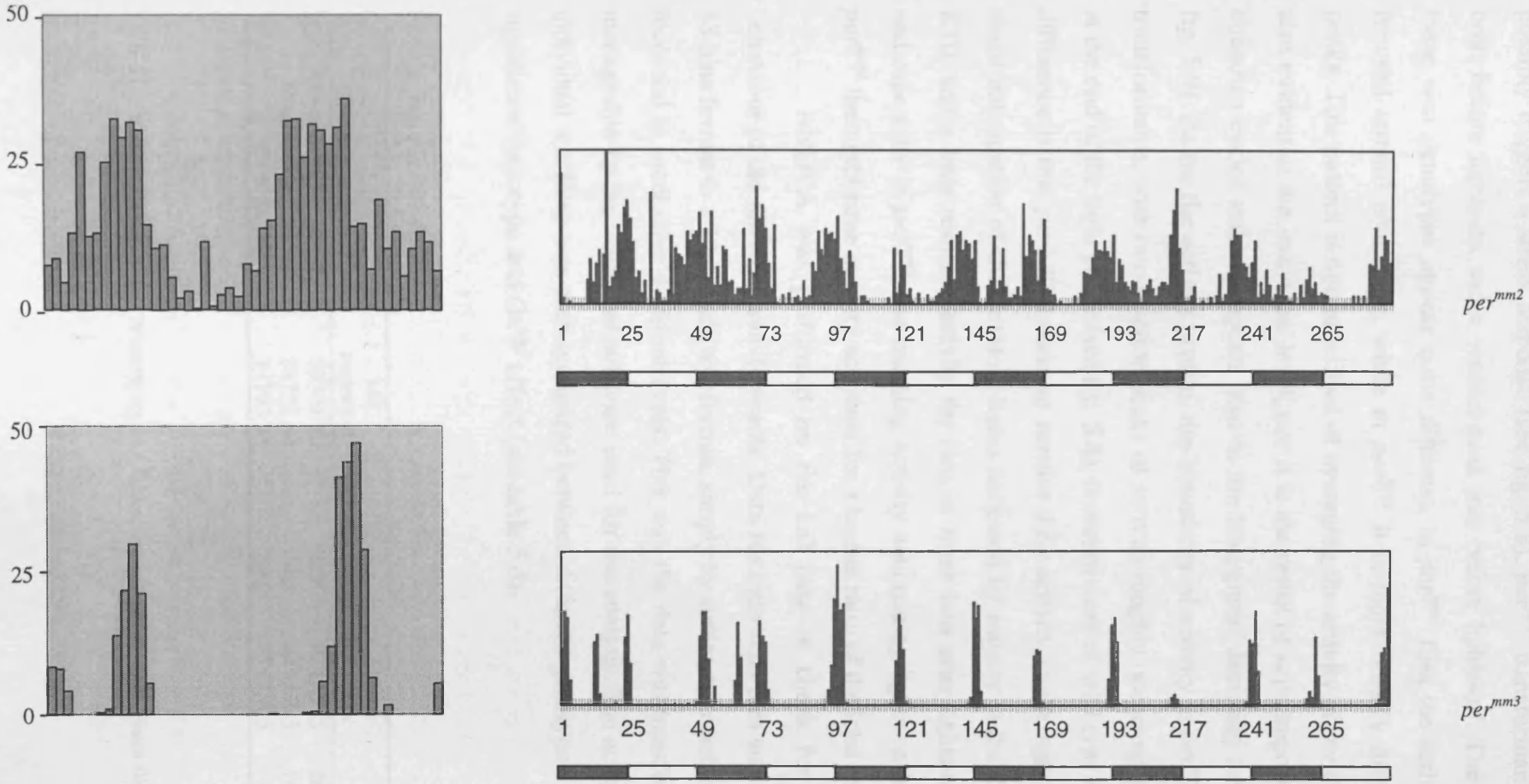


Fig. 5.9. A typical example of *per*^{mm2} (top) and *per*^{mm3} (bottom) activity pattern in LD 12:12 conditions. On the left are the charts displaying the superimposed activity of six circadian cycles, to give a 24 h period (the bars on the X-axis correspond to a 30 min. bin while the Y-axis represent the number of activity events). The histograms of the total activity throughout the six days of recording are shown on the right.

In per^{mm2} the peaks occur 1-2 h before lights-on and at lights-off, when the offset of light possibly triggers a stress response (see fig. 5.8). per^{mm3} transformants have a peak one hour before lights-on and a second peak just before lights-off. The activity profile for these two genotypes appear quite different, in per^{mm3} flies the activity is more tightly focused around peak time, while in per^{mm2} it is more broadly distributed around the peaks. This pattern is not an artifact of averaging the activity of several flies because it is also evident at the individual level, nor it is the result of superimposing together several circadian cycles since it appears also in the histograms depicting the daily activity (see fig. 5.9). As for the activity profile, the bimodality of activity in per^{mm2}-37 and per^{mm3}-19 transformants, with two discrete peaks of activity roughly occurring at the beginning and at the end of the light phase (see fig. 5.8), is reminiscent of wild type *D. melanogaster*. A difference is that per^{+} flies seem to restrict their activity to the light phase, with only a slight anticipation of the onset of lights indicated by activity in the three bins preceding ZT0, and a little residual activity for two or three bins after lights-off, while in per^{mm3}, and especially in per^{mm2}, the morning activity anticipating lights-on is notably higher. In per^{mm2} the night-time activity accounts for a higher ratio of the total locomotor activity.

ANOVA was performed on the LD data to check for differences in the behaviour of the different transformants. Data for individual flies was converted from the 48 bins format to a 12 "windows" format, simply by summing together the activity values recorded in successive 4-bin intervals. This way the data was transformed into a format manageable by the statistical software used for the analysis. The activity associated with individual windows was then compared between different genotypes. The ANOVA gave significant Genotype and GxW effect (see table 5.6).

Effect	df numerator	MS numerator	df denominator	MS denominator	F	p-level
Genotype (G)	3	79301.70	888	2747.975	28.85823	**.000000**
Window (W)	11	38752.84	888	2747.975	14.10233	**.000000**
GxW	33	14793.71	888	2747.975	5.38349	**.000000**

Table 5.6. Effects of Genotype and Window on the LD wrapped activity of transformant flies.

Genotype-Window (GxW) interactions were then tested in pairwise planned comparisons. No significant interactions were found between per^{mm1} and per^{mm2} (p = .4341) and between per^{mm2} and per^{mm3} (p = .2425), while in all the other pairings a significant difference was detected (p at least < .0068). I also substituted the raw data with the arcsin of the normalised activity in order to remove differences in the overall activity levels. The arcsin transformation reduces the interdependence of the ratio data. In these analysis, the results are the same except that the per^{mm2} - per^{mm3} interaction becomes significant, suggesting a difference between the two chimaeric genotypes.

Visual exploration of the LD activity patterns depicted in figure 5.8 shows that, although significantly different in the ANOVA, per^{+} 13.2 and per^{mm3} display a similarity in their activity profile, with most of the fly's total activity concentrated around the peaks. In contrast, the activity of per^{mm1} and per^{mm2} notably differs from that of wild type, being more evenly distributed throughout the circadian cycle, with a higher proportion of night-time activity.

DISCUSSION

Rescue of circadian rhythmicity of locomotor activity

Plasmid pMPS1, which contains the whole *per* coding sequence of *D. pseudoobscura* fused to the *D. melanogaster* 5' non coding region, rescues only weakly the arrhythmic phenotype of *D. melanogaster* per^{01} mutants (Petersen *et al.*, 1988). Using our own statistical analysis about 45% of per^{mps1} transformant flies are rhythmic (Peixoto *et al.*, 1998), compared to Petersen *et al.*'s analysis which only found a 10% rescue. Irrespective of the statistical method used, rescue due to the *D. pseudoobscura per* is poor, which is not surprising given the divergence detected in the *per* gene product between these two species (Colot *et al.*, 1988). Most of the differences observed between the Per proteins, lies in the C-terminal half of the sequence, but the results obtained with the chimaeric construct pMPS3 (Peixoto *et al.*, 1998), which encodes the

first half of the Per protein of *D. melanogaster* and the second half of *D. pseudoobscura*, clearly shows that the high level of divergence at the C-terminal half is not the reason for the poor rescue obtained with the *D. pseudoobscura* gene. Consequently attention is drawn to the *D. pseudoobscura* N terminus in which is located the PAS domain, the region of Per which shares similarity with some transcription factors (Nambu *et al.*, 1991; Hoffman *et al.*, 1991; Burbach *et al*, 1992) and which is implicated in protein-protein interactions with the Per partner, Tim (Gekakis *et al.*, 1995; Zeng *et al.*, 1996; Myers *et al.*, 1996; Lee *et al.*, 1996; Saez and Young, 1996).

The surprising result obtained with the per^{mm1} *Musca per* construct is consistent with this interpretation. The construct encodes the whole *Musca* Per protein and restores fully the periodicity in the locomotor activity of per^{01} flies (see table 5.2). The newly established rhythm in per^{mm1} transformants is about two hours shorter than that of both *D. melanogaster* per^{+} and *M. domestica* itself (see Chapter 4). The divergence between the second half of the *D. melanogaster* and *Musca* proteins is even higher than that between the C-terminal half of *D. melanogaster* and *D. pseudoobscura* Per (see Chapter 3). Surprisingly though, the PAS domain of *Musca* is more similar to that of *D. melanogaster* than the latter is to *D. pseudoobscura* (see Chapter 3). The role of PAS as a dimerization domain, promoting interaction with the partner molecule Tim, seems to be firmly established (Huang *et al.*, 1993; Zeng *et al.*, 1996; Myers *et al.*, 1996; Lee *et al.*, 1996). The ability of the *Musca* Per protein to direct efficient rescue of rhythmicity in *D. melanogaster* aperiodic mutants in contrast to *D. pseudoobscura* Per suggests that PAS-mediated, Per-Tim interactions can take place in an almost normal fashion between *Musca* Per and the host *D. melanogaster* partner, Tim, in per^{mm1} flies. It follows that because of the higher level of divergence displayed by the *D. pseudoobscura* PAS from *D. melanogaster*, the interaction between Per and Tim could be much less efficient in per^{mps1} transformants. This simple but compelling explanation may represent an example of intermolecular coevolution between the Per and Tim interacting domains. The cloning and sequencing of the *D. pseudoobscura* and *Musca tim* gene will support or refute this idea. Theoretically, if coevolution is indeed occurring we might expect the phylogenetic tree of the Tim dimerisation domain to be similar to that for PAS, with *Musca* clustering more closely to *D. melanogaster* than *D. pseudoobscura*. In addition, direct investigation

of intermolecular interactions between different species PAS and Tim using immunochemical methods could also be brought to bear on this question.

The shorter period obtained in per^{mm1} flies (see table 5.2) might be due to stronger interactions between *D. melanogaster* Tim with *Musca* PAS compared to its conspecific partner. In this case the Per-Tim complex would reach physiological concentration sooner and possibly translocate earlier into the nucleus (Curtin *et al.*, 1995). The corresponding per^{mps1} transformants have longer period (Petersen *et al.*, 1988; Peixoto *et al.*, 1998), suggesting weaker interactions. However it must be noted that the *A. pernyi* gene can mediate rescue of circadian rhythmicity in *D. melanogaster*, albeit poorly (only 23% of the transformants in one line show the rhythmic phenotype), and that the transformant's period is 21 h (Levine *et al.*, 1995). Similarily, the period of per^{01} flies carrying one copy of the *D. yakuba per* gene is about 22 h long (M. Couchman, personal communication). Different mechanisms could account for these similar behaviours. For example the *A. pernyi* protein (see fig. 3.9a) contains a number of amino acid changes in the "short period domain", a region which surrounds the per^{S} site and in which mutations consistently shorten the periodicity (Rutila *et al.*, 1992; Baylies *et al.*, 1992). Whether or not mutations in the short period domain affect PAS-Tim interactions, has not been investigated.

Another possibility is that the removal of the large intron 2 from the transgene (for details of per^{mm1} construction, see Chapter 2) affects the processing of the primary transcript by accelerating it. Indeed the peak of the per^{mm1} transcript's cycle is advanced about 4 h with respect to wild type per^{+} (see fig. 6.3 and Hardin *et al.*, 1990) but this could be due, indirectly, to the speeding up of the clock in per^{mm1} flies rather than indicating different kinetics of RNA processing. Biochemical analysis of the per^{mm1} transformants are required in order to address these hypothesis.

Free running locomotor activity periods were subjected to ANOVA. The between-lines variability appears to be significant for genotypes per^{+} 13.2 and per^{mm1} (see table 5.3). Therefore these lines were not initially pooled within their respective genotype. There are not substantial differences in the result of the statistical test when line per^{mm1}-34c is used instead of per^{mm1}-34a (compare table 5.4 and 5.5), so despite a possible position effect in one of the two lines, they are equally informative when compared to the other genotypes. Rather different is the situation regarding the four

control lines. Three of them appear to behave in a comparable way while the fourth (line 116a) is statistically different. ANOVAs have been performed using both groups as a representative sample for genotype per^{+} 13.2 (lines 2a, 17a and 34a, and line 116a) and the results are contradictory. In this case, more confidence is placed in the results obtained when pooled lines 2a, 17a and 34a are used in the test. Line 116a, behaves anomalously with respect to the other "sister lines", possibly indicating a position effect which diminishes the line's value as a control.

Peixoto *et al.* (1998) reported strong indications of an intramolecular coevolution between the Thr-Gly repeat and the ca. 60 amino acids fragment located at its immediate 5' terminus. The dramatic improvement in rescue ability of per^{mps3} compared to per^{mps2} was not replicated with the equivalent *D. melanogaster/Musca* chimaeric constructs per^{mm3} and per^{mm2}. The level of rescue of the two constructs is approximately equivalent and so is the period length of the transformants (see table 5.2). There is a slight difference in the temperature compensation in which per^{mm3} is more similar to the per^{+} 13.2 transformants (see fig. 5.6). The period of per^{mm2} transformants increases by raising the temperature from 25° to 29° C, while the same temperature change has the effect of decreasing the period length in per^{mm3}, and in per^{+} 13.2 transformants (fig. 5.6). The temperature effect depicted in figure 5.6 (slight fall in period length between 25° and 29° C) seems to be a characteristic of the *D. melanogaster* circadian behaviour since it is seen also in many wild strains of *D. melanogaster* (Sawyer *et al.*, 1997). The fact that per^{mm3} flies have a more wild type-like behaviour than per^{mm2} transformants, is in accordance with Peixoto's findings, but it is difficult to say how much of this small difference might be attributed to a real Genotype effect rather than being an artefact due to position effect in these lines. It must be noted though, that since the number of mutations in the flanking region correlates with the logarithm of the length of the repeat region, we would expect a highly mutated flanking region for the pair *D. melanogaster-D. pseudoobscura* (repeat lengths 66-209, respectively), and a more similar flanking region for the *D. melanogaster-Musca* (66-9) pairing. Therefore it is possible that the structural divergence of the flanking fragments in the latter pair is not so high as to cause dramatic effects at the functional level. Unfortunately the results obtained with the *D. melanogaster/Musca* transformants are not completely comparable to those obtained with the *D. melanogaster/pseudoobscura*

chimaeric genes. The chimaeric junction of the per^{mm2} transformants was placed four amino acids upstream with respect to the junction in per^{mps2}. Two of these four amino acids are mutated in *Musca* with respect to *D. melanogaster*, therefore the possibility that the replacement of these four amino acids of N-terminal *D. melanogaster* flanking sequence with the equivalent *Musca* fragment had improved the functionality of the chimaeric protein cannot be ruled out. In other words, if the per^{mm2} junction was in an identical position to that of per^{mps2}, a stronger defect may have occurred although it seems quite unlikely that any disruption would be as severe as that observed in per^{mps2}.

Huang *et al.* (1995) reported that the correct balance of intra- (between PAS and a region located downstream, the so-called C-domain) and inter- (between different PAS domains) molecular interactions, maintain temperature compensation. Both intra- and intermolecular interactions increase in strength with an increase in temperature but, because they do so to the same extent, there will not be a net effect of temperature on the ratio between the two different kinds of interactions. Mutations in either of these two Per domains, may change the the temperature constant of one type of interactions, with noticeable consequences on overall temperature compensation. Two considerations suggest that the C-domain interacting with PAS could include the Thr-Gly region or a closely associated segment: first, the PAS domain interacts weakly with the Thr-Gly region in a yeast two-hybrid system (I. Townson, personal communication) and second, mutations in the Thr-Gly region affect the temperature compensation of the circadian clock (Sawyer *et al.*, 1997; Peixoto *et al.*, 1998). Perhaps the correct orientation of the Thr-Gly domain with respect to PAS would be guaranteed by the N-terminal flanking fragment, therefore a three way coevolution between PAS, Thr-Gly and the flanking fragment could be envisaged. Assuming that the PAS domain is the major determinant of whether an interspecific transformed Per protein rescues rhythmicity, then the result obtained with per^{mm1} transformants indicates that the structure of the *Musca* and *D. melanogaster* PAS region could be quite similar (see also phylogeny in Chapter 3). Thus, any interaction between the *D. melanogaster* PAS domain and the *Musca* Thr-Gly region could be relatively normal, thereby giving reasonably good rescue in per^{mm2} chimaeric transformants.

Pattern of activity

The pattern of circadian activity displayed by the MM series of transformants can also reveal something about the functionality of the various chimaeric proteins. per^{mps1} restores circadian rhythms quite poorly, yet it seems to give to the transformants an activity profile similar to that characteristic of *D. pseudoobscura* when on a *D. melanogaster* per^{+} background, suggesting a dominant effect of the transgene (Petersen *et al.*, 1988). A more detailed analysis of the DD pattern of activity of *D. melanogaster* and *D. melanogaster/D. pseudoobscura* chimaeric transformants in a per^{01} background has been carried out by J. M. Hennessy. The activity profile of *D. melanogaster* wild type as well as per^{01} transformants carrying either the *D. melanogaster* per^{+} or the *D. melanogaster-pseudoobscura* per^{mps1} and per^{mps3} chimaeric constructs (per^{mps2} flies do not show rhythmic behaviour) were examined at different temperatures. The results clearly show that the profile of wild type *D. melanogaster* at 18°, consists of a single peak of activity occurring 3-4 hours after subjective dawn, becoming bimodal at 25°, when a shift of this peak towards later hours of the subjective day occurs (1-2 hours before the beginning of the subjective night) and with the appearance of a weaker peak at the subjective dawn. A further shift of the evening peak is observed at 29°. The same patterns of activity is encountered in the per^{+} and the per^{mps3} transformants, while *D. pseudoobscura* flies and per^{mps1} transgenics display a pattern of activity completely unimodal at the three temperatures analysed (Petersen *et al.*, 1988; Hennessy, personal communication). The N-terminal half of the Per protein thereby seems to be the focus for the determination of the species-specific pattern of activity.

A situation reminiscent of per^{mps1} is observed with the per^{mm1} chimaeric construct, which rescues the rhythmic behaviour of per^{01} flies and imparts to the transformants a pattern of DD locomotor activity which mimics the one of *Musca*. It is quite difficult to say though, whether the unimodal profile characteristic of per^{mm1} behaviour (see fig. 5.7) is the result of the *Musca* Per protein conveying to the transformants a *Musca*-like locomotor activity, or rather is simply the outcome of a disruption of the normal activity of *D. melanogaster*. Perhaps this is the most likely explanation since, when analysed in LD, the activity of per^{mm1} flies is quite different from that of *Musca*.

The behaviour of the two other transformant genotypes in DD is similar to that displayed by wild type *D. melanogaster*. The flies locomotor activity profile is unimodal at 18° C and becomes bimodal by raising the temperature. At 25° several flies start to be active in the morning, but they are not necessarily synchronised. The result is a widening of the existing peak on the plot of the average activity from different flies, but at 29° the morning peak is evident in most of the flies, making it visible also in the plot of the group. Therefore, even though a faint difference between per^{mm2} and per^{mm3} transformants is appreciable in the plots of activity at 25°, it is difficult to envisage a real difference in the activity profile between the two genotypes.

Surprisingly, a clear difference between the two genotypes is apparent in LD, when more homogeneity of behaviour would be expected given the twice-daily resetting operated by the switching on and off of the lights. In these conditions, per^{mm2} shows broader peaks of activity, while in per^{mm3} flies the peaks are more narrow, much more similar in fact to those of wild-type *D. melanogaster*. It appears that the relatively small sequence difference between per^{mm2} and per^{mm3} chimaeric genes, although not representing the difference between functionality and non-functionality as in the case of the corresponding *mps* constructs (Peixoto *et al.*, 1998), is nonetheless causing some disruptions in normal Per functions. In per^{mm2}, the *Musca* Thr-Gly repeat and the adjacent *D. melanogaster* 5' flanking sequence code for a protein which rescues the arrhythmicity of per^{01} to wild-type values, but which imparts to the host fly a different activity pattern, compared to that typical of wild-type *D. melanogaster* (see fig. 5.7). In the case of the per^{mm3} gene, the replacement of the *D. melanogaster* 5' flanking sequence with that of *Musca* results in a profile of activity nearly identical to that of wild-type *D. melanogaster*. These results thus support the intramolecular coevolution theory proposed by Peixoto *et al.*, (1998) in that normal functioning of a chimaeric *per* transgene appears to be re-established when Thr-Gly repeat and adjacent 5' sequence derive from the same species.

It also appears from the transformants LD data, that a host-specific first half of the chimaeric protein is required for expression of a "normal" profile of activity. Possibly this part of the *per* gene carries species-specific behavioural instructions, or perhaps the delicate interactions between this portion of Per and other components of the output machinery are altered in heterospecific interactions. The latter hypothesis seems more

likely, especially after comparing the LD profile of activity of *Musca* and per^{mm1} transformants (fig. 4.4 and 5.8). In the former, light seems to trigger the locomotor activity. Flies become active with the onset of light and reduce their movements at lights-off. In the *D. melanogaster* per^{mm1} transformants, the activity anticipates the onset of lights of several hours, and most of the locomotor activity seems to occur during the night.

Rather dramatic is the difference between the per^{mm1} and the per^{mm2-3} transformants in DD. It is difficult to say that the *Musca* Per protein confers to the transformants the behaviour typical of the donor organism, rather than simply indicate an abnormal functionality of the *Musca* protein in the *D. melanogaster* host. Nevertheless the patterns of activity of per^{mm1} flies show that one single protein can change complex aspects of behaviour, such as the daily profile of locomotor activity. The functions assumed by Per, must therefore be more complex than that of simply regulating the period of the circadian oscillator, since it appears to be involved in the determination of the behavioural output patterns.

CHAPTER 6

Spatial and temporal expression of the Per protein in *M. domestica* and in *D. melanogaster* transformants

Spatial and temporal expression of the Per protein in *M. domestica* and in *D. melanogaster* transformants

INTRODUCTION

Spatial and temporal expression of the Per protein has been studied in *D. melanogaster* by a variety of methods, including *in situ* hybridisation (James *et al.*, 1986), immuno-histochemistry (Siwicki *et al.*, 1988) and reporter protein staining (Liu *et al.*, 1988; Zerr *et al.*, 1990). An antibody raised against a synthetic peptide from the 14 amino acids surrounding the per^S site (Per S peptide - amino acids 605-618), detects Per expression in various tissues of the fly at different developmental stages. Thus, anti-S immunoreactivity is present in the embryo at the level of the ventral nerve cord, in the lateral brain of the pupa, and in the head, the thoracic nervous system, the gut and the ovary of the adult (Siwicki *et al.*, 1988; Zerr *et al.*, 1990). A variety of structures in the fly head show Per expression; strong staining is detected in the photoreceptors, glial cells throughout the central brain and two groups of neurons lateral to the protocerebral neuropil (ventral lateral neurons and dorsal lateral neurons, LNv and LNd). Most efforts were directed at analysing head expression because it is here that the locomotor activity pacemaker is located (Konopka *et al.*, 1983; Ewer *et al.*, 1992; Konopka *et al.*, 1996). Immunostaining is most prominent in flies sectioned during the night and is dramatically reduced during the day (Siwicki *et al.*, 1988; Zerr *et al.*, 1990). The peak of Per expression in all the head structures, occurs towards the end of the subjective night, while the minimum level is recorded around ZT12 (Zerr *et al.*, 1990).

The spatial location of Per is also changing rhythmically. For the first part of its accumulation phase Per is confined to the cytoplasm, after which it enters the nucleus during a restricted portion of the circadian cycle, at about ZT18 (Curtin *et al.*, 1995). Cycling of expression is maintained when the flies are kept in constant darkness.

In an elegant experiment which made use of per^{01}/per^+ mosaic flies, Ewer *et al.* (1992) showed that correct expression of Per in the central brain (as indicated by a lacZ

reporter) is sufficient for the generation of locomotor activity rhythms. Both glial and neuronal cells express Per in the central brain and glial expression alone can mediate weak and long period (over 30 h) rhythms. However robust 24 h periods were observed when Per was present in at least some of the lateral neurons. Mosaic analysis could not determine whether lateral neurons expression is sufficient for normal rhythmicity, or whether glial expression is also necessary. Subsequently Frisch *et al.* (1994) isolated a behaviourally rhythmic transgenic line which only expressed *per* in the lateral neurons. This finding has been corroborated by anatomical studies of the arrhythmic *disconnected* (*disco*) mutant (Dushay *et al.*, 1989). In these flies the LN are absent (Zerr *et al.*, 1990), but in the sporadic cases where *disco* flies appear to be rhythmic, such rhythmicity correlated with the presence of a single LNv (Helfrich-Forster and Engelmann, 1995). The Per protein has been shown to translocate into the nucleus in all the structures in which it is expressed, with the exception of the ovaries (Liu *et al.*, 1992). The level of *per* mRNA does not oscillate in ovaries (Hardin, 1994), where its translation product is confined to the cytoplasm, consistent with the negative feedback loop hypothesis (Hardin *et al.*, 1990). Finally, *per* mRNA cycling (and so, presumably, Per protein cycling) was shown to occur in the male body, even though in free-running conditions the amplitude of the oscillation is much reduced in comparison to LD cycling, or to DD cycling in heads (Hardin, 1994).

Surprisingly, a rather different picture emerges from the study of *per* expression in the lepidopteran *Antheraea pernyi*, the giant silkmoth. In this insect, *per* transcript and protein expression in the central brain is confined to 8 neurons, four in each hemisphere (Sauman and Reppert, 1996). Anti-Per and Tim staining appear to oscillate in a circadian fashion, showing a peak at ZT16-22 and low levels at ZT4-8 of a LD 17:7 cycle (Sauman and Reppert, 1996). Expression of Per (as well as Tim) is restricted to the cytoplasm of these 8 neurons, none being detected in the nucleus at any time point. *per* mRNA oscillates in the silkmoth brain in phase with the protein, with high expression being recorded at ZT14-22 and low expression at ZT4-8. No *per* expression was detected in the central brain outside the 8 Per staining neurons (Sauman and Reppert, 1996). In contrast with the molecular cycles in *D. melanogaster*, no phase delay is observed between transcript and protein cycles in *A. pernyi*. A second interesting feature of the cycle in *A. pernyi* is the oscillation of an antisense *per* transcript cycling in anti-

phase to the sense *per* mRNA in the same 8 cells which display Per protein expression (Sauman and Reppert, 1996).

Per expression is also detected in the photoreceptors of the silkmoth, where Per protein oscillates with low expression between ZT12-18 and high expression between ZT20-8, and the RNA oscillates with a peak occurring at ZT16-18 and a trough at ZT6-10 (Sauman and Reppert, 1996). Nuclear localisation of Per is observed in the photoreceptor cells (Sauman and Reppert, 1996). Therefore the silkmoth shows typical *D. melanogaster*-like cycling of *per* mRNA and protein in the photoreceptors, but a completely different pattern of regulation in the brain. Even more surprising is the pattern of *per* expression in the lepidopteran embryo (Sauman *et al.*, 1996). Per and Tim are detected in four pairs of cells in the dorsolateral brain, in a location corresponding to that of the eight Per-expressing neurons of the adult brain. The staining is cytoplasmic and does not vary in intensity during the course of the day. Per expression was also detected in the fat body and in the midgut epithelium. In the former, immunoreactivity to Per is cytoplasmic and constant, while in the midgut Per is localised within the nuclei and appears to oscillate. Tim staining was not detected in fat body and midgut epithelium at any time (Sauman *et al.*, 1996). Since a circadian clock is active in the lepidopteran embryo and Per is one of its key components (injections of a *per* anti-sense oligonucleotide disrupt the circadian gating of the pharate larva hatching, Sauman *et al.*, 1996), it follows that a functional circadian clock is provided by non-cycling Per and Tim coexpressed in the same cell (in the embryo's brain), or by non-cycling Per without Tim (in the fat body), or by cycling Per without Tim (in the midgut). Whatever of these tissues we assume to be the focus of the circadian eclosion pacemaker, clearly a different mechanism from the fly's negative feedback loop must be responsible for the functioning of this oscillator in the silkmoth embryo.

The function of the *per* gene in the housefly has not yet been determined. *Musca per* is able to restore robust locomotor activity rhythms to per^{01} arrhythmic mutant *D. melanogaster* (see chapter 5) and this indicates that *per* may retain basic clock functions in *Musca*. It is thus important to investigate whether *per* expression in *Musca* shows similar features to its *Drosophila* homologue.

METHODS

Crosses

Because of the large number of flies required for the experiments described in this chapter, it has been more practical to prepare transformant lines in which the *Musca* transgene is permanently placed on a per^{01} background, rather than use the male progeny from a cross to per^{01} females, as was the case in the locomotor activity experiments.

The crosses used to prepare these stable lines are depicted in the diagram represented in figure 6.1. A final cross was used to prepare the per^{01}; per^{mml}-34a line, where the flies balanced for the transgene were interbred and only their progeny lacking the balancer were collected. This way per^{01}; per^{mml}-34a transformants are homozygous for the chimaeric transgene, while the line per^{01}; per^{mml}-34c, which is homozygous lethal, carries a single copy of the per^{mml} gene.

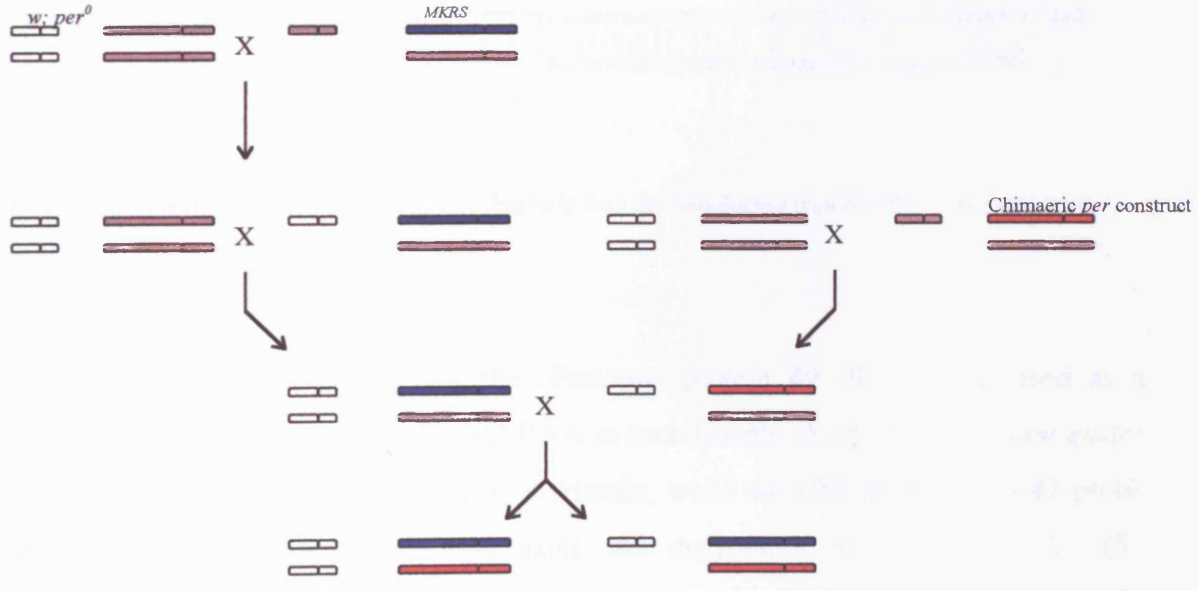


Figure 6.1. Scheme of the crosses performed for the formation of stable per^{01} transformant lines. Only the X and III chromosomes are shown. Wild type, unmarked chromosomes are represented by gray lines, in white is the *w; per*^{01} chromosome, in blue the *MKRS* balancer and in red a chimaeric construct-bearing III chromosome.

