Epigenetic Inheritance of Active Chromatin After Removal of the Main Transactivator

Giacomo Cavalli[1,2] and Renato Paro[1]*

The *Drosophila* Polycomb and trithorax group proteins act through chromosomal elements such as *Fab-7* to maintain repressed or active gene expression, respectively. A *Fab-7* element is switched from a silenced to a mitotically heritable active state by an embryonic pulse of transcription. Here, histone H4 hyperacetylation was found to be associated with *Fab-7* after activation, suggesting that H4 hyperacetylation may be a heritable epigenetic tag of the activated element. Activated *Fab-7* enables transcription of a gene even after withdrawal of the primary transcription factor. This feature may allow epigenetic maintenance of active states of developmental genes after decay of their early embryonic regulators.

During early embryogenesis, spatially restricted expression patterns of *Drosophila* homeotic genes are established by a regulatory cascade involving the products of segmentation genes (*1*, *2*). Subsequently, Polycomb (PcG) and trithorax (trxG) group proteins maintain these patterns (*3*, *4*) at the level of chromatin structure (*5*, *6*) by acting through overlapping cis regulatory elements (*7*–*13*). We previously reported that one of these elements, the *Fab-7* DNA from the bithorax complex (*14*–*16*), could be specifically switched into a repressing or an activating mode during embryogenesis (*17*). Once established, both regulatory modes are mitotically inherited throughout development and at a certain frequency through meiosis. For this reason, *Fab-7* was termed a cellular memory module (CMM). The *Fab-7*–dependent transmission of a chromatin state permissive for gene expression through multiple cell divisions suggested that an active process is involved in the maintenance of open chromatin. Local hyperacetylation of histones H3 and H4 in fission yeast centromeres can be epigenetically inherited through mitosis and meiosis (*18*). Here, we tested the roles of PcG and trxG proteins in maintaining the *Fab-7* state and whether epigenetic control involves similar histone modifications in *Drosophila*.

Fab-7–dependent chromosomal memory of silent or open chromatin states was previously described in transgenic *Drosophila* lines such as FLW-1 and FLFW-1 (*19*). These lines carry a heat shock–inducible GAL4 driver (hsp70-GAL4) regulating a GAL4-dependent *lacZ* reporter (UAS-lacZ) flanked by *Fab-7* and the mini-*white* gene. Silencing imposed by *Fab-7* on the flanking reporter genes was dependent on the components of the PcG, as heterozygous mutant PcG genes showed a relief of *white* gene repression (Fig. 1A). Conversely, *white* gene activity required the trxG, as heterozygous mutations in the different members tested (Fig. 1A) resulted in a down-regulation of expression. A GAL4 pulse during embryogenesis can impose a mitotically stable reprogramming of the *Fab-7* CMM from a silenced to an open chromatin state (*17*). The maintenance of the activated *Fab-7* state was dependent on *trithorax* (*trx*) but not on *Polycomb* (*Pc*) (Fig. 1, B to D). In a heterozygous Pc^- background, *Fab-7* can be switched by a GAL4 pulse and be stably maintained, resulting in strong *white* expression (Fig. 1C; HSE). In contrast, a trx^- mutation completely abolished the mitotic transmission (Fig. 1D).

This antagonistic role of the PcG and trxG suggests a competitive interaction of the protein components regulating either the repressed or the active state of *Fab-7*. Indeed, counteracting PcG-mediated silencing by a transient GAL4 pulse during the third instar larval stage induced displacement of Polycomb (PC) protein from a transgene carrying *Fab-7* in polytene chromosomes (Fig. 2, G to I) (*9*). Displacement of PC is accompanied by strong UAS-lacZ expression in salivary glands. However, after the disappearance of GAL4, PC protein was rebound to *Fab-7* and UAS-lacZ became repressed again (Fig. 2I). To assess whether the epigenetically activated *Fab-7* state correlates with a permanent loss of PcG proteins from the chromatin template, we administered a strong GAL4 induction pulse at embryogenesis in the FLFW-1 line. Polytene chromosomes of third instar larvae were immunostained with antibodies directed against PcG proteins. Surprisingly, all of

