23. Simpson, R. B. *Nucleic Acids Res.* **8**, 759-766 (1980).
24. Yu, X. M. & Reznikoff, W. S. *Nucleic Acids Res.* **12**, 5449-5464 (1984).
25. Weber, I. T. & Steitz, T. A. *Proc. natn. Acad. Sci. U.S.A.* **81**, 3973-3977 (1984).
26. Schmitz, A. *Nucleic Acids Res.* **9**, 277-292 (1981).
27. Ogata, R. T. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **74**, 4973-4976 (1977).
28. Becker, M. & Wang, J. C. *Nature* **309**, 682-687 (1984).
29. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 497-560 (1980).

Cell-type-specific contacts to immunoglobulin enhancers in nuclei

George M. Church*, Anne Ephrussi†, Walter Gilbert‡ & Susumu Tonegawa†

* Department of Anatomy, University of California, San Francisco, California 94143, USA
† Center for Cancer Research and Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA
‡ Biogen Research Corporation, 14 Cambridge Center, Cambridge, Massachusetts 02142, USA

The introns separating the variable and constant regions of active immunoglobulin genes contain tissue-specific transcriptional enhancer elements[1-3], DNA segments which act in *cis* in an orientation- and distance-independent (up to a few kilobases (kb)) manner to enhance transcription initiation at adjacent promoters[4-9]. The immunoglobulin heavy-chain enhancer is active only in lymphoid cells: in transfection assays it is capable of controlling in *cis* transcription from the simian virus 40 (SV40) T-antigen, rabbit β-globin and immunoglobulin gene promoters up to at least 2 kb away[1-3]. Genetic deletion analysis suggests that a region of as few as 140 base pairs (bp) is sufficient for the enhancement effect[1,2]. These functional characteristics and DNA sequences are conserved between mouse and man[10-15]. However, it is not known whether tissue-specific proteins bind to the enhancer. Proteins that interact with DNA at specific sequences can prevent or enhance the reactions of individual guanines or adenines with dimethyl sulphate (DMS)[16], and this property has been used to display the DNA contacts of various regulatory proteins[16-26]. Here we apply this DMS strategy in experiments involving single-copy genes within intact mammalian nuclei using genomic sequencing[27].

Nuclei were isolated from tissue culture cells and treated with DMS at 20 °C (Fig. 2*A* legend). The DNA was then isolated, cleaved with *Eco*RI, electrophoresed, after appropriate chemical treatment, on a denaturing sequencing gel, and the lanes of DNA were electrophoretically transferred and crosslinked to nylon membranes using ultraviolet light. Hybridization with a short single-stranded probe (Fig. 1) complementary to the terminal 110 nucleotides of one strand produced the image of a DNA sequencing ladder extending from the 3′ or 5′ end of the appropriate *Eco*RI site in the genome.

Figure 2*A* shows the genomic sequence pattern for the upper strand of the enhancer region of J558L, an immunoglobulin A (IgA) messenger RNA-producing myeloma, M104E, an IgM mRNA-producing myeloma, and L-cells. Figure 2*A*, lanes *f*, *g*, display the pattern produced by DMS treatment of J55L nuclei for 1 and 2 min, respectively. A band is present for every guanine (or cluster of guanines) expected from the known sequence, from positions 483 to 551; bands do not appear at any other position. The band at the position of the open circle in Fig. 2*A*, lane *g* (nuclei), is light (protected from DMS) relative to the band at the corresponding position in lane *h* (isolated DNA). The band just below this is darker (enhanced DMS reactivity) in the myeloma nuclei relative to the isolated DNA. Figure 4 shows the positions of these effects in the enhancer sequence, at guanine 404 (protected) and at the guanine doublet 406-407 (enhanced); the 406-407 doublet was not resolved. The IgM-producing MOPC-104E myeloma ascites tumour displays a similar pattern of protected and enhanced guanines (Fig. 2*A*, lanes *j*, *k*), while nuclei from fibroblastic non-lymphoid L-cells[28] do not show this effect (Fig. 2*A*, lanes *n*, *o*).