RNase protection

RNase protection assay was performed as described in chapter 2. The probe used was an antisense 210 bp *Bgl*I-*Bam*HI fragment from the *Musca per* gene (3696-3905 of the sequence represented in fig. 3.8). This fragment maps upstream to the beginning of the Thr-Gly region and contains the per^S site. Despite being localised in one of the regions of high homology of the Per protein (see chapter 3), the probe is able to discern between *D. melanogaster* and *Musca per* transcripts because of the high sensitivity of the RNase protection assay to mismatches between probe and transcript. Several substitutions are indeed observed between the *Musca* probe and the equivalent *D. melanogaster* fragment (see fig. 6.2).

```
Musca     1 AGATCTCAAATTAGAGCTACCAAATGAAAATGAACTAACCGTTTCCGAACGTGACTCGGTTATGTTGGGT 70
            ||||||||  | |||||  ||  | ||  ||  ||  | |||||  ||  ||  ||  |||   ||  |||  | ||
D.mel.      GGATCTCAAGCTGGAGCTGCCGCACGAGAACGAGTTGACCGTCTCGGAGCGGGACAGCGTGATGCTCGGC

         71 GAAATTTCACCACATCACGACTACTACGACAGTAAAAGTTCCATTGAGACTCCGCCATCGTACAATCAAT 140
            ||  |||||  ||  ||  ||||||||||  ||||||||  |||||||  |||||  |||||    |||||  ||
            GAGATTTCGCCGCACCACGACTACTATGACAGTAAGAGTTCCACCGAGACGCCGCCCAGCTACAACCAGC

        141 TGAATTACAATGAAAATTTGCAACGTTTTTTTAACAGCAAGCCTGTTACAGCACCCGTAGAAACGGATCC 210
            |  ||  ||  ||||  ||  |||  ||||||||  ||||||||||  ||  ||  ||  ||  |  ||     |||||
            TAAACTATAACGAGAACCTGCTGCGTTTTTTCAACAGCAAGCCGGTCACGGCGCCGGCGGAGCTCGATCC
```

Figure 6.2. Alignment of the *Bgl*I-*Bam*HI *Musca* fragment with the homologous region of the *D. melanogaster* gene.

Antisense RNA coding for the ribosomal protein 49 (RP49) was used as a measure of the relative amount of total RNA in each sample. A 650 bp *D. melanogaster RP49* subclone was kindly provided by P. Hardin, while the 100 bp *Musca RP49* probe was amplified from genomic DNA using the degenerate primers *RP49-5'* (5'-TGGMGIAARCCIAARGGIGTCGACAA-3') and *RP49-3'-2* (5'-TTRTGIACIARRAAYTTYTTGGATCCIGT-3'). The PCR product was then cut with the restriction endonucleases *Sal*I and *Bam*HI (the restriction sites are underlined in the primer sequence) and cloned into pBluescript II KS+ (Stratagene). The two degenerate primers have been designed in a region of the RP49 protein which shows a high degree

of conservation between the available sequences of *D. melanogaster* (O'Connell and Rosbash, 1984), *D. pseudoobscura* (Segarra and Aguade, 1993) and *D. subobscura* (Aguade, 1988).

The 300 bp antisense *per* probe (210 bp *per* fragment + 90 bp of the vector polylinker) was transcribed from the T3 promoter of the vector after linearising the plasmid with *Not*I. The 112 bp *D. melanogaster RP49* probe (57 bp *RP49* + 55 bp of the polylinker) was prepared from the *Pvu*II linearised plasmid with T3 RNA polymerase, and the 150 bp *Musca RP49* probe (100 bp *RP49* + 50 bp of polylinker) was synthesized from the *Sal*I cut plasmid using T7 RNA polymerase.

Quantitation was performed by using a phosphorimager (Molecular Dynamics) and Image Quant (Molecular Dynamics) software.

Immunohistochemistry

Musca flies were entrained in individual vials on a 12:12 LD schedule at 25° constant temperature. Flies were collected at the appropriate ZT time and their brains immediately dissected in physiological saline. The brains were then fixed in 4% formaldehyde for 24 h and then stored in 0.1M phosphate buffer (pH 7.4). Prior to sectioning, the brains were transferred to 30% sucrose for at least two hours. The brains were embedded in OCT mounting media and sectioned (10 μm) at -20° C. The slides with the brain sections were rinsed in PO_4 buffer and then pre-incubated for 1 h in PBS + 0.1% BSA + 0.3% Triton-X100 + 3% goat serum. Antibody incubation was performed for 1 h at room temperature with the proper dilution of 1° antibody in PBS + 0.1% BSA + 0.03% Triton-X100 + 3% goat serum. After three 10 min. washes in PBS + 0.1% BSA + 0.03% Triton-X100, the slides were incubated with the fluorophore-conjugated 2° antibody (1 h at RT) and again washed (3x) with PBS + 0.03% Triton-X100. Sections were then incubated for 10 minutes in PBS + 2 μg/ml DAPI (4,6-diamidino-2-phenylindole, dilactate), rinsed 10 min. in PBS and mounted in fluoromount medium + 25 mg/ml DABCO (1,4-diazabicyclo [2.2.2]-octane).

The work was initially carried on in the laboratory of Kathy Siwicki (Swarthmore College, Philadelphia, USA) with the help of J. Noveral. Subsequently the sectioning and immunostaining were performed in Leicester in collaboration with H. Dobbs.

The 1° antibodies used in the different phases of this work were:
a) ab mix. It consists of a 2:1 mixture of the monoclonal antibodies 10C3 and 12H9. 10C3 was raised by immunising a mouse against two fragments of *D. melanogaster* Per (fragment A, from the N-terminus of the protein to the end of the PAS region, and fragment B, from the Per-S region to the C-terminus) plus a fragment of *Manduca sexta* Per (from the beginning of the PAS region till the Per-S region). 10C3 is believed to recognise an epitope localised within the Per-S region, since it is able to recognise both *D. melanogaster* fragment B and *M. sexta* fragment. The monoclonal antibody 12H9 was raised against the second half of the *D. melanogaster* Per protein.
b) α-Per 978. This is a polyclonal antibody raised against the whole *D. melanogaster* Per protein expressed in a baculovirus system, kindly provided by J. C. Hall.

2° antibodies; a goat anti-mouse Cy3-conjugated antibody was used for the part of the work performed in Philadelphia, while a goat anti-rabbit FITC-conjugated antibody was employed in Leicester.

Western blotting

Protein electrophoresis and blotting was carried out as described in chapter 2. For the immunodetection of Per, the rabbit polyclonal α-Per 978 was used, while the rat polyclonal α-Tim 209, a gift from M. Young, was employed for immunodetection of the Tim protein. The working dilution for α-Per 978 was 1:10000, while for α-Tim 209 a 1:1000 dilution ratio was employed. Horseradish peroxidase-conjugated secondary antibodies were purchased from Sigma.

For quantification, the film exposed to the chemiluminescent detection reagents was digitised on a scanner, and the image analysed with the NIH Image (Research Services Branch of the National Institutes of Health) software.

Expression of GST-Per fusion protein

A portion of the *Musca per* gene, from nucleotide position 3067 to 3997 of the sequence in fig. 3.8, was amplified from *Musca* cDNA using primer *muexp-u* (5'-TGAATTCGCTATCCCGAAAAAC-3') and primer *muexp-d* (5'-AGGTCGACTGCCTCCACTATCC-3'), yielding a 868 nucleotide fragment. Some mismatches have been incorporated into the primers in order to provide suitable cloning sites (*Eco*RI and *Sal*I, underlined) and one extra nucleotide has been added in the upstream primer *muexp-u* for correcting the frame of the *Musca* fragment which was subsequently subcloned into pGEX-4T3 (Pharmacia). The plasmid was used to transform *E. coli* cells strain Top10F' (Invitrogen). Sequence analysis of the clone revealed that an A to G point mutation was introduced in the PCR amplification, causing a Thr to Ala change at amino acid position 399.

Expression of the glutathione S-transferase (GST) - *Musca* Per fusion was accomplished by growing the bacterial cells to an A_{600} = 0.5 and subsequently inducing for variable time with 1 mM IPTG. Cells were then pelletted and resuspended in PBS + 1% Triton X-100 to give a final A_{600} = 20. Ten µl aliquots were run on an SDS-PAGE and transferred on nitrocellulose.

Immunoprecipitation of Per complexes

200 µl of α-Per 978 antiserum were covalently bound to 110 µl of protein G-coupled Gammabind Sepharose (Pharmacia) as in Harlow and Lane (1988). The antibody and the Sepharose beads were incubated for 1 h at room temperature under gentle rotation. Beads were then washed twice with 1.5 ml of 0.2 M sodium borate (pH 9.0) and resuspended in 1.5 ml of 0.2 M sodium borate (pH 9.0). After the addition of 8 mg of dimethyl pimelimidate, the beads suspension was mixed for 30 min. at room temperature. The slurry was subsequently washed once in 0.2 M ethanolamine (pH 8.0), incubated with gentle mixing for 2 h at room temperature and finally washed once in PBS.

Two aliquots of the immobilised α-Per were incubated overnight at 4°C, with equivalent amounts (in term of total protein) of head protein extracts from per^{01}; per^{mm1}-34c collected at ZT6 and ZT18.

After incubation, the protein extract (referred to as D Ex, depleted extract) was removed by centrifugation and the beads were washed twice with 1.5 ml of PBS and rinsed with 0.5 ml of pre-elution buffer (10 mM PO_4 pH 8.0). The proteins were then eluted with 0.3 ml of elution buffer (0.1 M glycine pH 2.2), TCA precipitated and loaded onto an electrophoresis gel.

RESULTS

Expression of *per* RNA in heads of *D. melanogaster* transformants and *M. domestica*

RNA fragments protected by the *Musca per* antisense RNA probe fluctuate in abundance in *D. melanogaster* per^{mm1} transformants. *Musca per* RNA cycling was observed both in transformants carrying a copy of the wild-type *per* gene (per^{+}; per^{mm1}) and in flies where the transgene is in a per^{01} background (per^{01}; per^{mm1}). Levels of *per* were low at ZT0 (lights on), started to rise soon after this time and reached a peak at about ZT12 (see fig. 6.3). After ZT12, *Musca per* expression started to decline. Both lines show this general trend of *per* expression, although peak *per* levels in per^{+}; per^{mm1} transformants (expressed as a ratio *per/RP49*) are about twice those observed in per^{01}; per^{mm1} flies, while trough values are roughly similar in the two lines[1] (see fig. 6.4). The amplitude of the cycle is therefore 5-fold in per^{+}; per^{mm1} and 2.5-fold in per^{01}; per^{mm1} flies.

Compared to wild-type per^{+} cycling, where the trough and peaks are recorded at ZT4 and ZT15, respectively (Hardin *et al.*, 1990), the pattern of *Musca per* expression in per^{mm1} transformants is shifted towards earlier times, a situation reminiscent of per^{S} mutants (Hardin *et al.*, 1990).

[1] Comparison of the absolute levels of *per* transcript between the two lines has been made possible by the fact that they were both assayed in a single experiment, therefore the same specific activity and concentration of probes were employed.

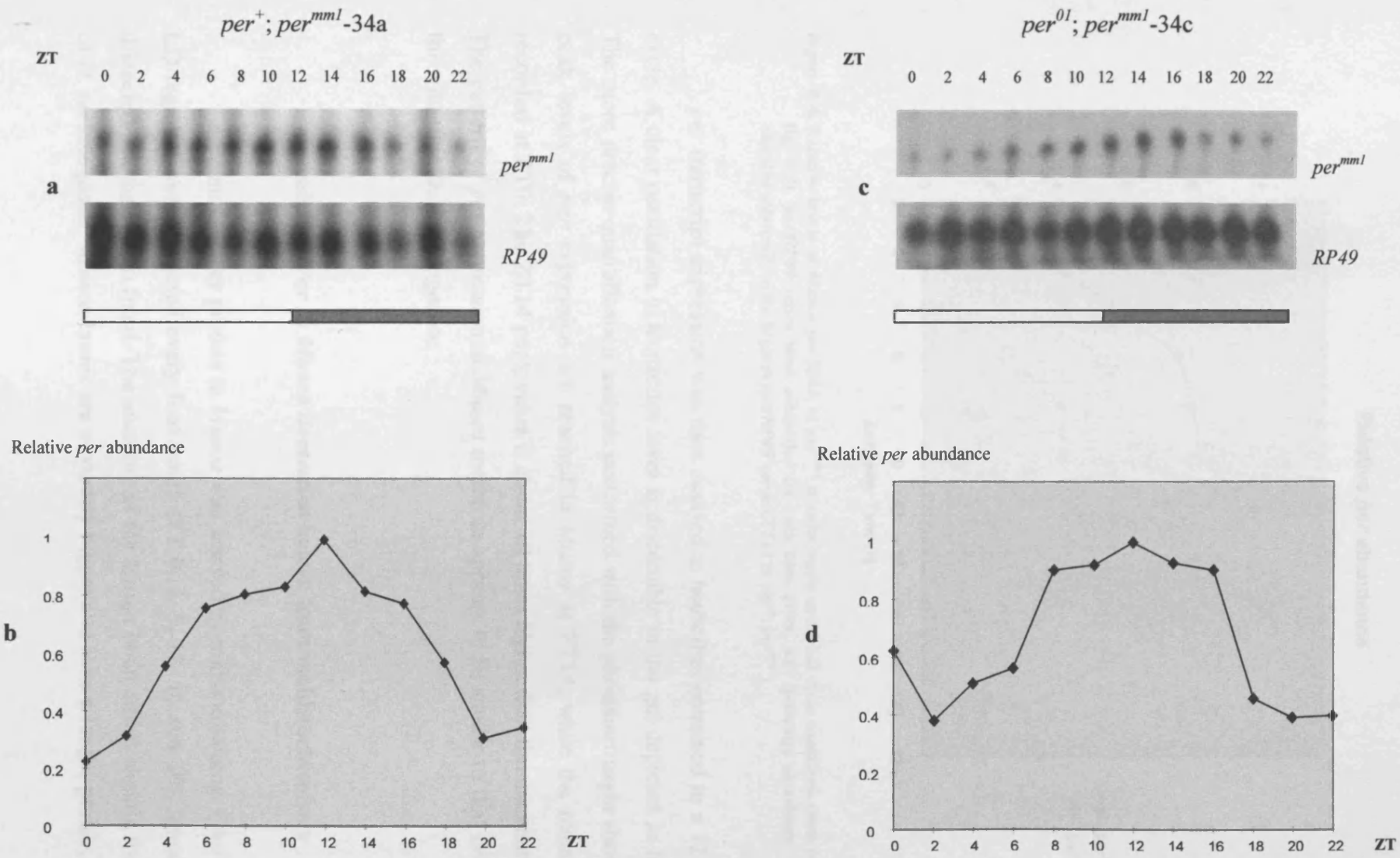


Figure 6.3. RNase protection assay (RPA) of the per^{Mml} gene on *D. melanogaster* transgenic lines. **a**, **c**: autoradiograph of the RPA of per^{+}; per^{Mml}-34a and per^{01}; per^{Mml}-34c, respectively. *per* levels appear on the top while *RP49* levels are on the bottom. **b**, **d**: quantitation of the RPA displayed. The *per/RP49* ratio was calculated for each time point and the highest value was arbitrarily assigned the value 1.

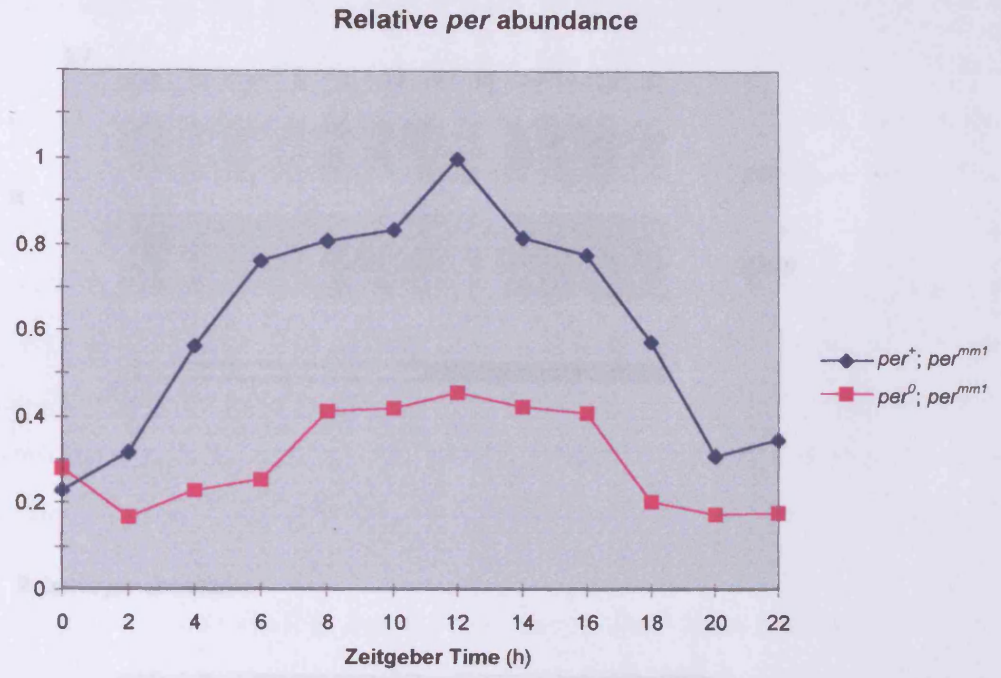


Figure 6.4. Relative levels of *Musca per* RNA in per^{mm1} transformants under LD cycle conditions (data from fig. 6.3). *per/RP49* ratios were calculated for each time point, and transcript abundance was adjusted relatively to the highest *per/RP49* value (ZT12 in per^{+}; per^{mm1})

per transcript expression was then assayed in houseflies entrained to a 12:12 LD cycle. A clear oscillation in transcript level is discernible in the gel depicted in fig. 6.5. The more precise quantification analysis performed with the phosphorimager shows that peak levels of *per* expression are reached in *Musca* at ZT14, while the minimum is recorded at ZT0. The ZT14 peak value is about 10 times higher than the minimum level. The pattern of *per* expression in *Musca* therefore appears to be similar to that described for wild-type *D. melanogaster*.

Expression of Per in *Musca domestica* heads: immunohistochemistry

Expression of Per protein in *Musca* was assayed by immunostaining. Flies from a LD regimen were collected every four hours (ZT 0, 4, 8, 12, 16 and 20), immediately dissected and their brain fixed. The anatomy of the *Musca* brain differs slightly from that of *D. melanogaster*. *Musca* brains are obviously bigger and have a longer profile, due to

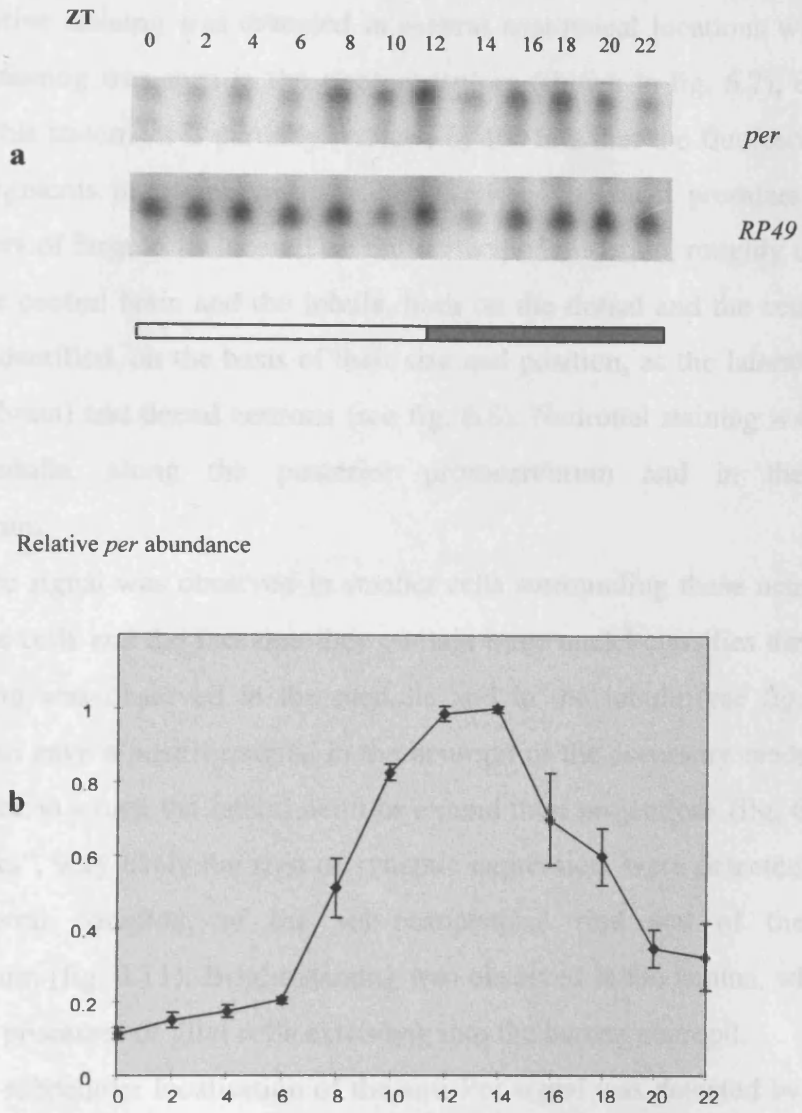


Figure 6.5. Levels of *per* transcript in *Musca*. **a**: RNase protection of the *Musca per* transcript; top: *per*; bottom: *RP49*. **b**: the quantitation of two independent RPA assays is reported as an average of the two set of values; error bars are shown in the chart.

a large extension of the optic lobes, relative to the fruitfly (see fig. 6.6). The various brain structures though, are easily comparable between the two different species.

Positive staining was detected in several anatomical locations within the *Musca* brain. No staining was seen in the photoreceptors (shown in fig. 6.7), even though the validity of this statement is partially reduced by the fact that the fluorescence emitted by the photopigments interferes with the observation. The most prominent signal was in some clusters of large cells located on the surface of the brain, roughly on the boundary between the central brain and the lobula, both on the dorsal and the ventral side, which have been identified, on the basis of their size and position, as the lateral (on the ventral side of the brain) and dorsal neurons (see fig. 6.8). Neuronal staining was also observed in the medulla, along the posterior protocerebrum and in the ventro-lateral protocerebrum.

More signal was observed in smaller cells surrounding these neurons. The small size of these cells and the fact that they contain large nuclei classifies them as glial cells. Glial staining was observed in the medulla and in the lobula (see fig. 6.9). Anti-Per antibody also gave a positive signal in the neuropil of the accessory medulla, a region of the optic lobe in which the lateral neurons extend their projections (fig. 6.10), and many small "sparks", very likely the sign of synaptic expression, were detected in the neuropil of the central complex, of the sub-oesophageal rind and of the ventro-lateral protocerebrum (fig. 6.11). Bright staining was observed in the lamina, where it might be localized in processes of glial cells extending into the lamina neuropil.

The subcellular localisation of the anti-Per signal was detected by contra-staining the sections with DAPI, which specifically stains nuclei and which emits fluorescence at a different wavelength than the fluorophore conjugated to the 2° antibody. Comparison of the two patterns of staining, shows that the signal given by anti-Per is mostly cytoplasmic. In contrast with the situation described for *D. melanogaster*, nuclear expression can be excluded for the *Musca* neurons, where the nuclei are either visible as a dark shadow within the bright positive signal, or are missing from the plane of the section, giving a uniform anti-Per signal but no DAPI staining. In the glial cells, Per staining is mostly cytoplasmic but nuclear staining is observed in some large glial cells at the edge of the medulla at ZT4 and ZT8 (see fig. 6.12).

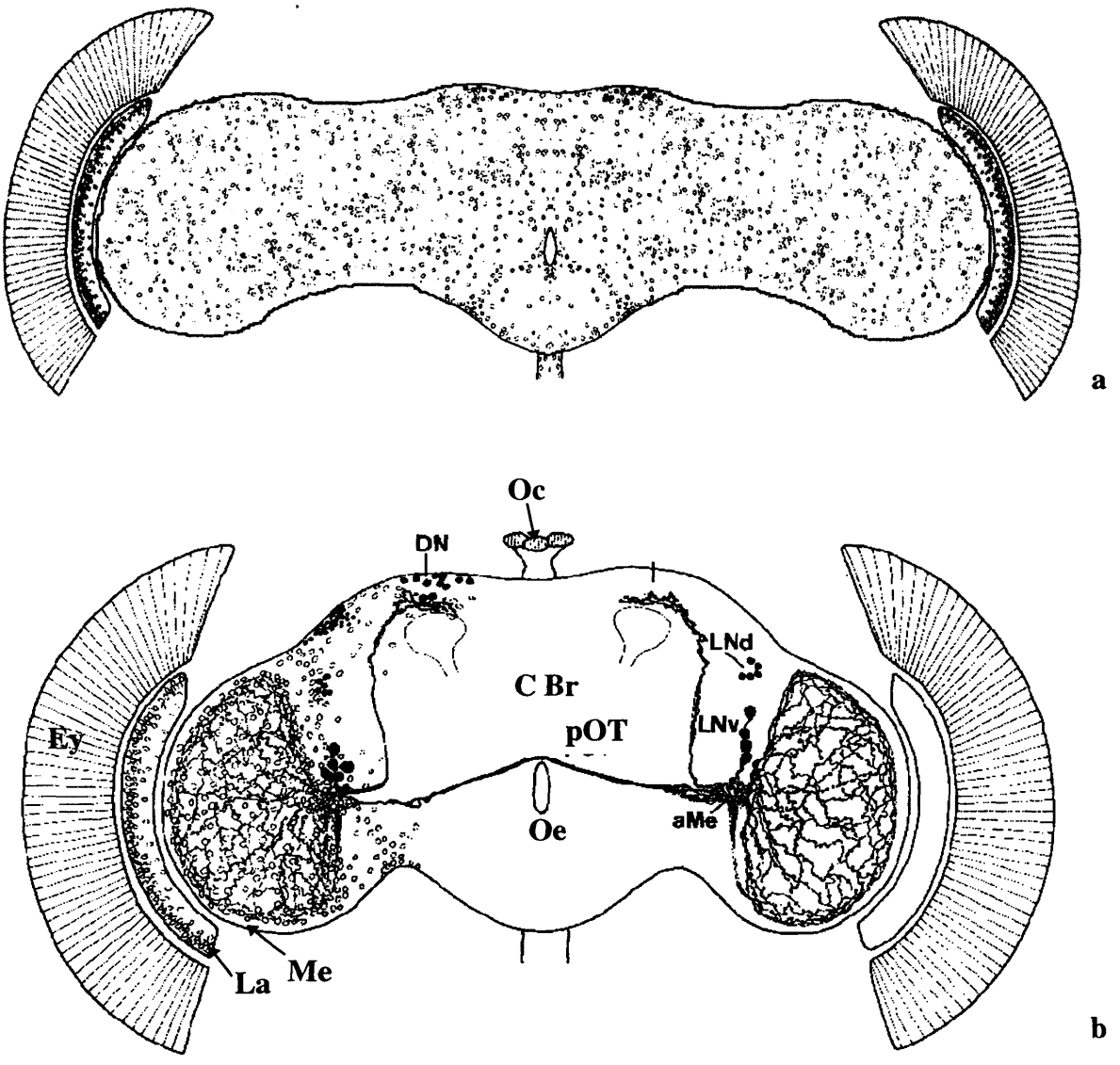


Figure 6.6. An anterior frontal view of the adult brains of *Musca* (**a**) and *D. melanogaster* (**b**). *per* expression is found in several locations of the *D. melanogaster* brain. In the compound eye (**Ey**) and ocelli (**Oc**); in glial cells dispersed throughout the whole brain, and in two groups of neurons, lateral neurons and dorsal neurons. The lateral neurons are located between central brain (**C Br**) and optic lobes and can be divided into two clusters, dorsal lateral neurons (**LNd**) and ventral lateral neurons (**LNv**); the dorsal neurons (**DN**) are located posteriorly on the dorsal side of the brain. The LNv project arborizations in the optic lobe and on the edge of the medulla (**Me**), in proximity to *per*-expressing glia, where they can potentially receive signals from the photoreceptors. Furthermore, they connect both emispheres via fibers in the posterior optic tract (**pOT**). The LNv also send arborizations to the accessory medulla (**aMe**), an appendage of the medulla which stems from the larval optic neuropil. Other structures shown are: **La**: lamina; **Oe**: oesophagus. Drawings are not in scale (modified from Helfrich-Forster, 1996).

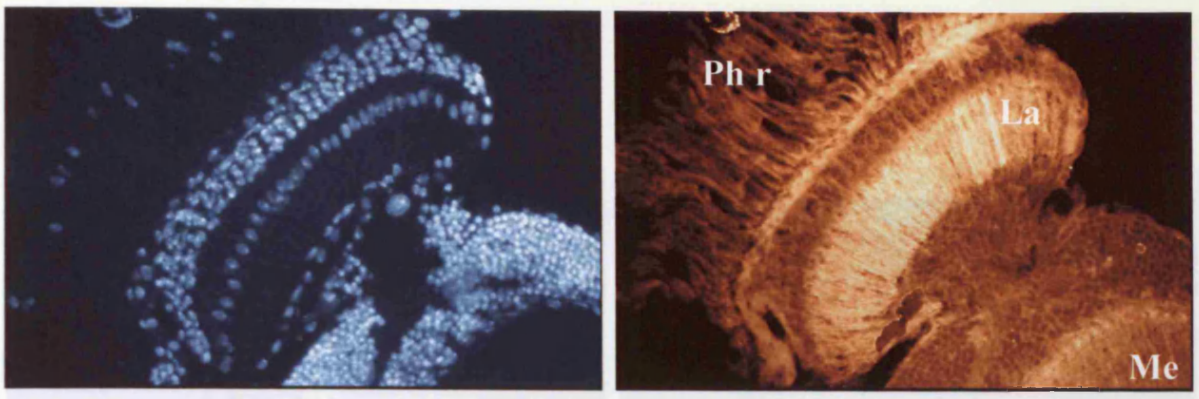


Figure 6.7. Staining observed in photoreceptors and lamina of a housefly collected at ZT12. Strong staining is detected in the lamina but no anti-Per signal is found in the photoreceptor cells. On the left is the DAPI staining which marks the photoreceptor nuclei. On the right, the same frame is seen with a different filter in order to visualise the antibody. **Ph r**: photoreceptors; **La**: lamina; **Me**: medulla. (400X).
Control slides processed with the secondary antibody only showed staining associated with the photoreceptors and none at the other structures detected with the monoclonal mix (data not shown). No attempt was made to preadsorb the primary antibody against the Per protein.