A

Mutation	Effect	Strength
PcG		
Pc^{XT109}	Derepression	+
ph^{410}	Derepression	+++
Psc^{1}	Derepression	+++
Psc^{Arp}	Derepression	+
Pcl^{T1}	Derepression	++
pho^{1}	Derepression	++
trxG		
trx^{E2}	Repression	++
Brm^{2}	Repression	+++
mor^{1}	Repression	++
Trl^{62}	Repression	+
Trl^{13C}	No Effect	——

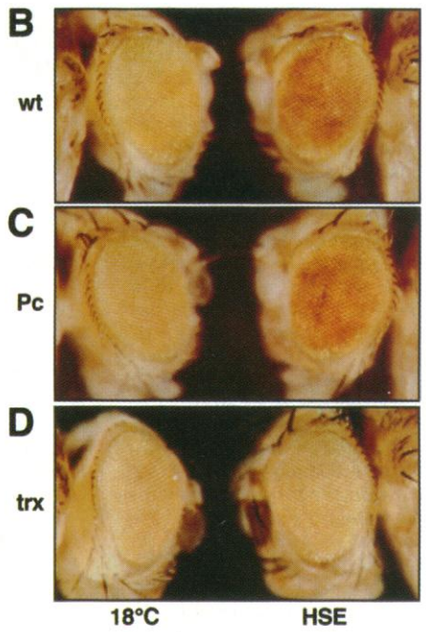

Fig. 1. Role of PcG and trxG genes in *Fab-7*–mediated epigenetic inheritance. (**A**) The eye color of flies of the 5F24 25,2 line (*19*) was compared with that of flies carrying the same *Fab-7* transgene insertion and heterozygous mutations in genes of the PcG or the trxG, as indicated. Darker eye color with respect to wild-type 5F24 25,2 was interpreted as derepression in the mutant background, whereas lighter eyes indicated repression. The relative strength of the effect is indicated on the right. (**B** to **D**) 5F24 25,2 embryos carrying one copy of $hsGAL4^{7-1}$ in the second chromosome and the heterozygous mutations Pc^{XT109} or trx^{E2} were either grown at 18°C or activated by an embryonic pulse of GAL4 (HSE). The extent of transmission of the activated state was measured by comparing the eye color of the HSE with that of the 18°C F_0 progeny 1 day after hatching. The trx^{E2} mutation prevented transmission of the *Fab-7*–activated state. At 18°C, all eyes show a similar color, because of background pigmentation caused by the mini-*white* gene present in the $hsGAL4^{7-1}$ transgene. In its absence, Pc^{XT109} flies show darker eyes, and trx^{E2} show lighter eyes than their wild-type (wt) siblings.

[1]Zentrum für Molekulare Biologie (ZMBH), Universität of Heidelberg, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany. [2]Institut de Génétique Humaine, 141, rue de la Cardonille, 34396 Montpellier, France.

*To whom correspondence should be addressed. E-mail: paro@sun0.urz.uni-heidelberg.de