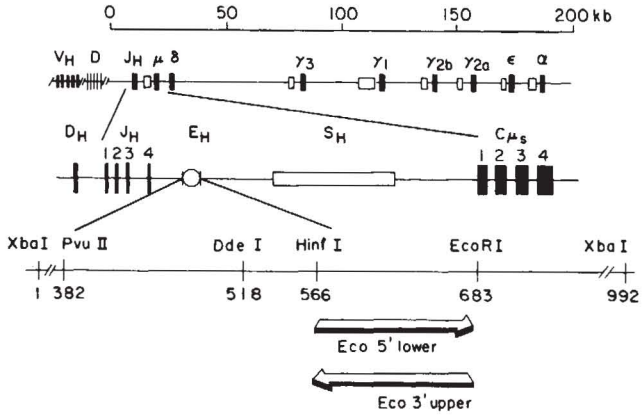

Fig. 1 Immunoglobulin heavy-chain gene structure and DNA probe strategy. The top line displays an overview of the locus, including variable, diversity and joining regions (respectively V_H, D_H, J_H) and eight constant-region exon clusters (μ-α, dark boxes). The enhancer region (E_H, open circle) is shown relative to flanking exons (D_H, J_H, C_μ, dark boxes) and switch regions (S_H, open boxes) present in the unrearranged germline configurations. Beneath it is an expansion of the enhancer region showing the positions of restriction sites used relative to the *Xba*I site. The sense of the single-stranded ^{32}P-labelled DNA probes is indicated by large arrows pointing 5′ to 3′. The probe names refer to the type of sequence they will produce. For example, *Eco* 3′ upper will produce the sequence ladder expected of 3′ end-labelling the *Eco*RI site. The 112-bp *Eco*RI/*Hin*fI fragment cloned in both orientations into *Sma*I-cut M13 mp8 phage vector provided the single-stranded templates for probe syntheses as described elsewhere[27].

Figure 2*B* shows the opposite strand obtained by reprobing the same membrane with the *Eco* 5′ lower probe. No effects were seen in the region near 404-407 on the other strand. However, 125 bp away, guanine 530 shows enhanced reactivity relative to isolated DNA or L-cell nuclei. Three guanines are indicated in Fig. 2*B* (positions 531, 533, 534) which showed DMS protection in Fig. 3*A* (lanes *a*, *d*, *f*, *m*, *n*) and Fig. 3*D* (lanes *a*, *e*-*h*). The two clusters of affected bases, centred about positions 404 and 533 (Fig. 4), show similar (inverted) sequence and protection patterns.

The two myelomas J558L and MOPC-104E produce high levels of immunoglobulin heavy-chain mRNA. There is only one allele of the enhancer in MOPC-104E; the other is deleted (Y. Kurosawa and S.T., unpublished). In J558L, we see the average of two allelic enhancers in two different chromosomal locations; one is producing heavy-chain mRNA, the other has been translocated to within the c-*myc* oncogene 5′ noncoding region[29]. In L-cells, both enhancer alleles should be in the germline configuration.

The effects observed cannot reflect the inherent variations in the DMS reactivity of guanines in double-stranded DNA or variations due to DNA sequence polymorphisms or somatic mutations, because DMS and piperidine treatments were performed on isolated DNA from each cell line in the same type of buffer as used for the nuclei (Fig. 2*A*, *B*, lanes *a*, *h*, *l*). Could the changes in DMS reactivity be the consequence of transcription opening up short stretches of single-stranded DNA? Certain guanines, such as the middle guanine of a 5′ GGA 3′ sequence, have a weak reaction in native DNA which can be increased in denatured DNA[20]. The RNA polymerase increases the reaction of a guanine in the *lac* promoter by this mechanism[20]. Reactions in denaturing formamide for the enhancer guanines assayed here[36] result in slight enhancements relative to native DNA at four 5′ GGA 3′ sequences, but not at the sites we find enhanced in nuclei. Thus, DNA secondary structural changes do not explain these effects.