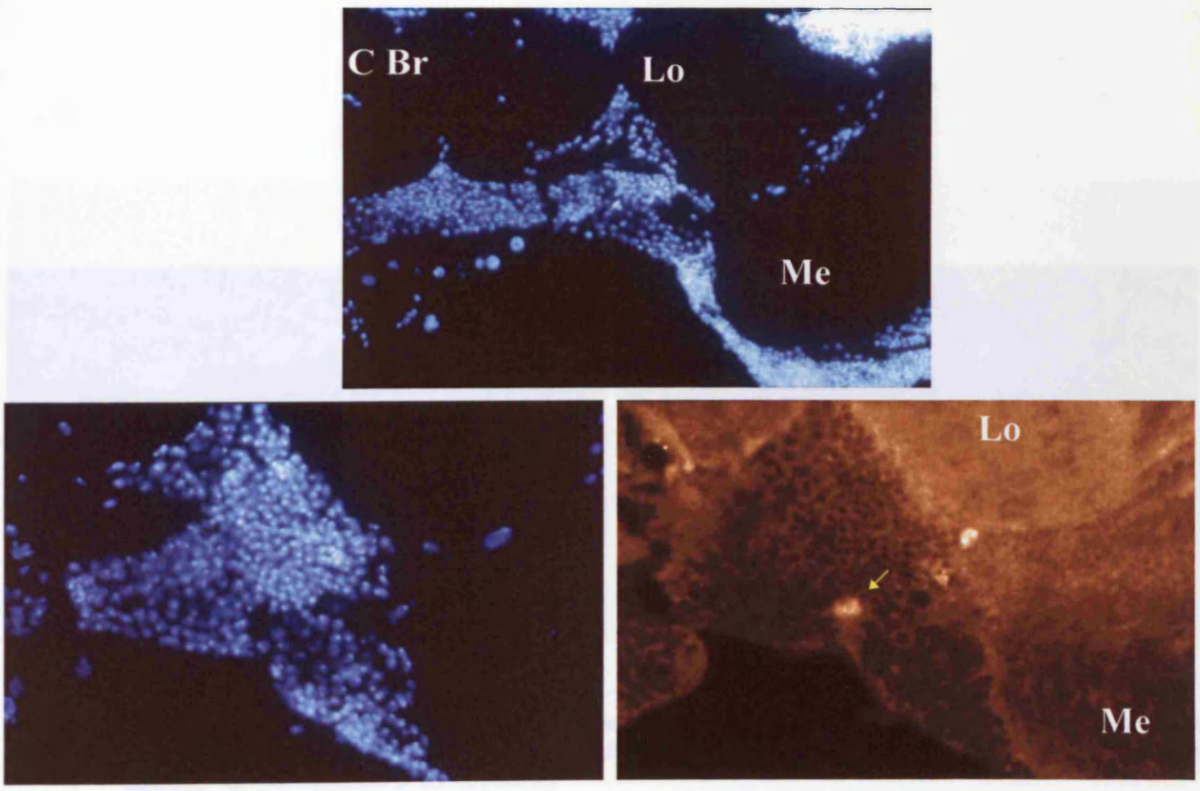


Figure 6.8. Section showing lateral neuron positive staining at ZT12 (yellow arrow). Notice that the Cy3 fluorescence is emitted in a region devoid of nuclei (no DAPI staining).**C Br**: central brain: **Lo**: lobula; **Me**: medulla. The lower magnification shows the location of the lateral neurons in respect to other major structures of the brain (bottom images: 400X; top: 200X).

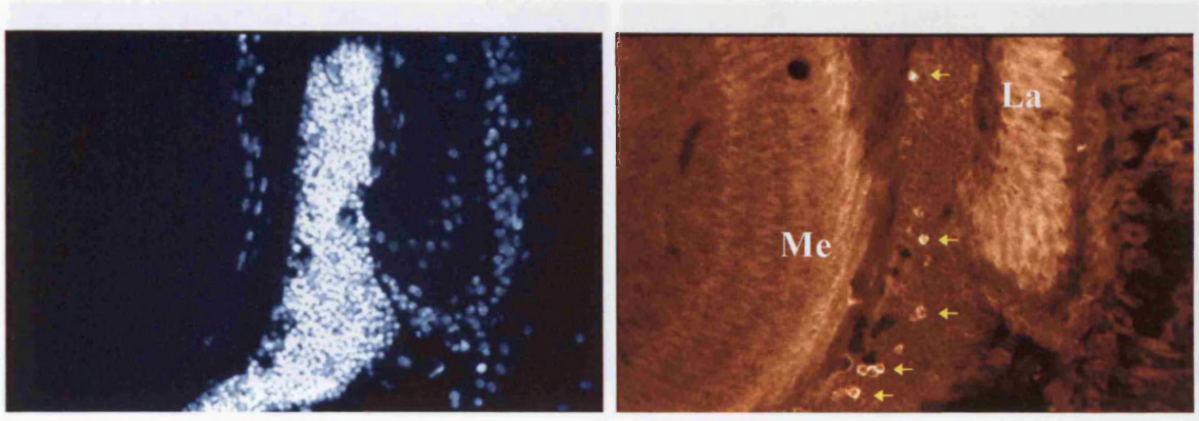


Figure 6.9. Glial staining at ZT0 (yellow arrows). The dense concentration of nuclei visible in DAPI correspond to glial cells between the lamina and the medulla. Positive antibody staining is detected in several glial cells, but in none is the staining nuclear (note the characteristic "doughnut" shape of the Cy3 signal). **La**: lamina; **Me**: medulla. Left: DAPI staining; right: Cy3 staining (400X).

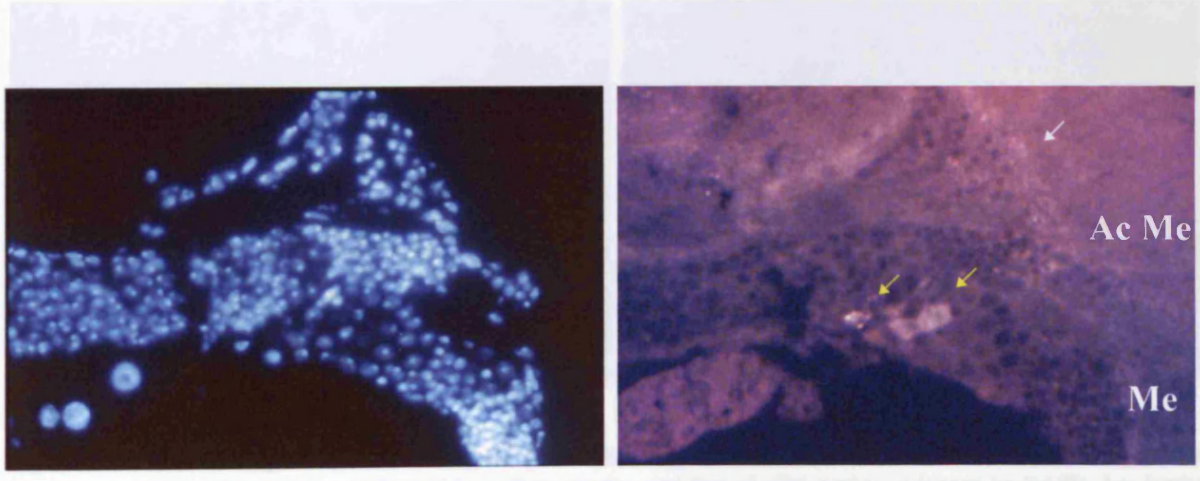


Figure 6.10. Positive signal detected in the neuropil of the accessory medulla (white arrow), a region of the fly brain where the LN extend their projections. Also visible Lateral Neuron staining (yellow arrows). **Me**: medulla; **Ac Me**: accessory medulla. Sample collected at ZT12 (400X).

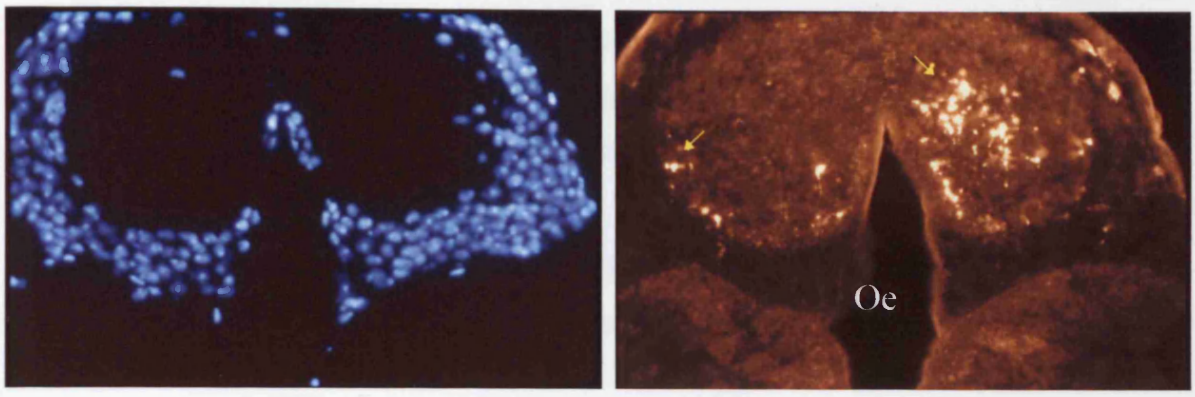


Figure 6.11. Synaptic Per expression in the central brain neuropil at ZT20. **Oe**: oesophagus. Left: DAPI staining; right: Cy3 staining (400X).

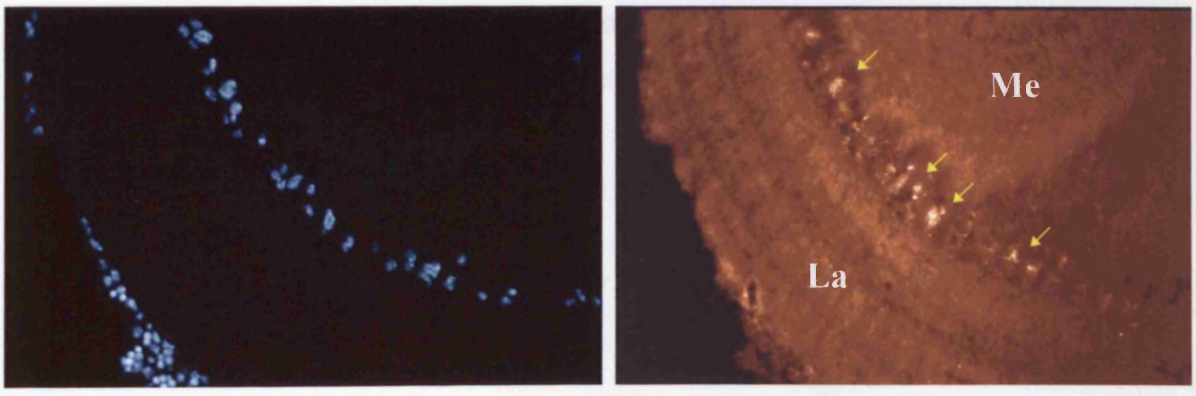


Figure 6.12. Glial staining at ZT4. Nuclear staining is visible in glial cells between lamina and medulla (confront position of positive Per staining, indicated by yellow arrows, and that of glial nuclei evidenced by DAPI). **La**: lamina; **Me**: medulla. Left: DAPI staining; right: Cy3 staining (400X).

A second feature of *Musca per* expression strongly contrasting with the one in *D. melanogaster* is the temporal pattern. The sectioned samples were collected at various times in the LD cycle in order to cover the whole circadian cycle but no differences in signal intensity were detected in the sections of brains dissected at different time points. Nuclear translocation of the Per protein seems to occur only in a subset of glial cells, at a ZT comprised between 0 and 4, while in *D. melanogaster* translocation take place at ZT18 (Curtin *et al.*, 1995).

All the results reported were obtained with the 10C3 + 12H9 antibody mix. Attempt to replicate them using α-Per 978 have so far been unsuccessful.

α-Per 978 recognises the *Musca* Per protein

A 289 amino acid fragment of the *Musca* Per protein was expressed as a fusion to GST. Immunoblotting of this fragment shows that α-Per 978 gives a positive signal which increases upon IPTG induction of the fusion peptide, while no signal is detected in controls expressing only the GST fragment (see fig. 6.13). The antibody raised against the whole *D. melanogaster* Per protein is therefore able to recognise a portion of the *Musca* homologous protein between amino acid 347 and 636, and including the second PAS repeat and the per^S site.

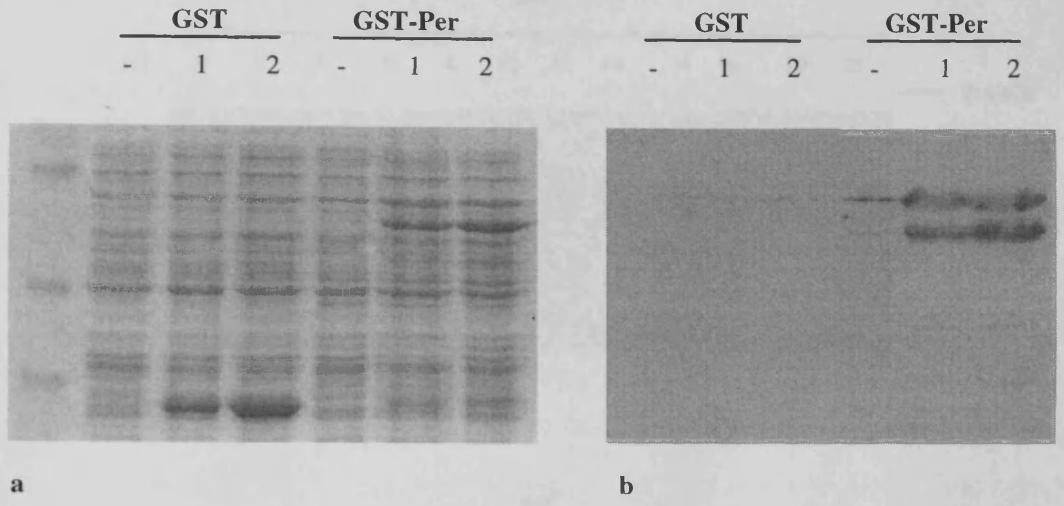


Figure 6.13. **a**: SDS-PAGE of bacterial proteins from Top10F' strain transformed with the pGEX-4T3 vector (GST) or with the pGEX-*Musca* Per construct (GST-Per). -, 1, 2: extracts from cells harvested prior, 1h and 2h after the IPTG induction, respectively. **b**: immunoblot of the gel shown in **a**, probed with α-Per 978.

Levels of Per and Tim expression in *D. melanogaster* transformant and *Musca domestica* heads: immunoblotting

Levels of Per protein were also assayed by immunoblotting. Protein extracts were obtained from fly's heads collected every two hours of a 12:12 LD cycle, run on a SDS-PAGE and blotted on nitrocellulose membrane. α-Per 978 recognises one band in per^+ flies, running with an apparent molecular weight of 180 kDa (see fig. 6.15a). The intensity of this band cycles as a function of the Zeitgeber Time of the sample. The minimum level is recorded at ZT10 while the maximum expression occurs at ZT20. The change in mobility due to phosphorylation (Edery *et al.*, 1994) is readily apparent. An additional band is detected in the 130-150 kDa region. This band, although of lower intensity than the former, cycles with the same phase as the 180 kDa band. The reported expression of Tim resembles that of Per with the exception that the decline of Tim in the light phase of an LD cycle procedes more rapidly than that of Per (Zeng *et al.*, 1996). In per^+ controls Tim reaches the maximum expression at ZT20, and the minimum at ZT6, (fig. 6.15b-c). Degradation of Tim follows a steeper profile respect to that of Per in the early phase of the circadian cycle, with the minimum level anticipating that of Per by about 4 hours (fig. 6.15c).

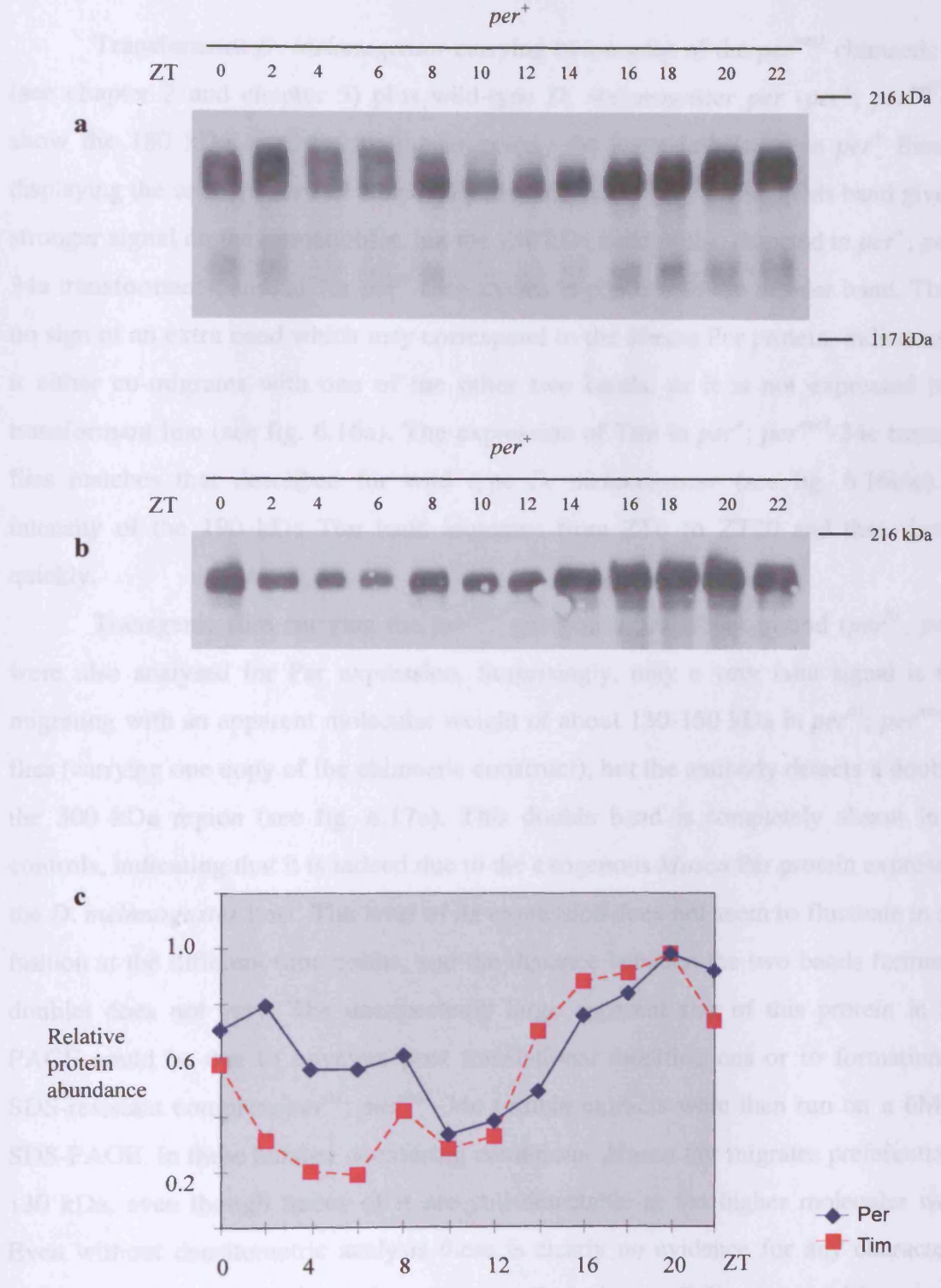


Figure 6.15. Levels of the circadian proteins Per and Tim in wild-type *D. melanogaster* flies. **a**: immunoblot of total head extract analysed with α-Per antiserum. **b**: the same samples assayed with α-Tim antibody. **c**: quantitation of the data displayed in a and b. Relative refers to the highest amount detected for each of the two proteins.

Transformant *D. melanogaster* carrying two copies of the per^{mml} chimaeric gene (see chapter 2 and chapter 5) plus wild-type *D. melanogaster per* (per^{+}; per^{mml}-34a) show the 180 kDa band fluctuating in exactly the same fashion as in per^{+} flies, and displaying the same pattern of temporal phosphorylation (fig. 6.16a). This band gives the stronger signal on the immunoblot, but the 130 kDa band is also detected in per^{+}; per^{mml}-34a transformants, and, as for per^{+} flies, cycles in phase with the heavier band. There is no sign of an extra band which may correspond to the *Musca* Per protein, indicating that it either co-migrates with one of the other two bands, or it is not expressed in this transformant line (see fig. 6.16a). The expression of Tim in per^{+}; per^{mml}-34c transgenic flies matches that described for wild type *D. melanogaster* (see fig. 6.16b-c). The intensity of the 190 kDa Tim band increases from ZT6 to ZT20 and then dampens quickly.

Transgenic flies carrying the per^{mml} gene on a per^{01} background (per^{01}; per^{mml}) were also analysed for Per expression. Surprisingly, only a very faint signal is found migrating with an apparent molecular weight of about 130-150 kDa in per^{01}; per^{mml}-34c flies (carrying one copy of the chimaeric construct), but the antibody detects a doublet in the 300 kDa region (see fig. 6.17a). This double band is completely absent in per^{+} controls, indicating that it is indeed due to the exogenous *Musca* Per protein expressed in the *D. melanogaster* host. The level of its expression does not seem to fluctuate in cyclic fashion at the different time points, and the distance between the two bands forming the doublet does not vary. The unexpectedly large apparent size of this protein in SDS-PAGE could be due to covalent post translational modifications or to formation of a SDS-resistant complex. per^{01}; per^{mml}-34c protein extracts were then run on a 6M urea SDS-PAGE. In these harsher denaturing conditions, *Musca* Per migrates preferentially at 130 kDa, even though traces of it are still detectable at the higher molecular weight. Even without densitometric analysis there is clearly no evidence for any characteristic cycling (see fig. 6.18). This result seems to exclude the possibility that the *Musca* protein is subjected to extensive post translational modifications, but it is still not clear whether the increase in molecular weight is due to self-aggregation or to formation of an heterocomplex. No Tim protein is detected in the 300 kDa region, probably indicating the exclusion of this protein from the complex (see fig. 6.17b).

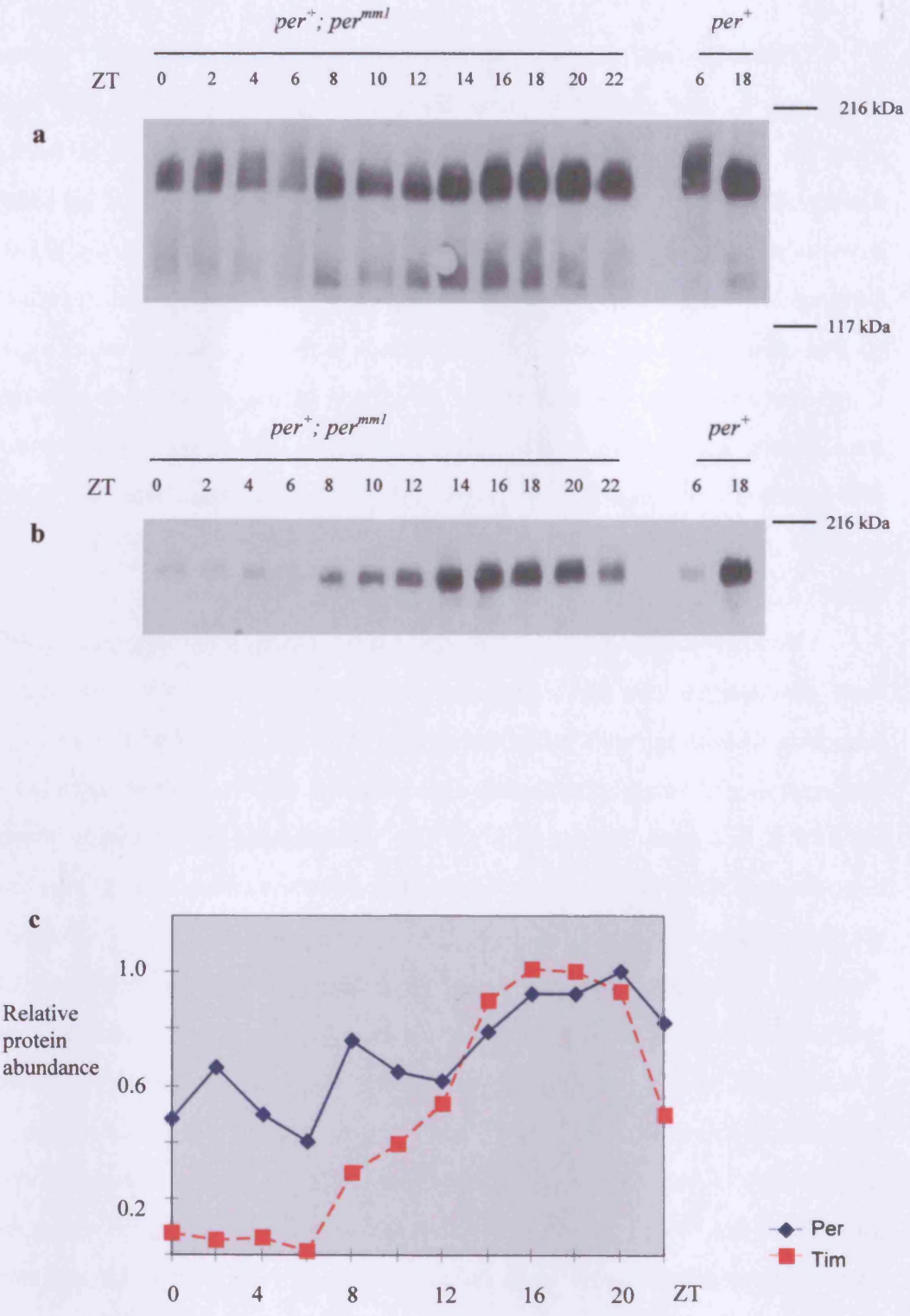


Figure 6.16. Levels of Per and Tim in per^{+}; per^{mm1}-34a transformants. Total head extracts were blotted and analysed with **a**: α-Per and **b**: α-Tim antibodies. **c**: densitometry analysis of the data shown in a and b.

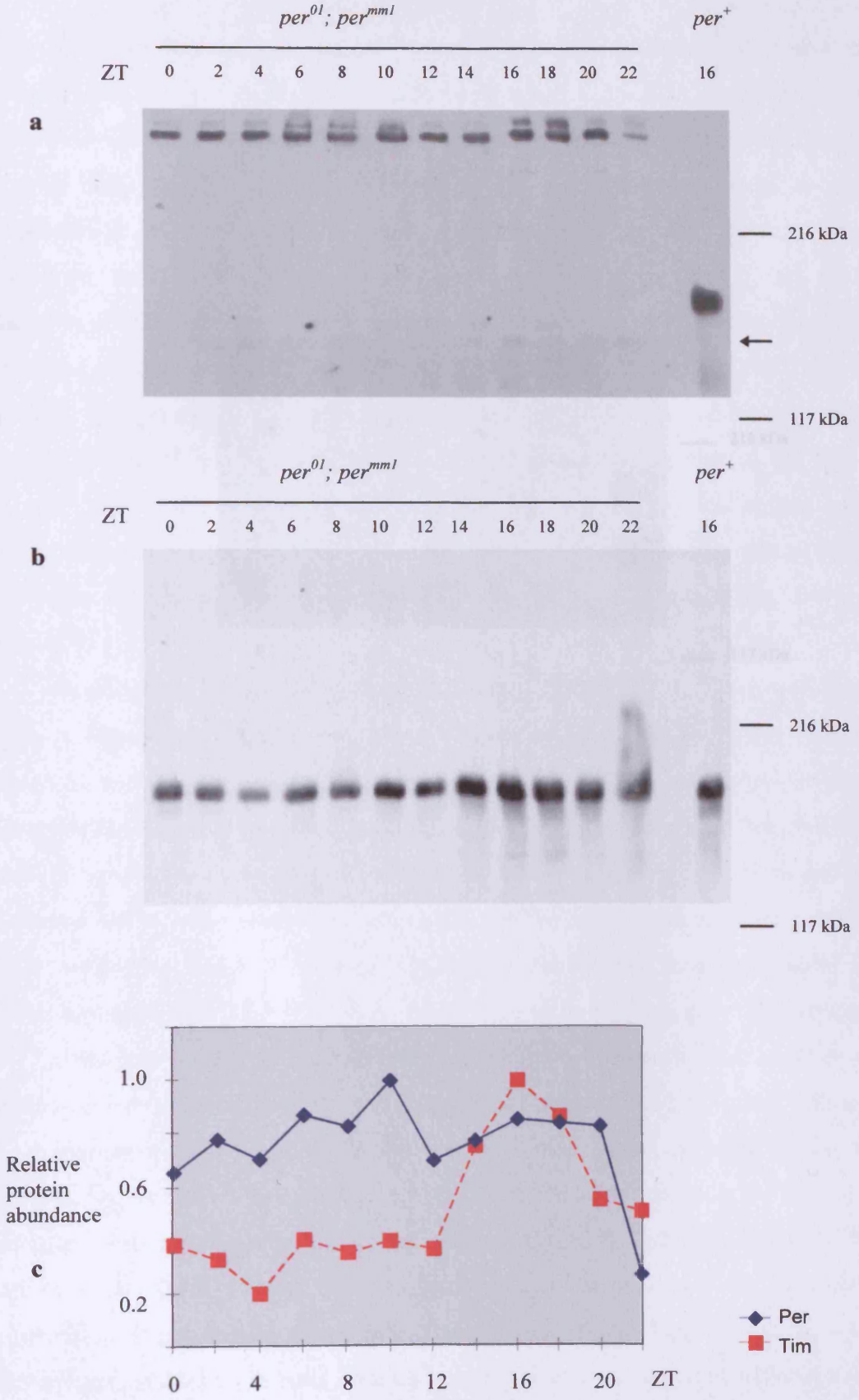


Figure 6.17. Quantitation of Per and Tim in per^{0}; per^{mm1}-34c transformant flies, carrying one dose of the transgene. **a**: immunoblot showing the levels of Per protein. Arrow denotes the position of the 150 Kda band. **b**: levels of Tim as visualised by α-Tim antibody. The protein amounts are visualised in **c**.