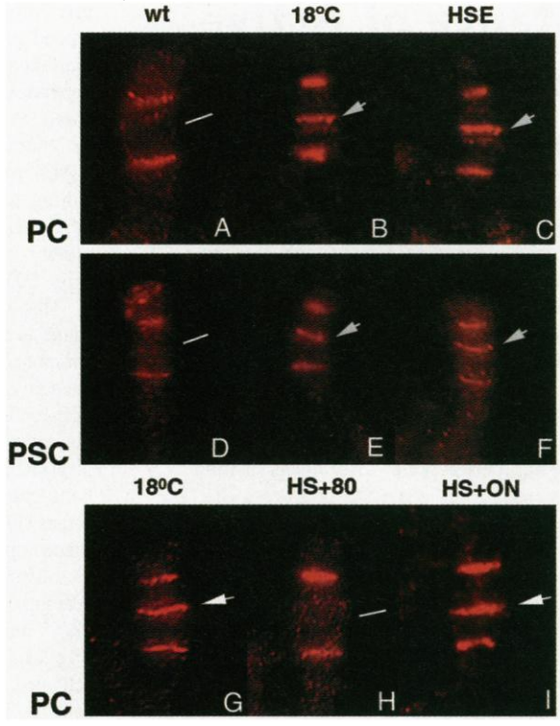

Fig. 2. Maintenance of the active state in the presence of PcG proteins. (**A** to **F**) Binding of the PcG proteins PC and PSC at *Fab-7* at 61C9 in different functional states was analyzed. The position of integration of the *Fab-7*–containing transgene is indicated by a tick in the absence of PcG protein binding or by an arrow in the case of PcG binding. PC and PSC proteins colocalize in two endogenous bands at cytological position 61C1,2 and 61F2 (A and D). In the FLFW-1 line grown at 18°C, both proteins bind at *Fab-7* (61C9) without GAL4 induction (B and E), as well as upon a GAL4 pulse at embryogenesis (HSE) (C and F). (**G** to **I**) A time course of PC displacement upon a GAL4 pulse in third instar larvae is shown. PC is bound at the transgene at 18°C (G), it is displaced upon a 45-min heat shock followed by 80 min of recovery at 18°C (H)(HS+80), and it reassociates to the transgene upon a heat shock pulse at mid third instar larvae followed by an overnight recovery at 18°C (I) (HS+ON).

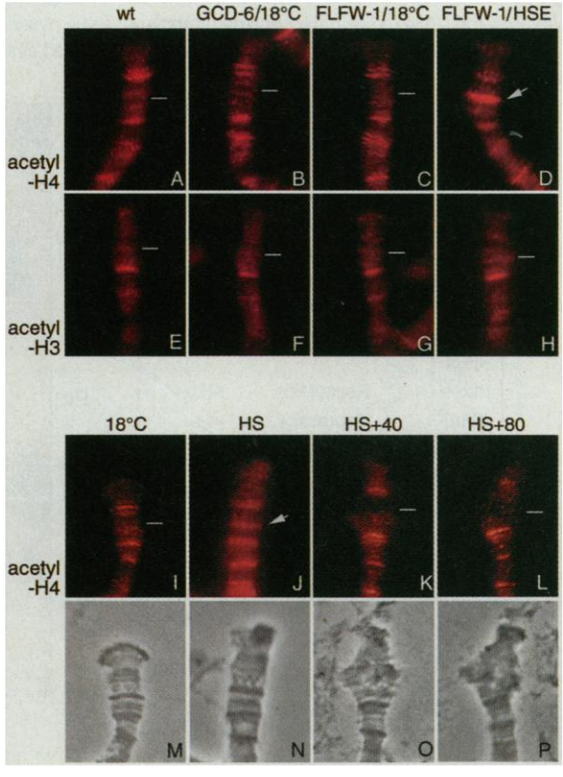

Fig. 3. Hyperacetylation of H4 marks epigenetically maintained open states of *Fab-7*. (**A** to **H**), Tetra-acetylated histone H3 and H4 were visualized on polytene chromosomes by immunostaining. Strains (*19*) and growth conditions were as indicated. The position of the *Fab-7*–containing transgene is indicated by a tick (when no histone hyperacetylation was detected) or by an arrow (in the presence of histone hyperacetylation). Hyperacetylation of histone H4 is observed at the transgene in larvae derived from embryos submitted to a GAL4 activation pulse (D) but not in the absence of GAL4 activation pulses (A to C). H3 hyperacetylation is not observed in any condition (E to H). (**I** to **P**) Time course of histone H4 hyperacetylation during GAL4-mediated activation by a heat shock pulse to third instar larvae. Immunostainings (I to L) and phase contrast (M to P) are shown. HS is a 45-min heat shock at 37°C; HS+40 and HS+80 are a 45-min heat shock followed by 40 or 80 min of recovery at 18°C, respectively.

the PcG proteins tested, PC and *Posterior Sex Combs* (PSC) (Fig. 2, A to F), *Polyhomeotic* (PH), and *Polycomb-like* (PCL) (*20*), were still strongly bound to the *Fab-7* transgene irrespective of the epigenetic state. Thus, an epigenetically activated state can be stably propagated in the presence of the protein components of the PcG. These data support previous observations that demonstrated binding of PC at cytological sites containing potentially active genes in polytene chromosomes (*21*) and binding of PH and PSC proteins at an actively transcribed gene in *Drosophila* Schneider cells (*22*). It has been reported that certain PcG genes may function as activators in specific tissues and at specific developmental times by genetic analyses (*23*, *24*). Although we did not observe a role for PC protein in the maintenance of the activated state of *Fab-7* (see Fig. 1C), it may be possible that other PcG proteins are involved in this process.

