Very long charge runs in systemic lupus erythematosus-associated autoantigens

(rheumatic disease/nuclear autoantigen/epitope)

VOLKER BRENDEL*, JAN DOHLMAN†, B. EDWIN BLAISDELL*, AND SAMUEL KARLIN*

*Department of Mathematics, Stanford University, Stanford, CA 94305; and †Department of Medicine, Division of Clinical Immunology and Rheumatology, University of Alabama Medical Center, Birmingham, AL 35294

Contributed by Samuel Karlin, November 13, 1990

ABSTRACT **Systemic lupus erythematosus and other chronic systemic autoimmune diseases are associated with circulating autoantibodies reactive with a limited set of mostly nuclear proteins. Using rigorous statistical methods we have identified segments of highly significant charge concentration in the majority of the characteristic nuclear and cytoplasmic autoantigens. Extremely long runs of charged residues, including some sequences of >20 consecutive charged residues (purely acidic or mixed basic and acidic), occur in about a third of these proteins, whereas equivalent runs are found in <3% of other mammalian proteins. The other sequences have less extreme charge clusters, the type and location of which are often conserved between several otherwise nonsimilar antigens. We propose that supercharged surfaces render the targeted host proteins strongly immunogenic and that antinuclear antibody profiles might result from chronic exposure to intracellular contents, possibly in conjunction with crossreactive viral products. The limited number of potential systemic autoantigens may partly be due to the rarity of requisite charge properties.**

Patients with systemic lupus erythematosus (SLE), mixed connective tissue disease (MCTD), Sjögren syndrome (SS), scleroderma variants, and a small number of other rheumatic autoimmune diseases develop autoantibodies directed against a variety of disease-specific cellular components. Many targets of the autoimmune response have been identified, consisting of DNA and more than 30 mostly chromatin- and RNA-associated proteins [including histones, CENP-B, DNA topoisomerase I, small nuclear ribonucleoproteins (snRNPs); for review see ref. 1]. How these proteins become autoantigenic and why exactly these and not other proteins have remained unclear. Equally intriguing is the fact that the usual patient has several different disease-specific autoantibodies, but not necessarily the same for different patients with equal clinical diagnosis (1).

One striking feature of several of the autoantigens is the presence of long charge runs, occurring, for example, in the CREST autoantigen CENP-B (2), in the DSC autoantigen DNA topoisomerase I (3), and in the SS autoantigen Ro-calreticulin (4). This is of note since it is a classical result that sufficiently charged synthetic peptides can elicit potent antibody responses (5). Charge interactions are also known to be important to many protein–nucleic acid and protein–protein interactions such as sequence-specific DNA binding (6), transcriptional activation (7), or receptor binding (8). Statistically significant clusters of charged residues are characteristic of eukaryotic regulatory proteins, including transcription and replication factors, developmental control proteins, high molecular weight heat shock proteins (hsp), and many G protein-coupled receptors (9–11). Also regulatory and surface proteins of eukaryotic DNA viruses frequently contain charge clusters and/or runs and periodic patterns of charge (12–14). By contrast, cellular proteins generally lacking distinctive charge configurations include cytoplasmic structural proteins (myosins, actins), ribosomal proteins (with the exception of the acidic proteins P0, P1, P2; see below), and most enzymes (hydrolases, oxidoreductases, transferases, nontransmembrane kinases).

Here we report a systematic analysis of charge properties of rheumatic disease-associated autoantigens. We have applied previously developed statistical methods (11) to the primary sequences, identifying significant clusters (locally high concentrations of charge), (contiguous) runs, and periodic patterns of charged residues (every second or third residue). Our results show that the set of autoantigens is extraordinarily rich in significant charge configurations, featuring several of the longest charge runs found in any proteins. We propose that the unusual charge properties of these proteins are related to their antigenic potential.