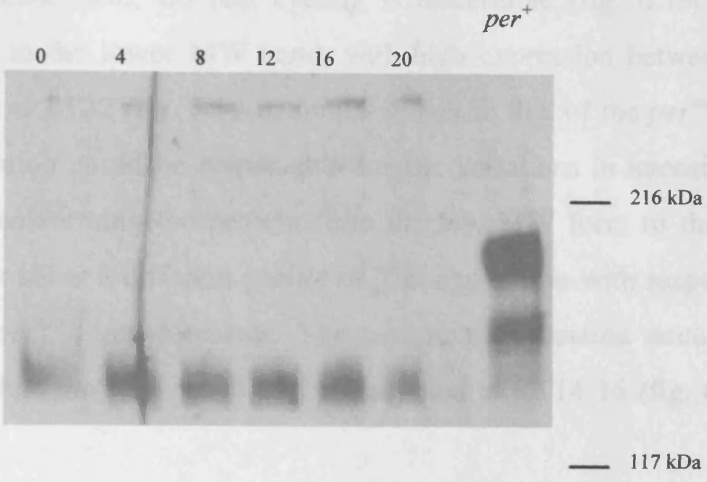


Figure 6.18. Immunoblot of protein extract from per^0; per^{mm1}-34c transformant flies. Protein samples were obtained from fly heads with the same procedure as for samples shown in fig. 6.17, but were run on a 6M urea PAGE

When protein extracts of per^{01}; per^{mm1}-34a flies (carrying two doses of the transgene) are obtained in extraction buffer with added 0.1% SDS, little signal is found in the doublet in the 300 kDa region, while the intensity of the 130-150 kDa region increases (fig. 6.19a). A band is detected in the 130 kDa region, with an intensity comparable to that of the 150 kDa signal. Although there seems to be variation in the intensity of the 150 kDa band, no real cycling is discernible (fig. 6.19c), while an oscillation is detected in the lower MW band, with high expression between ZT8 and ZT14, and a minimum at ZT22 (fig. 6.19d), similar in fact to that of the per^{mm1} transcript (fig. 6.3). Phosphorylation could be responsible for the variations in intensity recorded for these bands, by transforming the protein from the low MW form to the high MW one. per^{01}; per^{mm1} flies show a different profile of Tim expression with respect to either wild type and per^{+}; per^{mm1} transformants. The minimum expression occurs in per^{01}; per^{mm1} flies at ZT4, while the peak level is now recorded at ZT14-16 (fig. 6.17b-c and 6.19b-c).

Immunoblots were then performed on per^{01}; per^{mm2} flies. These transformants display a behavioural rhythm very close to that of wild type *D. melanogaster* (see chapter 5) and the chimaeric Per protein responsible for the rescue of the rhythmic phenotype derives its N terminus from the wild type *D. melanogaster* Per, fused to the *Musca* C-terminal half (see chapter 2). It appears that the presence of a *D. melanogaster* N-terminal half is sufficient to drive a circadian oscillation in chimaeric protein levels (fig. 6.20a), while the *Musca* C terminus is responsible for the faster migration of the chimaeric protein on SDS-PAGE (with respect to the *D. melanogaster* Per protein); the Per^{mm2} chimaeric protein runs on acrylamide gel at the same position as *Musca* Per. Extensive temporal phosphorylation also appears to take place in per^{01}; per^{mm2} flies.

Immunoblots of housefly protein extracts, show a strong signal in the region corresponding to 130 kDa (see fig. 6.21a) but there is no evidence for the rhythmic oscillation of its quantity (fig. 6.21d), in agreement with the immunostaining of *Musca* brain sections. Total protein extracts from *Musca* thoraces also fail to show any rhythmicity of expression of Per protein (fig. 6.21b). The anti Tim antibody also detects some antigens in *Musca*. A band appears on the blot with an apparent molecular weight of about 200 kDa, slightly larger than the *D. melanogaster* Tim protein. The immunoblot shows that expression of this protein oscillates in a circadian fashion. The peak appears

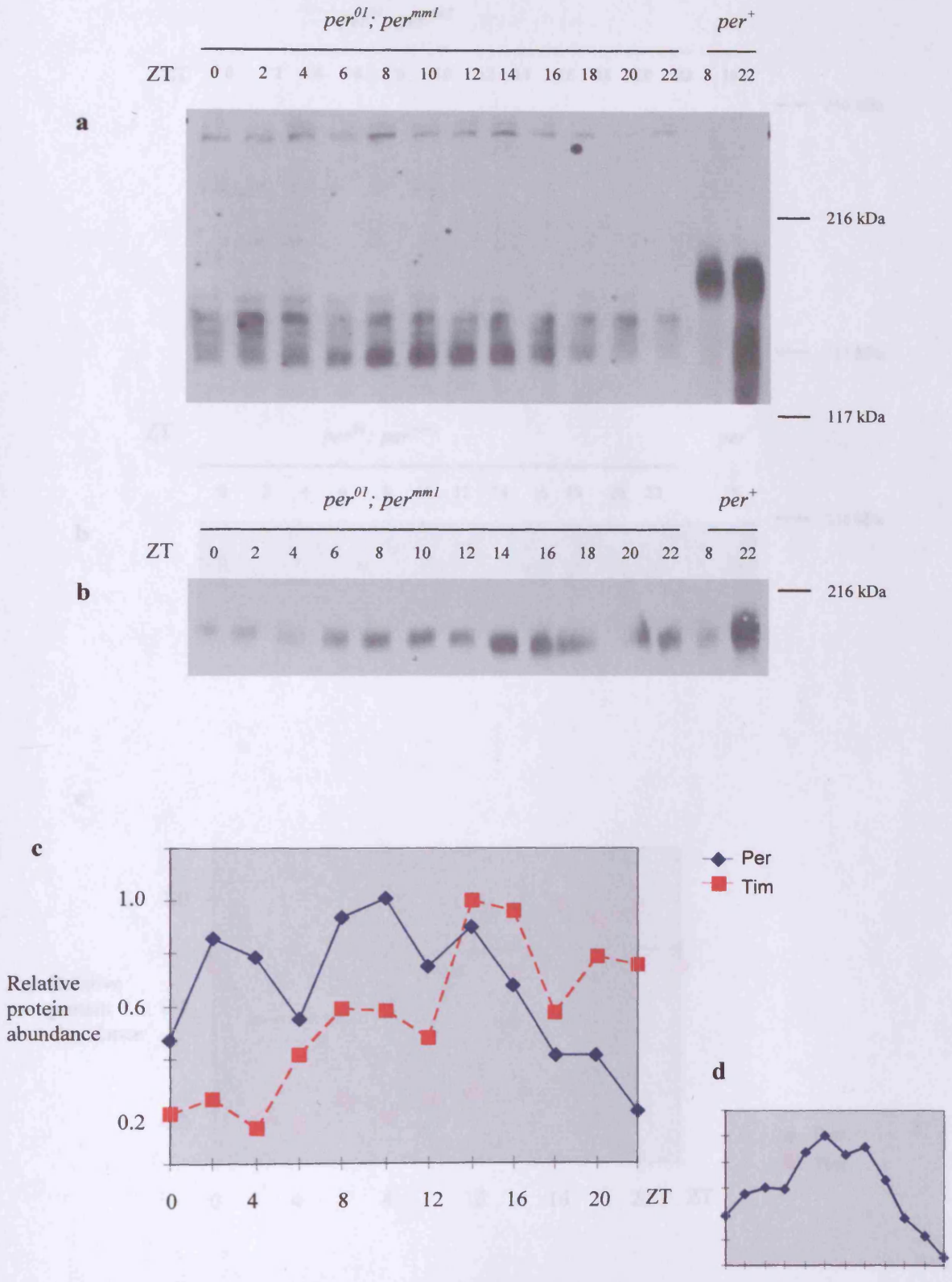


Figure 6.19. Oscillation in the levels of Per and Tim in per^{0}; per^{mm1}-34a transformants carrying two doses of the per^{mm1} transgene. **a**: immunoblot probed with α-Per antibody. **b**: an equivalent filter tested with α-Tim antiserum. **c**: graphical representation of the amount of Per and Tim proteins. The amount of Per is calculated as the sum of the 130 and 150 kDa bands. **d**: visualisation of the level of the 130 kDa Per band.

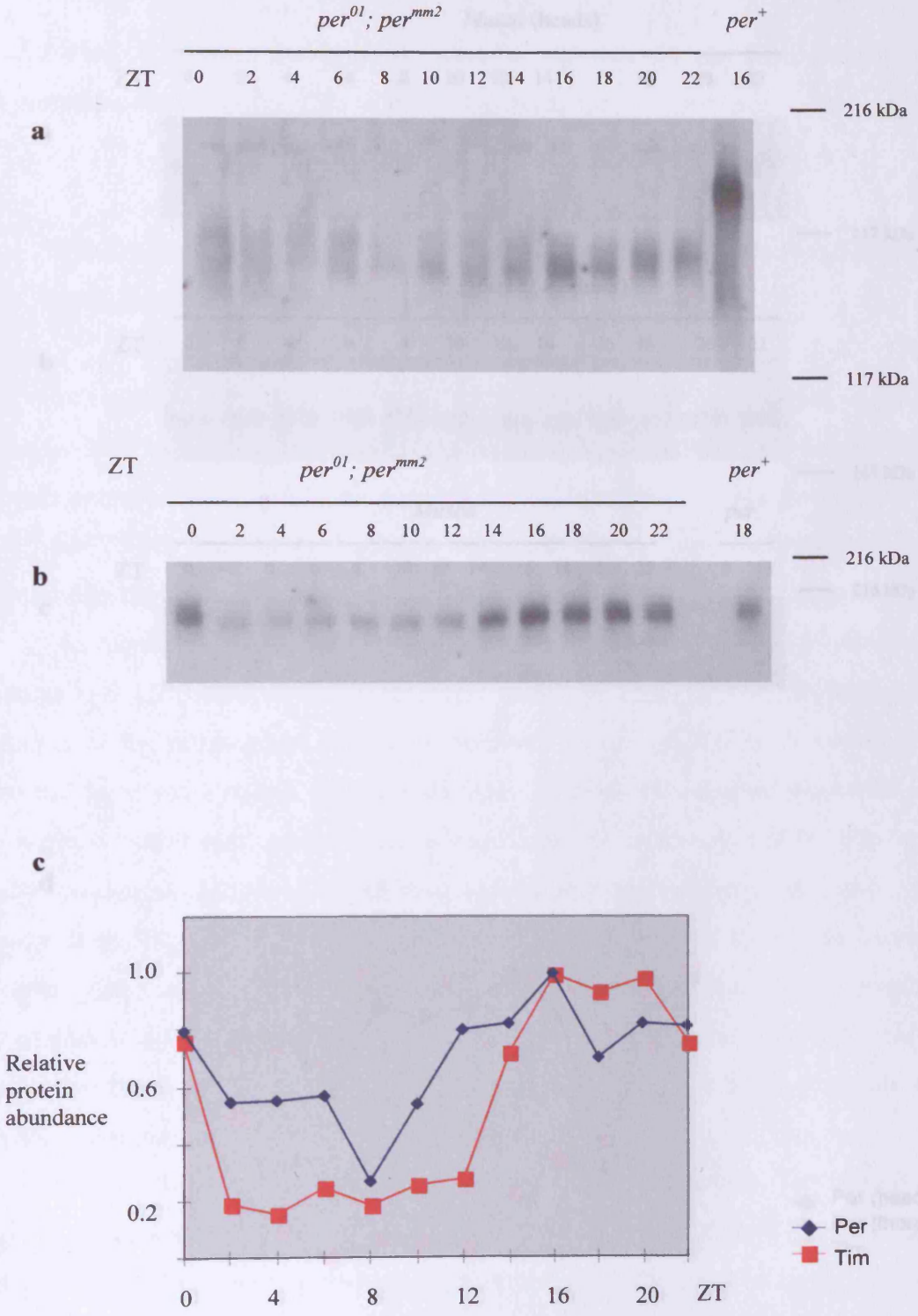


Figure 6.20. Quantitation of Per and Tim in per^{0}; per^{mm2} transformant flies. **a**: immunoblot showing the levels of Per protein. **b**: levels of Tim as detected by immunoblot. Quantification of protein amounts is reported in **c**.

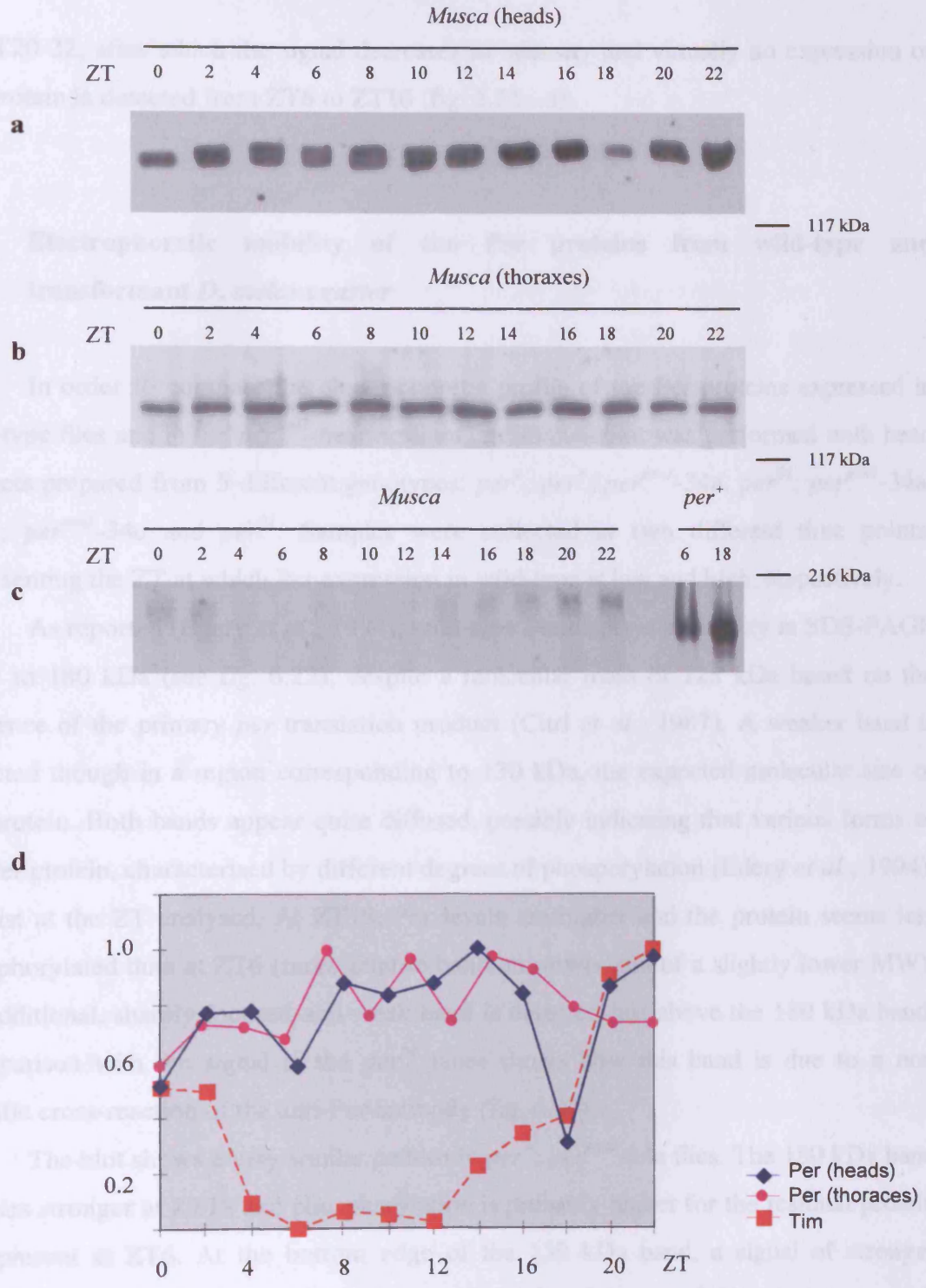


Figure 6.21. **a** and **b**: immunoblots of *Musca* total head and thorax extract, respectively, collected at various Zeitgeber times and analysed with anti-Per. **c**: immunoblot of head extracts probed with anti-Tim. The quantiation of these data is shown in **d**.

at ZT20-22, after which the signal decreases in intensity and virtually no expression of the protein is detected from ZT6 to ZT10 (fig. 6.21c-d).

Electrophoretic mobility of the Per proteins from wild-type and transformant *D. melanogaster*

In order to compare the electrophoretic profile of the Per proteins expressed in wild-type flies and in the per^{mml}-bearing lines, an immunoblot was performed with head extracts prepared from 5 different genotypes: per^{+}, per^{+}; per^{mml}-34a, per^{01}; per^{mml}-34a, per^{01}; per^{mml}-34c and per^{01}. Samples were collected at two different time points, representing the ZT at which Per expression in wild-type is low and high, respectively.

As reported (Edery *et al.*, 1994), wild-type Per displays a mobility in SDS-PAGE close to 180 kDa (see fig. 6.22), despite a molecular mass of 128 kDa based on the sequence of the primary *per* translation product (Citri *et al.*, 1987). A weaker band is detected though in a region corresponding to 130 kDa, the expected molecular size of the protein. Both bands appear quite diffused, possibly indicating that various forms of the Per protein, characterised by different degrees of phosporylation (Edery *et al.*, 1994), coexist at the ZT analysed. At ZT18, Per levels are higher and the protein seems less phosphorylated than at ZT6 (more intense band, narrower and of a slightly lower MW). An additional, sharply focused and weak band is detected just above the 180 kDa band. Comparison with the signal in the per^{01} lanes shows how this band is due to a non specific cross-reaction of the anti-Per antibody (fig. 6.22).

The blot shows a very similar pattern in per^{+}; per^{mml}-34a flies. The 180 kDa band appears stronger at ZT18 and phosphorylation is probably higher for the residual protein still present at ZT6. At the bottom edge of the 130 kDa band, a signal of stronger intensity is discernible. Compared to the equivalent band from per^{+} flies, it appears as if an additional component is present in this band in per^{+}; per^{mml}-34a transformants (fig. 6.22).

A rather diverse picture emerges in per^{01}; per^{mml}-34a and per^{01}; per^{mml}-34c flies. Here no sign of the 180 kDa band is observed and all the signal is focused in the 130-150 kDa region (formation of the 300 kDa aggregate was minimised by addition of 0.1%

SDS to the protein extraction buffer). At ZT18 the signal is evenly distributed on a relatively wide region, comparable to the 130 kDa band found in per^{+}-bearing flies, and at ZT8 it is mainly localised on a much narrower area, corresponding to the low MW edge of the wide 130 kDa band (see fig. 6.22).

Given that only two time-points were used for each genotype, no attempt of quantification was made. Figure 6.22 though, shows quite clearly that in per^{+} and per^{+}; per^{mm1}-34a Per expression is higher at ZT18, while such a definitive statement can not be made for the per^{01}; per^{mm1} lines. It appears rather clearly though, that a temporal change in the phosphorylation state of Per occurs in the per^{01}; per^{mm1} transformants, with a more phosphorylated form (higher molecular weight) predominant at the later stage of the circadian cycle (see fig. 6.22).

Per-Tim interactions

The ability of the *Musca* Period protein to mediate interactions with *D. melanogaster* Tim has been assayed by immunoaffinity. Comparable amounts of head extracts collected from per^{01}; per^{mm1}-34c transformant flies at ZT6 and ZT18, were incubated with the immobilised antibody. The proteins non-covalently bound to the resin (E Ex) were then eluted, TCA precipitated and run on SDS-PAGE, as was fresh total extracts (T Ex), and an equal volume of total extracts subjected to overnight incubation with the antibody beads (D Ex). After blotting, immunodetection with anti-Tim revealed a band of approximatively 190 kDa (T Ex and E Ex, fig. 6.23). The band was present in both ZT samples, even though it was noticeably stronger at ZT18 compared to ZT6. In the depleted extracts (D Ex), the Tim band is dramatically reduced in intensity with respect to the total extract samples. This could indicate that incubation with α-Per removes most of the Tim protein from the extract; the majority of Tim should therefore be complexed with Per. Alternatively, lack of Tim from the depleted extracts could possibly indicate degradation of the Tim protein during the overnight incubation with the bound antibody. Immunodetection using anti-Per was not very informative, possibly due to contamination by rabbit antibodies in the samples.

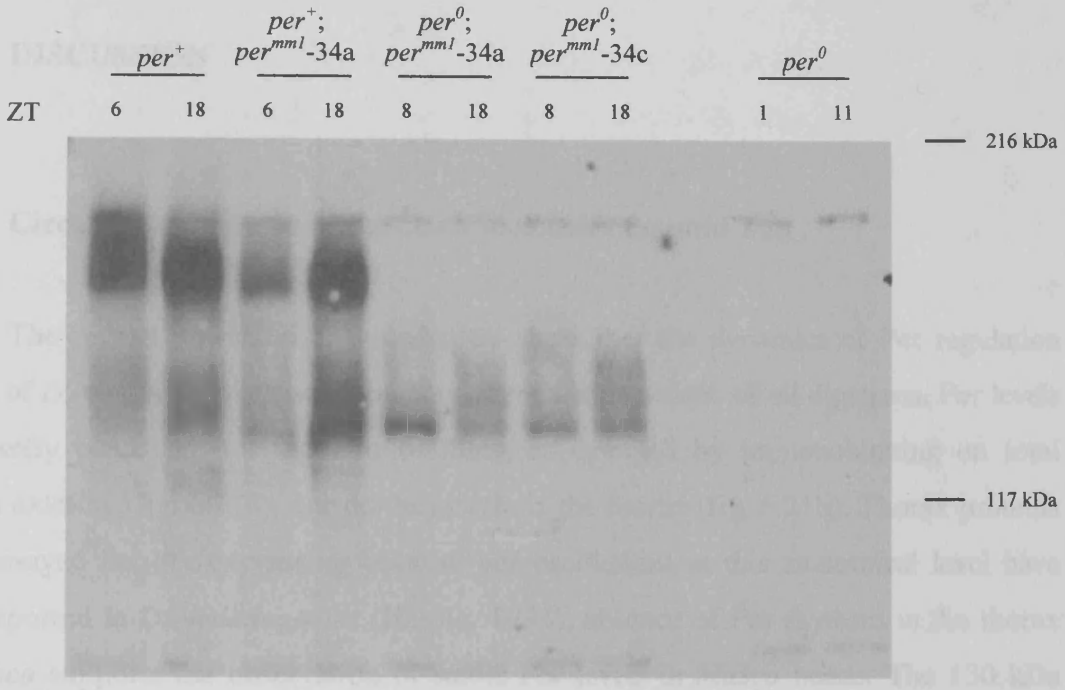


Figure 6.22. Immunoblot of total head extracts from different lines of *D. melanogaster* tested with α-Per. Compare the different electrophoretic mobility between the wild-type *D. melanogaster* Per protein (high MW) and that of the *Musca* Per expressed by the transgenic lines. Different mobility is also observed within the same line at different ZT (lowly and highly phosphorylated forms).

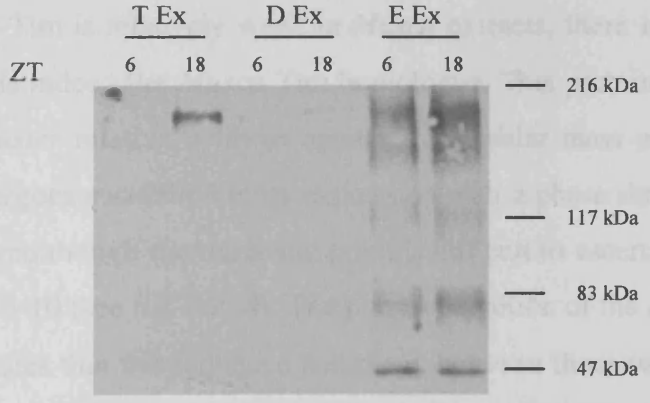


Figure 6.23. *Musca* Per can interact with *D. melanogaster* Tim in *per*^{01}; *per*^{mm1}-34c transformants. Total head protein extract was immunoprecipitated with immobilised α-Per and several fractions were used for an immunoblot probed with α-Tim antibody. T Ex: total head extract. D Ex: extract depleted of Tim protein by incubation with α-Per immobilised on protein G-coupled agarose beads. E Ex: immunoprecipitated proteins eluted from the α-Per–agarose matrix.

DISCUSSION

Circadian expression of the clock molecules Per and Tim

The results presented in this chapter show that the dynamics of Per regulation typical of *D. melanogaster* cannot be considered characteristic of all dipterans. Per levels in housefly heads do not seem to oscillate, as detected by immunoblotting on total protein extracts (fig. 6.21a), nor do they cycle in the thorax (fig. 6.21b). Thorax proteins were assayed for Per expression because *per* oscillations at this anatomical level have been reported in *D. melanogaster* (Hardin, 1994); absence of Per rhythms in the thorax of *Musca* supports the observation of stable Per levels in *Musca* heads. The 130 kDa band recognised by α-Per 978 represents almost certainly the *Musca* Per protein, given the ability of this antibody to detect a GST-*Musca* Per fusion protein expressed in *E. coli* (fig. 6.13), and the same 130 kDa band appearing in extracts derived from per^{01}; per^{mm1} transformants (fig. 6.17a and 6.19a) but absent in per^{01} mutants (fig. 6.22). The fact that Per does not seem to cycle in *Musca* heads does not prevent a circadian periodicity in locomotor activity, or in the molecular oscillation of Tim in LD (fig. 6.21c). Although the signal given by α-Tim is relatively weak in *Musca* extracts, there is little doubt that the band highlighted is indeed the *Musca* Tim homologue. This protein is slightly larger than the *D. melanogaster* relative, with an apparent molecular mass of about 200 kDa (fig. 6.21c), and undergoes oscillation in its expression with a phase similar to that of *D. melanogaster* Tim, even though the minimum point is difficult to ascertain given the low signal recorded at ZT6-10 (see fig. 6.21d). The poor recognition of the *Musca* protein by α-Tim probably indicates that the sequence homology between these two homologues is lower than that between the Per proteins from the corresponding species. However, the relatively high working dilution for α-Per (1:10,000) compared to α-Tim (1;1000) antisera, may reflect an abundance of different immunoglobulins directed against various epitopes of the Per protein, and a relatively small number of different immunoglobulins in α-Tim antiserum. In this case there would be more chance that an immunoglobulin recognises an evolutionarily conserved epitope on *Musca* Per than on *Musca* Tim.

The difference in the regulation of Per between *Musca* and *D. melanogaster* might reflect a significative diversity in the time-keeping mechanisms of these two dipterans. In order to investigate whether this is indeed the case, the expression of transgenic Per has been determined in transformant flies. Surprisingly, per^{01} flies carrying the chimaeric gene per^{mm1}, which consists of the 5' non-coding regulatory region from *D. melanogaster per* fused to the *Musca* gene, display a similar non-cycling expression of the Per protein in one of the lines analysed (see fig. 6.17a, c) and erratic levels in the other (fig. 6.19a, c). Indication for a cycling expression of a fast migrating form of the protein (possibly representing the unphosphorylated form) was detected (fig. 6.19d). This possible cycling pattern dispalys a considerable phase advance with respect to the wild type, which does not correlate with the Tim cycle from the same transformant flies (fig. 6.19b, c). The ability of these transgenic flies to measure time though, is demonstrated by rhythmicity in their locomotor activity profiles (chapter 5) and in the oscillating level of Tim expression. Apparently, basic *per* activity is rescued without the need for the cycling expression of Per protein. In the present discussion, it has been assumed that the cycle in Tim expression is an indication of restored rhythmicity at the molecular level, but it must be said that oscillations in Tim levels are observed also in per^{01} mutants where this oscillation is driven by the light-dark cycle as light induces Tim degradation (Zeng *et al.*, 1996; Myers *et al.*, 1996; Lee *et al.*, 1996). However the amplitude of this Tim cycle in per^{01} is strongly reduced compared to wild type (Zeng *et al.*, 1996). In my study, the amplitude of Tim oscillation in per^{01}; per^{mm1} transformants is similar to that of wild type *D. melanogaster* (compare fig. 6.17b-c and 6.19b-c with fig. 6.15b-c), suggesting that it is a real clock-controlled phenomenon rather than an indirect effect of light. The 2 h advance characteristic of Tim cycling in per^{01}; per^{mm1} flies correlates surprisingly well with the two hours reduction in the length of the circadian cycle typical of these transformants (see chapter 5). A similar advance of the peak of expression has been described for the LD cycling of both *per* RNA (Hardin *et al.*, 1990) and protein (Edery *et al.*, 1994) in per^{S} mutants.

In order to see whether differential expression levels of Per between wild-type *D. melanogaster* and per^{01}; per^{mm1} transformants are regulated at a pre- or post-translational stage, levels of the *per* transcript in the transformant flies have been determined. The chimaeric *per* transcript has been shown to undergo circadian oscillation (fig. 6.3). This

does not present a great surprise since the 5' non-coding regulatory region of this transgene, which is responsible for the oscillation of *per* transcript in a rhythmic fly (Zwiebel *et al.*, 1991; Hardin *et al.*, 1992a; Hao *et al.*, 1997) is derived from the *D. melanogaster* gene. The peak of expression of per^{mm1} is advanced with respect to that of wild-type flies. This is not entirely unexpected given that per^{mm1} transformants, as in the case of per^{S} mutants, have a short circadian periodicity (described in chapter 5), and it correlates well with the advance observed in the expression peak of Tim cycling. While basal per levels are comparable in the two lines, peak levels in per^{+}; per^{mm1}-34a are twice as high as those recorded for per^{01}; per^{mm1}-34c (fig. 6.4). This is possibly due to the fact that two copies of the per^{mm1} transgene are present in per^{+}; per^{mm1}-34a and only one copy is carried by the other transgenic line, but it cannot be excluded that the per^{+} gene can drive per^{mm1} cycle at a higher amplitude. This alternative has not been further investigated. Rather more surprising is the observed oscillation of *per* transcript in *Musca* (fig. 6.5). Since no cycling in Per protein is detected in this dipteran, a similar constitutive expression would be expected also for the transcript, given the proposed feedback relationship between the two products of the *per* gene (Hardin *et al.*, 1990). This molecular oscillation, for which a real biological significance is difficult to envisage since protein levels appear to be constant, has phase and amplitude matching those described for *D. melanogaster* (Hardin *et al.*, 1990).

Given that per^{01}; per^{mm1} transformant flies are rhythmic at the behavioural and molecular level, and per^{mm1} transcript level fluctuates on a circadian basis, the apparent constitutive expression of housefly Per must be related to the properties of the *Musca* protein rather than to a dysfunction of the clock. Dembinska *et al.* (1997) have shown that temporal proteolysis contributes significantly to the cycling expression of Per, so for example, a higher intrinsic stability of the housefly protein compared to that of the *D. melanogaster* Per, could alter the expression profile. In this respect one must remember that the *D. melanogaster* protein contains three PEST sequences while only one such domain is found in *Musca* Per (see table 3.2), possibly suggesting a slower rate of degradation for the *Musca* protein.

Figure 6.22 shows that two Per bands are detected in immunoblots in per^{+} flies while only the lower MW band is observed in per^{01}; per^{mm1} transformants. Given that this band roughly corresponds to the calculated MW for the *D. melanogaster* Per protein, it

is possible that the 180 kDa band, which is the main product in per^{+} flies, derives by modification of the Per protein soon after its translation. Such a modification would increase the protein's MW by about 50 kDa. Moreover, temporal phosphorylation occurs, which results in a reduction of the electrophoretic mobility of Per. The post-translational modification does not occur on the *Musca* Per protein, and therefore its size does not undergo the dramatic increase observed in per^{+}. The same is true for the Per protein expressed by per^{01}; per^{mm2} transformants. In these transformants Per cycles in a wild-type like fashion, but only the 150 kDa band is visible on the immunoblot (fig. 6.20), indicating that the initial hypothetical post-translational modification is not necessary for correct cycling to occur, and that the target of the modification is likely to reside in the C-terminal half of the protein, while the target for the proteolysis is possibly located in the N-terminal half. Unlike the undefined post-translational modification, phosphorylation occurs on the *Musca* protein in per^{01}; per^{mm1} flies, causing the 130 kDa band to shift towards heavier MW (fig. 6.19a). The fact that a stronger 130 kDa band is detected at ZT8-14, coinciding with high levels of per^{mm1} transcript, may indicate that there is no temporal delay between the accumulation of mRNA and newly synthesised, unphosphorylated protein.