If it is not the removal of PcG repressors on the template, what is the epigenetic tag that marks the activated *Fab-7* state? A single embryonic GAL4 pulse was administered to FLFW-1 embryos, and histone acetylation of the *Fab-7* transgene as a possible mark was analyzed by immunostaining polytene chromosomes (*25*) of third instar larvae with specific antibodies against the tetra-acetylated form of H4 and H3 histones (Fig. 3, A to H). Hyperacetylation of histone H4 was detected at the *Fab-7* transgene location in larvae derived from activated embryos, but not from control embryos raised at 18°C (compare Fig. 3D with Fig. 3C). Similarly, no H4 hyperacetylation at *Fab-7* was detected in control larvae that either did not carry the transgene (Fig. 3A, wt) or carried the target transgene in the absence of the hsp70-GAL4 driver (GCD-6 line; Fig. 3B). Furthermore, a line carrying the hsGAL4 driver and a control transgene with UAS-lacZ but no *Fab-7* (*19*) did not show increased acetylation upon an embryonic pulse of GAL4 (*26*). In contrast to H4, no hyperacetylation of histone H3 could be detected in activated FLFW-1 individuals (Fig. 3, E to H).

To determine whether H4 histone acetylation also marks a difference between embryonically activated heritable *Fab-7* states and larval transiently activated states, we analyzed H4 acetylation in a time course of GAL4-mediated transient transcriptional activation of the *Fab-7* transgene in third instar larvae (Fig. 3, I to P). In contrast to the hyperacetylation observed after GAL4 activation at embryogenesis, H4 hyperacetylation of the transgene was observed only weakly and transiently (Fig. 3J) at the end of the heat shock induction of GAL4 in third instar larvae. Forty minutes after the end of the heat shock pulse, H4 hyperacetylation was no longer detected (Fig. 3, K and L), even though the presence of a small chromosomal puff (Fig. 3, O and P) and the recruitment of large amounts of RNA polymerase II indicate continuing transcription (*27*). Therefore, the absence of maintenance of the hyperacetylated states correlates with the failure to reprogram the chromatin of the transgene. This suggests that postembryonic tissues are more refractory to histone acetylation of silent *Fab-7* templates.

Patterning transcription factors, like the

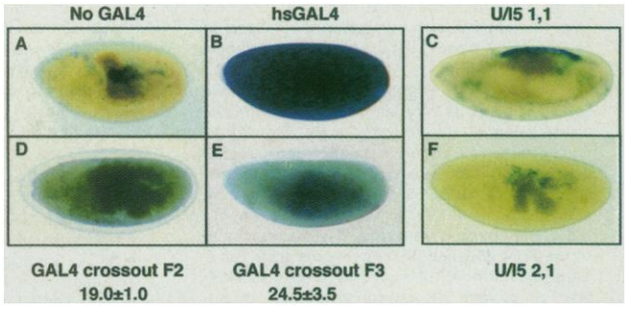

Fig. 4. Inheritance of the active state of *Fab-7* maintains expression of *lacZ* in the absence of the specific GAL4 transactivator. β-Gal staining of embryos is shown. (**A**) 5F24 25,2 line, carrying no GAL4 transgene. (**B**) Strong homogeneous expression is observed in the FLW-1 line upon a 45-min heat shock followed by 2 hours of recovery. (**D** and **E**) GAL4 cross-out F_2 and F_3 generations, derived from the FLW-1 line activated by a GAL4 pulse followed by crossing out of the GAL4 transgene (GAL4 cross-out F_1) and recrossing either once (D) or twice (E). A weak but homogeneous β-Gal expression is observed in all tissues. Below each panel, the mean percentage of embryos showing the depicted pattern ± standard deviation of two independent experiments is indicated. (**C** and **F**) Control lines U/l5 1,1 (C) and U/l5 2,1 (F), which carry the p/Ul5 construct, containing neither *Fab-7* nor GAL4. In these lines, most of the embryos are either not stained or show a very weak, variegated staining pattern. Only about 2% of the embryos showed a more homogeneous UAS-lacZ expression resembling the one in (D) and (E).