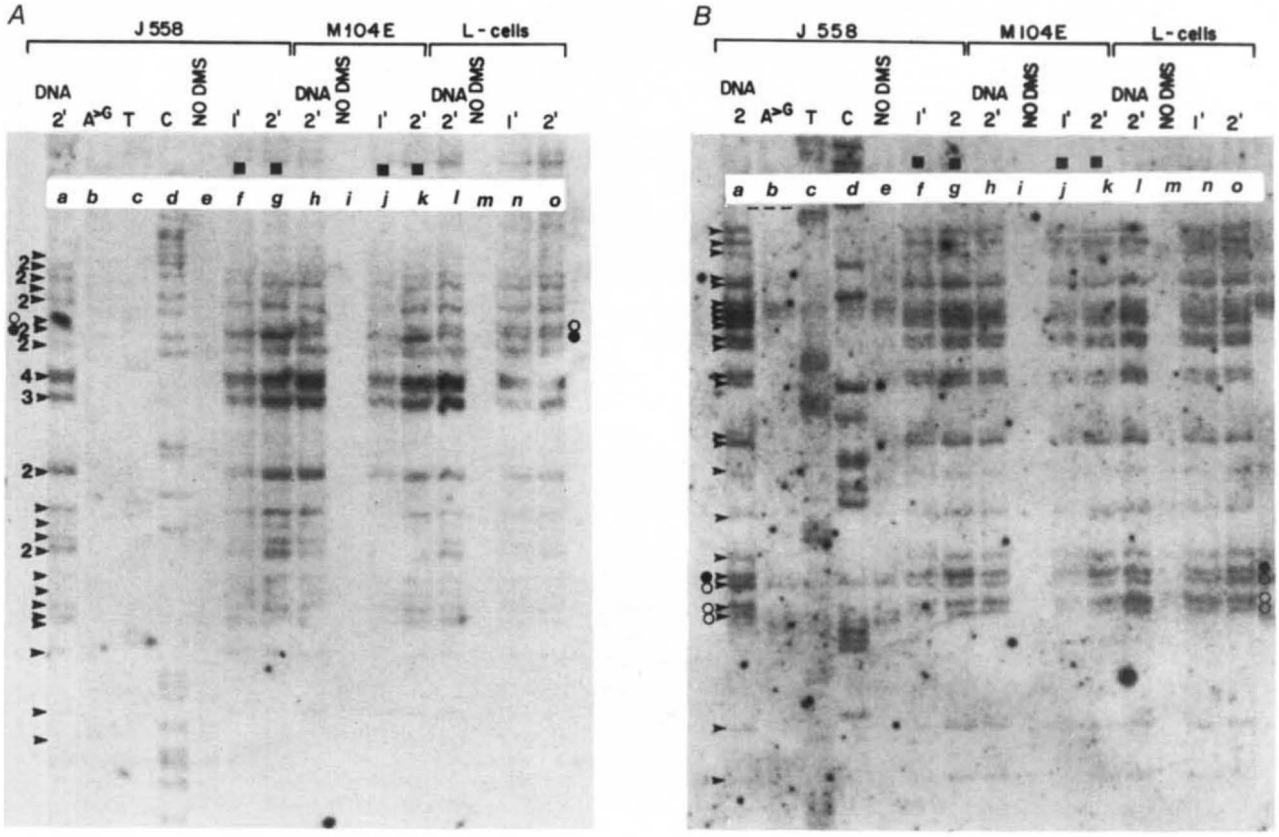

Fig. 2 *A*, DMS reactions in myeloma and fibroblast nuclei. Lanes *a–g*, DNA from J558L myeloma cells[1] (α, λ producers); lanes *h–k*, DNA from MOPC-104E myeloma cells[32] (μ, λ producers); lanes *l–o*, DNA from L-cells (tk^- $aprt^-$)[28] (no immunoglobulin production). Nuclei were prepared by Dounce homogenization with the loose pestle at 4 °C in 250 mM sucrose, 60 mM KCl, 15 mM NaCl, 3 mM $MgCl_2$, 15 mM Tris *p*H 7.5, 0.5 mM β-mercaptoethanol, 20 μM EGTA, 0.05% Triton X-100, followed by addition of 2 M sucrose to a final concentration of 500 mM and pelleting at 800*g* for 5 min. The loose nuclear pellet is easily resuspended in the same buffer lacking Triton, transferred away from the lower packed cell and debris pellet, then repelleted and resuspended at 10^8 nuclei ml^{-1}. Tris buffer and β-mercaptoethanol do not interfere with the DNA reaction with DMS. In lanes *f, j, n*, nuclei were incubated with 0.5% DMS (50 mM) at 20 °C for 1 min. In lanes *g, k, o*, nuclei were incubated with 0.5% DMS at 20 °C for 2 min. Lanes *a, h, l* are control DMS reactions on isolated DNA. In lanes *e, i, m*, control nuclei were incubated without DMS at 20 °C for 2 min. Lanes *b–d* are A > G, T and C reactions[33–35], respectively, using isolated J558L DNA. The reactions were analysed by cleavage with *Eco*RI enzyme, ethanol precipitation and the direct genomic sequencing procedure[27], loading 15–30 μg of DNA per lane. The lanes were transferred and crosslinked to a nylon membrane and probed with the *Eco* 3′ upper probe described in Fig. 1. The arrows indicate positions of all guanines. Positions 1–190 (see Figs 1, 5) were assayed. Shown here are 483 (topmost arrow) to 551 (bottom). Clusters of two, three or four guanines which were not well resolved are indicated by numbers beside the arrows. Guanines are indicated which are either protected (○) or enhanced (●) in a lymphoid-specific fashion. ■, Those lanes where these altered reactivities are observed. MOPC-104E has only a single copy per cell of the region displayed; we detect 1 fg (2×10^4 molecules) of probe complementary DNA in each band (there are 6×10^{-18} molecules per lane). *B*, After autoradiography, the membrane used in *A* was stripped of the first probe and reprobed with the *Eco* 5′ lower probe. The lanes and symbols are as in *A*. The topmost arrow indicates position 384, the bottom-most indicates position 494.