Despite this difference in the dynamics of Per regulation between *Musca* and *Drosophila*, *Musca* Per appears able to mediate protein-protein interaction with *D. melanogaster* Tim in per^{01}; per^{mm1} transformants (see figure 6.23). Given that Per-Tim interactions occur also at the level of the PAS region (Gekakis *et al.*, 1995), the ability of *Musca* Per to bind to *D. melanogaster* Tim was not entirely unexpected, on the basis of the phylogeny presented in chapter 3.

Immunocytochemistry of Per

The detection of Per protein in total extracts has a limitation, since it would fail to detect subtle changes in protein level occurring in restricted areas of the brain. It is quite reasonable to expect that the pacemaker of *Musca* resides in a small structure within the brain, as for *D. melanogaster*, where it has been localised in the lateral neurons (Ewer *et al.*, 1992; Frisch *et al.*, 1994). In order to clarify whether the dynamics

of Per expression was differentially regulated in the pacemaker structures from that in the rest of the brain, Per protein expression in *Musca* was investigated by immunocytochemistry. Patterns of anti-Per expression in the *Musca* brain occur in structures equivalent to those of the better characterised *D. melanogaster*. In particular, strong antibody staining is observed in the *Musca* "lateral" neurons (fig. 6.8), which correspond in location to the presumptive circadian pacemaker of *D. melanogaster*. For this reason a certain amount of confidence can be placed in the fact that the monoclonal antibodies detect the housefly's Per protein. Unfortunately, the lack of a per^0 mutant in *Musca* does not allow confirmation of the specificity of the antibody mix labelling, and there is not an absolute certainty that the monoclonal antibodies are indeed detecting the Per protein. Replication of these results with another anti-Per antibody, would further support this view. Investigations are currently being carried out with α-Per 978.

The temporal expression of the *Musca* protein does not match that of *Drosophila*. Per protein in the *Musca* brain is confined to the cytoplasm, with the noticeable exception of a small subset of glial cells located at the edge of the medulla (fig. 6.12), and its level is apparently constant, showing no sign of the robust rhythmicity characteristic of *D. melanogaster* (Siwicki *et al.*, 1988; Zerr *et al.*, 1990). The only sign of rhythmicity in *Musca* is the nuclearisation of Per in several glial cells, occurring during the middle of the light phase, but no variation in the expression level of the protein is recorded. These findings correlate well with the results obtained by Western blotting on total head extracts, in which a different antibody was utilised.

Implications for the clock model

The study of Per expression pattern in *Musca* and in the per^{01}; per^{mm1} transformant lines reveals interesting new insights about the mechanism underlying the function of the circadian clock. More brain sections from *Musca* and transformant flies have to be immuno-labelled and analysed before a definitive picture will emerge, but the preliminary results show that rhythmicity does not require the circadian cycling of Per protein. In other words, the role of Per as state variable must be reconsidered. In the circadian pacemaker, time is inferred by analogy, by measuring the level of a certain state

variable, like sand in a hourglass. In the model developed by Hardin *et al.* (1990) Per represents the state variable able to "store" the temporal information. Later work recognises the existence of a second variable, Tim which, oscillating with approximately the same phase and period as Per, carries the same information (Sehgal *et al.*, 1995; Zeng *et al.*, 1996). Evidence from *Musca* and the per^{mm1} transformants seem to suggest that the timing information would be "carried" by a different molecule, possibly Tim. This view is in agreement with the major role played by Tim in the resetting of the endogenous clock by photic stimuli. Tim is in fact the target of a light-dependent degradation pathway (Zeng *et al.*, 1996; Myers *et al.*, 1996; Lee *et al.*, 1996), whereas Per levels are only indirectly reduced by light by a decrease in its stability caused by the disappearance of Tim.

A second heretical consideration is that Per may not be directly involved in a negative feedback loop. Since dramatic variations in levels of Per protein on a circadian basis are not observed in either *Musca* and per^{01}; per^{mm1} transformants, the detected *per* mRNA daily fluctuation must occur independently of the levels of its own translation product, so that Per protein does not repress the rate of *per* transcription. The results presented here are more in line with a model where the *per* promoter is under the control of the circadian clock. Per is necessary for the clock to tick, but it does not directly act on its own promoter. On the other hand, absence of Per arrests the clock and all clock controlled functions, among which is the oscillation of *per* transcription. Zeng *et al.* (1994) have shown that constitutive expression of Per driven by a *ninaE* promoter (O'Tousa *et al.*, 1985) reduces the amplitude of transcription cycling of the endogenous *per* gene. Their work though, is limited to the fly's eye. Furthermore, in their experiment, Per protein was expressed at a level considerably higher than physiological Per concentrations, with possible consequences on cell homeostasis. This is possibly a quite different situation from that of the normal pacemaker.

A clear demonstration of the negative feedback loop model has therefore yet to be presented, and meanwhile the model fails to give a satisfactory answer to several points. For example, it does not explain why *per* levels start to decrease at about ZT14-15 (Hardin *et al.*, 1990), considerably earlier than the Per-Tim complex nuclear entry (ZT18, Curtin *et al.*, 1995) and therefore before a transcriptional feedback can be effected by Per on its own promoter, nor does it explain why in an arrhythmic per^{01}

mutant, levels of *per* mRNA are intermediate between the peak and trough values of a per^{+} fly (Hardin *et al.*, 1990) rather than being stably maintained at peak values, given the lack of repression from the Per protein. Additional elements must therefore be implicated in the cycling pattern of the *per* messenger; Stanewsky *et al.* (1997) show that sequences located within the first 2/3 of the *per* coding sequence are responsible for the diminished *per* level of per^{01} mutants, while So and Rosbash (1997) demonstrate that the cycle in *per* mRNA level is the outcome of an oscillating transcription activity combined with a circadian regulation of *per* half-life (longer during the accumulation phase of the mRNA curve, becoming shorter in the declining phase).

Also, the rescue of rhythmicity in locomotor activity, obtained with a heat-shock promoter driving expression of a 7.2 kb *per* gene fragment (*hsp-per*, Ewer *et al.*, 1988; 1990) can hardly be accommodated into the feedback loop model. A later study of per^{01}; *hsp-per* transgenic flies found that they exhibited circadian oscillations of Per protein (Frisch *et al.*, 1994). In the same work the authors also report cycling expression for a promoterless 7.2 kb *per* transcript, albeit with a lower amplitude than wild type, and show that the promoterless *per* can drive circadian oscillations of the per^{01} transcript and Per protein (Frisch *et al.*, 1994). The explanation given by Frisch *et al.* is that the Per protein can still operate the transcriptional feedback loop on the 7.2 kb derived gene. In view of the experiments described by So and Rosbash (1997), it appears that *per* mRNA cycling in per^{01}; promoterless-*per* flies is caused by the temporal regulation of *per* transcript stability rather than by an effect on transcription. Similarily, temporally controlled Per proteolysis (Dembinska *et al.*, 1997) could play a substantial role in the circadian oscillation of Per protein in per^{01}; *hsp-per* transformants.

It appears evident that a simple negative feedback loop, with only Per at its core is inadequate to explain the characteristics of the *per*/Per molecular cycle, and a variety of additional elements and mechanisms are being discovered. What is still unclear is whether these mechanisms directly require some sort of signalling from Per, or only require a functional timekeeping device. The presence of a functional Per protein in the per^{01} mutants (from the *hsp-*per^{+} transgene) partially restores clock functions and all rhythmic phenotypes, including clock-dependent RNA and protein synthesis/degradation. In per^{01}; per^{mml} flies, circadian oscillations of the *per* transcript occur apparently without extensive cycling of the Per protein, opening up the possibility that Per feedback on its

own transcript is, to say the least, a redundant feature for the molecular oscillation of the *per* mRNA.

As J. Hall has put it in one of his reviews (Hall, 1996), the negative feeedback loop has become a dogma, and several authors are going through a considerable effort of mental contortion in order to make their results fit the model.

CHAPTER 7

Conclusion

Conclusion

In this thesis, I have presented an analysis of the *per* locus of the housefly *M. domestica*. The function of the gene has not been directly determined in the housefly, nevertheless there is a concrete body of evidence for the *Musca* gene being a functional as well as a structural homologue of the *D. melanogaster per* locus. Several transformant *D. melanogaster* lines have been generated which contain the housefly *per* coding sequence, or chimaeric constructs in which different portions of the *D. melanogaster* gene have been replaced by the *Musca* equivalent sequences. All the transgenes tested have been successful in restoring circadian locomotor activity in per^{01} *D. melanogaster* mutants. The possibility that the product of the *per* gene has lost in *Musca* its role in timekeeping in favour of other physiological functions though, cannot be completely ruled out and will be investigated in the near future by injecting anti-Per antibody into the fly's brain, a technique successfully used by Cymborowski *et al.* (1996) to investigate the role of S-antigen (arrestin), an element of the phototransduction pathway, in mediating light entrainment in the fly *Calliphora vicina*. For this reason a *Musca*-specific anti-Per antibody is currently under preparation in our laboratory.

At first glance, the cloning of the first *per* homologue from a non-drosophilid did not unveil any particular surprise. The overall similarity between the *Musca* and any other *Drosophila per* gene is lower than among *Drosophila* species, and sequence comparisons have shown that the same division between conserved and non-conserved blocks first adopted for the *Drosophila per* homologues (Colot *et al.*, 1988) is still valid for the *Musca* gene. However a more detailed sequence analysis revealed that a 250 amino acid region of the *Musca* conceptual translation product is more similar to the *D. melanogaster* corresponding fragment, than the same region from the phylogenetically closer *D. pseudoobscura* or *D. virilis*. Interestingly, this region corresponds to the PAS domain, a fragment of Per involved in protein dimerisation with Per functional partner, the clock protein Tim (Huang *et al.*, 1995; Gekakis *et al.*, 1995). This observation would

not be very informative if it was not corroborated by some functional data. per^{01} *D. melanogaster* expressing the *Musca per* coding sequence (per^{01}; per^{mm1} transformants) display rhythmic behaviour in locomotor activity tests. The locomotor activity has been monitored at different temperatures and in different transformant lines, and the percentage of rhythmic flies was always higher than 76%, a figure close to the rescue of circadian periodicity obtained with a conspecific *D. melanogaster* per^{+}-bearing transformation construct (see chapter 5). In contrast, per^{01} mutants carrying a copy of the *D. pseudoobscura per* gene (per^{01}; per^{mps1} transformants) display periodic locomotor activity in about 45% of the flies tested, and with much longer and weaker periodicity (Peixoto *et al.*, 1998).

Therefore, despite the overall similarity among the *Drosophila* Per proteins being higher, *Musca* Per with its highly conserved PAS region is more effective in rescuing the per^{01} behavioural phenotype than the *D. pseudoobscura* homologue. Per-Tim dimerisation occurs through the PAS domain (Gekakis *et al.*, 1995; Saez and Young, 1996) and is a definitive requirement for clock functions (Zeng *et al.*, 1996; Myers *et al.*, 1996; Lee *et al.*, 1996). These observations strongly suggest that, in per^{mm1} transformants, the dimerisation domain of *D. melanogaster* Tim interacts with *Musca* PAS at a level comparable to that reached with its physiological partner, the wild-type *D. melanogaster* Per. Conversely, the *D. pseudoobscura*-PAS / Tim interactions in per^{mps1} transformants would be much weaker. In turn we can deduce that the dimerisation domain from *D. melanogaster* Tim is probably structurally more similar to that of *Musca* than it is to *D. pseudoobscura*, since interacting portions of two proteins must be subjected to co-evolutionary pressure.

Intramolecular coevolution between two portions of the Per protein has been proposed (Peixoto *et al.*, 1993; Nielsen *et al.*, 1994) and later demonstrated by Peixoto *et al.* (1998). The repetitive domain localised in the middle of the protein, the so called Thr-Gly region, is evolutionarily coupled to an adjacent 5' fragment of about 60 amino acids. In addition, Huang *et al.* (1995) demonstrated that PAS interacts with Per fragment 524-685 (called the C-domain), which contains this coevolving 5' flanking fragment, and concluded that this intramolecular interaction was involved in the temperature compensation of the clock. From these indications, the picture emerges of a rather complex interplay between the dimerisation domain of Tim, PAS, the C-domain

and the Thr-Gly region, where the structure of the former three and the length of the latter represent variables of the same equation which are linked together. Since the repeat region has undergone extensive expansion in *D. pseudoobscura*, it seems very plausible that this has affected the evolution of the PAS domain and possibly of the interacting Tim dimerisation domain.

This comparative analysis of the *per* gene from different species thus provides us with an interesting example of intermolecular coevolution based on functional and phylogenetic data. In order to further investigate this hypothesis, our laboratory has cloned the *D. pseudoobscura tim* gene (E. Rosato, personal communication) and is now trying to obtain the *Musca tim* homologue. Another possible route consists in analysing interactions between *in vitro* translated PAS regions from these three different species and the *D. melanogaster* Tim dimerisation domain or using heterospecific PAS and Tim fragments in yeast two hybrids assays.

The behavioural analysis of per^{mm1} transformants confirms that the role of the *per* gene goes further than simply determining the pace of the fly's clock. *per* controls also the pattern of locomotor activity, influencing the time of the day in which the fly is more active and that in which it is resting. Incidentally, in constant darkness per^{mm1} transgenic flies display a pattern of activity reminiscent of that of *Musca* in the same experimental conditions and clearly different from that typical of wild-type *D. melanogaster.* It would therefore be very tempting, in the trail of Petersen *et al.*, (1988), to speculate that *per* confers species-specificity of behaviour, but the analysis of transformants in light-dark conditions shows that this may not be correct. In fact per^{mm1} flies in LD display a rather peculiar activity pattern, with the activity concentrated during the night and very little or no activity during the light phase of the cycle. This pattern is totally different from that of the same flies in free running conditions and from that of *Musca* in both LD and DD, where most of the activity occurs during the subjective day. This result is nevertheless equally exciting, since it demonstrates that one single gene can greatly influence the distribution of the total locomotor activity throughout the circadian cycle, controlling at which time of the day or night the fly is active or resting.

The *Musca per* gene has unexpectedly revealed another facet, the implications of which could lead to a reconsideration of the current model of the circadian pacemaker first developed by Hardin *et al.* (1990) and described in chapter 1. For Hardin *et al.* and

for the general consensus of opinion since 1990, the out-of-phase cycling oscillations in the amount of *per* transcript and protein are the core of the circadian clock. A negative feedback mechanism down-regulates *per* transcription when Per levels are high, and degradation of Per removes the repression with consequent re-initiation of *per* transcription. The model has been further enriched in order to integrate the findings of Per phosphorylation (Edery *et al.*, 1994) and Per nuclear translocation (Curtin *et al.*, 1995), of the existance of a second clock gene, *tim* (Vosshall *et al.*, 1994; Sehgal *et al.*, 1994) and the interactions between Per and Tim (Zeng *et al.*, 1996; Myers *et al.*, 1996).

Feedback has long been thought to play a major part in time keeping mechanisms (Goodwin, 1965), therefore the characterisation of the temporal pattern of expression of the *per* gene products was a strong indication that *per* is implicated in a negative feedback loop. Furthermore, as *per* was the only clock gene identified in *Drosophila* at that time, any model of pacemaker mechanism was focused around this gene. Thus, the *per*-centric feedback loop model was born out of expectations as well as observations. Further support for the *per* feedback model came from the work of Aronson *et al.* (1994), who demonstrated the negative control of the Frq protein over *frq* transcription, and showed that a negative feedback loop of the *frq* mRNA and protein constitute the *Neurospora crassa* circadian oscillator. Zeng *et al.* (1994) have shown that the constitutive overexpression of *per* driven by the *ninaE* promoter (which is active in photoreceptor cells, O'Tousa *et al.*, 1985) reduces the amplitude of the endogenous *per* mRNA oscillations in the eye of transgenic flies compared to wild-type, but the direct repression of Per on the transcription of its own RNA has not been demonstrated in *Drosophila* brains. As Sauman and Reppert (1996) point out, what is true in one tissue is not necessarily true in another. Therefore the peculiar pattern of *per* expression observed in per^{01}; per^{mm1} transformants (see chapter 6) is not contradicting any compelling direct evidence for the *per* negative feedback loop, but is at odds with the theoretical model which has been partially borrowed from the analogous *Neurospora* oscillator (Aronson *et al.*, 1994).

To establish whether Per is part of an autoregulatory feedback loop which constitutes the endogenous oscillator, or whether *per* cycling is simply an output feature of the circadian oscillator, has proven a difficult task. The distinction between causes and their effects is not always immediate, and normally the temporal succession of the events

helps in the discrimination. Unfortunately a cycling system does not offer such help and a different approach must be used. The use of transgenic flies expressing *per* from a constitutive heterologous promoter has been attempted in several instances (Ewer *et al.*, 1988; 1990; Frisch *et al.*, 1994; Vosshall and Young, 1995) with the result that the flies are able to generate weak rhythms in locomotor activity. However sequences located within the transcribed portion of *per* are able to drive rhythmic expression of the *per* RNA (Stanewsky *et al.*, 1997) so in principle it is not possible to exclude negative feedback of Per through these elements, and indeed cycling of *per* RNA and protein has been reported for these "constitutive" transgenes (Frisch *et al.*, 1994; Vosshall and Young, 1995). If a rhythmic *per* mutant could be discovered in which constitutive levels of Per protein were maintained throughout the 24 h cycle in the key pacemaker cells, then this would present a strong empirical argument against the *per*-based negative feedback loop.

per^{mm1} transcript abundance cycles in a circadian fashion in a per^{01} host (fig. 6.3) with a somewhat reduced amplitude and a phase advance with respect to wild-type message. This oscillation closely resembles that of a luciferase reporter fused to the *per* promoter (*plo-luc*, Stanewsky *et al.*, 1997), and is in agreement with the claim of Stanewsky *et al.* that additional sequences located within the *per* coding region (and therefore missing in *plo-luc* and possibly in per^{mm1}) are contributing to the cycling of *per* transcript. The oscillations of per^{mm1} transcript though, is not sufficient to drive cycling protein expression. It could be argued that the amplitude of RNA oscillation is not high enough to ensure a corresponding oscillation for the protein, since proteins synthesised from cycling RNAs show a reduction in amplitude when compared to the transcript (Wood, 1995) but, since temporally controlled modification of Per supposedly is an important element of Per's own dynamics (Dembinska *et al.*, 1997), it seems more plausible that the constant expression of the *Musca* Per protein is related to differential post-translational processing with respect to wild-type Per. For example, the half-life of the *Musca* Per protein may be longer than that of the *D. melanogaster* homologue due to different degradation signals. Temporally controlled post-translational modifications are possibly more important than the level of *per* transcript in regulating expression of the Per protein, explaining how Per can oscillate in transgenes where *per* transcript displays low amplitude cycles (like promoterless *per* [Frisch *et al.*, 1994]) or does not cycle at all

(*glass-per* [Vosshall and Young, 1995], crippled *ninaE* promoter [Cheng and Hardin, 1998]). Whatever the reason, it appears that the oscillation in Per abundance is not necessary for clock functions, given the robustly rhythmic behaviour of per^{01}; per^{mml} transformants. In agreement with this statement, Stanewsky *et al.* (1997) recently reported that oscillations of *plo-luc* and of another reporter containing the *per* promoter plus about two thirds of the coding sequence fused to the luciferase gene (BG-*luc*), which can be followed in single individuals as luciferase activity, dampen to arrhythmicity after few days in DD conditions, yet the flies retain locomotor rhythmicity for at least three weeks in constant darkness. Although BG-*luc* is not able to mediate the rhythmic locomotor phenotype, it seems to be a faithful reporter of Per expression (Dembinska *et al.*, 1997; Stanewsky *et al.*, 1997). The implications of this finding are quite important for defining the role of *per* within the pacemaker. It appears that the timing state of the system cannot be described by the concentration of Per protein, and therefore Per cannot be considered as a state variable of the clock. Nonetheless, the importance of Per for circadian rhythmicity is well documented. Expression of Per protein is necessary for rhythmicity, but cycling Per expression does not seem to be an absolute requirement. This seems to indicate that *per* codes for an accessory component of the pacemaker machinery rather than for the oscillating core. So why does *D. melanogaster* display cyclic expression of both transcript and protein? A possible answer is that this is a "safety mechanism". If we imagine that Per-Tim interactions are necessary for nuclear translocation and that the role of Per is restricted to Tim nuclearisation, it follows that a constant level of Per protein would not affect the pacemaker if Tim expression is temporally regulated. Regulation of Per expression would ensure that Per-Tim complex formation and successive nuclearisation always occur at the proper time, even in the presence of an anomalous Tim expression. *Drosophila*, but not *Musca*, could therefore have developed this safeguard mechanism as an additional control over the temporal gating of Tim nuclear translocation.

Clearly *Musca* is a promising additional model system for studying circadian molecules, together with silkmoth, *Neurospora*, Cyanobacteria, Gonyaulax and now the mouse, in which *per* has recently been cloned (Sun *et al.*, 1997; Tei *et al.*, 1997). The results described in this thesis may cast some doubt on the generality of the negative

feedback model as it relates to *Musca*, and this should be viewed as a challenge to the current dogma which pervades the field.

APPENDIX

Big Flies, Small Repeats: The "Thr-Gly" Region of the *Period* Gene in *Diptera*

J. Nielsen,[*] *A. A. Peixoto,*[*] *A. Piccin,*[*] *R. Costa,*[†] *C. P. Kyriacou,*[*] *and D. Chalmers*[*]

[*]Department of Genetics, University of Leicester and [†]Dipartmento di Biologia, Universita di Padova

The region of the clock gene *period* (*per*) that encodes a repetitive tract of threonine-glycine (Thr-Gly) pairs has been compared between Dipteran species both within and outside the Drosophilidae. All the non-Drosophilidae sequences in this region are short and present a remarkably stable picture compared to the Drosophilidae, in which the region is much larger and extremely variable, both in size and composition. The accelerated evolution in the repetitive region of the Drosophilidae appears to be mainly due to an expansion of two ancestral repeats, one encoding a Thr-Gly dipeptide and the other a pentapeptide rich in serine, glycine, and asparagine or threonine. In some drosophilids the expansion involves a duplication of the pentapeptide sequence, but in *Drosophila pseudoobscura* both the dipeptide and the pentapeptide repeats are present in larger numbers. In the nondrosophilids, however, the pentapeptide sequence is represented by one copy and the dipeptide by two copies. These observations fulfill some of the predictions of recent theoretical models that have simulated the evolution of repetitive sequences.

Introduction

The *period* (*per*) gene in *Drosophila* determines biological rhythmicity in the circadian (24 h), ultradian (< 24 h), and infradian (>24 h) time domains (see Kyriacou and Hall 1994 for a review). Mutations in the gene shorten, lengthen, or obliterate circadian pupal-adult eclosion and adult locomotor activity cycles (Konopka and Benzer 1971) and have a corresponding effect on the 60-s male lovesong cycle (Kyriacou and Hall 1980, 1989; Kyriacou et al. 1990*a*) and the 10-d egg-to-adult developmental cycle at 25°C (Kyriacou et al. 1990*b*). The *per* gene encodes a large ~120 KD protein and contains two regions of special interest. The first of these is the so-called PAS-encoding region, which includes two 51 amino acid imperfect repeat motifs. PAS regions are also found in a number of other *Drosophila* and mammalian genes (Hoffman et al. 1991; Nambu et al. 1991; Burbach et al. 1992). Recent experiments suggest that *per* may associate with other PAS-containing peptides via the PAS region (Huang et al. 1993), and the current view is that *per* encodes a protein that acts as part of a transcriptional network (Kyriacou 1994). Consistent with this scenario is the observation that *per* is predominantly a nuclear protein (Liu et al. 1992) and that the *per* protein negatively regulates its own mRNA cycling (Hardin et al. 1990).

The second region of interest within *per* encodes a long uninterrupted stretch of threonine-glycine (Thr-Gly) pairs in *Drosophila melanogaster* (Jackson et al. 1986; Citri et al. 1987). This region evolves relatively quickly, but Thr-Gly regions are also found in other *Sophophoran* species, for example, in *Drosophila pseudoobscura*, in which a less extensive Thr-Gly region is also associated with many copies of a degenerate 5 amino acid repeat (Colot et al. 1988) that is polymorphic in length (Costa et al. 1991). Within the *melanogaster* subgroup, all eight species have Thr-Gly repeats (Peixoto et al. 1992), and within *D. melanogaster*, the number of repeats is polymorphic, both in laboratory (Yu et al. 1987) and natural (Costa et al. 1991, 1992) populations. The same is true for the sibling species *Drosophila simulans* (Rosato et al., submitted), but in *D. melanogaster*, the Thr-Gly length polymorphism shows a strong and statistically significant latitudinal cline throughout Europe and North Africa (Costa et al. 1992). This raises the intriguing possibility that the length polymorphism may be under some kind of selection, presumably thermal. This view is supported to some extent by the finding that the in vitro removal of the uninterrupted repeat in *D. melanogaster* gives a temperature-sensitive circadian locomotor phenotype (Ewer et al. 1990). This Thr-Gly deletion also produces shorter-than-normal lovesong

Key words: *Drosophila*, *period* gene, repetitive DNA, biological rhythms.

Address for correspondence and reprints: D. Chalmers, Department of Genetics, University of Leicester, Leicester LE1 7RH, United Kingdom.

Mol. Biol. Evol. 11(6):839-853. 1994.

0737-4038/94/1106-0003$02.00

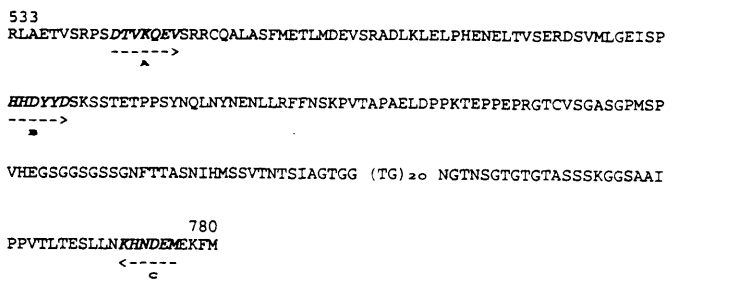


FIG. 1.—The positioning of the primers A, B, and C within the Thr-Gly domain of the *per* gene. The amino acid sequence is taken from Citri et al. (1987), and the uninterrupted stretch of 20 Thr-Gly repeats are represented in parentheses.

cycles that resemble those of *D. simulans* (Kyriacou and Hall 1986; Yu et al. 1987). Chimeric *per* genes between *D. melanogaster* and *D. simulans* reveal that the species specificity of the lovesong cycle (Kyriacou and Hall 1980, 1986) maps to the Thr-Gly repeat plus the immediate flanking regions (Wheeler et al. 1991). Thus experimental manipulations of the repeat and its immediate flanking environment produce changes in both the circadian and lovesong behavioral phenotypes. It is therefore of considerable further interest that a comparison of the "repeat" region and these flanking areas in species of the *Sophophora* and *Drosophila* subgenera suggested a possible coevolution of the length of the repeat with its adjacent sequences (Peixoto et al. 1993).

In this study, we compare the *per* sequences from the "Thr-Gly" region for species outside the *Drosophila* genus, indeed outside the family of Drosophilidae. These sequences reveal a surprisingly stable region of the gene.

Material and Methods

Flies

The following flies were used in this study: *Beris vallata* (soldierfly), *Syritta pipiens* (hoverfly), *Musca domestica* (housefly), *Loxocera albisata, Rhagoletis completa* (a true fruitfly), *Drosophila pictiventris, Zaprionus tuberculatus,* and *Drosophila picticornis. Beris vallata* and *S. pipiens* were provided by J. Mously, New Walk Museum, Leicester. *Rhagoletis completa* were provided by the University of Padova. The Drosophilidae were obtained from the American Drosophila Stock Center, Bowling Green State University, Ohio. The taxonomy of the flies is described in the Appendix.

The *Drosophila melanogaster, Drosophila pseudoobscura,* and *Drosophila virilis* DNA sequences were obtained from the literature (Colot et al. 1988).

DNA Extraction

Fly DNA for PCR reactions was extracted using a modification of the method described by Gloor and Engels (1990). Single flies were ground in 50–400 µl buffer (10 mM Tris-HCl pH 8.2, 1mM EDTA, 25 mM NaCl, 200 µg proteinase K/ml). The homogenate was left at 37°C for 1 h and heated to 95°C for 2 min. The DNA was extracted with phenol:chloroform:isomyl alchohol (24:24:2), ethanol precipitated and resuspended in 10–80 µl H_2O. In each 20 µl PCR reaction 1–4 µl were used (the optimal volume for PCR reactions varies from fly to fly dependent on, among other things, the size of the fly and how long it had been dead).

PCR Amplification, Primers, Cloning, and DNA Sequencing

PCR amplification was performed with degenerate primers. The following primers were used (Y = 50% C and 50% T, R = 50% A and 50% G, N – 25% of each nucleotide):

A. 5′-GAYACNGTNAARCARGARGT-3′ (543–549),
B. 5′-CAYCAYGAYTAYTATGA-3′ (598–603), and
C. 5′-CATYTCRTCRTTRTGYTT-3′ (771–776).

The primer positions in parentheses refer to the corresponding *Drosophila melanogaster* amino acid sequence published by Citri et al. (1987). Figure 1 shows the positions of the various primers in relation to the *D. melanogaster* protein sequence. For all flies except *Beris vallata* (where primers A and C were used), DNA was amplified using primers B and C in 30 cycles (95°C 1 min, 50°C 1 min, and 70°C for 2 min) on a Perkin Elmer Cetus theromocycler and electrophoresed on a 3.5% low-melt agarose gel. The main band in the right size range was extracted and reamplified using primers B and C with the same DNA sequence but containing an additional *Eco*RI (GAATTC) site at the 5′ end. The reamplified band was gel purified, extracted, and cut with *Eco*RI before being ligated into pUC18 at the *Eco*RI site, and transformed into DH5a-cells. Maxi preparations were carried out on appropriate clones, and the DNA sequences were determined by the dideoxy chain termination method (Sanger et al. 1977) using instructions and reagents from a T7-polymerase kit (Pharmacia) and a Sequenase kit (United States Biochemical). All sequences were confirmed from two independent clones to eliminate errors caused by misincorporation of nucleotides by Taq-polymerase. Discrepancies arising were resolved by sequencing a third independent clone. The "Thr-Gly" region sequence from *Musca* matched the DNA sequence that was later obtained from a lambda clone, containing the complete *Musca per* transcription unit. The complete *Musca per* sequence will be reported elsewhere (A. Piccin, R. Costa, C. P. Kyriacou, and D. Chalmers, unpublished data). Consequently, we are confident that the sequences we have obtained by PCR represent the Thr-Gly regions in the *per* orthologues of these larger Dipterans.

DNA Sequence Analysis

Sequence handling, alignment, and translation into amino acid sequence was performed with the help of the University of Wisconsin Genetic Computer Package (UWGCG) version 7.0 programs (Devereux et al. 1984). The multiple alignment program CLUSTAL V (Higgins et al. 1992) was used to help alignment where there was evidence of deletion/insertion events. We also searched for periodicity in the sequences of region var2 in *Drosophila pseudoobscura,* by using the autocorrelation procedure in the SPSS statistical package (SPSS 1990).

