Association of Charge Clusters with Functional Domains of Cellular Transcription Factors

Volker Brendel; Samuel Karlin

Proceedings of the National Academy of Sciences of the United States of America, Volume 86, Issue 15 (Aug. 1, 1989), 5698-5702.

Stable URL:
http://links.jstor.org/sici?sici=0027-8424%2819890801%2986%3A15%3C5698%3AAOCCWF%3E2.0.CO%3B2-B

Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 5698–5702, August 1989
Biochemistry

Association of charge clusters with functional domains of cellular transcription factors

(DNA-binding domains/activating domains/"zinc finger"/"leucine zipper")

VOLKER BRENDEL AND SAMUEL KARLIN*

Department of Mathematics, Stanford University, Stanford, CA 94305

Contributed by Samuel Karlin, April 19, 1989

ABSTRACT **Using rigorous statistical methods, we have identified and evaluated unusual properties of the distribution of charged residues within the sequences of eukaryotic cellular transcription factors. Virtually all transcription factors, including GAL4, c-Jun, C/EBP, CREB, Oct-1, Oct-2, Sp1, Egr-1, CTF-1, steroid and thyroid hormone receptors, and others, carry one or more highly significant charge clusters. For the most part these clusters (conserved within families of homologous proteins) are of positive net charge but contain also substantial numbers of acidic residues. Predominantly basic charge clusters are often, but not exclusively, associated with DNA-binding domains, and vice versa. Negative charge clusters of note occur only in the yeast protein PHO4 and in the proteins encoded at the *Drosophila* loci zeste (z) and *knrl*. This dearth of statistically significant negative charge clusters raises questions with respect to the generality of acidic activation domains. A number of sequences (Oct-1, Oct-2, zeste, Dhr23, E75, and knrl) contain multiple charge clusters together with one or more significantly long uncharged regions. The occurrence of multiple charge clusters is a rare phenomenon (found in less than 3% of all proteins, mainly in *Drosophila* developmental control proteins and in transactivators of eukaryotic DNA viruses). Most of the proteins with zinc-binding "fingers" carry a mixed charge cluster centered at the zinc-finger motif preceded by a long uncharged stretch, suggestive of a modular structure for these proteins.**

Eukaryotic transcription control involves a variety of regulatory proteins that bind specifically to promoter and enhancer elements and associate with RNA polymerase II and/or other factors in elaborate complexes (1, 2). It has been suggested that regions of concentrated charge in the proteins involved in these complexes play an important role in mediating the various protein–nucleic acid and protein–protein interactions. For several transcription factors, including the yeast proteins GAL4 and GCN4, mammalian "zinc-finger" proteins like Sp1 and Krox-20, and homeodomain-containing transcription factors (for example, Oct-2 and Pit-1), the DNA-binding domains are associated with basic regions. GAL4 and GCN4 additionally contain one or more separate activating domains, which are acidic (3, 4). By contrast, Sp1 has no acidic regions but contains long uncharged regions that apparently harbor the activation function and may contribute to greater affinity for DNA (5). Several structural motifs are implicated in DNA binding (for review, see ref. 6). The characteristics of transactivating domains have remained unclear except for the quoted predominance of net negative charge in some cases and the complete absence of charged residues in other cases.

We have been generally interested in the distribution of charged residues along protein sequences. Previously we established methods to discern statistically significant charge clusters (defined as 25- to 75-residue segments high in specific charge content) as well as significantly long runs of certain charge types and special periodic patterns of charged residues (7). Application of these methods to a large number of viral proteins revealed a striking richness of significant charge configurations in transactivators and other regulatory proteins of mainly eukaryotic DNA viruses (7–9). A paramount example is the Epstein–Barr virus (EBV)-encoded nuclear antigen EBNA1, a protein required for replication and maintenance of viral genomes in latently infected cells (10). The distribution of the charged residues in this protein distinguishes four separate charge clusters, two of negative and two of positive sign; moreover, the positive charge clusters center on long tandem reiterations of (+, 0),† and the C-terminal negative cluster carries, with two mismatches, the periodic charge pattern $(G, -, -)_7$; all these distributional features are statistically significant in the sense that less than 1% of random sequences of the same amino acid composition would contain any one of these structures (7). Significant charge configurations also occur in all other primary nuclear antigens of EBV expressed in the latent state as well as in a strong transactivator of the lytic cycle (pBMLF1), in the immediate early transactivators ICP0 and ICP4 of herpes simplex virus type 1, p62 and p63 of varicella–zoster virus, and IE1 of cytomegalovirus; in the E1A 32-kDa protein of adenovirus; in the large tumor (T) antigens of papovaviruses; and in E1 and E7 of papillomavirus (7–9). All of these proteins are involved in the regulation of early viral transcription or viral replication and in some cases in oncogenic transformation.