The dark bands interpreted as enhanced guanines are not caused by crossreacting or contaminating restriction fragments of the appropriate lengths, as no enhanced bands of equal length appear on the opposite strand. Also, *Eco*RI low-specificity nicking, cell-type-specific endogenous nuclease or depurination activities are ruled out by the omission of DMS from the incubations in Fig. 2*A*, *B* (lanes *e, i, m*). Figure 3*A*, lanes *c, e, h, k*, which omits the piperidine step, demonstrates controls for any DMS-activated cell-type-specific nucleases. The tissue-specific effects on DMS reactivity are best explained by molecules (proteins) making contacts near guanine N-7 atoms in the major groove of the enhancer DNA. These experiments do not prove that the binding of such molecules is the cause, rather than an effect, of the transcriptional activity.

To provide additional controls and as a step towards isolating the molecules responsible for the protection patterns, the stabilities of the putative DNA–protein interactions were monitored as a function of buffer composition and ionic strength. Figure 3*A* shows that the cell-type-specific patterns at guanines 530–534 in the J558L myeloma are resistant to modest alterations in the DMS reaction protocol, for example a fourfold higher Tris concentration (lanes *a–o*), 0 °C (lanes *a, b*), 150 mM monovalent cation concentration (lanes *f, g*) or 6 mM EGTA (lanes *m–o*). The controls in these experiments use the non-immunoglobulin-expressing[30] Friend virus-transformed MEL[31] erythroleukaemic cells, which are non-adherent and closer than L-cells in developmental lineage to lymphoid cells. The interaction responsible for the altered DMS reactivity at the lower site is sensitive to 250 mM monovalent cation concentration (Fig. 3*A*, lane *i*) and higher (lane *l*). The upper site at 404–407 shows the same sensitivity on reprobing the membrane in Fig. 3*B*. The striking difference between the enhanced and the protected guanines visible at 75 and 150 mM (Fig. 3*A*, *B*, *d, f, m, n*) is eliminated at 250 and 500 mM salt (Fig. 3*A*, *B*, lanes *i, l*). Note the similarity of the myeloma and erythroid patterns at high salt concentrations (Fig. 3*A*, *B*, lanes *i, j*). Figure 3*C* shows that small shifts in the lymphoid pattern of residues 404–407 towards the pattern seen with isolated DNA do not begin until about 225 mM salt. The protection pattern of residues 530–534 is lost at the same concentrations on the other strand (not shown). Some loss can occasionally be seen at both sites at concentrations as low as 75 mM (Fig. 3*D*, lane *e*). The patterns are clearest in