METHODS

Data Sets. Our primary data set consists of the sequenced protein targets of autoantibodies from patients with SLE, MCTD, SS, or scleroderma variants (CREST, DSC). The proteins discussed in the 1989 review by Tan (1) are histones H1 [National Biomedical Research Foundation Protein Identification Resource (NBRF), protein database name HSHU1B], H2A (HSHUA1), H2B (HSHUB1), H3 (HSHU3), and H4 (HSHU4); Sm antigens U-B (ref. 15), U-D (A27668), and U-E (S01900); nuclear ribonucleoprotein (RNP) antigens U1-A (S01497), U2-B″ (A25910), U1-C (S01387), and U1-70kD (A25707); SS-A/Ro antigens Ro60kD (A31760) and calreticulin (ref. 4); SS-B/La (A31888); Ku-70kD (ref. 16) and Ku-86kD (A32626); proliferating cell nuclear antigen (PCNA) (WMHUET); ribosomal proteins P0 (A27125), P1 (B27125), and P2 (C27125); hsp 90kD (S06898); DNA topoisomerase I (A30887); CENP-B (A27272); fibrillarin (ref. 17); and the nuclear lamins A (VEHULA), B (mouse sequence from ref. 18), and C (VEHULC). In addition, we have included nonhistone chromosomal proteins HMG-1 (S02826; see ref. 19) and HMG-17 (A27777); neurofilament proteins L (mouse sequence A25227; see ref. 20), M (A27864), and H (S00979); heat shock cognate protein hsc70 (A27077; see ref. 21), ubiquitin (A22005; see ref. 22); nucleolar protein B23 (S06926; see refs. 23, 24); and the recently

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.

Abbreviations: CREST, calcinosis, Raynaud phenomenon, esophageal dysmotility, sclerodactyly, and telangiectasia; DSC, diffuse scleroderma; RNP, ribonucleoprotein; hnRNP, heterogeneous nuclear RNP; hsc, heat shock cognate protein; hsp, heat shock protein; MCTD, mixed connective tissue disease; NBRF, National Biomedical Research Foundation Protein Identification Resource; PCNA, proliferating cell nuclear antigen; SLE, systemic lupus erythematosus; snRNP, small nuclear ribonucleoprotein; SS, Sjögren syndrome; MHC, major histocompatibility complex.

identified SLE antigens Ki (ref. 25), c-MYC (TVHUM; see ref. 26), and heterogeneous nuclear RNP-A (hnRNP-A) (S04617; see ref. 27). This adds up to a total of 39 sequences, or 34 distinct sequences if we pool respectively U1-A with U2-B″, the lamins, and the ribosomal proteins (highly similar in the charge regions).

As a control set we have used the entire NBRF protein database, Release 25.0 (June 1990), which consists of 17,731 protein sequence files. Altogether the database contains about 1000 distinct mammalian proteins (of lengths at least 120 residues, no two sequences being more than about 25% identical; cf. ref. 28).

Statistical Evaluation of Charge Runs. The significance of charge runs within a protein was estimated by reference to a random sampling model. Details of the model have been reviewed elsewhere (11). In brief, for a sequence of length N and a letter occurring with probability λ (here chosen as the fraction of specified charge in the protein) the probability of observing a run of this letter of length exceeding $L = \log N/(-\log \lambda) + z$ is at most $1 - \exp\{-(1 - \lambda)\lambda^z\}$. Setting this probability equal to 0.0001 we obtain z and L corresponding to the length of runs significant at the 0.01% level. The significance level was chosen very conservatively to accommodate the problem of multiple comparisons when searching a large database. Thus only 1 in 10,000 sequences would be expected to contain a run of the prescribed length merely due to chance fluctuations. Formulas for estimating the significance of runs with errors (intervening noncharged residues) are given in ref. 11. For a protein sequence of length 300–1000 residues and letter frequency 11.5% (average content of either lysine plus arginine or glutamate plus aspartate) the minimal significant length at the 0.01% significance level is 7 with no errors, 9 with one error, and 10 with two errors.

Charge clusters refer to short segments (25–75 residues in length) of specific high charge content (either positive, negative, or mixed, the latter involving a high number of both basic and acidic residues). Unlike runs, in a cluster the charged residues will not necessarily be contiguous. The procedure to identify statistically significant charge clusters in a protein sequence is reviewed in ref. 11. The clusters displayed in Table 2 are significant at the 1% level.

RESULTS

Charge Runs. Statistically significant (0.01% level) charge runs in the rheumatic disease-associated autoantigens are displayed in Table 1. Three of the 34 sequences contain extremely long acidic runs, and 7 contain very long mixed charge runs with basic to acidic residue ratio varying from 2:1 in DNA topoisomerase I to 1:5 in neurofilament triplet L protein. The runs in U1-70kD are primarily +, − alternations (the + residues being exclusively arginine, whereas the − residues are about equally glutamate or aspartate), but the other mixed charge runs appear unpatterned. The charge regions in U1-70kD occur carboxyl to the RNA binding domain, an arrangement that is conserved in U1-A and Ro-60kD (Table 2). The $(+-)_{10}$ run in U1-A (Table 2) is perfectly maintained in U2-B″, although half of the residues entail D → E or K → R substitutions, pointing to a role for charge rather than for specific amino acid identity in this region. The charge runs in CENP-B are within the region of the major autoantigenic epitope (Table 1), but for topoisomerase I (Table 1) and Ku-70kD (Table 2) the major autoantigenic epitopes do not coincide with the charge segments. Several of the other sequences contain charge runs of lengths just at the boundary of our strict significance level: lamins A and C contain the acidic run EDDEDEDGDD, lamin B has E_8, c-MYC bears EEEQEDEEE, La antigen has DDEHDEHDE, and long acidic and mixed charge runs also occur in neurofilament proteins M and H (not shown). It is noteworthy that the acidic run in c-MYC (residues 253–261) is part of an autoantigenic truncated c-MYC p42 but is not part of a truncated c-MYC p23, which is not autoantigenic (26).