Our objective here is to provide a more complete analysis of the charge distribution in transcriptional regulators. To this end we have screened, in addition to viral proteins, more than 2500 primarily human, mouse, *Drosophila*, yeast, and *Escherichia coli* proteins for significant charge configurations, emphasizing clusters, runs, and periodic patterns involving either or both charge types. Significance was determined as described previously (7, 11). The data show that significant charge configurations abound among cellular proteins involved in regulatory functions, including nuclear transcription factors, steroid and thyroid hormone receptors, nuclear protooncogene products, developmental control proteins, high molecular weight heat shock proteins, and transmembrane proteins like the voltage-gated ion channels, growth factor and neurotransmitter receptors, and opsins, but are uncommon for the bulk of enzymes and constitutive proteins. We shall focus in this paper on nuclear transcription factors

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: EBV, Epstein–Barr virus.

*To whom reprint requests should be addressed.

†Protein sequences are displayed in the standard one-letter code. + stands for R, K, or H; − for D or E; and 0 for all other amino acids. Some authors treat H as uncharged; however, due to the low frequency of H in most proteins, this does not affect the qualitative outcome of our type of analysis.

and hormone receptors. Heat shock proteins are discussed in ref. 11; the detailed results for the other classes of proteins will be discussed elsewhere.

Nuclear Transcription Factors

Some 20 protein factors involved in the regulation of eukaryotic transcription have been identified and sequenced. Four groupings may be distinguished: (*i*) the GCN4-related proteins, characterized by a similar DNA-binding domain toward the C terminus; (*ii*) transcription factors containing multiple zinc-finger sequences; (*iii*) transcription factors with a POU domain (12); and (*iv*) factors of more individual profile, including GAL4, C/EBP, PHO4, zeste and others.

Significant charge clusters occur in virtually all of these proteins while, in contrast to many viral transactivators, there occur almost no significant runs or periodic patterns of charge (PHO4 is an exception; see below). In Fig. 1 we ascertained positive (overwhelmingly basic), negative (overwhelmingly acidic), and mixed charge clusters (involving substantial numbers of charged residues of both signs, where the relative numbers of basic to acidic components may be balanced or skewed). Also shown are significantly long uncharged segments. The characterization of shared and contrasting tendencies and the statistics with respect to occurrence, multiplicity, type, and location of the charge clusters suggest diagnostic sequence features that can provide insights into protein function, structure, and classification.

The DNA-binding domains of GAL4, the GCN4-related proteins, the homeobox part of the POU domain, and the zinc-finger sequences are all highly charged, delimiting by our analysis positive charge clusters or mixed charged clusters with basic residues predominant. Long uncharged segments are frequent in the zinc-finger-containing proteins and also occur in a number of the other transcription factors. CREB, Oct-1, Oct-2, and the zeste gene product carry multiple regions of concentrated charge, a rare attribute found in less than 3% of more than 2500 proteins examined. Significant negative charge clusters occur only in PHO4 and the zeste protein, in sharp contrast to most transactivator proteins of eukaryotic DNA viruses, which commonly manifest negative charge clusters (7–9).

Most of the charged region that entails the DNA-binding function is considerably conserved among GCN4, CPC1, and the Jun/AP-1 family of transcriptional activators (13, 14). When aligned, these three proteins match at 14 (4 R, 4 K, 1 E, 2 A, 1 N, 2 L) of 32 residues in this region, with the strongest conservation occurring among the positively charged residues. A heptad 4-repeat of leucines [$(X_6L)_4$; a "leucine zipper"] starting near the carboxyl end of the charge cluster in conjunction with the basic region is hypothesized to facilitate homo- and heterodimerization of these proteins as α-helices with a parallel orientation characteristic of a coiled-coil conformation (15). The leucine 4-repeat also occurs in C/EBP, Myc, and c-Fos (15), in Fra-1, the protein encoded by the *fos*-related immediate-early gene *fra-1* (16), and in CREB (17) but is not found in CPC1. In c-Jun, GCN4, and C/EBP the leucine repeat is immediately carboxyl to the positive charge cluster in the fourth quartile, whereas in Fos and Fra-1 the leucine repeat is fully contained in the centrally located charge cluster of cumulative balanced charge. The leucine repeat in Myc occurs near the C terminus, substantially distant from the centrally located (acidic) charge cluster. The sets of intervening residues between the leucine components among these proteins show highly varied composition (especially at positions 3 and 4) entailing no semblance of conservation. The extended region is highly charged, predominantly of positive sign, which may be the more decisive ingredient that underlies dimerization inde-