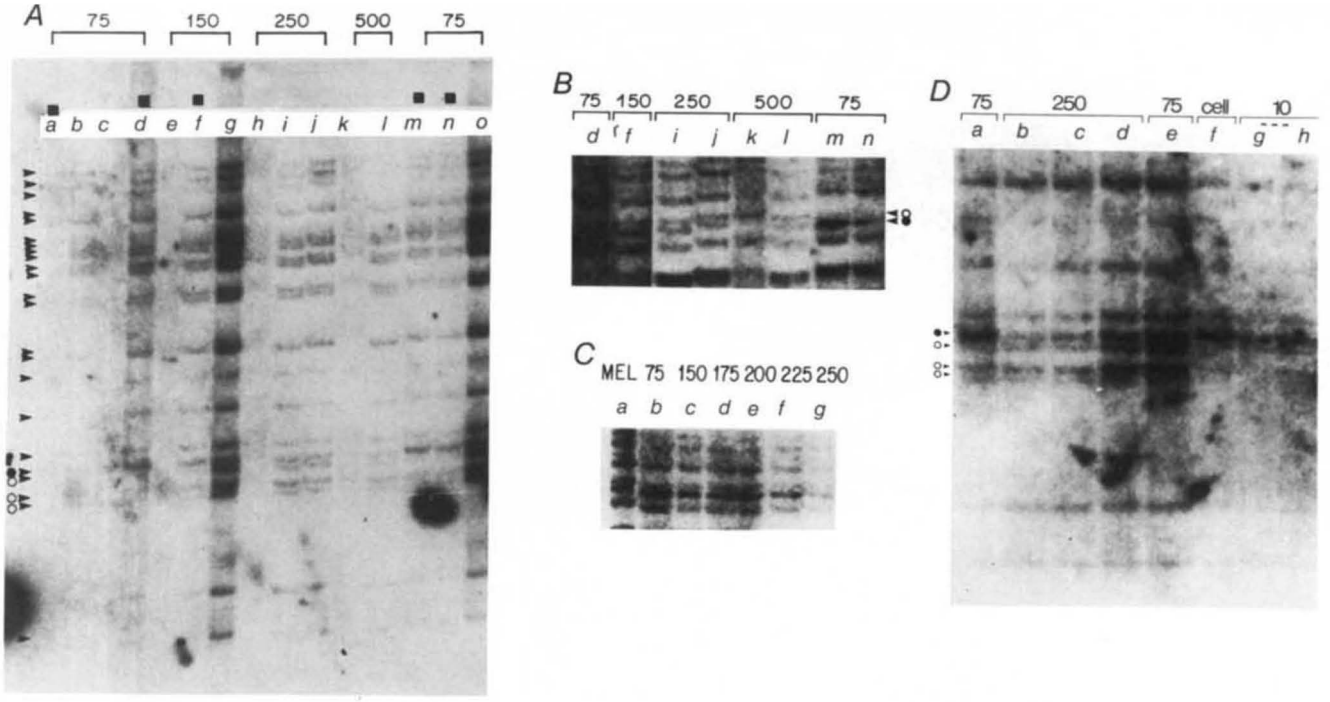

Fig. 3 *A*, Effect of NaCl on DMS reaction patterns in lymphoid and erythroid nuclei. Nuclei isolation conditions were as for Fig. 2. Lanes *a*, *b*, nuclei were incubated with 0.5% DMS at 4 °C for 8 min. Lanes *c–o*, nuclei were incubated with 0.5% DMS at 20 °C for 4 min. Lanes *a*, *c–f*, *h–i*, *k–n*, DNA from J558L cell nuclei. Lanes *b*, *g*, *j*, *o*, DNA from MEL-D2-745-JG1-PC4 erythroleukaemic[31] cell nuclei. Lanes *c*, *e*, *h*, *k*, controls where piperidine cleavage at modified guanine bases was omitted. The Na^+ plus K^+ concentrations were as follows: 75 mM in lanes *a–d* and *m–o*; 150 mM in lanes *e–g*; 250 mM in lanes *h–j*; 500 mM in lanes *k–l*. The basic reaction buffer (75 mM Na^+ plus K^+) consisted of (in mM) 250 sucrose, 60 KCl, 15 NaCl, 3 $MgCl_2$, 60 Tris *p*H 8.2, 0.5 β-mercaptoethanol and 20 μM EGTA. Other solutions contained higher NaCl concentrations. The solutions used in lanes *m–o* contained 6 mM EGTA. *B*, After autoradiography, the membrane was stripped of the first probe and reprobed with the *Eco* 3′ upper probe. The lanes and symbols are as for *A*. The enhancement/protection symbols are at positions 404–407. *C*, As for *A* and *B*, adjusting only NaCl concentrations to achieve total monovalent cation concentrations indicated at the head of each lane (in mM). The site at 404–407 is shown using the *Eco* 3′ upper probe. *D*, The lower strand of guanines near site 531–534, showing effects of very low salt and transient high salt treatment. DMS reactions for lanes *a* and *e* were done in standard (75 mM) conditions as in previous figures. Lane *b*, DMS reaction at 250 mM salt; lane *c*, at 75 mM after dilution from 250 mM; lane *d*, at 75 mM after washing nuclei at 250 mM; lane *f*, DMS reaction on cells in growth medium[36]. Lanes *g*, *h*, nuclei at 10 mM NaCl, 0.05% Triton X-100 with 10 mM Tris-HCl *p*H 7.5 and 50 mM Tris *p*H 8.2, respectively.

very low salt, 10 mM (lanes *g*, *h*) and these are comparable to what is seen in whole cells treated with DMS[36]. The salt suppression of the tissue-specific reaction patterns observed might be due directly to titration of electrostatic interactions or indirectly to changes in DNA conformation or to the major release of unrelated chromosomal proteins. In any case, these data support our suggestion that the two sites have similar properties, including salt elution.

The similarities in DNA sequence and DMS reactivity pattern between the two putative binding sites noted above could represent binding by a single recognition protein. If this protein were necessary for a lymphoid-specific function, we would predict that similar DNA sequences would be found in other known immunoglobulin enhancers and homologous regions in other species. A human heavy-chain enhancer which functions in mouse lymphoid cells lies within a 279-bp *Alu*I fragment[13] and contains two homologous sequences at positions similar to those in the mouse gene (Fig. 5). A mouse κ enhancer is contained on a 475-bp *Alu*I fragment between the κ joining (J_κ) and constant (C_κ) region. Within this region, a 130-bp stretch is conserved among mice, rabbits and man[11], and contains two homologous 10-bp sequences in all three species. Outside this region there are two further matches. Figure 5 derives a consensus sequence, GCCAGGTGGC, from these comparisons.