products of many segmentation genes, act only shortly on their downstream genes during early *Drosophila* development, whereas the PcG/trxG memory system subsequently maintains the embryonically programmed patterns (*3*). For this reason, we tested whether embryonically activated *Fab-7* can maintain expression of the reporter gene *lacZ* in the absence of the primary transcription factor GAL4 (*28*). The leakiness of the hsp70 promoter prevented the complete disappearance of GAL4 protein during single fly development. To overcome this problem, we made use of the fact that activated *Fab-7* can be efficiently propagated through meiosis in the line FLW-1 (*17*). This allowed us to cross out the GAL4 driver to test *lacZ* expression in the complete absence of GAL4 in subsequent generations (*28*). Upon crossing out GAL4 in activated flies, 20 to 25% of the GAL4-less embryos showed substantial levels of homogeneous β-galactosidase (β-Gal) expression in all embryonic cells in two consecutive generations (Fig. 4, D and E). This percentage correlates well with the fraction of adults showing meiotically stable *white* derepression (*17*). Unfortunately, it was not possible to also test for the meiotic inheritance of H4 hyperacetylated states because of a staining pattern with endogenous bands at the insertion site of the transgene in the FLW-1 line. However, the functional analysis demonstrates that epigenetic inheritance of an active *Fab-7* chromatin state results in transcriptional activity of the UAS-lacZ reporter even in the absence of the GAL4 transactivator.

A weak expression of *lacZ* in the absence of GAL4 may arise from a heritable loss of PcG-mediated repression, thereby neutralizing the silencing ability of *Fab-7* and consequently reflecting the ground state of a nonrepressed chromatin template. If this were the case, it may be expected that in flies carrying an UAS-lacZ construct without *Fab-7* (pU/l5) (*9*, *17*), a similar weak homogeneous β-Gal staining pattern would be observed in all embryos in the absence of GAL4. To test this point, we analyzed β-Gal staining in two independent lines carrying the pU/l5 construct but no GAL4 driver: the U/l5 1,1 and U/l5 2,1 transgenic lines (*9*). In both cases, most of the embryos were not stained or were stained in a weakly variegated fashion in random cells (Fig. 4, C and F). This strongly suggests that meiotic inheritance of *Fab-7* CMM-activated states does not simply reflect lifting of PcG-mediated silencing but rather the inheritance of an active chromatin state, which is competent for transcriptional activation.

In a recent report, Cosma *et al.* (*29*) measured the time and interdependence of transcription and chromatin remodeling factor recruitment to the yeast HO promoter. They found that the transcription factor Swi5p only fleetingly binds to the HO promoter before the recruitment of remodeling factors such as Swi/Snf (related to the *Drosophila* trxG) and SAGA take over. This suggests that Swi5p might only be transiently necessary for activation. Our results support and extend these findings by showing in a functional manner that trxG protein complexes recruited at a CMM relieve the requirement for the activating factor for transcriptional maintenance. We identify hyperacetylation of histone H4 as an epigenetic mark for the activated *Fab-7* state. Unlike the short-lived H4 hyperacetylation induced by transient gene activation at late developmental stages, the mark set at embryonic stages is mitotically stable and inheritable. An important maintenance function of the PcG and trxG protein complexes at CMMs might be to protect epigenetic marks from erasure.