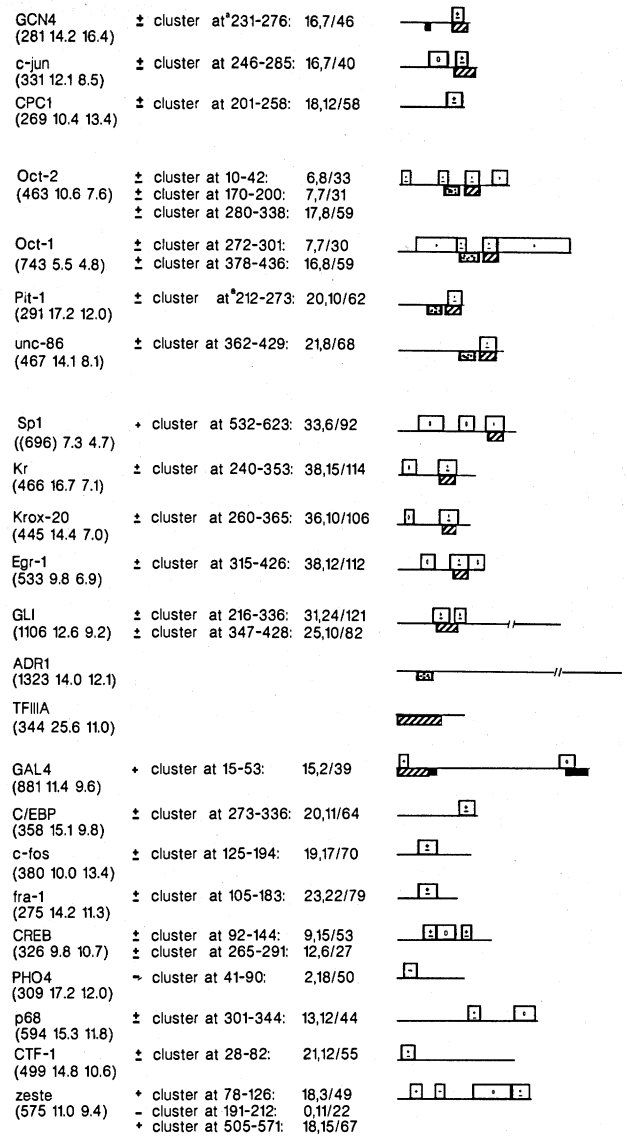

FIG. 1. Significant charge clusters in cellular transcription factors. Below the name of each protein are given the number of residues and the percentage of positively and negatively charged residues in the protein, respectively. Each charge cluster is identified by its coordinates in the protein, the number of positive and negative charges, and the length of the segment. Schematic representations of the proteins are drawn approximately to scale. Charge clusters and long uncharged regions are indicated by open boxes. The DNA-binding (hatched; residues 1–147) and activation (solid; residues 148–196 and 768–881) domains of GAL4 are according to ref. 4 and those of GCN4 (residues 222–281 and 107–125, respectively) according to ref. 3. Positions of the POU (dotted) and homeo (hatched) domains are as follows (12): Oct-2, 187–253 and 281–340; Pit-1, 132–198 and 214–273; Oct-1, 280–354 and 379–438; unc-86, 273–342 and 363–422. The hatched boxes in Sp1, Krüppel (Kr), Krox-20, Egr-1, GLI, TFIIIA, and ADR1 correspond to zinc-finger segments. The coordinates for the uncharged regions are as follows: c-Jun, 123–208; CREB, 145–232; Oct-2, 399–463; Oct-1, 81–271 and 437–743; Sp1, 130–268 and 364–455; Kr, 31–98; Krox-20, 50–109; Egr-1, 150–212 and 427–519; GAL4, 742–813; p68, 510–594; zeste, 337–501. Residues 231–257 (231–260) of GCN4 are largely conserved in c-Jun (residues 258–284; 56% identity) and CPC1 (residues 221–250; 70% identity). Also, the charge clusters of Krox-20 and Egr-1 correspond to highly conserved sequences (residues 260–358 and 318–416, respectively; 93% identity). No localized charge clusters occur in TFIIIA or ADR1. Our criteria for identifying significant charge configurations in a protein sequence have been described (7, 11). Superscript [a] (see GCN4 and Pit-1) indicates that the charge cluster is not statistically significant at the 1% level, but of note for other reasons; see text.