The charge runs displayed in Table 1 are not only highly significant with respect to the charge composition of the individual proteins but are in fact among the longest found in the entire database. Among mammalian proteins equivalently long acidic runs occur only in nucleolin, upstream binding factor, mouse Hox1.1, and N-myc (Table 3) as well as in rabbit ryanodine receptor, rat parathymosin, amyloid A4-751 protein, rat sialoprotein, calsequestrin, bovine dopamine- and cAMP-regulated neuronal phosphoprotein, and mouse early T-lymphocyte activation 1 protein (unpublished data). Long mixed charge runs occur in mouse major histocompatibility complex (MHC) class III 42-kDa polypeptide, mouse Hox3.1, Ig-δ heavy chain C1–C2 hinge region, and mouse fos-B (Table 3) and in spermine binding proteins, rat chromogranin A, Ca^{2+}-transporting ATPase, and rat glycine receptor 48-kDa strychnine-binding chain. Thus, among about 1000 distinct mammalian proteins (NBRF database) ordered by length of their longest charge run, the rheumatic disease-associated antigens take 3 of the top 15 spots on the acidic chart and feature 7 of the 15 longest mixed charge runs.

Charge Clusters. If highly charged segments are an important characteristic of autoantigens then the sequences without very long charge runs might still contain such segments, yet of different statistical description. Table 2 gives charge clusters identified in six of the other sequences. Here cluster refers to short segments with a statistically significant high proportion of charged residues (see *Methods*). Interestingly, although U-B, U1-A, U1-C, and U2-B″ are not similar on the amino acid level (with the exception of the terminal regions of U1-A and U2-B″), they do conserve their charge configuration consisting of a mixed charge cluster (with basic residues outnumbering acidic residues in a ratio of 2:1) followed by a significantly long uncharged stretch. Multiple significant charge clusters as in hsc70 (and CENP-B, U1-70kD, DNA topoisomerase I; Table 1) are rare, occurring in less than about 4% of mammalian proteins (data not shown).

Long (+0) reiterations are evident at the amino terminus of histone H2A.1 (residues 3–21: RGKQGGKARAKAKTRSSRA; pattern mismatches underlined) and in the carboxyl-terminal domain of the SS-B/La antigen aminoproximal to the charge run reported above (residues 328–345: KWKSKGRRFKGKGKGNKA). Particularly outstanding is the highly significant $(GR)_9$ pattern in the nuclear RNA-associated protein U-D (not shown). It is of note that long GR runs also occur in the Epstein–Barr virus nuclear antigens EBNA1 and EBNA2 and in the herpes simplex virus type 1 immediate-early regulatory protein IE63 (UL-54) (ref. 12; see *Discussion*). (G, R)-rich regions are further found at the carboxyl termini of nucleolin (34) and of the scleroderma autoantigen fibrillarin, which contains eight repeats of the motif GXRGGF (17).

Several of the autoantigens without statistically significant charge configurations in their primary structure are highly charged overall—e.g., the histones (all >30% charged residues), Ki (>30% charge), HMG-17 (>40% charge). Thus, these proteins might well contain runs or clusters of charge in their folded state, although these regions would not show up in primary sequence analysis.