Results

For descriptive purposes we have divided the sequence into five regions: conserved areas 1, 2, and 3 (cons1, cons2, cons3) and variable areas 1 and 2 (var1, var2). The conserved areas are easily aligned (see figs. 2 and 3), whereas the variable regions cannot be aligned for all species. The boundary between cons1 and var1 (figs. 2 and 3) is also the boundary used by Colot et al. (1988) to define the margin between a conserved and variable region of *per*. The boundary between var1 and cons2 defines the beginning of the region studied earlier (Peixoto et al. 1992, 1993) in the *melanogaster* subgroup and the *Sophophora* and *Drosophila* subgenera, respectively. The divisions between var2 and the flanking regions cons2 and cons3 are also in the same positions as in previous studies (Peixoto et al. 1992, 1993).

Conserved Regions

Figure 2 gives the amino acid sequence of the three conserved regions, which consist mostly of small neutral

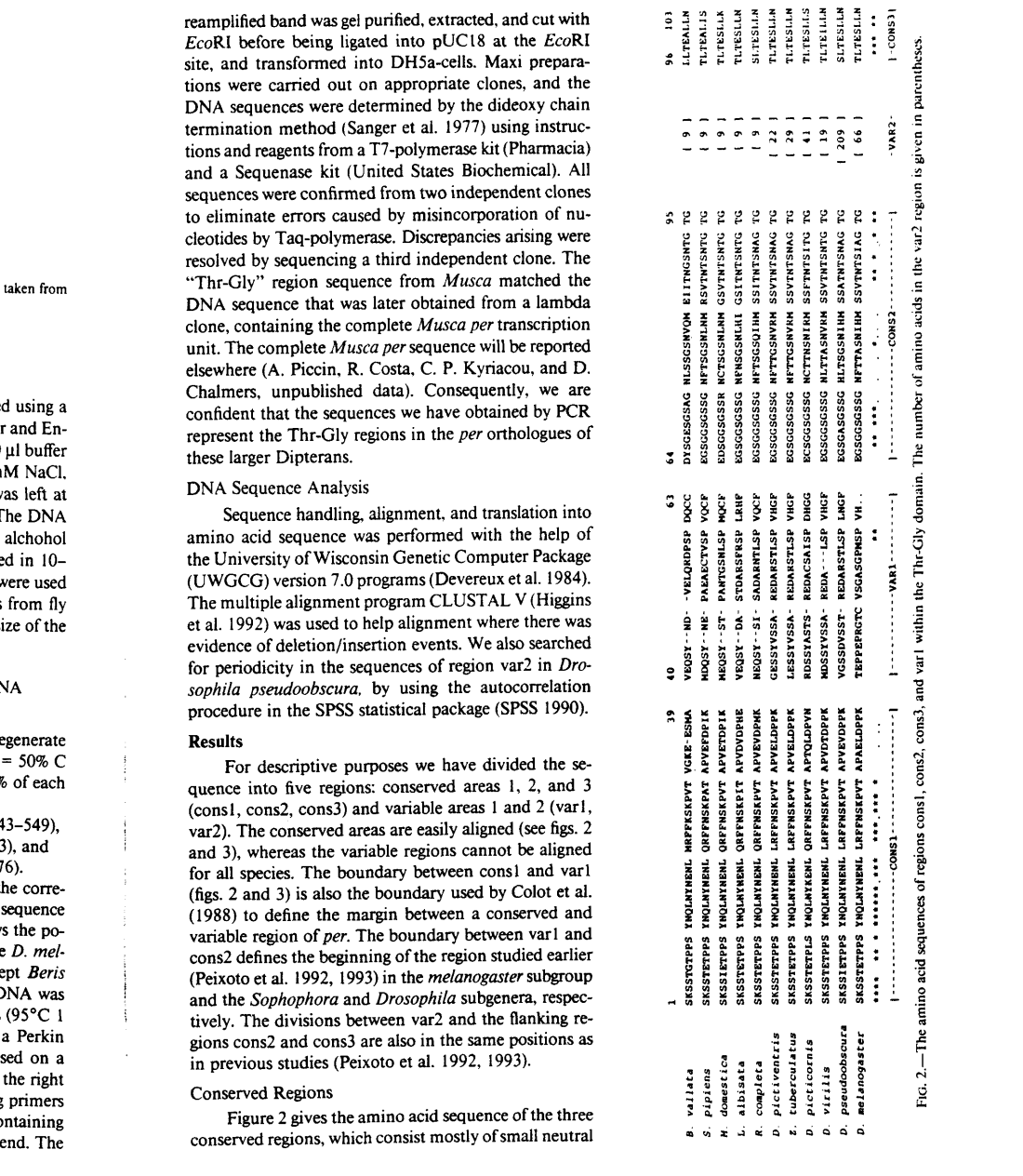


FIG. 2.—The amino acid sequences of regions cons1, cons2, cons3, and var1 within the Thr-Gly domain. The number of amino acids in the var2 region is given in parentheses.

amino acids except for the conserved positions between amino acids 10–25 in cons1 where these are mostly acidic residues interspersed with aromatic amino acids. A striking feature of the 79 positions in regions cons1, cons2, and cons3 (fig. 2) is that of the 42 completely conserved positions, 10 are serines. Serines also intrude into 7 of the remaining 37 nonconserved positions. Of the 12 conserved positions in cons2, 4 are glycines, with a further 3 nonconserved positions predominantly involving glycine. Ser-Gly pairs are particularly prominent in cons2, resembling the Ser-Gly pairs observed in mammalian proteoglycans (Ruoslahti 1988).

The DNA sequences in the conserved region 1, 2, and 3 (see fig. 3) were used to construct a phylogenetic tree using Kimura's (1980) two parameter distance and the neighbor-joining method (Saitou and Nei 1987) available in the CLUSTAL V package. Most bootstrap values (Felsenstein 1985*a*) fell below 50% except for one, which reached a level of 98%. This branch point placed *Rhagoletis completa* as the sister group of the Drosophilidae (fig. 4). We also generated a tree based on the protein sequence of regions cons1, 2, and 3, but this again produced low bootstrap values (data not shown). Given these low bootstrap values, we cannot place much confidence in our phylogeny. Possible reasons for this appear in the Discussion.

Variable Regions

Var1

The amino acid sequences of the region var1 are also illustrated in figure 2 and present a mixture of short stretches of two to three amino acids, which permits an alignment to be made. Deletions, insertions, and point mutations give this region the appearance of a rapidly evolving area (fig. 3). The *Drosophila melanogaster* sequence in particular stands out clearly from the others. It has a duplication of 9 nucleotides (CGGAGCCGC perfect repeat) beginning at nucleotide position 119 of var1, which interestingly is out of frame with the amino acid sequence, and a deletion of six bases at the end of the region. In both the *D. melanogaster* and *Drosophila pseudoobscura* sequences other slippage-like events appear to have taken place to account for the present constitution. for example, the sequence CAGTGG is duplicated (positions 150–155, 159–164) in *D. melanogaster*, as is the sequence GCAGCAC (positions 137–143, 161–167) in *D. pseudoobscura*. However, in both species, these duplications are out of frame with their amino acid sequences.

FIG. 4.—A phylogenetic tree based on the nucleotide sequences of regions cons1, cons2, and cons3. See text for details.

Var2

The hexapeptide.—Variable region 2 (var2) consists of mainly small, neutral amino acids (fig. 5), and the 11 species divide themselves into two distinct groups, the non-Drosophilidae having a short nine amino acid sequence beginning Thr-Ser-Ser (TSS) but with an expansion of this region within the Drosophilidae. In the outgroup *Beris vallata*, we note the hexapeptide TSSGSF, which is found in slightly different forms in all the nondrosophilids. In *Drosophila pictiventris*, the hexapeptide TSSAGG may have originally duplicated, lost both serines giving TAGG, and mutated to NGGG (ACNGCN—AA$\frac{C}{T}$GGN). This may have been followed by a duplication of the NG encoding sequence to give NGNGGG, and after further point mutations and slippage events, the sequence NGNGGSGA may have been produced. Within this sequence is embedded the pentapeptide sequence NGGSG, and derivatives of this sequence assume considerable significance below.

The pentapeptide.—Clearly, at least one of the two serines within the TSSGSF hexapeptide has been lost in the other drosophilids to create the TSGGG sequence we see in *Zaprionus tuberculatus*, for example. This sequence appears to have duplicated and mutated in *tuberculatus* to give the adjacent NSAGG pentapeptide, and traces of the duplication event can also be found in *Drosophila picticornis*, although the degeneracy of the second repeat is high, which perhaps suggests a deletion of the alanine from an NSAGG motif to NSGG (fig. 5). In *Drosophila virilis*, while a second pentapeptide NSASG follows the TSGGE, the second copy is not an obvious DNA repeat of the first. From figure 5 can be seen a number of repetitive motifs suggesting that a combination of slippage and point mutations may have generated the sequence NSASGSSN. Irrespective of how these pentapeptide and hexapeptide duplications occurred, it is clear from these drosophilids that by one means or another, a second pentapeptide, beginning NS or NG, has been derived from a pentapeptide beginning

FIG. 5.—The nucleotide sequences of the var2 region. The amino acids represented by the nucleotides are given below the sequences. The arrows depict repeats, which may be perfect or degenerated. The *Drosophila pseudoobscura* and *Drosophila virilis* sequences are taken from Colot et al. (1988) and the *Drosophila melanogaster* sequence from Citri et al. (1987).

Table 1
Dissimilarity Index of Adjacent versus Alternate Pentapeptides (a and b types) in *Drosophila pseudoobscura* Region var2

													Mean	SEM
5′ Region														
	a1	a2	b1	a3	b2	a4	b3	a5	a6	a7	b4	b5		
Adjacent . . .		4	8	9	6	5	4	5	2	3	7	5	5.27	.63
Alternate . . .			6	2	3	4	1	3	1	4	6	5	3.50	.50
Middle Region														
	b6	b7	b8	b9	b10	b11	b12	b13	b14	b15	b16			
Adjacent . . .	4	2	4	4	5	3	1	4	4	1	4		3.27	.41
Alternate . . .	2	2	2	5	2	3	3	4	2	5	3		3.00	.36
3′ Region														
	b15	b18	b19	b20	b21	a8	b22	b23	b24	a9	a10			
Adjacent . . .	3	4	2	3	2	8	7	3	6	6	7		4.64	.16
Alternate . . .	5	4	4	5	6	4	7	6	8	5			5.40	.43

NOTE.—The top line represents the number of nucleotides that are different between adjacent pentapeptides, and the lower line is the dissimilarity between alternative pentapeptide units. The boxes represent the likely units of duplication; for example, a3b2 probably duplicated from a2b1, b5 from b4, and b13b14 from b11 and b12. The dotted line box around a5a6 reflects the uncertainty in the derivation of unit a6, which may have duplicated from a5 or b3. The mean ± SEM of this index for adjacent and alternate repeat units for the 5′, middle, and 3′ regions is also provided (see text).

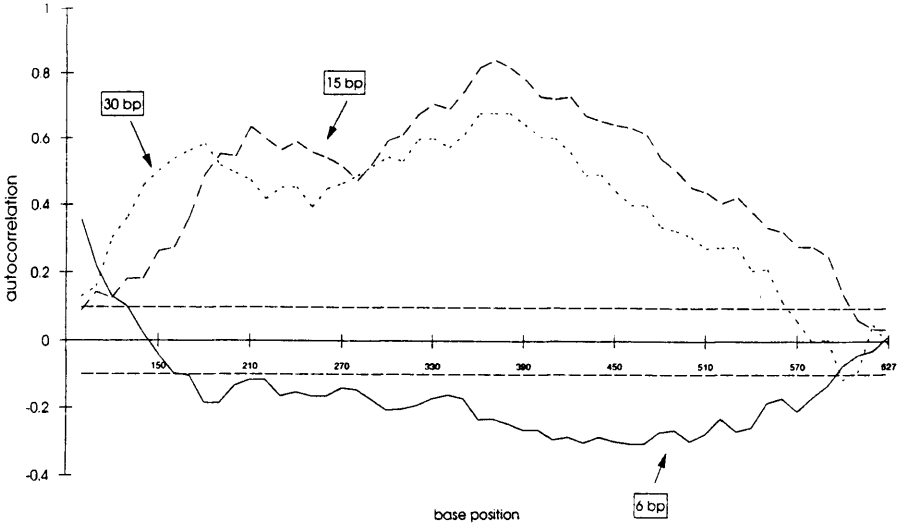


FIG. 6.—Autocorrelation analysis for the Adenine nucleotide of *Drosophila pseudoobscura* region var2. Autocorrelation was carried out by using a 100-bp window (see text). Three highly significant periodicities were observed corresponding to 6-, 15-, and 30-bp repeats, and their autocorrelation coefficients are plotted as the window moves through the sequence (see text). The dotted lines on the positive and negative side of $r = 0$ represent the 95% confidence limits.

TS. In both *D. pictiventris* and *D. virilis*, however, duplications within the N $\frac{S}{G}$ XXX pentapeptide have changed its structure slightly. As in *D. pictiventris*, it is also clear that slippage involving a single codon GNN may also have occurred in *D. picticornis* and *Z. tuberculatus*, which gave 11 copies in the former species and 10 in the latter (see fig. 5).

The Thr-Gly repeats and Drosophila pseudoobscura.—Just upstream of region var2 are the last few residues of region cons2 that have either one or two Thr-Gly pairs (figs. 2 and 3). The last four amino acids of region cons2 are either TGTG or AGTG (fig. 2), the alanine representing a single mutation of the ACN threonine codon (fig. 3). However, together with the first amino acid of region var2, be it a threonine or a glycine (in *Drosophila melanogaster* only), at least two Thr-Gly (or Gly-Thr) pairs are found at the interface between regions cons2 and var2 (except in *D. melanogaster*, but see below).

In *Drosophila pseudoobscura*, the Thr-Gly repeat has expanded in region var2 to give eight degenerate copies (nine including the first Thr-Gly pair in region cons2), which are interrupted briefly by a degenerate duplication of a threonine codon. Within this repetitive array emerges a TSGTG pentapeptide labeled *a1* in figure 5. This pentapeptide is similar to the TSGCG motif found in *Drosophila picticornis*, for example. Following this pentapeptide is another dipeptide TV and then the sequence TCASG (labeled *a2*) appears, which is again similar to the TSGXX pentapeptide observed in the other species. The following five amino acid residues NMDAN (labeled *b1*) may represent a mutated form of the NS $\frac{A}{G}$ XX pentapeptide found in the other drosophilids, such as NSGGK of *Drosophila picticornis*.

By comparing the number of nucleotide changes between adjacent pentapeptide units and alternating pentapeptide units, it is clear from table 1 that the greatest similarity is in the alternating pentapeptides at the beginning of the repeat array. In figure 5 and table 1, the two types of pentapeptide repeats have been labeled as either type a or type b to represent their possible derivations from either the TSGXX or NS $\frac{A}{G}$ XX encoding units, respectively. From table 1, it appears that the first few duplications involve the decapeptide unit, with the alternating pentapeptides being more similar to each other than the adjacent units. By dividing the pentapeptide array into a 5′, middle, and 3′ region, and computing the average nucleotide dissimilarity between adjacent and alternative pentapeptide units in these three regions, respectively, we obtained means of 5.27 versus 3.5 for the 5′ region, 3.27 versus 3.0 for the middle section, and 4.64 versus 5.4 for the 3′ region. These scores reflect the changing pattern of duplications along the repeat array. In the 5′ region alternate repeats are less dissimilar than adjacent ones, which suggests that the decapeptide is the unit of duplication. In the middle region, it is difficult to ascertain whether the duplications involve decapeptide or pentapeptide units because of the predominance of b-type repeats. In the 3′ region duplications of adjacent pentapeptides are slightly more common than decapeptides. Of further interest, is that the repeats in the middle region are more similar to each other than the repeats in the flanking regions (see table 1). As the dissimilarity data are not statistically independent, it is not appropriate to test for significant differences between the three regions.

For a more rigorous examination, we took each of the four nucleotides in turn in the region var2 of *D. pseudoobscura* and for each of the 627 positions scored the presence of that nucleotide as 1 and the absence as 0. We then used an autocorrelation procedure (SPSS 1990) and applied it to the first 100 bp of the sequence. The autocorrelation searches for periodicity in the data by correlating scores that are one position apart (lag 1), two positions apart (lag 2), and so on. Any periodicity in the data is reflected as a significant autocorrelation coefficient for a particular lag. We then moved the 100 bp "window" downstream by 10 bp and repeated the procedure, and we continued doing this until we had reached the end of the sequence, in all generating 53 autocorrelations for each nucleotide. By examining the data, it was clear that the highest correlations fell at lags of 6, 15, and 30, representing periodicities of 6, 15, and 30 bp. However, the relative strength of these correlations changed as the 100-bp window was moved in the 3′ direction. Figure 6 shows the adenine autocorrelation coefficients for the 6, 15, and 30-bp periodicities. Similar results are observed for G, C, and T (data not shown). Note how in the 5′ region, a 6-bp repeat shows the strongest correlation coefficient. As this 100-bp window moves downstream, the 30-bp repeat then shows the most prominent coefficients, until finally the 15-bp repeat predominates. These changes in the autocorrelation coefficient reflect the changing pattern of repeats in this region, with the Thr-Gly encoding hexamer giving way to the decapeptide encoding repeat (30 bp), followed by the 15-bp repeat. Note also that the correlation coeffi-

cients reach their highest values in the middle region with the 15-bp repeats, which thus reveals their greater homogeneity within this segment. These results confirm our preliminary observations from table 1 and our interpretation of the putative duplication events that have occurred in this region.

The Thr-Gly repeat and Drosophila melanogaster.—Finally, in *Drosophila melanogaster*, the Thr-Gly encoding hexamers have expanded dramatically. Note that *melanogaster* is the only species in which region var2 begins with a glycine rather than a threonine. This appears to be due to a duplication of the Gly-Thr-Gly encoding sequence that represents the last three amino acids of area cons2 (see figs. 2 and 3), which thereby produces $(Gly\text{-}Thr\text{-}Gly)_2$, and giving the interruption in the alternating Thr-Gly pattern (fig.5). Similar duplications of this Gly-Thr-Gly sequence are observed in *Drosophila orena* and *Drosophila erecta*, also from the *D. melanogaster* subgroup, but in both these species an additional duplication produces the sequences $(Gly\text{-}Thr\text{-}Gly)_3$ (Peixoto et al. 1992).

After this initial duplication in *D. melanogaster*, the Thr-Gly encoding hexamer expands to give at least 20 copies of the uninterrupted Thr-Gly dipeptide. As the ACNGGN repeat degenerates toward the end of the sequence, the pentapeptide NSGTG appears, which looks similar to the b-type pentapeptides of *Drosophila pseudoobscura*. This appears to be the only copy, although if we assume that the next Thr-Gly pair is a duplication of the last two residues at the end of the NSGTG sequence, the next 10 amino acids also appear to provide two pentapeptides, the second of which, KGGSA, bears a passing resemblance to unit b24 in *D. pseudoobscura*. Perhaps the immediately preceding pentapeptide TASSS is a degenerate form of the a-type unit of *D. pseudoobscura*.

Did the Thr-Gly repeat generate the hexa- and pentapeptide repeats?—It is noticeable that as the Thr-Gly repeat arrays degenerate in *Drosophila melanogaster* and *Drosophila pseudoobscura*, it is possible to derive both pentapeptide types (see unit a1, in *D. pseudoobscura* and unit b in *D. melanogaster;* fig. 5). Thus it is conceivable that the original hexapeptide sequence in the larger flies might have been initially generated by an ancient and perhaps modest expansion of the Thr-Gly repeat. Thus the sequence TSSGSA from *Rhagoletis completa*, for example, may have originally been derived from an expansion of the Thr-Gly repeat. Duplication and mutation of the threonine codon ACN to a serine codon, TCN, followed by a duplication will generate the residues TSS. The following glycine residue would have made up the glycine partner of the preceding threonine, and the following SA pair may represent another degenerate form of the ACNGGN Thr-Gly encoding hexamer. With this scenario, slippage within the Thr-Gly repeats may have initially generated the hexapeptide, which subsequently produced the pentapeptide by deletion of one of the serine residues.

Discussion

Previous reports have shown that within the *Drosophila* genus, the region corresponding to the *Drosophila melanogaster* Thr-Gly repeat has undergone rapid evolution, both in amino acid composition and in length (Colot et al. 1988; Costa et al. 1991; Peixoto et al. 1992, 1993). Our sequences from outside the genus reveal that at least two Thr-Gly pairs are encoded at the interface of regions cons1 and var2 in all species examined. Therefore, it would seem that long stretches of Thr-Gly repeats are not required for any function that is conserved within the Diptera. The Thr-Gly region plays a prominent role in the history of the development of ideas about *per* gene function, because the *melanogaster* Thr-Gly repeat is similar to the Ser-Gly repeats of mammalian proteoglycans (Ruoslahti 1988), and for some time *per* was believed to encode a proteoglycan (see Kyriacou and Hall, in press, for a review). However, the *per*-proteoglycan hypothesis does not appear to have passed the test of time (see Saez et al. 1992; Siwicki et al. 1992, Flint et al. 1993).

The corresponding var2 region within the other species of the Drosophilidae that we have examined contains other repetitive motifs. *Drosophila picticornis*, for example, shows evidence of slippage in a trinucleotide repeat and mutation to different amino acids including pairs or triplets of alanines, prolines, or glycines. Evidence of an even more degenerate form of this repeat can be seen in *Zaprionus tuberculatus*, which is thought to be phylogenetically remote from *D. picticornis* (Grimaldi 1990; DeSalle 1992; DeSalle and Grimaldi 1992). Outside the Drosophilidae, the construction of region var2 is dramatically different. We have detected no length variation in this region in the five species examined, and within this nine amino acid stretch is embedded a degenerate pentapeptide motif Thr-Ser-Gly-X-X. In the drosophilids *Drosophila pictiventris, D. picticornis, Z. tuberculatus*, and *Drosophila virilis*, this pentapeptide, which we call the "a" type, is followed by another that nearly always begins with an asparagine residue. This second pentapeptide (the "b" type) has apparently evolved by an initial duplication of the first TSGXX pentapeptide, followed in some species by further slippage within this unit. In *D. pictiventris*, however, an original hexapeptide encoding unit appears to have duplicated and, after further slippage events and point mutations, generated the b-type asparagine-encoding repeat unit. Thus by a series of duplications and point mutations, two pentapeptides, one encoding TSGXX, the other $NS\frac{A}{G}XX$, evolved in tandem.

Large numbers of both the a- and b-type forms of these pentapeptide units predominate within the var2 region in *Drosophila pseudoobscura*. An initial analysis revealed that the repeats in the middle section were more similar to each other than those at the ends (see table 1). This suggests that the slippage or unequal crossover events that are generating these repeat arrays are predominantly affecting the middle section of the array, and consequently the b-type repeats in this section may be more recent acquisitions. This is confirmed by comparisons among four *D. pseudoobscura* strains in which insertions and deletions of pentapeptide encoding units occur only in the region between units b4 and b15, representing largely the middle region (Costa et al. 1991). This kind of picture also emerges in the organization of the MS32 and MS1 minisatellite loci in the primates, in which the repeat units in the central positions of the array are more homogeneous than at the ends (Gray and Jeffreys 1991).

By applying an autocorrelation procedure to the *D. pseudoobscura* var2 sequence, we observed that in the 5′ region the 6-bp repeat showed the most significant periodic pattern, followed by the 30-bp repeat, and finally the 15-bp array as we moved in the 3′ direction. This reflected the changing pattern of duplication events in the array and confirmed our initial analyses (table 1)—that duplications of the decapeptide encoding units (a + b, 30 bp) had occurred in the 5′ region. Furthermore, as the dissimilarity in the 5′ region for alternate repeat units is 3.5 ± 0.5 nucleotides and is close to the value of 3.0 ± 0.36 for the middle region, then it would seem that both the middle and 5′ regions are of more recent origin than the 3′ region, where the dissimilarity index between adjacent or alternative units is generally higher (table 1). This pattern is reminiscent of the organization of the involucrin gene, in which the coding repeats appear to be added sequentially in the 5′ direction, with all anthropoid lineages sharing a common ancestral 3′ region (Djian and Green 1991; Philips et al. 1991; Green and Djian 1992). Thus the autocorrelation procedure has provided a new and elegant analysis of the structure of the repeat and has confirmed and extended our initial analysis using adjacent and alternate repeat dissimilarity (see table 1). It is somewhat surprising that this simple statistical test has not been used before to analyze repetitive DNA sequences. In fact, a rather different application involving autocorrelation has recently been used to analyze spatial patterns of DNA sequences (Barbujani and Bertorelle, submitted).

As the *D. melanogaster* and *D. pseudoobscura* Thr-Gly repeats degenerate at their 3′ ends, it was possible to derive both a and b pentapeptide units (fig. 5). This suggested that the sequence resembling the Thr-Gly repeat may possibly predate the origin of the hexa- and pentapeptide sequence. This reconstruction would imply that there may have been at least two Thr-Gly expansions, the first that generated the hexapeptide and the second that gave the more dramatic pattern seen in the *D. melanogaster* and *D. pseudoobscura* lineages. If this scenario is correct, then we might expect to find dipterans, or indeed insects from other orders, that have maintained Thr-Gly pairs but do not have the larger repeat units.

Theoretical models of repetitive DNA evolution (Gray and Jeffreys 1991; Harding et al. 1992) initially require two tandem repeats in order to allow for mispairing and the subsequent generation of high copy number of these repeats by unequal exchanges. After a certain period of expansion followed by a consolidation time or "dwell," in which large numbers of repeats are maintained, there is a contraction to a small copy number and eventually a collapse to a single stable copy. For example, in great apes, Old and New World monkeys, the MS32 noncoding hypervariable minisatellite appears stable and monomorphic, having only three to four repeat units. In humans, however, MS32 shows extreme variation in copy number (Gray and Jeffreys 1991). Computer simulations of minisatellite evolution using the best estimates of mutation rate and assuming a direct relationship between allele length revealed that only one in 250 lineages would attain a phase of large numbers (> 200) of repeat units (Gray and Jeffreys 1991). Similar results were obtained by Harding et al. (1992) using different mutational models. In both studies, most lineages maintained about two copies of the repeat, which eventually collapsed to a single, stable copy. However, it must be remembered that these models were developed for noncoding repeats and assume selective neutrality. For a coding sequence we would expect that some constraints must apply so that the repeat does not expand to an enormous copy number. Nevertheless, these models suggest that perhaps the Drosophilidae represent a rare lineage in which the expansion of the ancestral repeat unit or units has occurred, even though the general rule will be that this region of *per* will be short and nonvariable in most families.

We therefore suggest that two adjacent ancestral regions, one encoding a Thr-Gly dipeptide and another encoding a degenerate pentapeptide, for example,

TSGGG (consensus from *Z. tuberculatus, D. picticornis,* and *D. virilis*), expanded in copy number, in a reasonably independent fashion in the lineage that gave rise to the *Drosophila*. In some of these species, a modest and imperfect expansion of the five amino acid repeat occurred, whereas in other species such as *D. pseudoobscura* this repeat expanded more dramatically. *Drosophila pseudoobscura* is the only species we have analyzed that shows expansions of both the pentapeptide repeat and the dipeptide repeat. *Drosophila melanogaster*, on the other hand, shows a dramatic expansion of the dipeptide Thr-Gly repeat. It is extremely difficult to generate any kind of phylogeny based on these repeats, as we have attempted to do in previous publications based on the simple Thr-Gly encoding array in the *D. melanogaster* subgroup of species (Costa et al. 1991; Peixoto et al. 1992). However, in our reconstruction of the evolution of region var2, it appeared that *D. pictiventris*, by sharing the hexapeptide sequence with the nondrosophilids, may represent the initial split from the lineage giving rise to the other drosophilids. However, our formal phylogeny based on the more conserved regions surrounding region var2 could not confirm this and provided little resolution for the relationships between these species. The branching of the hoverfly (*Syritta pipiens*) was consistent with phylogenies based on morphological characters (see, e.g., McAlpine 1989; Wada 1991), but inside the Schizophora (includes the four remaining families), the phylogeny was debatable. This study places the Tephritidae as a sister group to the Drosophilidae, whereas McAlpine (1989) places it with the Psilidae and Griffiths (1972) places it with Muscidae. Our bootstrap values for this step were 98%. The relative branch points of the Drosophilidae cannot be resolved with confidence given the low bootstrap values of these branches within this family. For example, our phylogeny places *D. melanogaster* closer to another species, *Zaprionus*, rather than to its relatives *D. virilis* and *D. pseudoobscura* (Throckmorton 1975). Reasons for this could be stochastic in that the sequences involved are too short or that this region is under some kind of selection.

In fact, it has been suggested (Peixoto et al. 1993) that length variation in the var2 repetitive region may drive compensatory mutations in the flanking regions within the genus *Drosophila*, or *vice versa*. This idea was based on the observation (Peixoto et al. 1993) that the correlation between the protein divergence in the regions immediately flanking the repeat (cons2 + 3) and the length difference in the repetitive region (var2) between pairs of species was highly significant, whereas the correlation of length difference with the third position change found in the same flanking region was not. This difference between the two correlations persisted even when a number of corrections and other controls were performed. This intriguing result suggested that the correlation between the protein divergence in the flanking regions and the length difference was not simply proportional to the length of time since any two species had a common ancestor, because a similar correlation should have been observed with the synonymous position. Thus perhaps high mutation pressure in the form of slippage-driven events in the repeat had driven compensatory changes in the flanking regions. Alternatively, perhaps changes in the flanking region had driven length changes within the repeat.

To further examine this relationship, we correlated the pairwise length difference in region var2 within the drosophilids, with the protein divergence in the upstream region, cons1, which is a region further upstream to that used elsewhere (Peixoto et al. 1993). The correlation was nonsignificant (Spearman's $\rho = 0.11$, $n = 15$), whereas the correlation between the length of region var2 with the flanking region cons2 + 3 was significant ($\rho = 0.62$, $n = 15$, $P < 0.05$). Thus, it is difficult to argue that the significant positive correlation between the immediately flanking regions (cons2 + 3) and the length difference is simply an inevitable consequence of the time elapsed since a common ancestor, because no significant correlation was observed with the upstream region cons1. This result therefore provides some further support for the view that length changes in the repeat region are associated with "coevolutionary" changes in the flanking regions. However, our correlations must be treated with caution as no correction was made for phylogeny (see Felsenstein 1985*b*). The critical experiments to determine whether the length of region var2 has coevolved with the immediate flanking region will require an analysis of hybrid *per* genes between different *Drosophila* species in which the exact positions of the chimeric junctions involving the Thr-Gly repeat and the flanking sequences are manipulated.

Thus in the nondrosophilids we have a surprisingly stable picture concerning both of the repeats, which is puzzling because we have experimental evidence that the long Thr-Gly repeat in *D. melanogaster* plays a role in the "fine-tuning" of the temperature stability of the circadian phenotype (L. Sawyer and A. Peixoto, unpublished data). These observations may help to explain the European clinal distribution of the length polymorphism of the repeat in *D. melanogaster* (Costa et al. 1992). Perhaps, then, the longer repeats in the *Drosophila* species (Colot et al. 1988; Costa et al. 1991, 1992; Peixoto et al. 1992, 1993) and the variability in length both between and within these species have been recruited to play a role in the thermostability of the clock. If this does represent an additional adaptation, it is important to note that it has apparently arisen from the expansion of the repetitive region by "genome turnover events" such as slippage and unequal crossover, two mechanisms that play a pivotal role in the more general phenomenon of "molecular drive" (see, e.g., Dover 1989, 1993). However, our observations in the larger Dipterans then raise the question of how these flies can adapt their clock to the changes in temperature that they will inevitably experience within their geographical ranges. Temperature compensation is a cardinal feature of true clock phenotypes (Pittendrigh 1954), and if the long repeats in *D. melanogaster* serve this function, then how does the *Musca domestica per* protein, for example, compensate for such environmental challenges? Perhaps the answer will be in an examination of the sequence flanking the Thr-Gly region from *Musca* strains collected at different latitudes.