pendent of the leucine residues. It is interesting that the essential region effecting the Fos–Jun complex corresponds to the central mixed charge cluster of Fos (18). Parenthetically, the probability of finding a 4-unit heptad leucine repeat in random amino acid sequences of lengths 300, 500, and 1000 residues and 10% leucine content is about 0.03, 0.04, and 0.09, respectively, large enough as caveats in interpretations of leucine repeats.

CREB is distinguished by two mixed charge clusters separated by a long uncharged region (Fig. 1). The uncharged region is abundant with glutamine and threonine. The second charge cluster may correspond to the DNA-binding domain, although the proposed sequence similarity of this region to the DNA-binding domains of c-Jun and GCN4 (17) would seem to be tenuous. Also the stated acidity of the N terminus is hardly striking.

Oct-2 contains three highly charged segments and a long uncharged C terminus rich in proline and glycine (Fig. 1). The third charge cluster is a homeo-like domain and, together with a close upstream region, forms a bipartite DNA-binding domain (POU domain), conserved in Pit-1, Oct-1, and unc-86 (ref. 12 and references therein). Both Oct-2 and Oct-1 are human promoter-binding proteins that recognize the same octanucleotide. However, whereas Oct-1 is ubiquitous, Oct-2 appears to be B-cell-specific (12). The different specificities of Oct-1 and Oct-2 should reside outside the conserved POU domain. Indeed, Oct-1 and Oct-2 have very different charge structures apart from the POU domain. Oct-1 lacks a third charge cluster and its POU domain is flanked by a glutamine-rich region on the N-terminal side and by an alanine-, serine-, and threonine-rich region on the C-terminal side.

The human transcription factor Sp1 has a highly uneven distribution of charges. The N-terminal 470 residues of the 696C product (19) contain as few as 8 basic and 10 acidic residues, whereas the 92-residue segment 532–623 contains 33 positive and 6 negative residues. The charge cluster centers on three tandem zinc-finger motifs responsible for DNA binding (19). The long uncharged region abounds with serine, threonine, and glutamine residues. This region increases the affinity of Sp1 for DNA in high salt concentration (5). The glutamine-rich portions as well as residues flanking the zinc-finger domain are considered important for transcriptional activation (5). Interestingly, CPC1 also contains a glutamine-rich region, including the period-2 run $(QX)_8$ = QAQAQVQTQPQTQTQT, preceding the charge cluster at the C terminus (13). The association of a highly charged zinc-finger domain with a long uncharged region is also evident in the *Drosophila* Krüppel protein, the mouse Krox-20 protein (20), and the murine (and human) early growth-response gene product Egr-1 (21) but not in the human Krüppel-family protein GLI (Fig. 1). The uncharged regions of Krüppel protein and Krox-20 are rich in nonhydrophobic residues and may be in open coil formation suitable for providing flexibility for domain interactions. TFIIIA, which contains nine tandem zinc-finger sequences (22), is high in charge (25.6% basic residues and 11.0% acidic residues) distributed rather evenly over the whole range of the protein and thus does not display any local clustering. Also yeast ADR1 (23) does not show any significant charge clusters; its zinc-finger sequence segment consists of only two contiguous motifs whereas the other zinc-finger proteins have three or more tandem motifs.

The two activating segments of GAL4 are the most acidic portions of this protein (4), but by our statistical criteria (7, 11) these do not qualify as significant charge clusters (in contrast to the DNA-binding domain, which includes a significant positive charge cluster; Fig. 1). While the activation segments are high in acidic residues, the concentration is not unusually high in the sense that one would expect to find regions of similar concentration in random sequences of the same composition. Similarly, the transcriptional activation function of GCN4 is associated with an internal acidic region (3) that by statistical criteria does not display a significantly high concentration of acidic residues.