DISCUSSION

Supercharge in Nuclear Autoantigens. As a set of sequences, the protein targets of autoantibodies of patients suffering from systemic autoimmune diseases stand out relative to statistical expectation and relative to comparative sets of other proteins: they display a very high occurrence of highly charged segments, either in the form of very long runs

Table 1. Very long charge runs in human autoantigens targeted in SLE, SS, or scleroderma

Protein [file; length f(KRH) f(DE)]*	(+, −, 0), location,† sequence	Disease association, comments
Acidic runs		
Major centromere autoantigen CENP-B (A27272; 594 11.8 21.2)	(0, 48, 12) at 401: $E_{14}GEGE_6GE_6G_2EGE_2LGE_4VE_3$-GDVDSDE$_5DE_2$ (0, 27, 4) at 503: ED(SD)$_2$SE$_3$D$_2$E$_3$D$_2$ED-ED$_5$E$_2$DGDE	CREST; α-satellite DNA binding; major autoantigenic epitope within 448–594 (ref. 29)
Nonhistone chromosomal protein HMG-1 (S02826; 215 25.1 26.0)	(0, 30, 0) at 186: E$_4$DE$_2$DE$_2$DE$_5$DE$_2$DEDE$_3$D$_4$E	SLE
Nucleolar phosphoprotein B23 (S06926; 280 16.1 21.8)	(0, 26, 2) at 147: DED$_6$E$_2$D$_3$ED$_6$FD$_2$E$_2$AE$_2$	Scleroderma; nucleophosmin
Mixed charge runs		
U1-70kD (A25707; 614 25.1 18.6)	(45, 23, 11) at 407: HR(DR)$_3$(ER)$_2$RERSRERDKERE-R$_3$(SR)$_2$DR$_4$(SR)$_2$DKE$_2$R$_3$SRERS(KD)$_2$RDRKR$_2$-S$_2$RSRERAR$_2$(ER)$_2$KE$_2$ (24, 17, 7) at 521: E$_2$KGR(DR)$_2$ER$_3$SHRSERER$_3$(DR)$_5$-EHKRGERGSERGRDE	MCTD, SLE; nuclear matrix protein; R, 19.1%; RNA binding domain at 271–363 (ref. 30)
DNA topoisomerase I (A30887; 765 26.1 18.6)	(43, 20, 9) at 22: (HK)$_2$DKHKDREHRHKEHK$_2$EKD-REKSKHSNSEHKDSEK$_2$HK(EK)$_2$TKHKDGS$_2$EK-HKDKHKDRDKEKRKE$_2$K	DSC; K, 17.5%; second charge cluster at 137 (30, 23, 21); major autoantigenic epitope within 657–765 (ref. 31)
Calreticulin precursor (ref. 4; 417 13.7 26.1)	(11, 36, 4) at 358: KDKQDE$_2$QRLKE$_4$DK$_2$RKE$_4$AED-KED$_2$EDK(DE)$_2$EDE$_2$DKE$_2$DE$_3$D	SS, SLE (Ro antigen)
hsp 90kD (S06898; 730 16.8 20.5)	(17, 30, 10) at 221: EKERDKEVSD$_2$EAE$_2$KEDKE$_3$-(KE)$_2$EKESEDKPEIEDVGSDE$_4$K$_2$DGDK$_6$	SLE
Neurofilament triplet L protein (A25227; 543 14.7 22.5)	(7, 35, 12) at 478: E$_4$(KE)$_2$EGE$_4$GAE$_4$A$_2$KDESED-TKE$_4$G$_2$EGE$_3$DTKESE$_4$K$_2$E$_2$	SLE; cytoskeleton; the displayed sequence is from mouse
60s ribosomal phosphoprotein P2 (C27125; 115 10.4 18.3)	(4, 11, 2) at 92: E$_2$K$_2$DEK$_2$E$_2$SE$_2$SD$_3$	SLE; the charge run is preceded by an (A, G, P)-rich run at 34–89 and is conserved in P0, P1, and across species
U1-A (S01497; 282 14.2 8.2)	(7, 3, 0) at 103: ERDRKREKRK	SLE, MCTD; snRNP; in the homologous U2-B″ the charge run is DKEKKKEKKK

*Sequence names refer to entries in the NBRF protein database, Release 25.0. For each sequence are displayed its number of residues and its percentage of basic (K, R, H) and acidic (E, D) residues, respectively.

†Each run is characterized by a triplet (x, y, z), where x is the number of basic residues, y is the number of acidic residues, and z is the number of all other residues. The location of the run within the protein is identified by the position of the first residue of the run.

of charged residues or in the form of charge clusters (including some proportion of uncharged residues; Tables 1 and 2). We emphasize that the long charge runs in these proteins are in the extreme tail of the theoretical and the empirical distribution of charge runs (see *Results*) and constitute in that sense supercharge tracts. Highly charged synthetic peptides are known to induce a strong immune response (5). Could the supercharge in the nuclear autoantigens act in a similar way?