We interpret the broad variety of deletion studies[1,2,10,13] as showing that other sequences as well as the consensus sequences are necessary for enhancer action. A single 10-bp immunoglobulin consensus element alone is probably not sufficient for enhancement because single examples of similar sequences (Fig. 5) occur in pBR322 and SV40, while these vectors in transfection assays have little or no lymphoid-specific transcription background. If the 10-bp element does contribute to lymphoid-specific transcriptional enhancement, one must be cautious in interpreting the deletion data. For example, the 140-bp *Pvu*II/*Dde*I enhancer fragment[1] inserted 80 bp from the pBR322 match could represent an unanticipated duplication of the 10-bp elements synergistically producing some enhancement activity.

The significance of the reiterated consensus sequence can be tested by examining the DMS reactivity of the other immunoglobulin enhancers in various cell types and by analysing the effects of point mutations on DMA protection/enhancement patterns and transcription levels after transfection. The DMS reactivity patterns provide assays for isolating enhancer binding molecules.

In conclusion, we have detected a factor that binds to an enhancer element in a tissue- and sequence-specific fashion.

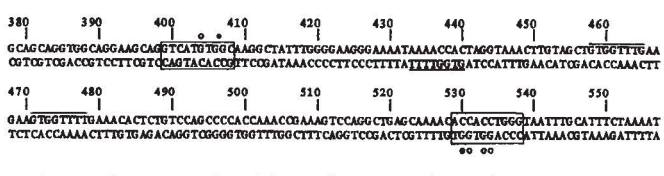

Fig. 4 Summary of positions of G nucleotides with altered DMS reactivities within the immunoglobulin heavy-chain enhancer region. Altered DMS reactivities of guanines specific for lymphoid nuclei are indicated for protected (○) and enhanced (●), relative to isolated DNA. Thin lines indicate homologies to viral enhancer core sequences[38] noted in previous work[1,2,10–15]. Numbering is from the *Xba*I site as in Fig. 1. Sequence data are from refs 1, 2. Boxes enclose 10-bp sequences aligned in Fig. 5.

locus			5'													3'	match	reference
									○		●							
mouse	H	404	›	A	G	G	T	C	A	T	G	T	G	G	C	A A G	8	1,2
										○	○		○	●				
mouse	H	533	‹	T	A	C	C	C	A	G	G	T	G	G	T	G T T	8	1,2
mouse	H	353	‹	A	G	T	C	A	A	G	A	T	G	G	C	C G A	7	1,2
mouse	H	385	‹	G	C	A	G	C	A	G	C	T	G	G	C	A G G	7	1,2
human	H	89	›	A	C	A	G	C	A	G	G	T	G	G	C	A G G	8	12-14
human	H	226	‹	T	G	T	C	C	A	G	G	T	G	T	T	G T T	7	12-14
mouse	k	77	‹	C	T	G	C	C	A	G	A	T	G	G	C	C T C	9	11
human	k	77	‹	C	T	G	C	C	A	G	G	T	G	G	C	C T C	10	11
rabbit	k	77	‹	C	T	G	C	C	A	G	A	T	G	G	C	C C C	8	11
mouse	k	129	›	C	A	G	G	C	A	G	G	T	G	G	C	C C A	9	10,11,15
human	k	129	›	T	A	G	G	C	A	G	G	T	G	G	C	C A A	9	11
rabbit	k	129	›	G	A	G	G	C	A	G	G	T	G	A	C	C C A	8	11
mouse	k	180	›	G	T	C	C	C	A	T	G	T	G	G	T	T A C	7	10,11,15
rabbit	k	199	›	C	G	G	C	C	A	G	G	T	G	C	A	G G T	8	11,37
Ig enhancers						G	C	C	A	G	G	T	G	G	C			
frequency:				36	57	93	79	100	79	50	29							
					29	57	100	71	100	64	29							
viral enhancers									G	T	G	G	A T	A T	A T	G	4	38
sv40 T		4073	‹	T	T	G	C	C	A	G	G	T	G	G	G	T T A	9	39,40
pBR322		4281	‹	A	C	G	T	C	A	G	G	T	G	G	C	A C T	9	41

Fig. 5 A consensus sequence for immunoglobulin intron enhancers. The position of guanines with altered DMS reactivity in the mouse heavy-chain enhancer in nuclei are indicated: ○, protected; ●, enhanced. Beneath the consensus sequence is the percentage of the matches found at that nucleotide position for the 14 enhancer segments aligned as shown. A weak similarity with a viral consensus[39] is also shown. The positions of the centres of the sequences are numbered in accordance with the references given. The symbols ⟨ , ⟩ indicate whether the sequence shown has been inverted relative to the original report or not, respectively. The column marked 'match' indicates the number of matches to the 10-bp consensus for the enhancer segment on that line.