Yeast PHO4 is a 309-residue protein thought to act as a transcriptional activator of phosphatase genes in *Saccharomyces cerevisiae* (24). The sequence reveals a significant acidic cluster close to the N terminus that contains the period-2 charge pattern $(-, 0)_6$, including a glutamate-asparagine reiteration, ENGNENENEQDS (positions 60–71); and centrally in the protein occurs a formidable proline-lysine repeat, PTPKPKPKPKQYPK (positions 179–192). Both reiterations include, in the main, residue types (asparagine and proline) that are not conducive to secondary structure formation, and these segments would therefore be expected to be in open coil formation, possibly interacting with each other or with other targets (see refs. 7 and 8).

p68, thought to be a DNA or RNA helicase involved in replication, transcription, or RNA processing (25), features a mixed charge cluster and a significantly uncharged C terminus (only 3 charged residues over a length of 85 amino acids). CTF-1, of the CTF/NF-1 family of cellular hexanucleotide GCCAAT-binding proteins implicated in activation of both transcription and replication in eukaryotes (26), contains a pronounced mixed charge cluster, occurring near the N terminus.

The *Drosophila* transcription factor encoded at the zeste locus involves an unusual charge distribution featuring multiple charge clusters of positive, negative, and mixed sign, respectively, as well as a long uncharged region. The uncharged region is abundant with runs of glutamine, alanine, and alternating glutamine and alanine. While functional domains of the zeste protein have not been delineated, it is known to regulate in *Drosophila* embryogenesis *Ubx*, white, and DPP (decapentaplegic complex) gene expression and is thought to interact with other protein factors in mediating transactivation and transvection (27).

Steroid and Thyroid Hormone Receptors

The steroid and thyroid hormone receptors comprise a family of structurally related transcription factors that activate or repress transcription of particular sets of genes in response to specific binding of their associated hormones (for review, see ref. 28). All the proteins of this family consists of a hypervariable N-terminal domain, a central DNA-binding domain, a hinge region, and a C-terminal hormone-binding domain (28–30). At least 18 representatives of this family have been sequenced, and all of them contain significant charge clusters (Fig. 2). There is an interesting difference in charge distribution between the steroid hormone receptor family and the thyroid hormone receptor family. While all eight steroid hormone receptors conserve a *positive charge cluster* overlapping the second zinc-finger motif in the DNA-binding domain and extending into the adjacent hinge region, the thyroid hormone receptors and related proteins conserve a *balanced mixed charge cluster* that is fully contained in the hinge region.

Two transactivation domains have been identified (31) in the human glucocorticoid receptor (hGR; Fig. 2). Both these domains are of net negative charge, but their degree of acidity is not statistically significant.

The *Drosophila* steroid hormone receptor-related protein sequences of the *knrl* gene product (32), Dhr23 (ecdysone receptor; sequence provided by M. Koelle), and E75 (unknown ligand) contain multiple charge clusters interspersed with several long uncharged segments. The N-terminal half of the positive charge cluster in the *knrl* sequence is part of the conserved DNA-binding domain of the vertebrate steroid and thyroid hormone receptors. Based on this similarity and on