An involvement of mixed charge runs in recognition of protein autoantigens in SLE is supported by the mapping of the antigenic determinant on the ribosomal proteins P0, P1, and P2 to their homologous carboxyl-terminal regions, based on reactivity to the equivalent segment (KKEEKKEESEE-EDED) of eL12 of *Artemia salina* (35). The serine residues (see Table 1) are probably phosphorylated, further increasing the acidity of these segments (36). A subset of SLE antisera has also been shown to recognize the mixed charge run in U1-70kD, ERKRR (37). The importance of nonspecific charge–charge interactions is further suggested by the frequently observed reactivities in SLE to several other charge-rich structures such as proteoglycans, nucleic acids, and negatively charged phospholipids (38) and by the crossreactivities between them (DNA–proteoglycans, DNA–phospholipids; refs. 39 and 40). It is clear, on the other hand, that charge properties cannot account for all of the observations. For example, anti-H2B autoantibodies recognize sites within the peptide PAKSAPAPKK (residues 3–12) but not in the highly charged region between residues 9–23, which corresponds to an accessible surface in chromatin (41). Possibly a more generally conserved feature of the autoantigens consists of extended surface structures that are charge-, proline-, or glycine-rich (e.g., U-B, U-D, hsc70, fibrillarin).

Charge Configurations in Other Proteins. Statistically significant charge configurations are not exclusive to the nuclear autoantigens analyzed here and we shall discuss some such occurrences in viral and human proteins that might have a bearing on autoimmune phenomena.

Epstein–Barr virus, known to be associated with autoimmune phenomena (42, 43), harbors several proteins with remarkable charge configurations (12). The major nuclear antigens of the latent state all contain multiple charge clusters and periodic patterns of charge, including long GR runs in EBNA1 and EBNA2. Significant charge runs are very prominent in the proteins of human cytomegalovirus (but with the exception of a mixed charge run in the DNA polymerase of herpes simplex virus not in the other human herpesviruses), including long acidic runs in the phosphorylated matrix proteins UL32 (pp150), UL82 (pp65), and UL99 (pp28), and in UL100, UL116, and US26 as well as long mixed charge runs in UL25, UL37, and UL84 (data not shown). Adenovirus hexon protein contains a $(-)_{16}$ run; the core antigen of hepatitis B contains a positive charge cluster, the surface antigen contains a mixed charge cluster, and the DNA polymerase contains a negative cluster (13). The retroviral $p30^{gag}$ protein shares sequence similarity with several autoantigens including U1-70kD and topoisomerase I (31). The region of similarity in the viral protein is immediately aminoproximal to a remarkably strong mixed charge run, counting 16 basic and 16 acidic residues within a stretch of 36 (13).

Table 2. Charge clusters in human autoantigens targeted in SLE, SS, or scleroderma

Protein [file; length f(KRH) f(DE)]*	Charge configuration†	Disease association, comments
hsc70 (A27077; 646 13.8 14.7)	+ cluster 246–272: 14, 3/ 27 +− cluster 506–535: 9, 10/ 30 0 run 602–642 [(G, P)-rich]	SLE; clathrin uncoating ATPase; hsp70s are adjuvant arthritis (and possibly rheumatoid arthritis) antigens
U-B (ref. 15; 285 12.3 5.6)	+− cluster 44–65: 8, 6/ 22 0 run 113–179 [(A, G, P)-rich]	SLE; snRNP; G, 15.1%; P, 17.2%
U1-A (S01497; 282 14.2 8.2)	+− cluster 88–114: 11, 5/ 27 0 run 157–207 (G-rich)	MCTD, SLE; snRNP; the charge cluster is conserved in the U2-B″ homolog; RNA binding site consensus at 52–59
U1-C (S01387; 160 13.8 5.6)	+− cluster 15–44: 10, 5/ 30 0 run 62–122 [(G, M, P)-rich]	MCTD, SLE; snRNP; P, 25.0%
SS-A/Ro60kD (A31760; 538 15.6 12.5)	+− cluster 207–257: 15, 13/ 51	SS, SLE; nuclear; RNA Y1–Y5 associated; RNA binding site consensus at 92–161
Ku-70kD (ref. 16; 609 16.7 15.8)	+− cluster 225–254: 10, 9/ 30	SLE; Ku-70 and Ku-86 bind to DNA as heterodimer; major autoantigenic epitope within 410–609

*See footnote * in Table 1.

†Charge clusters are distinguished as either positive (+), negative (−), or mixed (+−), the latter involving substantial numbers of basic and acidic residues. The location of each charge cluster is given by the positions of its first and last residues followed by the number of basic and acidic residues/cluster length. 0 runs refer to significantly long uncharged segments.