References and Notes

1. G. Gellon and W. McGinnis, *Bioessays* **20**, 116 (1998).
2. P. W. Ingham and A. Martinez Arias, *Cell* **68**, 221 (1992).
3. R. Paro and P. J. Harte, in *Epigenetic Mechanisms of Gene Regulation*, V. E. A. Russo, R. A. Martienssen, A. R. Riggs, Eds. (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1996), pp. 507–528.
4. V. Pirrotta, *Cell* **93**, 333 (1998).
5. Z. Shao *et al.*, *Cell* **98**, 37 (1999).
6. A. Breiling, E. Bonte, S. Ferrari, P. B. Becker, R. Paro, *Mol. Cell. Biol.*, in press.
7. J. Simon, A. Chiang, W. Bender, M. J. Shimell, M. O'Connor, *Dev. Biol.* **158**, 131 (1993).
8. A. Busturia and M. Bienz, *EMBO J.* **12**, 1415 (1993).
9. D. Zink and R. Paro, *EMBO J.* **14**, 5660 (1995).
10. V. Chinwalla, E. P. Jane, P. J. Harte, *EMBO J.* **14**, 2056 (1995).
11. H. Strutt, G. Cavalli, R. Paro, *EMBO J.* **16**, 3621 (1997).
12. V. Orlando, E. P. Jane, V. Chinwalla, P. J. Harte, R. Paro, *EMBO J.* **17**, 5141 (1998).
13. S. Tillib *et al.*, *Mol. Cell. Biol.* **19**, 5189 (1999).
14. K. Hagstrom, M. Müller, P. Schedl, *Genes Dev.* **10**, 3202 (1996).
15. ———, *Genetics* **146**, 1365 (1997).
16. J. Mihaly, I. Hogga, J. Gausz, H. Gyurkovics, F. Karch, *Development* **124**, 1809 (1997).
17. G. Cavalli and R. Paro, *Cell* **93**, 505 (1998).
18. K. Ekwall, T. Olsson, B. M. Turner, R. C. Allshire, *Cell* **91**, 1021 (1997).
19. Strains used in this study were OregonR (wild type) and transgenic lines carrying *Fab-7* constructs. U/l5 1,1 and U/l5 2,1, which carry the pU/l5 transgene, were described in (*9*). The LW-1 line is U/l5 1,1 carrying additionally hsGAL4$^{7-1}$ balanced over the CyO chromosome. The GCD-6 line containing the p24F6 construct at cytological location 61C9 and the FLFW-1 line, which carries the same transgene and hsGAL4$^{7-1}$ balanced over the CyO chromosome, were described in (*17*). The 5F4 25,2 line, carrying the p5F24 construct at cytological location 13F, was described in (*9*). The FLW-1 line, carrying the same construct and hsGAL4$^{7-1}$ balanced over the CyO chromosome, was described in (*17*).
20. In addition to PC and PSC, binding of PH and PCL protein to *Fab-7* in the FLFW-1 line grown at 18°C (controls) or after an embryonic pulse of GAL4 was studied. All four proteins bound to the construct irrespective of the epigenetic state of *Fab-7*. Binding was indistinguishable in activated compared with control templates.
21. B. Zink and R. Paro, *Nature* **337**, 468 (1989).
22. H. Strutt and R. Paro, *Mol. Cell. Biol.* **17**, 6773 (1997).
23. M.-O. Fauvarque, V. Zuber, J.-M. Dura, *Mech. Dev.* **52**, 343 (1995).
24. D. LaJeunesse and A. Shearn, *Development* **122**, 2189 (1996).
25. Immunostaining of tetra-acetylated histones on polytene chromosomes used antibodies from Upstate Biotechnology. For heat shock induction of GAL4, embryos were collected at 18°C in standard apple juice–agar plates and heat shocked at 37°C by immersion in a water bath for 45 min. After heat shock, embryos were transferred to 18°C in fresh tubes containing standard *Drosophila* medium, allowed to develop into third instar larvae, and used for polytene chromosome studies. Late wandering third instar larvae were collected and heat shocked at 37°C for 45 min in a 1.5-ml Eppendorf tube. In all studies with GAL4 induction, memory of activated chromatin states was monitored by inspection of adult eye pigmentation as described in (*17*). In the FLFW-1 line, derepressed flies have red eyes, as opposed to controls, which have yellow eyes at hatching. Male salivary glands were used for experiments on polytene chromosomes, because activation is stronger in males than in females (a strong heat shock pulse of GAL4 results in up to 80% fully red-eyed males). Activation of *lacZ* expression upon GAL4 pulses was monitored by β-Gal staining of salivary glands. At 18°C, most salivary glands show a variegated staining pattern. Upon administration of a GAL4 pulse to third instar larvae, all glands stained strongly and homogeneously. When the GAL4 pulse was given to embryos, the large majority of salivary glands of third instar larvae stained homogeneously, although severalfold less strongly than upon a pulse at third instar larvae. Only in a few glands was a variegated pattern retained. All analyses of polytene chromosomes were performed in two to five independent experiments, and 10 to 20 chromosomes were inspected in each