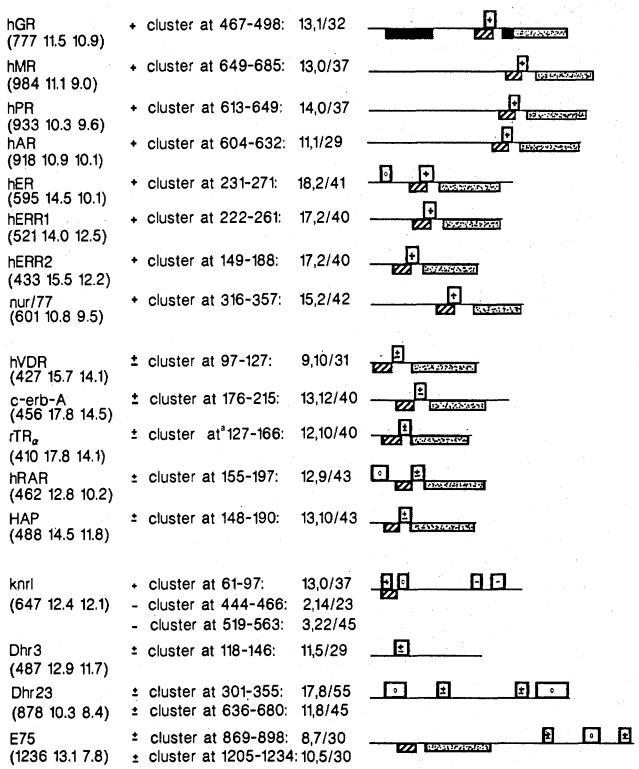

FIG. 2. Significant charge clusters in steroid and thyroid hormone receptors. Entries are as explained in the legend to Fig. 1. The DNA-binding (hatched) and ligand-binding (dotted) domains are mostly according to ref. 28 and the transactivation domains (solid) in hGR according to ref. 31. The coordinates for the uncharged regions are as follows: hER, 62–111; hRAR, 1–61; knrl, 116–158; Dhr23, 69–158 and 728–857; E75, 1045–1107. The charge clusters of hGR, hMR, hPR, hAR, hER, hERR1, and hERR2 (human glucocorticoid, mineralocorticoid, progesterone, androgen, estrogen, and estrogen-related receptors) and that of nur/77 (growth factor-inducible protein identified by cDNA cloning; ref. 30) entail highly conserved sequences (more than 65% amino acid identity). Equally conserved are the charge clusters of c-ErbA and rat thyroid hormone receptor type α (rTRα) and the charge clusters of human retinoic acid receptor (hRAR) and the hepatocellular carcinoma-related receptor HAP. Residues 61–80 of knrl are 55% conserved compared with 467–486 of hGR. hVDR, human vitamin D_3 receptor. Superscript [a] indicates that the charge cluster of rTRα is not statistically significant at the 1% level, but of note for other reasons; see text.

the spatial and temporal patterns of its expression, *knrl* is hypothesized to be an early regulatory gene, possibly encoding a constitutive transcriptional regulator functioning without a ligand (32). The C terminus of the *knrl* gene product contains two very strong acidic charge clusters (Fig. 2). The second long uncharged region of E75 is rich in serine reiterations including the sequence $S_{17}TS_2NCS_4AS_2$, and the N-terminal uncharged region is abundant with glutamine repeats. The second long uncharged stretch of Dhr23 contains the glutamine alternation $(Q, 0)_{21}$ where the 0 residue is mostly proline or leucine. Poorly conserved regions abundant in proline and glutamine lying between two distinct highly conserved functional domains have been proposed to be a flexible hinge, as in c-Jun and the mineralocorticoid, glucocorticoid, and estrogen receptors (33). They could help the protein to adopt conformations favorable to interaction with charged regions on other proteins in the formation of multimeric protein complexes.

Discussion

Our comparative sequence analysis of the distribution of charged residues in more than 2500 protein sequences from a wide assortment of eukaryotic and prokaryotic species strongly indicates a role for charge clusters in protein regulatory function. Most transactivators of eukaryotic DNA viruses carry one or more significant charge clusters in their primary sequence (7–9). Here we have shown that cellular transcription and replication factors, including steroid and thyroid hormone receptors, generally carry a single charge cluster, albeit not of an invariant sign and sometimes in conjunction with long uncharged regions or additional charge clusters. Significant charge clusters are also prominent in other classes of regulatory proteins, including high molecular weight heat shock proteins (11), nuclear protooncogene products and transforming proteins, a wide variety of transmembrane proteins (voltage-gated ion channels, growth factor and neurotransmitter receptors, opsins), and developmental control proteins (data not shown). By contrast, the majority of constitutive eukaryotic proteins, including globins, immunoglobins, and enzymes of various classes, as well as most prokaryotic proteins of any description hardly have any significant charge clusters as defined here.

Positive charge clusters or mixed charge clusters with a predominance of basic residues (2:1 ratio of basic to acidic residues) coincide with the DNA-binding domains of GAL4, the GCN4-related proteins, and the zinc-finger-containing transcription factors and with the homeobox-encoded DNA-binding domains. In the steroid hormone receptors there is a conserved positive charge cluster overlapping the carboxyl end of the DNA-binding domain, whereas the thyroid hormone receptors conserve a balanced mixed charge cluster downstream of the DNA-binding domain. The conservation of the charge cluster in the thyroid hormone receptor is especially remarkable, since, on the amino acid level, the hinge region is much less conserved than the other domains (28).