This supports the hypothesis that tissue-specific genes make products that act on enhancers to turn on sets of genes during differentiation.

We thank Virginia Folsom, Stephen Gillies, Herman Eisen and Vivien Igras for cell lines and help, Julian Banerji, Harry Nick, Ting Wu and Kai Zinn for discussions, Mary Weiss, Tom Maniatis, David Botstein and Jim Wang for comments on the manuscript. This work was supported by Biogen NV, NIH grants GM09541-22 (to W.G.), AI-17879-03 (to S.T.), predoctoral training grant 2T32CA09255 (to A.E.), and American Cancer Society grant ACS-NP-327A (to S.T.). G.M.C. is a Monsanto Fellow of the Life Sciences Research Foundation.

Received 24 October; accepted 20 December 1984.

1. Gillies, S. D., Morrison, S. L., Oi, V. T. & Tonegawa, S. *Cell* **33**, 717-728 (1983).
2. Banergi, J., Olson, L. & Schaffner, W. *Cell* **33**, 729-740 (1983).
3. Neuberger, M. S. *EMBO J.* **2**, 1373-1378 (1983).
4. Benoist, C. & Chambon, P. *Nature* **290**, 304-310 (1981).
5. de Villiers, J. & Schaffner, W. *Nucleic Acids Res.* **9**, 6251-6264 (1981).
6. Gruss, P., Dhar, R. & Khoury, G. *Proc. natn. Acad. Sci. U.S.A.* **8**, 943-947 (1981).
7. Moreau, P. *et al. Nucleic Acids Res.* **9**, 6047-6068 (1981).
8. Wasylyk, B., Wasylyk, C., Augereau, P. & Chambon, P. *Cell* **32**, 503-514 (1983).
9. Banerji, J., Rusconi, S. & Schaffner, W. *Cell* **27**, 299-308 (1981).
10. Picard, D. & Schaffner, W. *Nature* **307**, 80-82 (1984).
11. Emorine, L., Kuehl, M., Weir, L., Leder, P. & Max, E. E. *Nature* **304**, 447-449 (1983).
12. Rabbits, T. H., Forster, A., Baer, R. & Hamlyn, P. H. *Nature* **306**, 806-809 (1983).
13. Hayday, A. *et al. Nature* **307**, 334-340 (1984).
14. Mills, F. C., Fisher, L. M., Kuroda, R., Ford, A. M. & Gould, H. J. *Nature* **306**, 809-812 (1983).
15. Chung, S.-Y., Folsom, V. & Wooley, J. *Proc. natn. Acad. Sci. U.S.A.* **80**, 2427-2431 (1983).
16. Gilbert, W., Maxam, A. & Mirzabekov, A. *Alfred Benzon Symp.* **9**, 139-148 (Munksgaard, Copenhagen, 1976).
17. Johnson, A. D. *et al. Nature* **294**, 217-223 (1981).
18. Taniguchi, T., O'Neill, M. & De Crombrugghe, B. *Proc. natn. Acad. Sci. U.S.A.* **76**, 5090-5094 (1979).
19. Siebenlist, U., Simpson, R. B. & Gilbert, W. *Cell* **20**, 269-281 (1980).
20. Johnsrud, L. *Proc. natn. Acad. Sci. U.S.A.* **75**, 5314-5318 (1978).
21. Lu, A.-L., Jack, W. E. & Modrich, P. *J. biol. Chem.* **256**, 13200-13206 (1981).
22. Germino, J. & Bastia, D. *Cell* **34**, 124-134 (1983).
23. Brent, R. & Ptashne, M. *Proc. natn. Acad. Sci. U.S.A.* **78**, 4204-4208 (1981).
24. Kirkegaard, K. & Wang, J. *Cell* **23**, 721-729 (1981).
25. Karin, M. *et al. Nature* **308**, 513-519 (1984).
26. Myers, R. M., Rio, D. C., Robbins, A. K. & Tjian, R. *Cell* **25**, 373-384 (1981).
27. Church, G. M. & Gilbert, W. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991-1995 (1984).
28. Wigler, M. *et al. Proc. natn. Acad. Sci. U.S.A.* **76**, 1373-1376 (1979).
29. Gerondakis, S., Cory, S. & Adams, J. M. *Cell* **36**, 973-998 (1984).
30. Kemp, D. J., Harris, A. W., Cory, S. & Adams, J. M. *Proc. natn. Acad. Sci. U.S.A.* **77**, 2876-2880 (1980).
31. Volloch, V. & Housman, D. *J. Cell Biol.* **93**, 390 (1982).
32. McIntire, K. R., Asofsky, R. M., Potter, M. & Kuff, E. L. *Science* **150**, 361-363 (1965).
33. Maxam, A. M. & Gilbert, W. *Meth. Enzym.* **65**, 497-560 (1980).
34. Rubin, C. M. & Schmid, C. W. *Nucleic Acids Res.* **8**, 4613-4619 (1980).
35. Herr, W., Corbin, V. & Gilbert, W. *Nucleic Acids Res.* **10**, 6931-6944 (1983).
36. Ephrussi, A., Church, G., Tonegawa, S. & Gilbert, W. *Science* **227**, 134-140 (1985).
37. Emorine, L. & Max, E. E. *Nucleic Acids Res.* **11**, 8877-8890 (1983).
38. Weiher, H., Konig, M. & Gruss, P. *Science* **219**, 626-631 (1983).
39. Reddy, V. B. *et al. Science* **200**, 494-502 (1978).
40. Fiers, W. *et al. Nature* **273**, 113-119 (1978).
41. Sutcliffe, J. G. *Cold Spring Harb. Symp. quant. Biol.* **43**, 77-90 (1978).