experiment to ensure detection of many activated chromosomes in each case.
26. Immunostaining of polytene chromosomes with antibodies against tetra-acetylated histone H4 was carried out in the LW-1 line. This line contains the pU/l5 construct inserted at the 93B cytological location and hsGAL4$^{7-1}$ balanced over the CyO chromosome. Larvae raised from control embryos grown at 18°C or from embryos submitted to an embryonic GAL4 pulse showed the same pattern. In contrast, a transient hyperacetylation of histone H4 could be observed at 93B upon a 45-min 37°C heat shock at third instar larvae. The signal was lost 100 min after the end of the GAL4 pulse (*30*). Therefore, a line that does not show trxG-mediated inheritance of active chromatin states does not show stable H4 hyperacetylation.
27. Association of RNA polymerase II during transcriptional activation of the UAS-lacZ reporter in the *Fab-7* template of the FLFW-1 line was studied by immunostaining polytene chromosomes in a time course experiment of heat shock GAL4 induction. Strong recruitment of RNA polymerase II was observed 40 and 80 min after the end of a 45-min 37°C heat shock (*30*). This suggests that hyperacetylation at the same template could have been observed (note the absence of hyperacetylated H4 signal at *Fab-7* in Fig. 3, K and L) if it did occur to a substantial degree.
28. The line FLFW-1, carrying a construct with two *Fab-7* elements, is very inefficient in the meiotic transmission of the activated state. The FLW-1 line, with the *Fab-7* transgene p5F24 at cytological position 13F, showed effective meiotic transmission of active states. Flies of the FLW-1 line were activated by an embryonic pulse of GAL4. Twenty red-eyed F_0 females were crossed with thirty males of the 5F24 25,2 line (*9*), carrying the same *Fab-7* transgene but no GAL4-expressing construct. GAL4-less F_1 flies were selected for strong eye pigmentation (25 to 30% of the total progeny) and recrossed further to obtain F_2 and F_3. β-Gal staining of F_2 and F_3 embryos was performed as described by B. Zink, Y. Engström, W. J. Gehring, and R. Paro [*EMBO J.* **10**, 153 (1991)]. The staining pattern was compared with control FLW-1 embryos at 18°C or upon a 45-min heat shock followed by 2 hours of recovery or with embryos of the lines U/l5 1,1 and U/l5 2,1 grown at 18°C. Further proof of the absence of the hsp70-GAL4 transgene from activated FLW-1 flies was obtained by demonstrating that β-Gal staining of F_2 embryos from activated FLW-1 flies was now independent of a heat shock treatment. Furthermore, the absence of the GAL4 transgene was directly demonstrated by Southern (DNA) analysis of genomic DNA of activated FLW-1 flies (*30*).
29. M. P. Cosma, T. Tanaka, K. Nasmyth, *Cell* **97**, 299 (1999).
30. G. Cavalli and R. Paro, data not shown.
31. We thank F. Sauer, M. Méchali, L. Ringrose, and I. Chen for comments and critical reading of the manuscript; C. Grimaud for support in genetic analyses of PcG and trxG mutant strains; H. Ehret for technical support; and Y. Cully for photographic work. G.C. was supported by a European Molecular Biology Organization long-term fellowship. The work of G.C. is funded by grants from the Association pour la Recherche sur le Cancer and by the Fondation pour la Recherche Médicale. The work of R.P. is funded by grants from the Deutsche Forschungsgemeinschaft and the Human Frontier Science Program and by the Fond der Chemischen Industrie.

2 April 1999; accepted 16 September 1999

A Direct Estimate of the Human αβ T Cell Receptor Diversity

T. Petteri Arstila,*† Armanda Casrouge, Véronique Baron, Jos Even, Jean Kanellopoulos, Philippe Kourilsky

Generation and maintenance of an effective repertoire of T cell antigen receptors are essential to the immune system, yet the number of distinct T cell receptors (TCRs) expressed by the estimated 10^{12} T cells in the human body is not known. In this study, TCR gene amplification and sequencing showed that there are about 10^6 different β chains in the blood, each pairing, on the average, with at least 25 different α chains. In the memory subset, the diversity decreased to 1×10^5 to 2×10^5 different β chains, each pairing with only a single α chain. Thus, the naïve repertoire is highly diverse, whereas the memory compartment, here one-third of the T cell population, contributes less than 1 percent of the total diversity.