The roles of the charge clusters in the aforementioned cases may be varied. Transcription factors must have correct positive and negative selectivity for the various binding sites in promoters. Binding can be modulated by means that set the protein in the correct place, orientation, and conformation. It is conceivable that charge clusters may facilitate these processes. Many protein binding sites in DNA are dyad-symmetric, and the proteins that bind to these sites often bind cooperatively as homo- or heterodimers (6). Charge clusters and patterns could contribute to establishing sufficiently stable dimers. Ordinarily, cations neutralize the negatively charged phosphate backbone of DNA. Displacement of these cations may be a prerequisite of binding. Strong positive or mixed charge clusters in a protein might mediate the displacement of the cations and direct contact of the protein with the DNA.

What about negative charge clusters? Several experiments with mutant and hybrid proteins have suggested an important role for regions of concentrated negative charge in the transactivation functions of GAL4 (3) and GCN4 (4). On the other hand, it is clear that net negative charge is not a critical determinant of activating domains, since the charge level does not correlate strictly with the level of GCN4 activity (3) and some of the down-mutations of GAL4 constructs do not involve changes in net charge (4). Moreover, most of the zinc-finger proteins lack any distinctive acidic charge regions but contain long stretches of predominantly proline, alanine, and polar uncharged residues (asparagine, glutamine, serine, threonine); see below. While many of the immediately-early viral regulatory proteins [EBV nuclear antigens EBNA1- to -4 and transactivator pBMLF1; varicella–zoster virus p62 and p63; herpes simplex virus ICP0, ICP22, and ICP27 (but not VP16); polyoma middle-sized tumor (T) antigen; papilloma E1 and E7] as well as the oncogene product Myc and the anti-oncogene product RB contain significant negative charge clusters, such clusters do not occur in the transcription factors (with the exception of PHO4 and the zeste and

knrl gene products). In particular, none of the acidic regions in GAL4, GCN4, CREB, and the human glucocorticoid receptor stand out as unusually acidic, given the amino acid composition of these proteins, and do not qualify as statistically significant charge clusters. Given these relatively few cases of significant negative charge clusters among cellular transcription factors, one might wonder whether acidity really is a general and critical characteristic of activation domains.

Multiple significant charge clusters, in conjunction with one or more significantly long uncharged stretches preponderant with polar residues, occur in several of the immediate-early viral regulatory proteins (e.g., EBNA1 to -4 and pBMLF1 of EBV, ICP0 and ICP4 of herpes simplex virus), the transcription factors CREB, Oct-2, and zeste, the *Drosophila* hormone receptor-family proteins knrl, Dhr23, and E75, as well as in several *Drosophila* developmental regulatory proteins (for example, those encoded at the engrailed, paired, cut, deformed, *dsx*, and *per* loci). It would seem natural to posit that these different clusters within one protein correspond to different functional domains. The multiple charge clusters may be of the same type (as in CREB, Oct-1, and Oct-2) or of different types (as in zeste). In the *Drosophila* proteins generally one of the charge clusters corresponds to the homeodomain, while the other charge clusters may vary in type and positioning.

Intriguing also is the conjunction of long uncharged regions with a significant charge cluster, present for example in c-Jun, CREB, and zeste and in the zinc-finger proteins Sp1, Krüppel, Krox-20, and Egr-1. In Sp1 the uncharged region abundant with glutamine, threonine, and serine residues is implicated in increasing the DNA-binding affinity as well as in transactivation (5). Along these lines, the protooncogene products c-Jun and c-Ets are quite similar in charge distribution, displaying a significant positive charge cluster near the C terminus preceded by a significantly long uncharged stretch. Among the Jun and Ets protein families across species the significant charge cluster and uncharged run are part of the most conserved regions. In general, if the charge regions are the functionally important parts of these proteins, possibly the uncharged stretches play scaffold or hinge roles and provide flexibility to the tertiary conformation, orienting multiple binding domains to different partners in intermolecular complexes (9).