In summary, we have presented *per* gene sequences outside the *Drosophila* genus. The repetitive region of *per* is based on two different repetitive sequences, one encoding a pentapeptide and the other a dipeptide. Outside the Drosophilidae this region appears to be extremely stable, but within the Drosophilidae both types of repeats have undergone expansion. Our DNA sequence analysis coupled to interspecific transformation studies may reveal the functional significance (if any) of the different types of Thr-Gly regions found in these species.

Acknowledgments

We wish to acknowledge a grant from the Commission of The European Communities to R.C. and C.P.K., a Brazilian Conselito Nacional de Desenvolvimento Cientifico e Technologico scholarship to A.A.P., an ERASMUS studentship to J.N., an Agricultural and Fisheries Research Council grant to C.P.K., and a British Council/MURST award for Anglo-Italian cooperation. We also thank John Mousley, from the Leicester Museum, for providing *Beris vallata* and *Syritta pipiens*.

APPENDIX

The Classification of Flies Used in This Study

Order—Diptera
- Suborder—Brachycera
  - Family—Stratiomyidae—*Beris vallata*
- Suborder—Cyclorrhapha
  - Division—Aschiza
    - Family—Syrphidae—*Syritta pipiens*
  - Division—Schizophora
    - Section—Acalyptrate
      - Family—Psilidae—*Loxocera albisata*
      - Family—Tephritidae—*Rhagoletis completa*
      - Family—Drosophilidae
        - Genus—Hirtodrosophila—*Drosophila pictiventris*
        - *Genus—Zaprionus—Zaprionus tuberculatus*
        - Genus—Drosophila
          - Sub-genus—Sophophora—*Drosophila melanogaster*, *Drosphila pseudoobscura*
        - Genus—Drosophila
          - Subgenus Drosophila—*Drosophila virilis*
        - Hawaiian species, Picturewinged group—*Drosophila picticornis*
    - Section—Calyptrate
      - Family—Muscidae—*Musca domestica*

LITERATURE CITED

BARBUJANI, B., and G. BERTORELLE. Analysis of DNA diversity by spatial correlation (submitted).

BURBACH, K. M., A. POLAND, and C. A. BRADFIELD. 1992. Cloning of the Ah-receptor cDNA reveals a distinctive ligand activated transcription factor. Proc. Natl. Acad. Sci. USA **89**:8185–8189.

CITRI, Y., H. V. COLOT, A. C. JACQUIER, Q. YU, J. C. HALL, D. BALTIMORE, and M. ROSBASH. 1987. A family of unusually spliced and biologically active transcripts is encoded by a *Drosophila* clock gene. Nature **326**:42–47.

COLOT, H. V., J. C. HALL, and M. ROSBASH. 1988. Interspecific comparison of the period gene of *Drosophila* reveals large blocks of non-conserved coding DNA. EMBO J. **7**:3929–3937.

COSTA, R. A., A. A. PEIXOTO, G. BARBUJANI, and C. P. KYRIACOU. 1992. A latitudinal cline in a *Drosophila* clock gene. Proc. R. Soc. Lond. [Biol.] **250**:43–49

COSTA, R. A., A. A. PEIXOTO, J. R. THACKERAY, R. DALGLEISH, and C. P. KYRIACOU. 1991. Length polymorphism in the threonine-glycine encoding repeat region of the *period* gene in *Drosophila*. J. Mol. Evol. **32**:238–246.

DE SALLE, R. 1992. The origin and possible time of divergence of the Hawaiian Drosophilidae: evidence from DNA sequences. Mol. Biol. Evol. **9**:905–916.

DE SALLE, R., and D. GRIMALDI. 1992. Characters and the systematics of Drosophilidae. J. Hered. **83**:182-188.

DEVEREUX, J. P., P. HAEBERLI, and O. SMITHIES. 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–396.

DJIAN, P., and H. GREEN. 1991. Involucrin gene of tarsioids and other primates: alternatives in evolution of the segment of repeats. Proc. Natl. Acad. Sci. USA **88**:5321–5325.

DOVER, G. A. 1989. Slips, strings and species. Trends Genet. **5**:100–103.

———. 1993. Evolution of genetic redundancy for advanced players. Curr. Opin. Genet. Dev. **3**:902–910.

EWER J., M. HAMBLEN-COYLE, M. ROSBASH, and J. C. HALL. 1990. Requirement for *period* gene expression in the adult and not during development for locomotor activity rhythms of imaginal *Drosophila melanogaster*. J. Neurogenet. **7**:31–73.

FELSENSTEIN, J. 1985*a*. Confidence limits on phylogenies: an approach using bootstrap. Evolution **39**:783–791.

———. 1985*b*. Phylogenies and the comparative method. Am. Nat. **125**:1–15.

FLINT, K. K., M. ROSBASH, and J. C. HALL. 1993. Transfer of dye among salivary gland cells is not affected by genetic variations of the *period* clock gene in *Drosophila melanogaster*. J. Membrane Biol. **136**:333–342.

GLOOR, G., and W. ENGELS. 1990. Single fly DNA preps for PCR. Drosophila Inf. Newslett. **1** (January).

GRAY, I. C., and A. J. JEFFREYS. 1991. Evolutionary transience of hypervariable minisatellites in man and primates. Proc. R. Soc. Lond. [Biol.] **243**:241–250.

GREEN, H., and P. DJIAN. 1992. Consecutive actions of different gene-altering mechanisms in the evolution of Involucrin. Mol. Biol. Evol. **9**:977–1017.

GRIFFITHS, G. C. P. 1972. The phylogenetic classification of Diptera Cyclorrhapha, with special reference to the structure of the male postabdomen. Ser. Entomol. **8**:1–340.

GRIMALDI, D. 1990. A phylogenetic revised classification of genera in the Drosophilidae. Bull. Am. Mus. Nat. Hist. **197**: 1–139.

HARDIN, P. E., J. C. HALL, and M. ROSBASH. 1990. Feedback of the *Drosophila period* gene product on circadian cycling of its messenger RNA levels. Nature **345**:536–540.

HARDING, R. M., A. J. BOYCE, and J. B. CLEGG. 1992. The evolution of tandemly repetitive DNA: recombination rules. Genetics **132**:847–859.

HIGGINS, D. G., A. J. BLEASBY, and R. FUCHS. 1992. CLUSTAL V: improved software for multiple sequence alignment. Comp. Appl. Biosci. **8**:189–191.

HOFFMAN, E. C., H. REYES, F. F. CHU, F. SANDER, L. H. CONLEY, B. A. BROOKS, and O. HANKINSON. 1991. Cloning of a factor required for activity of the Ah (Dioxin) receptor. Science **252**:954–958.

HUANG, Z. J., I. EDERY, and M. ROSBASH. 1993. PAS is a dimerization domain common to *Drosophila* period and several transcription factors. Nature **364**:259–263.

JACKSON, F. R., T. A. BARGIELLO, S. H. YUN, and M. W. YOUNG. 1986. Product of the *period* locus of *Drosophila* shares homology with proteoglycans. Nature **320**:185–188.

KIMURA, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. **16**:111–120.

KONOPKA, R. J. and S. BENZER. 1971. Clock mutants of *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA **68**:2112–2116.

KYRIACOU, C. P. 1994. Clock research *per*ring along: it's about time! Trends Genet. **10**:69–71.

KYRIACOU, C. P., and J. C. HALL. 1980. Circadian rhythm mutations in *Drosophila melanogaster* affect short term fluctuations in the males courtship song. Proc. Natl. Acad. Sci. USA **77**:6929–6933.

———. 1986. Interspecific genetic control of courtship song production and reception in *Drosophila*. Science **232**:494–497.

———. 1989. Spectral analysis of *Drosophila* courtship song rhythms. Anim. Behav. **37**:850–859.

———. 1994. Genetic and molecular analysis of *Drosophila* behaviour. Adv. Genet. (in press).

KYRIACOU, C. P., M. VAN DER BERG, and J. C. HALL. 1990*a*. *Drosophila* courtship song cycles in normal and period mutant males revisited. Behav. Genet. **20**:617–644.

KYRIACOU, C. P., M. OLDROYD, M., J. WOOD, M. SHARP, and M. HILL. 1990*b*. Clock mutations alter developmental timing in *Drosophila*. Heredity **64**:395–401.

LIU, X., L. J. ZWIEBEL, D. HINTON, S. BENZER, J. C. HALL, and M. ROSBASH. 1992. The *period* gene encodes a predominantly nuclear protein in adult *Drosophila*. J. Neurosci. **12**:2735–2744.

MCALPINE, J. F. 1989. Phylogeny and classification of the Muscomorpha. Pp. 1397–1518 *in* J. F. MCALPINE, ed. Manual of Nearctic Diptera. Vol. **3**, Monograph No. **32**. Biosystematic Research Centre, Ottawa.

NAMBU, J. R., J. O. LEWIS, K. A. WHARTON, and S. T. CREWS. 1991. The *Drosophila* single minded gene encodes a Helix-Loop-Helix protein that acts as a master regulator of CNS midline development. Cell **67**:1157–1167.

PEIXOTO, A. A., S. CAMPESAN, R. COSTA, and C. P. KYRIACOU. 1993. Molecular evolution of a repetitive region within the *per* gene of *Drosophila*. Mol. Biol. Evol. **10**:127–139.

PEIXOTO, A. A., R. COSTA, J. C. WHEELER, J. C. HALL, and C. P. KYRIACOU. 1992. Evolution of the threonine-glycine repeat region of the period gene in the *melanogaster* species subgroup of *Drosophila*. J. Mol. Evol. **35**:711–719.

PHILLIPS, M., R. H. RICE, P. DJIAN, and H. GREEN. 1991. The Involucrin genes of the white-fronted Capuchin and Cottontop Tamarin: the Platyrrhine middle region. Mol. Biol. Evol. **8**:579–591.

PITTENDRIGH, C. S. 1954. On the clock systems controlling emergence time in *Drosophila*. Proc. Natl. Acad. Sci. USA **40**:1018–1029

ROSATO, E., A. A. PEIXOTO, G. BARBUJANI, R. COSTA, and C. P. KYRIACOU. Molecular polymorphism in the period gene of *Drosophila simulans* (submitted).

RUOSLAHTI, E., 1988. Structure and biology of proteoglycans. Annu. Rev. Cell Biol. **4**:229–255.

SAEZ, L., M. W. YOUNG, M. K. BAYLIES, G. GASIC, T. A. BARGIELLO, and D. C. SPRAY. 1992. *Per*—no link to gap junctions. Nature **360**:542.

SAITOU, N., M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

SANGER, F. F., S. NICKLEN, and A. R. COULSON. 1977. DNA sequencing with chain terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.

SIWICKI, K. K., K. FLINT, J. C. HALL, M. ROSBASH, and D. C. SPRAY. 1992. The *Drosophila period* gene and dye coupling in larval salivary glands: a re-evaluation. Biol. Bull. **183**:340–341.

SPSS Inc. 1990. SPSS/PC+ trends manual. SPSS, Chicago.

THROCKMORTON, L. H. 1975. The phylogeny, ecology and geography of Drosophila. Pp. 421–469 *in* R. C. KING, ed. Handbook of genetics. Vol. 3. Plenum, New York.

WADA, S. 1991. Morphologiche Indizien für das unmittelbare schwestergruppenuer-Laltnis der Schizophora mit der synphoidea (Aschiza) in der phylogenetischen Systematik der Cyclorrhapha. (Diptera: Brachicera) J. Nat. Hist. **25**:1531–1570.

WHEELER, D. A., C. P. KYRIACOU, M. GREENACRE, Q. YU, J. RUTILA, M. ROSBASH, and J. C. HALL. 1991. Molecular transfer of a species-specific courtship behaviour from *Drosophila simulans* to *Drosophila melanogaster*. Science **251**: 1082–1085.

YU, Q., H. V. COLOT, C. P. KYRIACOU, J. C. HALL, and M. ROSBASH. 1987. Behaviour modification by in vitro mutagenesis of a variable region within the period gene of *Drosophila*. Nature **326**:765-769.

PAUL SHARP, reviewing editor

Received March 15, 1994

Accepted June 16, 1994

BIBLIOGRAPHY

BIBLIOGRAPHY

Aguade M. (1988). Nucleotide-sequence comparison of the *rp49* gene region between *Drosophila subobscura* and *Drosophila melanogaster*. Mol. Biol. Evol. 5 : 433-441.

Aizawa M. and Yoshino H. (1986). Genetics of circadian eclosion rhythm in the house-fly, *Musca domestica*. Zoolog. Sci. 3 : 1102. Meeting abstract.

Amir S. and Stewart J. (1996). Resetting of the circadian clock by a conditioned stimulus. Nature 379 : 542-545.

Aronson B., Johnson K. and Dunlap J.C. 1991. The circadian clock locus *frequency*: a single ORF defines period length and temperature compensation. Proc. Natl. Acad. Sci. USA 91 : 7683-7687.

Aronson B.D., Johnson K.A., Loros J.J. and Dunlap J.C. (1994). Negative feedback defining a circadian clock: autoregulation of the clock gene *frequency*. Science 263 : 1578-1584.

Ashburner M. (1989). *Drosophila*. A laboratory handbook. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Aviv H. and Leder P. (1972). Purification of biologically active globin messanger RNA by cromatography on oligothymidylic acid-cellulose. Proc. Natl. Acad. Sci. USA 69 : 1408-1412.

Balakrishnan B. and Rodrigues V. (1991). The *Shaker* and *shaking-B* genes specify elements in the processing of gustatory informations in *Drosophila melanogaster*. J. Exp. Biol. 157 : 161-181.

Bargiello T.A. and Young M.W. (1984). Molecular genetics of a biological clock in *Drosophila*. Proc. Natl. Acad. Sci. USA. 81 : 2142-2146.

Bargiello T.A., Jackson F.R. and Young M.W. (1984). Restoration of circadian behavioural rhythms by gene transfer in *Drosophila*. Nature 312 : 752-754.

Bargiello T.A., Saez L., Bayles M.K., Gasik G., Young M.W. and Spray D.C. (1987). The *Drosophila* clock gene *per* affects intercellular junctional communication. Nature 328 : 686-691.

Baylies M.K., Bargiello T.A., Jackson F.R and Joung M.W. (1987). Changes in abundance or structure of the *per* gene product can alter periodicity of the *Drosophila* clock. Nature 326 : 390-392.

Baylies M.K., Vosshall L.B., Sehgal A. and Young M.W. (1992). New short period mutations of the *Drosophila* clock gene *per*. Neuron 9 : 575-581.

Ben-Shlomo R., Ritte U. and Nevo E. (1996). Circadian rhythm and the *per* ACNGGN repeat in the mole rat, *Spalax ehrenbergi*. Behav. Genet. 26 : 177-184.

Bertholf L.M. (1932). The extend of the spectrum for *Drosophila* and the distribution of stimulative in it. Z. Vergl. Physiol. 18 : 32-64.

Boudreaux H.B. (1978). Arthropod phylogeny with special reference to insects. Wiley-Interscience Press, New York, USA.

Bourdon M.A., Oldberg A. and Ruoslahti E. (1985). Molecular cloning and analysis of a Chondroitin sulfate proteoglycan cDNA. Proc. Natl. Acad. Sci. USA 82 : 1321-1325.

Brown J.K.M. (1994). Bootstrap hypotesis tests for evolutionary trees and other dendrograms. Proc. Natl. Acad. Sci. USA. 91 : 12293-12297.

Burbach K.M., Poland A. and Bradfield C.A. (1992). Cloning of the Ah-receptor cDNA reveals a distinctive ligand-activated transcription factor. Proc. Natl. Acad. Sci. USA 89 : 8185-8189.

Castiglione-Morelli M.A., Guantieri V., Villani V., Kyriacou C.P., Costa R. and Tamburro A.M. (1995). Conformational study of the Thr-Gly repeat in the *Drosophila* clock protein PERIOD. Proc. R. Soc. London 260 : 155-163

Cheng Y.Z. and HardinP.E. (1998). *Drosophila* photoreceptors contain an autonomous circadian oscillator that can function without period mRNA cycling. J. Neurosci. 18 : 741-750.

Chiba Y., Uki M., Kawasaki Y., Matsumoto A. and Tomioka K. (1993). Entrainability of circadian activity of the mosquito *Culex pipiens pallens* to 24-Hr temperature cycles with special reference to involvement of multiple oscillators. J. Biol. Rhythms 8 : 211-220.

Chomczynski P. and Sacchi N. (1987). Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. Anal. Biochem. 162 : 156-159.

Chou P.V. and Fasman G.D. (1974). Prediction of protein conformation. Biochemistry 13 : 222-245.

Citri Y., Colot H.V., Jacquier A.C., Yu Q., Hall J.C., Baltimore D. and Rosbash M. (1987). A family of unusually spliced biologically active transcripts encoded by a *Drosophila* clock gene. Nature 326 : 42-47.

Colot H.V., Hall J.H. and Rosbash M. (1988). Interspecific comparison of the *period* gene of *Drosophila* reveals large blocks of non-conserved coding DNA. EMBO J. 7 : 3929-3937.

Costa R.A., Peixoto A.A., Thackeray J.R., Dalgleish R. and Kyriacou C.P. (1991). Length polymophism in the threonine-glycine encoding repeat region of the *period* gene in *Drosophila*. J. Mol. Evol. 32 : 238-246.

Costa R.A., Peixoto A.A., Barbujani G. and Kyriacou C.P. (1992). A latitudinal cline in a *Drosophila* clock gene. Proc. R. Soc. London 250 : 43-49.

Cote G.G. and Brody S. (1977). Circadian rhytms in *Drosophila melanogaster*: analysis of period as a function of gene dosage at the *per* (*period*) locus. J. Theor. Biol. 121 : 487-503.

Coyne J.A. (1992). Genetics and speciation. Nature 355 : 511-515.

Courey A.J. and Tjian R. (1988). Analysis of SP1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell 55 : 887-898.

Cowan T.M. and Siegel R.W. (1986). *Drosophila* mutations that alter ionic conduction disrupt acquisition and retention of a conditioned odor avoidance response. J. Neurogenet. 3 : 187-201.

Cowling D.E. and Burnet B. (1981) Courtship songs and genetic control of their acoustic characteristics in sibling species of the *D. melanogaster* subgroup. Anim. Behav. 29 : 924-935.

Crews S.T., Thomas J.B. and Goodman C.S. (1988). The *Drosophila single-minded* gene encodes a nuclear protein with sequence similarity to the *per* gene product. Cell 52 : 143-151.

Crosthwaite S.K., Loros J.J. and Dunlap J.C. 1995. Light resetting of a circadian clock is mediated by a rapid increase in *frequency* transcript. Cell 81 : 1003-1012.

Crosthwaite S.K., Dunlap J.C. and Loros J.J. 1997. *Neurospora wc-1* and *wc-2*: transcription, photoresponses, and the origins of circadian rhythmicity. Science 276 : 763-769.

Cummings F.W. (1975). A biochemical model of the circadian clock. J. Theor. Biol. 55 : 455-470.

Curtin K.D., Huang Z.J. and Rosbash M. (1995). Temporally regulated nuclear entry of the *Drosophila period* protein contributes to the circadian clock. Neuron 14 : 365-372.

Cymborowski B., Hong S.F., McWatters H.G. and Saunders D.S. (1996). S-antigen antibody partially blocks entrainment and the effects of constant light on the circadian rhythm of locomotor activity of the adult blow fly, *Calliphora vicina*. J. Biol. Rhythms 11 : 68-74.

Dayhoff M.O. (1979). Atlas of Protein Sequence and Structure, Volume 5, Supplement 3, 1978. National Biomedical Research Foundation, Washington, D.C.

de Belle J.S. and Sokolowski M.B. (1989a). Rover/sitter foraging behaviour in *Drosophila melanogaster*. Genetic localization to chromosome-2L using compound autosomes. J. Insect Behav. 2 : 291-299.

de Belle J.S., Hilliker A.J. and Sokolowski M.B. (1989b). Genetic localization of *foraging* (*for*): a major gene for larval behaviour in *Drosophila melanogaster*. Genetics 123 : 157-164.

Dembinska M.E., Stanewsky R., Hall J.C. and Rosbash M. (1997). Evidence that temporal proteolysis contributes to Per protein cycling. J. Biol. Rhythms.

Devereux J., Haeberli P. and Smithies O. (1984). A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. 12 : 387-396.

Dover G.A. (1982). Molecular drive: a cohesive mode of species evolution. Nature 299 : 111-117.

Dover G.A. and Flavell R.B. (1984). Molecular coevolution: DNA divergence and maintenance of function. Cell 38 : 622-623.

Dover G.A. (1987). DNA turnover and the molecular clock. J. Mol. Evol. 26 : 47-58.

Dowse H.B., Dushay M.S., Hall J.C. and Ringo J.M. (1989). High resolution analysis of locomotor activity rhythms in *disconnected*, a visual system mutant of *Drosophila melanogaster*. Behav. Genet. 19 : 529-542.

Dunlap J.C. (1993). Genetic analysis of circadian clocks. Annu. Rev. Physiol. 55 : 683-728.

Dushay M.S., Rosbash M. and Hall J.C. (1989). The *disconnected* visual system mutations in *Drosophila melanogaster* drastically disrupt circadian rhythms. J. Biol. Rhythms 4 : 1-27.

Dushay M.S., Konopka R. J., Orr D., Greenacre M., Kyriacou C.P., Rosbash M. and Hall J.C. (1990). Phenotypic and genetic analysis of *Clock*, a new circadian rhythm mutant in *Drosophila melanogaster*. Genetics 125 : 557-558.

Dushay M.S., Rosbash M. and Hall J.C. (1992). Mapping the *Clock* rhythm mutation to the *period* locus of *Drosophila melanogaster* by germline transformation. J. Neurogenet. 8 : 173-179.

Edery I., Zwiebel L.J., Dembinska M.E. and Rosbash M. (1994). Temporal phosphorylation of the *Drosophila* period protein. Proc. Natl. Acad. Sci. USA 91 : 2260-2264.

Ema M., Suzuki M., Morita M., Hirose K., Sogawa K., Matsuda Y., Gotoh O., Saijoh Y., Fujii H., Hamada H. and Fujiikuriyama Y. (1996). cDNA cloning of a murine homolog of *drosophila single-minded*, its messenger-RNA expression in mouse development, and chromosome localization. Biochem. Biophys. Res. Commun. 218 : 588-594.

Emery I.F., Noveral J.M., Jamison C.F. and Siwicki K.K. (1997). Rhythms of *Drosophila period* gene expression in culture. Proc. Natl. Acad. Sci. USA 94 : 4092-4096.

Emili A., Greenblatt J. and Ingles J.C. (1994). Species-specific interaction of the glutamine-rich activation domains of SP1 with the TATA box-binding protein. Mol. Cell. Biol. 14 : 1582-1593.

Ehret C.F. and Trucco E. (1967). Molecular models for the circadian clock. I. The chronon concept. J. Theor. Biol. 15 : 240-262.

Ewer J., Rosbash M. and Hall J.C. (1988). An inducible promoter fused to the *period* gene of *Drosophila* conditionally rescues adult *per*-mutant arrhythmicity. Nature 333 : 82-84.

Ewer J., Hamblen-Coyle M., Rosbash M. and Hall J.C. (1990). Requirement for *period* gene expression in the adult and not during development for locomotor activity rhythms of imaginal *Drosophila melanogaster*. J. Neurogenet. 7 : 31-73.

Ewer J., Frisch B., Hamblen-Coyle M.J., Rosbash M. and Hall J.C. (1992). Expression of the *period* clock gene within different cell types in the brain of *Drosophila* adults and mosaic analysis of these cells' influence on circadian behavioral rhythms. J. Neurosci. 12 : 3321-3349.

Federoff N.V. (1979). On spacers. Cell 16 : 697-710.

Feinberg A.P. and Vogelstein B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal. Biochem. 137 : 6-13.

Feldman J.F. and Hoyle M.N. (1973). Isolation of circadian clock mutants of *Neurospora crassa*. Genetics 75 : 605-613.

Feliciello I. and Chinali G. (1993). A modified alkaline lysis method for the preparation of highly purified plasmid DNA from *Escherichia coli*. Anal. Biochem. 212 : 394-401.

Felsenstein J. (1985). Confidence limits in phylogenies: an approach using the bootstrap. Evolution 39 : 783-791.

Felsenstein J. (1988). Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. 22 : 521-565.

Fitch W.M. and Margoliash. E. (1967). Construction of phylogenetic trees. Science 155 : 279-284.

Flint K., Rosbash M. and Hall J.C. (1993). Transfer of dye among salivary gland cells is not affected by genetic variations of the *period* clock gene in *Drosophila melanogaster*. J. Membr. Biol. 136 :333-342.

Francis C. and Sargent M.L. 1979. Effects of temperature perturbations on circadian conidation in *Neurospora*. Plant Physiol. 64 : 1000-1009.

Franco M.G., Rubini P.G. and Vecchi M. (1982). Sex-determinants and their distribution in various populations of *Musca domestica* L. of western Europe. Genet. Res. Camb. 40 : 279-293.

Frisch B., Hardin P.E., Hamblen-Coyle M.J., Rosbash M. and Hall J.C. (1994). A promoterless *period* gene mediates behavioral rhythmicity and cyclical *per* expression in a restricted subset of the *Drosophila* nervous system. Neuron 12 : 555-570.

Gailey D.A. and Hall J.C. (1989) Behaviour and cytogenetics of *fruitless* in *Drosophila melanogaster*: different courtship defects caused by separate, closely linked lesions. Genetics 121 : 773-785.

Garceau N.Y., Liu Y., Loros J.J. and Dunlap J.C. 1997. Alternative initiation of translation and time-specific phosphorylation yeld multiple forms of the essential clock protein Frequency. Cell 89 : 469-476.

Giebultowicz J.M. and Hege D.M. (1997). Circadian clock in Malpighian tubules. Nature 386 : 664.

Gekakis N., Saez L., Delahaye-Brown A.M., Myers M., Sehgal A., Young M.W. and Weitz C.J. (1995). Isolation of *timeless* by Per protein interaction: defective interaction between *timeless* protein and long-period mutant Per^{L}. Science 270 : 811-815.

Gloor G. and Engels W. (1990). Single-fly DNA preps for PCR. Drosophila information newsletter vol 1.

Goldbeter A. (1995). A model for circadian oscillations in the *Drosophila period* protein (PER). Proc. R. Soc. London 261 : 319-324

Gooch V.D., Wehseler R.A. and Gross C.G. (1994). Temperature effects on the resetting of the phase of the *Neurospora* circadian rhythm. J. Biol. Rhythms 9 : 83-94.

Goodwin B.C. (1965). Oscillatory behavior in enzymatic control processes. In: Advances in enzyme regulation. Pergamon Press, Oxford, England.

Gray I. and Jeffreys A.J. (1991). Evolution transience of hypervariable minisatellites in man and primates. Proc. R. Soc. London. Biol. 243 : 241-253.

Gribskov M., Burgess R.R. and Devereux J. (1986) Pepplot, a protein secondary structure-analysis program for the UWGCG sequence-analysis software package. Nucleic Acids Res. 14 : 327-334.

Grummt I., Roth E. and Paule M.R.R. (1982). Ribosomal RNA transcription *in vitro* is species-specific. Nature 296 : 173-176.

Gwinner E. (1986). Internal rhythms in bird migration. Sci. Am. 254 : 76-84.

Hall J.C. and Rosbash M. (1987). Genetic and molecular analysis of biological rhythms. J. Biol. Rhythms 2 : 153-178.

Hall J.C. (1994). Pleiotropy of behavioral genes. In flexibility and constraint in behavioral systems (edited by Greenspan R.J. and kyriacou C.P.). Wiley Press, Chichester.

Hall J.C. (1995). Tripping along the trail to the molecular mechanisms of the biological clocks. Trends Neurosci. 18 : 230-240.

Hall J.C. (1996). Are cycling gene products as internal Zeitgebers no longer the Zeitgeist of chronobiology? Neuron 17 : 799-802.

Hamblen M., Zehring W.A., Kyriacou C.P., Reddy P., Yu Q., Wheeler D.A., Zwiebel L.J., Konopka R.J., Rosbash M. and Hall J.C. (1986). Germ-line transformation involving DNA from the *period* locus in *Drosophila melanogaster*: overlapping genominc fragments that restore circadian and ultradian rhythmicity to per^0 and per^- mutants. J. Neurogenet. 3 : 249-291.

Hamblen-Coyle M., Konopka R.J., Zweibel L.J., Colot H.V., Dowse H.B., Rosbash M. and Hall J.C. (1989). A new mutation at the *period* locus of *Drosophila melanogaster* with some novel effects on circadian rhythms. J. Neurogenet. 5 : 229-256.

Han S.Z. and Engelmann W. (1987). Azadirachtin affects the circadian rhythms of locomotion in *Leucophaea maderae*. J. Interdiscipl. Cycle Res. 18 : 71-81.

Hao H., Allen D.L. and Hardin P.E. (1997). A circadian enhancer mediates Per-dependent mRNA cycling in *Drosophila*. Mol. Cell Biol. 17 : 3687-3693.

Hardin P.E., Hall J.C. and Rosbash M. (1990). Feedback of the *Drosophila period* gene product on circadian cycling of its messenger RNA levels. Nature 343 : 536-540.

Hardin P.E., Hall J.C. and Rosbash M. (1992a). Circadian oscillations in *period* gene mRNA levels are transcriptionally regulated. Proc. Natl. Acad. Sci. USA 89 : 11711-11715.

Hardin P.E., Hall J.C. and Rosbash M. (1992b). Behavioural and molecular analysis suggest that circadian output is disrupted by *disconnected* mutants in *D. melanogaster*. EMBO J. 11 : 1-6.

Hardin P.E. (1994). Analysis of mRNA cycling in *Drosophila* head and body tissues indicates that body oscillators behave differently from head oscillators. Mol. Cell. Biol. 14 : 7211-7218.

Hardin P.E. and Siwicki K.K. (1995). The multiple roles of *per* in the *Drosophila* circadian clock. Semin. in Neurosci. 7 : 15-25.

Harding R.W. and Turner R.V. (1981). Photoregulation of the carotenoid biosynthetic pathways in *albino* and *white collar* mutants of *Neurospora crassa*. Plant Physiol. 68 : 745-749.

Harding R.M., Boyce A.J. and Clegg J.B. (1992). The evolution of tandemly repetitive DNA: recombination rules. Genetics 132 : 847-859.

Harlow E. and Lane D. (1988). Antibodies. A laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, 522.

Helfrich C., Cymborowski B. and Engelmann W. (1985). Circadian activity rhythm of the house fly continues after optic tract severance and lobectomy. Chronobiol. Int. 2 : 19-32.

Helfrich-Forster C. and Engelmann W. (1995). The occurrence of clear rhythmic locomotor activity in individual *disconnected* mutants of *Drosophila melanogaster* is accompanied by the presence of lateral neurons in these flies. The Physiologist 38 : 6.2.

Helfrich-Forster C. (1996). *Drosophila* rhythms: from brain to behavior. Semin. Cell Dev. Biol. 7 : 791-802.

Hennig W. (1981). Insect phylogeny. Pitman Press, Bath, England.