What about other human proteins bearing single or multiple charge clusters or long runs of charge? Previous analyses have shown that charge clusters are prevalent among nuclear and membrane-associated regulatory proteins (9, 10). Charge runs as extreme as those prevalent among the systemic autoantigens are very rare (Table 3; unpublished data). Among the plasma proteins, coagulation cofactors V and VIII and high molecular weight kininogen contain multiple charge clusters. Protein C, which limits excess thrombin production, features a mixed charge cluster. Reactivities to forms of these proteins might contribute to some of the coagulation abnormalities seen in lupus patients who manifest anti-phospholipid antibodies (lupus-like anticoagulants; ref. 44) and who are at risk for thrombotic episodes (45). Antinucleolar autoantibodies are characteristic of patients with scleroderma. If charge is a determining factor in autoimmune targeting we would expect these patients to exhibit antibodies also to nucleolin (which contains three acidic runs exceeding length 20; Table 3) and to the human upstream binding factor [with two acidic tracts $(—)_{21}$ and $(—)_{18}$; Table 3].

Anionic regions occur in many nuclear proteins, and physiological roles for these regions in their possible association with chromatin have been speculated (46). Could the reported prominence of charge runs among the autoantigens be merely a consequence of the predominantly nuclear location of the autoantigens? Two facts argue against this possibility. (*i*) Several of the characteristic autoantigens are nonnuclear yet have long charge runs (calreticulin, hsp90, the neurofilament proteins, and the acidic ribosomal proteins; Table 1). (*ii*) By no means do all nuclear proteins contain long charge runs (e.g., transcription factors often have charge clusters but no long charge runs; ref. 9). Thus the observation holds that the systemic autoimmune disease-associated autoantigens are distinguished as bearing some of the longest charge runs found in any proteins sequenced to date, nuclear or not. Experiments will have to show whether or not this correlation is of etiopathological significance.

Possible Mechanisms. It appears that the charge runs and clusters generally do not coincide with the major autoantigenic epitopes (compare Tables 1 and 2). Conceivably, low-affinity nonspecific charge interactions may cause an initial stimulation of antibodies, but less charged epitopes with greater antibody binding potential might dominate in the mature response (47–49). Such antibody stimulation may be

Table 3. Very long charge runs in some other mammalian proteins

Protein [file; length f(KRH) f(DE)]	(+, −, 0), location, sequence
Acidic runs	
Nucleolin (S04631; 707 16.4 24.9)	(0, 38, 0) at 234: $(ED)_2E_3D_2EDED_5ED_2ED_4ED_2E_{10}$
	(0, 25, 2) at 143: $EDSDE_3D_3SE_2DE_2D_2(ED)_4E$
	(0, 24, 1) at 185: $EDED_2ED_2ED_2ED_5E_2D_2SE_3$
Upstream binding factor (ref. 32; 764 20.7 21.1)	(0, 21, 0) at 678: $E_2D_2E_2DED_2(ED)_2E_4D_2E$
	(0, 25, 3) at 715: $EDESEDGDENE_2D_2ED(ED)_3(ED)_2$
Mouse Hox1.1 homeotic protein (A28329; 229 13.1 15.7)	(0, 16, 0) at 214: DE_{15}
Human N-myc (TVHUM2; 464 13.8 14.2)	(0, 13, 0) at 264: $D_2ED_2E_3DE_4$
Mixed charge runs	
Mouse MHC class III 42-kDa polypeptide (ref. 33; 375 21.6 18.7)	(32, 28, 4) at 184: $(RS)_2RDRSHDRS(RD)_3KE(RD)_7(KD)_2(RD)_4KE(RD)_5RERDRE$
Mouse Hox3.1 homeotic protein (S00548; 242 16.9 13.6)	(5, 17, 4) at 217: $RDE_2KVE_3GNE_4KE_4KE_2NKD$
Ig-δ C-region (DHHU; 383 12.5 9.7)	(6, 9, 1) at 140: $E_2K_3(EK)_2E_2QE_2RE$
Mouse fos-B (S04108; 338 9.5 12.4)	(6, 5, 1) at 153: $E_4KR_2VR_2ER$

See footnotes to Table 1.

enhanced by the multivalency of many of the antigens that incorporate multiple charge clusters or internal repeats (e.g., fibrillarin, U1-A, U1-B, U1-C) or that occur in large complexes (e.g., chromatin, P0–P1–P2 complexes, RNPs). Alternative mechanisms for charge-promoted autoantibody production are also possible. Mature antibody responses to nuclear autoantigens are influenced by MHC molecules and are dependent on T-cell help (50–54). The identified charge sequences may participate in more specific interactions with these polymorphic receptors to determine the magnitude of the response. They may also promote nonspecific cellular immune responses by their ability to facilitate uptake by antigen-presenting cells. This possibility is supported by the presence of charge-sensitive phagocytic responses in macrophages (55–57). By this mechanism charged proteins or nucleic acids may augment immune reaction to uncharged proteins to which they are complexed (e.g., Ku86-70).