We appreciate helpful discussions with our Stanford colleagues Drs. E. Blaisdell, K. Burtis, M. Cleary, G. Crabtree, M. Koelle, M. Krasno, and E. Mocarski. This work was supported in part by National Institutes of Health Grants GM10452-26 and GM39907-01 and National Science Foundation Grant MCS82-15131 to S.K. and by Sloan Foundation Grant B1987-2 to V.B.

1. Ptashne, M. (1986) *Nature (London)* **322,** 697–701.
2. Struhl, K. (1987) *Cell* **49,** 295–297.
3. Hope, I. A., Mahadevan, S. & Struhl, K. (1988) *Nature (London)* **333,** 635–640.
4. Ma, J. & Ptashne, M. (1987) *Cell* **51,** 113–119.
5. Courey, A. J. & Tjian, R. (1988) *Cell* **55,** 887–898.
6. Schleif, R. (1988) *Science* **241,** 1182–1187.
7. Karlin, S., Blaisdell, B. E., Mocarski, E. S. & Brendel, V. (1989) *J. Mol. Biol.* **205,** 165–178.
8. Karlin, S. & Brendel, V. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 9396–9400.
9. Blaisdell, B. E. & Karlin, S. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6637–6641.
10. Yates, J. L., Warren, N. & Sugden, B. (1985) *Nature (London)* **313,** 812–815.
11. Karlin, S., Blaisdell, B. E. & Brendel, V. (1989) *Methods Enzymol.*, in press.
12. Herr, W., Sturm, R. A., Clerc, R. G., Corcoran, L. M., Baltimore, D., Sharp, P. A., Ingraham, H. A., Rosenfeld, M. G., Finney, M., Ruvkun, G. & Horvitz, H. R. (1988) *Genes Dev.* **2,** 1513–1516.
13. Paluh, J. L., Orbach, M. J., Legerton, T. L. & Yanofsky, C. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 3728–3732.
14. Vogt, P. K., Bos, T. J. & Doolittle, R. F. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3316–3319.
15. Landschulz, W. H., Johnson, P. F. & McKnight, S. L. (1989) *Science* **243,** 1681–1688.
16. Cohen, D. R. & Curran, T. (1988) *Mol. Cell. Biol.* **8,** 2063–2069.
17. Hoeffler, J. P., Meyer, T. E., Yun, Y., Jameson, J. L. & Habener, J. F. (1988) *Science* **242,** 1430–1433.
18. Turner, R. & Tjian, R. (1989) *Science* **243,** 1689–1694.
19. Kadonaga, J. T., Carner, K. R., Masiarz, F. R. & Tjian, R. (1987) *Cell* **51,** 1079–1090.
20. Joseph, L. J., Le Beau, M. M., Jamieson, G. A., Jr., Acharya, S., Shows, T. B., Rowley, J. D. & Sukhatme, V. P. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7164–7168.
21. Christy, B. A., Lau, L. F. & Nathans, D. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7857–7861.
22. Ginsberg, A. M., King, B. O. & Roeder, R. G. (1984) *Cell* **39,** 479–489.
23. Hartshorne, T. A., Blumberg, H. & Young, E. T. (1986) *Nature (London)* **320,** 283–287.
24. Legrain, M., De Wilde, M. & Hilger, F. (1986) *Nucleic Acids Res.* **14,** 3059–3073.
25. Ford, M. J., Anton, I. A. & Lane, D. P. (1988) *Nature (London)* **332,** 736–738.
26. Santoro, C., Mermod, N., Andrews, P. C. & Tjian, R. (1988) *Nature (London)* **334,** 218–224.
27. Pirrotta, V., Manet, E., Hardon, E., Bickel, S. E. & Benson, M. (1987) *EMBO J.* **6,** 791–799.
28. Evans, R. M. (1988) *Science* **240,** 889–895.
29. Chang, C., Kokontis, J. & Liao, S. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 7211–7215.
30. Hazel, T. G., Nathans, D. & Lau, L. F. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8444–8448.
31. Hollenberg, S. M. & Evans, M. (1988) *Cell* **55,** 899–906.
32. Oro, A. E., Ong, E. S., Margolis, J. S., Posakony, J. W., McKeown, M. & Evans, R. M. (1988) *Nature (London)* **336,** 493–496.
33. Bohmann, D., Bos, T. J., Admon, A., Nishimura, T., Vogt, P. K. & Tjian, R. (1987) *Science* **238,** 1386–1392.