Similarity between the vaccinia virus 19K early protein and epidermal growth factor

A. H. Reisner

CSIRO, Division of Molecular Biology, North Ryde, New South Wales, PO Box 184, Australia 2113

An analysis of the 1,217-amino acid residue sequence of the precursor of mouse epidermal growth factor (mEGF)[1] revealed regions of considerable similarity with bovine factor X, a blood coagulation factor. Similarities of mEGF itself with factor X[2], pancreatic secretory trypsin inhibitor[3] and, most strikingly, transforming growth factor I (TGF-I)[4,5] have been observed. On the basis of the comparisons described here, it seems that the presumptive 140-residue 19K early protein (relative molecular mass (M_r) 19,000) of vaccinia virus[6] from residues 40–91 shows an overall identity of 36% (19/53 residues) with both mEGF and urogastrone (human epidermal growth factor, hEGF); a single deletion is assumed for vaccinia virus 19K protein which allows the six Cys residues (positions 45–80) to be aligned with those of mEGF or hEGF. This protein is encoded in the 10.3-kilobase (kb) inverted terminal repeat[6]. Because it is an early protein with an EGF-like central portion, the 19K vaccinia virus protein may have an autocrine function and may be required for DNA synthesis.

The double-stranded DNA genome of vaccinia virus consists of ~182,000–186,000 nucleotide base pairs (bp)[7,8] and can

a

```
mEGF        NSYPGCPSSYDGYCLNGGVCMHIESLDSYTCNCVIGYSGDRCQTRDLRWWELR

hEGF        NSDSECPLSHDGYCLHDGVCMYIEALDKYACNCVVGYIGERCQYRDLKWWELR      37/53

TGF-I       VVSFNKCPDSHTQYC-FHGTCRFLVQEEKPACVCHSGYVGVRCEHADLLA----    17/53 mEGF
                                                                       21/53 hEGF

Vaccinia    PAIRLCGPEGDGYCL-NGDCIHARDIDGMYCRCSHGYTGIRCQHVVLVDYQRSE    19/53
19K Prot
(40-92)
```

b

```
Vaccinia    MSMKYLMLLFAAMIIRSFADSGNAIETTSPEITNATTDI
19K Prot
Complete    PAIRLCGPEGDGYCLNGDCIHARDIDGMYCRCSHGYTGIRCQHVVLVDYQRSE

            MPNTTTSVIPSPGIMLVLVGIIIITCCLLSVVYRFTRRTKLPIQDMVVP
```

Fig. 1 *a*, Amino acid residue sequences for mEGF, urogastrone (hEGF), TGF-I and residues 40–92 of vaccinia virus 19K protein. The values on the right indicate the ratio of matches found with mEGF after alignment. Alignments were performed using the SEQA programme of Kanehisa *et al.*[18]. *b*, Complete sequence of the vaccinia virus 19K protein. Apart from the six Cys residues comparable with those in EGF, two additional ones are present at positions 118 and 119. Residues 6–15 and 106–123 are strongly hydrophobic.