We thank Drs. G. Air, P. Bucher, J. Hazelrig, W. Koopman, H. McDevitt, J. McGhee, and J. Parnes for discussions and comments on the manuscript. This work was supported by National Institutes of Health Training Grant 5T32AR0745007 to J.D. and National Institutes of Health Grants HG00335-03 and GM10452-26 to S.K.

1. Tan, E. M. (1989) *Adv. Immunol.* **44,** 92–151.
2. Earnshaw, W. C., Sullivan, K. F., Machlin, P. S., Cooke, C. A., Kaiser, D. A., Pollard, T. D., Rothfield, N. F. & Cleveland, D. W. (1987) *J. Cell Biol.* **104,** 817–829.
3. D'Arpa, P., Machlin, P. S., Ratrie, H., III, Rothfield, N. F., Cleveland, D. W. & Earnshaw, W. C. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2543–2547.
4. McCauliffe, D. P., Lux, F. A., Lieu, T.-S., Sanz, I., Hanke, J., Newkirk, M. M., Bachinski, L. L., Itoh, Y., Siciliano, M. J., Reichlin, M., Sontheimer, R. D. & Capra, J. D. (1990) *J. Clin. Invest.* **85,** 1379–1391.
5. Sela, M. (1969) *Science* **166,** 1365–1374.
6. Mitchell, P. J. & Tjian, R. (1989) *Science* **245,** 371–378.
7. Ma, J. & Ptashne, M. (1987) *Cell* **51,** 113–119.
8. Dohlman, J., De Loof, H. & Segrest, J. (1990) *Mol. Immunol.* **27,** 1009–1020.
9. Brendel, V. & Karlin, S. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 5698–5702.
10. Karlin, S. & Brendel, V. (1990) *Oncogene* **5,** 85–95.
11. Karlin, S., Blaisdell, B. E. & Brendel, V. (1990) *Methods Enzymol.* **183,** 388–402.
12. Blaisdell, B. E. & Karlin, S. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 6637–6641.
13. Karlin, S. & Brendel, V. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 9396–9400.
14. Karlin, S., Blaisdell, B. E., Mocarski, E. S. & Brendel, V. (1989) *J. Mol. Biol.* **205,** 165–177.
15. Ohosone, Y., Mimori, T., Griffith, A., Akizuki, M., Homma, M., Craft, J. & Hardin, J. A. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 4249–4253.
16. Reeves, W. H. & Sthoeger, Z. M. (1989) *J. Biol. Chem.* **264,** 5047–5052.
17. Lapeyre, B., Mariottini, P., Mathieu, C., Ferrer, P., Amaldi, F., Amalric, F. & Caizergues-Ferrer, M. (1990) *Mol. Cell. Biol.* **10,** 430–434.
18. Höger, T. H., Krohne, G. & Franke, W. W. (1988) *Eur. J. Cell Biol.* **47,** 283–290.
19. Bustin, M., Reisch, J., Einck, L. & Klippel, J. (1982) *Science* **215,** 1245–1247.
20. Kurki, P., Helve, T., Dahl, D. & Virtanen, I. (1986) *J. Rheumatol.* **13,** 69–73.
21. Minota, S., Cameron, B., Welch, W. J. & Winfield, J. B. (1988) *J. Exp. Med.* **168,** 1475–1480.
22. Muller, S., Briand, J.-P. & Van Regenmortel, M. H. V. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 8176–8180.
23. Pfeifle, J., Andere, F. A. & Franke, M. (1986) *Ann. Rheum. Dis.* **45,** 978–986.
24. Kindås-Mugge, I. (1989) *Biochem. Biophys. Res. Commun.* **163,** 1119–1127.
25. Nikaido, T., Shimada, K., Shibata, M., Hata, M., Sakamoto, M., Takasaki, Y., Sato, C., Takahashi, T. & Nishida, Y. (1990) *Clin. Exp. Immunol.* **79,** 209–214.
26. Yamauchi, T., Naoe, T., Kurosawa, Y., Shiku, H. & Yamada, K. (1990) *Immunology* **69,** 117–120.
27. Montecucco, C., Caporali, R., Negri, C., De Gennaro, F., Cerino, A., Bestagno, M., Cobianchi, F. & Astaldi-Ricotti, G. C. B. (1990) *Arthritis Rheum.* **33,** 180–185.