Henze K., Badr A., Wettern M., Cerff R. and Martin W. (1995) A nuclear gene of eubacterial origin in *Euglena gracilis* reflects cryptic endosymbioses during protist evolution. Proc. Natl. Acad. Sci. USA 92 : 9122-9126.

Hess B. and Boiteux A. (1971). Oscillatory phenomena in biochemistry. Annu. Rev. Biochem. 40 : 237-258.

Higgins D.G. and Sharp P.M. (1988). CLUSTAL: a package for performing multiple sequence alignments on a microcomputer. Gene 73 : 237-244

Hoffman E.C., Reyes H., Chu F.F., Sander F., Conley L.H., Brooks B.A. and Hankinson O. (1991). Cloning of a factor required for activity of the Ah (dioxin) receptor. Science 252 : 954-958.

Huang T.C. and Grobbelaar N. (1995). The circadian clock in the prokaryote *Synechococcus* RF-1. Microbiology 141 : 535-540.

Huang Z.J., Edery I. and Rosbash M. (1993). PAS is a dimerization domain common to *Drosophila* Period and several transcription factors. Nature 364 : 259-262.

Huang Z.J., Curtin K.D. and Rosbash M. (1995). PER protein interactions and temperature compensation of a circadian clock in *Drosophila*. Science 267 : 1169-1172.

Ishida N., Noji S., Ono K, Koyama E., Nohno T., Taniguchi S., Togunaga A., Fujii T. and Mitsui Y. (1991). Diurnal regulation of *per* repeat mRNA in the suprachiasmatic nucleus in rat brain. Neurosci. Lett. 122 : 113-116.

Ishida N., Matsui M., Nishimatsu S., Murakami K and Mitsui Y. (1994a). Molecular cloning of a gene under control of the circadian clock and light in the rodent SCN. Mol. Brain Res. 26 : 197-206.

Ishida N., Mitsui Y. and Tsukada M. (1994b). Molecular conformation and physical properties of poly(Gly-Thr)$_n$: a compound model for the *period* repetitive sequences. Mol. Biol. Life Sci. Adv. 13 : 107-111.

Iverson L.E., Tanouye M.A., Lester H.A., Davidson N. and Rudy B. (1988). A-type potassium channels expressed from *Shaker* locus cDNA. Proc. Natl. Acad. Sci. USA 85 : 5723-5727.

Jackson F.R., Bargiello T.A., Yun S.H., Young M.W. (1986). Product of *per* locus of *Drosophila* shares homology with proteoglycans. Nature 320 : 185-188.

James A.A., Ewer J., Reddy P., Hall J.C. and Rosbash M. (1986). Embrionic expression of the *period* clock gene in the central nervous system of *Drosophila melanogaster*. EMBO J. 5 : 2313-2320.

Jeffreys A.J., Wilson V., Neumann R. and Keyte J. (1988). Amplification of human minisatellites by polymerase chain reaction: towards DNA fingerprinting of single cells. Nucleic Acids Res. 16 : 10953-10971.

John B. and Miklos G. (1988). The eukaryote genome in development and evolution. Allen & Unwin Academic Press, London.

Jones M.D.R., Cubbin C.M. and Marsh D. (1972). Light-on effects and the question of bimodality in the circadian flight activity of the mosquito *anopheles gambiae*. J. Exp. Biol. 57 : 347-357.

Judd B.H., Shen M.W. and Kaufman T.C. (1972). The anatomy and function of a segment on the X chromosome of *Drosophila melanogaster*. Genetics 71 : 139-156.

Kalderon D., Roberts B.L., Richardson W.D. and Smith A.E. (1984). A short amino acid sequence able to specify nuclear location. Cell 39 : 499-509.

Kaplan W.D. and Trout W.E. (1969). The behaviour of four neurological mutants of *Drosophila*. Genetics 61 : 399-309.

Karlin S. and Burge C. (1996). Trinucleotide repeats and long homopeptides in genes and proteins associated with nervous system disease and development. Proc. Natl. Acad. Sci. USA 93 : 1560-1565.

Kimura M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16 : 111-120.

Kimura M. (1983). The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge.

King D.P., Zhao Y., Sangoram A.M., Wilsbacher L.D., Tanaka M., Antoch M.P., Steeves T.D.L., Hotz Vitaterna M., Kornhauser J.M., Lowrey P.L., Turek F.W. and Takahashi J. (1997). Positional cloning of the mouse circadian *clock* gene. Cell 89 : 641-653.

Klemenz R., Weber U. and Gehering W.J. (1987). The *white* gene as a marker in a new p element vector for gene transfer in *Drosophila*. Nucleic Acids Res. 15 : 3947-3959.

Kondo T., Strayer C.A., Kulkarni R.D., Taylor W., Ishiura M., Golden S.S. and Johnson C.H. (1993). Circadian rhythms in prokaryotes: luciferase as a reporter of circadian gene expression in cyanobacteria. Proc. Natl. Acad. Sci. USA 90 : 5672-5676.

Konopka R.J. and Benzer S. (1971). Clock mutants of *Drosophila melanogaster*. Life Sci. Adv. Series C 5 : 47-49.

Konopka R.J. (1972). Ph.D. thesis, California Institute of Technology, Pasadena, California, USA.

Konopka R.J., Wells S. and Lee T. (1983) Mosaic analysis of a *Drosophila* clock mutant. Mol. Gen. Genet. 190 : 284-288.

Konopka R.J. (1987). Neurogenetics of *Drosophila* circadian rhythms. In: Evolutionary genetics of invertebrate behaviour (edited by Huettel M.D.). Plenum press, New York, USA.

Konopka R.J., Pittendrigh C. and Orr D. (1989). Reciprocal behavior associated with altered homeostasis and photosensitivity of *Drosophila* clock mutants. J. Neurogenet. 6 : 1-10.

Konopka R.J., Kyriacou C.P. and Hall J.C. (1996). Mosaic analysis in the *Drosophila* CNS of circadian and courtship-song rhythms affected by a *period* clock mutation. J. Neurogenet. 11 : 117-139.

Kreitman M. (1991). Detecting selection at the level of DNA. In: Evolution at the molecular level. (Edited by Selander R.K., Clark A.G. and Whittam T.S.). Sinauer Ass. Inc. Publishers, Sunderland.

Kulkarni S.J. and Hall J.C. (1987). Behavioural and cytogenetic analysis of the *cacophony* courtship song mutant and interacting genetic variants in *Drosophila melanogaster*. Genetics 115 : 461-475.

Kulkarni S.J., Steinlauf A.F. and Hall J.C. (1989). The *dissonance* mutant of courtship song in *Drosophila melanogaster*: isolation, behaviour and cytogenetics. Genetics 118 : 267-285.

Kyriacou C.P. and Burnett B. (1979). Genetic analysis of the phototaxis near the upper limit of the visual spectrum of *Drosophila melanogaster*. Behav. Genet. 9 : 123-128.

Kyriacou C.P. and Hall J.C. (1980). Circadian rhythms mutations in *Drosophila melanogaster* affect short term fluctuations in the male's courtship song. Proc. Natl. Acad. Sci. USA 77 : 6929-6933.

Kyriacou C.P. and Hall J.C. (1982). The function of courtship song rhythms in *Drosophila*. Anim. Behav. 30 : 794-801.

Kyriacou C.P. and Hall J.C. (1986). Interspecific genetic control of courtship song production and reception in *Drosophila*. Science 232 : 494-497.

Kyriacou C.P. and Hall J.C. (1989). Spectral analysis of *Drosophila* courtship song rhythms. Anim. Behav. 37 : 850-859.

Kyriacou C.P., Oldroyd M., Wood J., Sharp M. and Hill M.. (1990). Clock mutations alter developmental timing in *Drosophila*. Heredity 64 : 395-401.

Lachaise D., Cariou M.L., David J.R., Lemeunier F., Tsacas L. and Ashburner M. (1988). Historical biogeography of the *Drosophila melanogaster* species subgroup. Evol. Biol. 22 : 159-225.

Lee C., Parikh V., Itsukaichi T., Bae K. and Edery I. (1996). Resetting the *Drosophila* clock by photic regulation of PER and a PER-TIM complex. Science 271 : 1740-1744.

Levine J.D., Sauman I., Imbalzano M., Reppert S.M. and Jackson F.R. (1995). Period protein from the giant silkmoth *Antheraea pernyi* functions as a circadian clock element in *Drosophila melanogaster*. Neuron 15 : 147-157.

Lewis M.T. and Feldman J.F. 1993. The putative Frq clock protein of *Neurospora crassa* contains sequence elements that suggest a nuclear transcriptional regulatory role. Protein Seq. Data Anal. 5 : 315-323.

Li H., Ko H.P., Whitlock J.P. (1996). Induction of Phosphoglycerate Kinase-1 gene-expression by hypoxia - Roles Of Arnt and HIF1-ALPHA. J. Biol. Chem. 271 : 21262-21267.

Lindebro M.C., Poellinger L. and Whitelaw M.C. (1995). Protein-protein interaction via PAS domains: role of the PAS domain in positive and negative regulation of the bHLH/PAS dioxin receptor-Arnt transcription factor complex. EMBO J. 14 : 3528-3539.

Linden H. and Macino G. (1997).White collar 2, a partner in blue-light signal transduction, controlling expression of light-regulated genes in *Neurospora crassa*. EMBO J. 16 : 98-109.

Lindsley D.L. and Zimm G.M. (1992). The genome of *Drosophila melanogaster*. Academic Press, San Diego USA.

Liu X., Lorenz L., Yu Q., Hall J.C. and Rosbash M. (1988). Spatial and temporal expression of the *period* gene in *Drosophila melanogaster*. Genes Dev. 2 : 228-238.

Liu X., Yu Q., Huang Z., Zwiebel L.J., Hall J.C. and Rosbash M. (1991). The strength and periodicity of *D. melanogaster* circadian rhythms are differently affected by alterations in *period* gene expression. Neuron 6 : 753-766.

Liu X., Zwiebel L.J., Hinton D., Benzer S., Hall J.C. and Rosbash M. (1992). The *period* gene encodes a predominantly nuclear protein in adult *Drosophila*. J. Neurosci. 12 : 2735-2744.

Liu Y., Garceau N.Y., Loros J.J. and Dunlap J.C. (1997). Thermally regulated transcriptional control of Frq mediates aspects of temperature responses in the *Neurospora* circadian clock. Cell 89 : 477-486.

Lukowitz W., Shroeder C., Glaser G., Huelskamp M. and Tautz D. (1994). Regulatory and coding regions of the segmentation gene *hunchback* are funtionally conserved between *Drosophila viris* and *Drosophila melanogaster*. Mech. Dev. 45 : 105-115.

Lutz F.E. (1932). Experiments with Orthoptera concerning diurnal rhythms. Am. Mus. Novitates 50 : 1-24.

Mahon A.C., Nambu J.R., Taussig R., Shyamala M., Roach A. and Scheller R.H. (1985). Structure and expression of the egg-laying hormone gene family in *Aplysia*. J. Neurosci. 5 : 1872-1880.

Malcrida A., Gasperi G., Biscaldi G.F. and Milani R. (1985). Functional significance of gene clusters in the housefly *Musca domestica* and in other Diptera. S.I.T.E. Atti 5 : 307-310.

McClung C.R., Fox B.A. and Dunlap J. C. (1989). *frequency*, a clock gene in *Neurospora* shares a sequence element with the *Drosophila* clock gene *period*. Nature 339 : 558-562.

McClung C.R. and Kay S.A. (1994). Circadian rhythms in the higher plant *Arabidopsis thaliana*. In: *Arabidopsis thaliana* (edited by Somerville C.S. and Meyerowitz E.). Cold Spring Harbor Press, Cold Spring Harbor. 615-637.

Morin L.P. (1994). The circadian visual-system. Brain Res. Rev. 19 : 102-127.

Milani R. (1967). The genetics of *Musca domestica* and of other muscaoid flies. In: Genetics of insect vectors of disease. (edited by Wright J.W. and Pal R.). Elsevier Press, Amsterdam.

Milani R., Rubini P.G. and Franco M.G. (1967). Sex determination in the housefly. Genetica Agraria 21 : 385-411.

Myers M.P., Wagner-Smith K., Wesley C.S., Young M. and Sehgal A. (1995). Positional cloning and sequence analysis of the *Drosophila* clock gene, *timeless*. Science 270 : 805-808.

Myers M.P., Wagner-Smith K., Rothenfluh-Hilfiker A. and Young M. (1996). Light-induced degradation of TIMELESS and entrainment of the *Drosophila* circadian clock. Science 271 : 1736-1740.

Nambu J.R., Lewis J.O., Wharton K.A. and Crews S.T. (1991). The *Drosophila single minded* gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. Cell 67 : 1157-1167.

Nambu J.R., Chen W., Hu S. and Crews S.T. (1996). The *Drosophila melanogaster similar* bHLH-PAS gene encodes a protein related to Human Hypoxia-Inducible Factor 1-alpha and *Drosophila single-minded*. Gene 172 : 249-254.

Nielsen J., Peixoto A.A., Piccin A., Costa R., Kyriacou C.P. and Chalmers D. (1994). Big flies, small repeats: the "Thr-Gly" region of the *period* gene in *Diptera*. Mol. Biol. Evol. 11 : 839-853.

Nishikimi M., Fukuyama R., Minoshima S., Shimizu N. and Yagi K. (1994). Cloning and chromosomal mapping of the human nonfunctional gene for l-gulono-gamma-lactone oxidase, the enzyme for l-ascorbic-acid biosynthesis missing in man. J. Biol. Chem. 269 : 13685-13688.

Njus D., Sulzman F.M. and Hastings J.W. (1974). Membrane model for the circadian clock. Nature 248 : 116-120.

O'Connell P. and Rosbash M. (1984). Sequence, structure and codon preference of the *Drosophila* ribosomal protein 49 gene. Nucleic Acids Res. 12 : 5495-5513.

O'Tousa J.E., Baehr W., Martin R.L., Hirsh J., Pak W.L. and Applebury M.L. (1985). The *Drosophila ninaE* gene encodes an opsin. Cell 40 : 839-850.

Padgett R.A., Grabowski P.J., Konarska M.M., Seiler S. and Sharp P.A. (1986). Splicing of messenger RNA precursors. Annu. Rev. Biochem. 55 : 1119-1150.

Parker A.H. (1962). Studies on the diurnal rhythms of the housefly, *Musca domestica* L., in a dry tropical environment. Acta Trop. 19 : 97-119.

Peixoto A.A., Costa R., Wheeler J.C., Hall J.C. and Kyriacou C.P. (1992). Evolution of the threonine-glycine repeat region in the *melanogaster* species subgroup of *Drosophila*. J. Mol. Evol. 35 : 711-719.

Peixoto A.A., Campesan S., Costa R. and Kyriacou C.P. (1993). Molecular evolution of a repetitive region within the *per* gene of *Drosophila*. Mol. Biol. Evol. 10 : 127-139.

Peixoto A.A., Hennessy J.M., Townson I, Hasan G., Posbash M., Costa R. and Kyriacou C.P. (1998). Molecular coevolution within a *Drosophila* clock gene. Proc. Natl. Acad. Sci. USA. 95 : 4475-4480.

Pelandakis M., Higgins D.G. and Solignac M. (1991). Molecular phylogeny of the subgenus *Sophophora* of *Drosophila* derived from large subunits of ribosomal RNA sequences. Genetica 84 : 87-94.

Petersen G., Hall J.C. and Rosbash M. (1988). The *period* gene of *Drosophila* carries species-specific behavioural instructions. EMBO J. 7 : 3939-3947.

Pittendrigh C.S., Bruce V.G., Rosenzweig N.S. and Rubin M.L. (1959). A biological clock in *Neurospora*. Nature 184 : 169-170.

Pittendrigh C.S. (1960). Circadian rhythms and the circadian organization of living systems. Cold Spring Harb. Symp. Quant. Biol. 25 : 159-184.

Pittendrigh C.S. (1967). Circadian systems, I. The driving oscillator and its assay in *Drosophila pseudoobsura*. Proc. Natl. Acad. Sci. USA 58 : 1672-1677.

Power J.M., Ringo J.M. and Dowse H.B. (1995). The effect of *period* mutations and light on the activity rhythms of *Drosophila melanogaster*. J. Biol. Rhythms 10 : 267-280.

Price J.L., Dembinska M.E., Young M.W. and Rosbash M. (1995). Suppression of period protein abundance and circadian cycling by the *Drosophila* clock mutation *timeless*. EMBO J. 14 : 4044-4049.

Pyza E. and Meinertzhagen I.A. (1995). Monopolar cell axons in the first optic neuropil of the housefly, *Musca domestica* (L), undergo daily fluctuations in diameter that have a circadian basis. J. Neurosci. 15 : 407-418.

Qiu J. and Hardin P.H. (1996). *per* mRNA cycling is locked to lights-off under photoperiodic conditions that support circadian feedback loop function. Mol. Cell. Biol. 16 : 4182-4188.

Quinn W.G. and Greenspan R.J. (1984). Learning and courtship in *Drosophila*: two stories with mutants. Annu. Rev. Neurosci. 21 : 67-93.

Ralph M.R. and Menaker M. (1988). A mutation of the circadian system in golden hamster. Science 241 : 1225-1227.

Rechsteiner M. (1988). Regulation of enzyme levels by proteolysis: the role of PEST regions. Adv. In Enzy. Regul. 27 : 135-151.

Reddy P., Zehring W.A., Wheeler D.A., Pirrotta V., Hadfield C., Hall J.C. and Rosbash M. (1984). Molecular analysis of the *period* locus in *Drosophila melanogaster* and identification of a transcript involved in biological rhythms. Cell 38 : 701-710.

Reddy P., Jacquier A.C., Abovich N., Petersen G and Rosbash M. (1986). The *period* locus of *D. melanogaster* codes for a proteoglycan. Cell 46 : 53-61.

Reppert S.M., Tsai T., Roca A.L. and Surman I. (1994). Cloning of a structural and functional homolog of the circadian clock gene *period* from the giant silkmoth *Antheraea pernyi*. Neuron 13 : 1167-1176.

Roberts D.H., Lehar J. and Dreher J.W. (1987). Time series with CLEAN. I. Derivation of a spectrum. Astron. J. 93 : 968-989.

Roberts S. (1956). Clock controlled activity rhythm in the fruit fly. Science 124 : 172.

Robertson H.M., Preston C.R., Phillis R.W., Johnsonchlitz D.M., Benz W.K. and Engels W.R. (1988). A stable genomic source of P element transposase in *Drosophila melanogaster*. Genetics 118 : 461-470.

Rodriguez I.R., Mazuruk K., Schoen T.J. and Chader G.J. (1994). Structural-analysis of the human hydroxyindole-o-methyltransferase gene - presence of 2 distinct promoters. J. Biol. Chem. 269 : 31969-31977.

Rogers S., Wells R. and Rechsteiner M. (1986). Amino acid sequences common to rapidly degraded proteins: the PEST hypothesis. Science 234 : 364-368.

Rosato E., Peixoto A.A., Barbujani G., Costa R. and Kyriacou C.P. (1994). Molecular polymorphism in the *period* gene of *Drosophila simulans*. Genetics 138 : 693-707.

Rosato E., Peixoto A.A., Gallippi A., Kyriacou C.P. and Costa R.(1996). Mutational mechanisms, phylogeny and evolution of a repetitive region within a clock gene of *D. melanogaster*. J. Mol. Evol. 42 : 392-408.

Ruoslahti E. (1988). Structure and biology of proteoglycans. Annu. Rev. Cell Biol. 4 : 229-255.

Russo C.A.M., Takezaki N. and Nei M. (1995).Molecular phylogeny and divergence times of Drosophilid species. Mol. Biol. Evol. 12 : 391-404.

Rutila J.E., Edery I., Hall J.C. and Rosbash M. (1992). The analysis of new short-period circadian-rhythm mutants suggests features of *Drosophila-melanogaster period* gene-function. J. Neurogen. 8 : 101-113.

Saez L. and Young M.W. (1996). Regulation of nuclear entry of the *Drosophila* clock proteins Period and Timeless. Neuron 17 : 911-920.

Sanger F., Nicklen S. and Coulson A.R. (1977). DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA. 74 : 5463-5467.

Sambrook J., Fritsch E.F. and Maniatis T. (1989). Molecular cloning. A laboratory manual. 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor.

Sargent M.L. and Kaltenborn S.H. 1972. Effect of medium composition and carbon dioxide on circadian conidation in *Neurospora*. Plant Physiol. 50 : 171-175.

Sauman I. and Reppert S.M. (1996). Circadian clock neurons in the silkmoth *Antheraea pernyi*: novel mechanisms of Period protein regulation. Neuron 17 : 889-900.

Sauman I., Tsai T., Roca A.L. and Reppert S.M. (1996). Period protein is necessary for circadian control of egg hatching behavior in the silkmoth *Antheraea pernyi*. Neuron 17 : 901-909.

Saunders D.S. (1982). Insect clocks. Pergamon Press, Oxford, England.

Sawyer L.A., Hennessy J.M., Peixoto A.A., Rosato E., Parkinson H., Costa R. and Kyriacou C.P. (1997). Natural variation in a *Drosophila* clock gene and temperature compensation. Science 278 : 2117-2120.

Schlotterer C., Hauser M.T., von Haeseler A. and Tautz D. (1994). Comparative evolutionary analysis of rDNA ITS regions in *Drosophila*. Mol. Bio. Evol. 11 : 513-522.

Segarra C. and Aguade M. (1993). Nucleotide divergence of the rp49 gene region between *Drosophila melanogaster* and 2 species of the *obscura* group of *Drosophila*. J. Mol. Evol. 36 : 243-248.

Sehgal A., Price J.L., Man B. and Young M.W. (1994). Loss of circadian behavioral rythm and *per* RNA oscillations in the *Drosophila* mutant *timeless*. Science 263 : 1603-1606.

Sehgal A., Rothenfluh-Hilfiker A., Hunter-Ensor M., Chen Y., Myers M.P. and Young M.W. (1995). Rhythmic expression of *timeless*: a basis for promoting circadian cycles in *period* gene autoregulation. Science 270 : 808-811.

Selkov E.E. (1975). Stabilization of energy charge, generation of oscillations and multiple steady states in energy metabolism as a result of purely stoichiometric regulation. Eur. J. Biochem. 59 : 151-157.

Sharp P.M., Burgess C.J., Cowe E., Lloyd A.T. and Mitchell K.J. (1992). Selective use of termination codons and variations in codon choice. In: Transfer RNA in protein synthesis (edited by Hatfield D.L., Byeong J.L. and Pirtle R.M.). CRC Press, Boca Raton, Florida. 397-425.

Siwicki K.K., Eastman C., Petersen G, Rosbash M. and Hall J.C. (1988). Antibodies to the *period* gene product of *Drosophila* reveal diverse tissue distribution and rhythmic changes in the visual system. Neuron 1 : 141-150.

Siwicki K.K., Strack S., Rosbash M., Hall J.C. and Jacklet J.W. (1989). An antibody to the *Drosophila* period protein recognises circadian pacemaker neurons in *Aplysia* and *Bulla*. Neuron 3 : 51-58.

Siwicki K.K., Flint K., Hall J.C., Rosbash M. and Spray D.C. (1992a). The *Drosophila period* gene and dye coupling in larval salivary glands: a re-evaluation. Biol. Bull. 183 : 340-341.

Siwicki K.K., Shwartz W.J. and Hall J.C. (1992b). An antibody to the *Drosophila* period protein labels antigens in the suprachiasmatic nucleus of the rat. J. Neurogenet. 8 : 33-42.

Smietanko A. and Engelmann W. (1989a). Lithium effects on circadian locomotor activity rhythm of *Musca domestica* under continuous light conditions. J. Interdiscipl. Cycle Res. 20 : 3-13.

Smietanko A. and Engelmann W. (1989b). Splitting of circadian rhythms of *Musca domestica* flies with azadirachtin. J. Interdiscipl. Cycle Res. 20 : 71-79.

Smith R.F. and Konopka R.J. (1982). Effects of dosage alterations at the *per* locus, on the period of the circadian clock of *Drosophila*. Mol. Gen. Genet. 185 : 30-36.

Sneath P.H.A. and Sokal R.R. (1973). Numerical taxonomy. W.H. Freeman and Co., San Francisco.

So W.V. and Rosbash M. (1997). Post-transcriptional regulation contributes to *Drosophila* clock gene mRNA cycling. EMBO J. 23 : 7146-7155.

Sommer R. and Tautz D. (1991). Segmentation gene expression in the housefly *Musca domestica*. Development 113 : 419-430.

Southern E.M. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 98 : 503-517.

Spradling A.C. (1986). P element-mediated transformation. In: *Drosophila*: a practical approach. (edited by Roberts D.B.). IRL press, Oxford, England. 175-197.

Stanewsky R., Jamison C.F., Plautz J.D., Kay S.A. and Hall J.C. (1997). Multiple circadian-regulated elements contribute to cycling *period* gene expression in *Drosophila*. EMBO J.

Steller H., Fischbach K.H. and Rubin G.M. (1987). *disconnected*: a locus required for neuronal pathway formation in the visual system of *Drosophila*. Cell 50 : 1139-1153.

Sun Z.S., Albrecht U., Zhuchenko O., Bailey J., Eichele G. and Lee C.C. (1997). *Rigui*, a putative mammalian ortholog of the *Drosophila period* gene. Cell 90 : 1003-1011.

Swanson H.I., Chan W.K. and Bradfield C.A. (1995). DNA binding specificities and pairing rules of the Ah receptor, ARNT and SIM proteins. J. Bio. Chem. 270 : 26292-26302.

Sweeney B.M. (1974). A physiological model for circadian rhythms from the *Acetabularia* rhythm paradoxes. Int. J. Chronobiol. 2 : 25-33.

Tautz D. and Sommer R.J. (1995). Evolution of segmentation genes in insects. Trends Genet. 11 : 23-27.

Takahashi J.S. (1993). Circadian-clock regulation of gene expression. Curr. Opin. Genet. Dev. 3 : 301-309.

Takahashi J.S. (1995). Molecular neurobiology and genetics of circadian-rhythms in mammals. Annu. Rev. Neurosci. 18 : 531-553.

Thackeray J.R. (1989). Ph.D. thesis, Leicester University, Leicester, UK.

Thackeray J.R. and Kyriacou C.P. (1990). Molecular evolution in the *Drosophila yakuba period* locus. J. Mol. Evol. 31 : 389-401.

Thomas J.B. and Wyman R.J. (1982). A mutation in *Drosophila* alters normal connectivity between two identified neurones. Nature 298 : 650-651.

Thomas J.B. and Wyman R.J. (1984). Mutations altering synaptic connectivity between identified neurons in *Drosophila*. J. Neurosci. 4 : 530-538.

Tei H., Okamura H., Shigeyoshi Y., Fukuhara C., Ozawa R., Hirose M and Sakaki Y. (1997). Circadian oscillation of a mammalian homologue of the *Drosophila period* gene. Nature 389 : 512-516.

Timpe L.C., Schwartz T.L., Tempel B.L., Papazian D.M., Jan Y.N. and Jan L.Y. (1988). Expression of functional potassium channels from *Shaker* in *Xenopus* oocytes. Nature 331 : 143-145.

Tomlinson A. and Ready D.F. (1986). Sevenless: a cell-specific homeotic mutation of the *Drosophila* eye. Science 231 : 400-402.

Tsutsumi C. (1973). Characteristics of the daily behavior and activity patterns of the adult housefly with special reference to time-keeping device. Jpn. J. Med. Sci. Biol. 26 : 119-141.

Towbin H., Staehelin T. and Gordon J. (1979). Elecrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. Proc. Natl. Acad. Sci. USA. 76 :4350-4354.

Vitaterna M.H., King D.P., Chang A.M., Kornhauser J.M., Lowrey P.L., McDonald J.D., Dove W.F., Pinto L.H., Turek F.W. and Takahashi J.S. (1994). Mutagenesis and mapping of a mouse gene, *clock*, essential for circadian behavior. Science 264 : 719-725.

Vosshall L.B., Price J.L., Sehgal A., Saez L. and Young M.W. (1994). Block in nuclear localization of *period* protein by a second clock mutation, *timeless*. Science 263 : 1606-1609.

Vosshall L.B. and Young M.W. (1995). Circadian rhythms in *Drosophila* can be driven by *period* expression in a restricted group of central brain cells. Neuron 15 : 345-360.

Watanabe A., Ikejima M., Suzuki N. and Shimada T. (1996). Genomic organization and expression of the human MSH3 gene. Genomics 31 : 311-318.

Wharton K.A., Yedvobnick B. and Finnerty V.G. (1985). *opa*: a novel family of transcribed repeats shared by the *notch* locus and other developmentally regulated loci in *D. melanogaster*. Cell 40 : 55-62.

Wheeler D.A., Kyriacou C.P., Greenacre M.L., Yu Q., Rutila J.E., Rosbash M. and Hall J.C. (1991). Molecular transfer of a species-specific behavior from *Drosophila simulans* to *Drosophila melanogaster*. Science 251 : 1082-1085.

Wilcock J. (1969). Gene action and behaviour. An evaluation of major gene pleiotropism. Psyhcol. Bull. 72 : 1-29.

Winfree A.T. (1967). Biological rhythms and the behavior of populations of coupled oscillators. J. Theor. Biol. 16 : 15-42.

Wood K.W. (1995). Marker proteins for gene expression. Curr. Opin. Biotech. 6 : 50-58.

Young M.W. and Judd B.H. (1978). Nonessential sequences, genes, and polytene chromosome bands of *Drosophila melanogaster*. Genetics 88 : 723-742.

Young M.W. (1992). New clock mutants in *Drosophila*. EMBO workshop on Molecular Chronobiology. Abstr. page 11.

Yu Q., Jacquier A.C., Citri Y., Hamblen M., Hall J.C. and Rosbash M. (1987 a). Molecular mapping of point mutations in the *period* gene that stop or speed up biological clocks in *Drosophila melanogaster*. Proc. Natl. Acad. Sci. USA 84 : 784-788.

Yu Q., Colot H.V., Kyriacou C.P., Hall J.C. and Rosbash M. (1987 b). Behaviour modification by *in vitro* mutagenesis of a variable region within the *period* gene of *Drosophila*. Nature 326 : 765-769.

Zehring W.A., Wheeler D.A., Reddy P., Konopka R.J., Kyriacou C.P., Rosbash M. and Hall J.C. (1984). P-element transformation with *period* locus DNA restores rhythmicity to mutant, arrhythmic *Drosophila melanogaster*. Cell 39 : 369-376.

Zeng H., Hardin P.E. and Rosbash M. (1994). Constitutive overexpression of the *Drosophila period* protein inhibits *period* mRNA cycling. EMBO J. 13 : 3590-3598.

Zeng H., Qian Z., Myers M.P. and Rosbash M. (1996). A light-entrainment mechanism for the *Drosophila* circadian clock. Nature 380 : 129-135.

Zerr D.M., Hall J.C., Rosbash M. and Siwicki K.K. (1990). Circadian rhythms of period protein immunoreactivity in the CNS and the visual system of *Drosophila*. J. Neurosci. 10 : 2749-2762.

Zuckerkandl E. and Pauling L. (1965). Evolutionary divergence and convergence in proteins. In: Evolving genes and proteins (edited by Bryson V. and Vogel H.J.). Academic Press, New York.

Zwiebel L.J., Hardin P.E., Liu X., Hall J.C. and Rosbash M. (1991). A post-transcriptional mechanism contributes to circadian cycling of a *per*-β-galactosidase fusion protein. Proc. Natl. Acad. Sci. USA 88 : 3882-3886.