Adaptive immunity is dependent on a genetic recombination machinery that assembles a diverse set of functional immunoglobulin or TCR genes from a pool of discontinuous gene segments. The available pool, for the human αβ TCR, consists of 42 variable (V) and 61 joining (J) segments in the α locus and 47 V, two diversity (D), and 13 J segments in the β locus. During the V_α-J_α or V_β-D_β-J_β rearrangement, nucleotide additions or deletions at the junctions add to the diversity (*1, 2*). As a result, most of the variation in each chain lies in the complementarity-determining region 3 (CDR3), which is encoded by the V(D)J junction and interacts with the antigenic peptide presented by the major histocompatibility complex molecule (*3*). A further diversifying factor is the pairing of an α chain to a β chain to form the TCR heterodimer. The potential diversity thus created has been calculated to be perhaps 10^{15} (*2*). Obviously, not all of it is used, but the T cell repertoire remains complex enough to have precluded any attempts to directly measure it. Here, we have analyzed a fraction of the repertoire in a way that allowed us to extrapolate the results to the whole repertoire (*4*).

To determine the diversity of the TCR β chain, we selected a V-gene family with a single member, $V_\beta18$, and studied its rearrangement to the $J_\beta1.4$ segment. Complementary DNA from 10^8 peripheral blood T cells from a healthy donor (male, between 20 and 30 years old) was amplified with $V_\beta18$- and $J_\beta1.4$-specific primers and separated on an acrylamide gel, producing a pattern of eight bands spaced by three nucleotides that correspond to in-frame TCR transcripts with different CDR3 lengths (Fig. 1A). The band corresponding to a CDR3 with a length of 12 amino acids was purified and cloned, and all of the different sequences within it were then identified by sequencing (Fig. 1B) (*5*). The number of different sequences, 17, was used to calculate the β-chain diversity in the sample (Table 1). The intensity of the bands in a V-J profile is proportional to the diversity within them (*6*), so the 12–amino acid band contained 9.3% of the total $V_\beta18$-$J_\beta1.4$ diversity. The $J_\beta1.4$ gene is used, on the average, by 3.0% of T cells (*7*), and the frequency of $V_\beta18^+$ cells in our donor was 0.8%. Together, these give an average frequency of $(0.093 \times 0.03 \times 0.008)/17 = 1.3 \times 10^{-6}$ for any one β sequence; that is, the sample contains 0.8×10^6 different β chains. This was confirmed by an analysis of the $V_\beta16$-$J_\beta2.2$ rearrangement in the same sample. The $V_\beta16$ family also consists of only one member, but unlike the $J_\beta1.4$ gene, the $J_\beta2.2$ gene may be rearranged to both D_β genes and thus has, in principle, more potential to generate diversity. We found 15 different sequences with a 13–amino acid CDR3 and a total diversity of 0.8×10^6 different β chains. In a second donor (female, 26 years old), the $V_\beta16$-$J_\beta2.2$, 13–amino acid band contained 14 different sequences, with the total diversity in this sample of 10^8 T cells being 1.2×10^6. Thus, analysis of three rearrangements, from two donors, produces estimates that are in close agreement and indicates that the repertoire in the blood consists of at least 10^6 different β chains.

To determine how well our sample reflected the whole diversity in blood, we analyzed the $V_\beta18$-$J_\beta1.4$ 12–amino acid rearrangement in two more samples of 10^8 T cells from the first donor. Overall, 69% of the distinct sequences were shared between the three samples. The second sample contained four previously unidentified sequences, and the third sample contained only one. This suggests that together these three samples contain most of the total β-chain diversity in the blood. Thus, there are probably less than

Unité de Biologie Moléculaire du Gène, INSERM U277, Institut Pasteur, 75724 Paris Cedex 15, France.

*Present address: Department of Virology, Haartman Institute, University of Helsinki, Haartmaninkatu 3, FIN-00014 Helsinki, Finland.

†To whom correspondence should be addressed. E-mail: petteri.arstila@helsinki.fi