28. Brendel, V. (1991) in *Advances in Mathematics and Computers in Medicine*, ed. Witten, M. (Pergamon, New York), Vol. 5, in press.
29. Earnshaw, W. C., Machlin, P. S., Bordwell, B. J., Rothfield, N. F. & Cleveland, D. W. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4979–4983.
30. Dreyfuss, G., Swanson, M. S. & Piñol-Roma, S. (1988) *Trends Biochem. Sci.* **1,** 86–91.
31. Maul, G. G., Jimenez, S. A., Riggs, E. & Ziemnicka-Kotula, D. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8492–8496.
32. Jantzen, H.-M., Admon, A., Bell, S. P. & Tjian, R. (1990) *Nature (London)* **344,** 830–836.
33. Lévi-Strauss, M., Carroll, M. C., Steinmetz, M. & Meo, T. (1988) *Science* **240,** 201–204.
34. Bourbon, H.-M., Lapeyre, B. & Amalric, F. (1988) *J. Mol. Biol.* **200,** 627–638.
35. Elkon, K., Skelly, S., Parnassa, A., Moller, W., Danho, W., Weissbach, H. & Brot, N. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 7419–7423.
36. Rich, B. E. & Steitz, J. A. (1987) *Mol. Cell. Biol.* **7,** 4065–4074.
37. Guldner, H. H., Netter, H. J., Szostecki, C., Jaeger, E. & Will, H. (1990) *J. Exp. Med.* **171,** 819–829.
38. Harris, E. N., Asherson, R. A. & Hughes, G. R. V. (1988) *Annu. Rev. Med.* **39,** 261–271.
39. Faaber, P., Capel, P. J. A., Rijke, G. P. M., Vierwinden, G., Van De Putte, L. B. A. & Koene, R. A. P. (1984) *Clin. Exp. Immunol.* **55,** 502–508.
40. Shoenfeld, Y., Rauch, J., Massicotte, H., Datta, S. K., Andre-Schwartz, J., Stollar, B. D. & Schwartz, R. S (1983) *N. Engl. J. Med.* **308,** 414–420.
41. Portanova, J. P., Cheronis, J. C., Blodgett, J. K. & Kotzin, B. L. (1990) *J. Immunol.* **144,** 4633–4640.
42. Rhodes, G., Rumpold, H., Kurki, P., Patrick, K. M., Carson, D. A. & Vaughan, J. H. (1987) *J. Exp. Med.* **165,** 1026–1040.
43. Birkenfeld, P., Haratz, N., Klein, G. & Sulitzeanu, D. (1990) *Clin. Immunol. Immunopathol.* **54,** 14–25.
44. Freyssinet, J.-M. & Cazenave, J.-P. (1987) *Thromb. Haemostasis* **58,** 679–681.
45. Hasselaar, P., Derksen, R. H. W. M., Blokzijl, L., Hessing, M., Nieuwenhuis, H. K., Bouma, B. N. & de Groot, P. G. (1989) *Ann. Rheum. Dis.* **48,** 933–940.
46. Earnshaw, W. C. (1987) *J. Cell Biol.* **105,** 1479–1482.
47. Chambers, J. C., Kenan, D., Martin, B. J. & Keene, J. D. (1988) *J. Biol. Chem.* **263,** 18043–18051.
48. Padlan, E. A. (1990) *Proteins: Struct. Funct. Genet.* **7,** 112–124.
49. Guldner, H. H., Netter, H. J., Szostecki, C., Lakomek, H. J. & Will, H. (1988) *J. Immunol.* **141,** 469–475.
50. Todd, J. A., Acha-Orbea, H., Bell, J. I., Chao, N., Fronek, Z., Jacob, C. O., McDermott, M., Sinha, A. A., Timmerman, L., Steinman, L. & McDevitt, H. O. (1988) *Science* **240,** 1003–1009.
51. Harley, J. B., Reichlin, M., Arnett, F. C., Alexander, E. L., Bias, W. B. & Provost, T. T. (1986) *Science* **232,** 1145–1147.
52. Datta, S. K., Patel, H. & Berry, D. (1987) *J. Exp. Med.* **165,** 1252–1268.
53. Ando, D. G., Sercarz, E. E. & Hahn, B. H. (1987) *J. Immunol.* **138,** 3185–3190.
54. Shores, E. W., Eisenberg, R. A. & Cohen, P. L. (1988) *J. Immunol.* **140,** 2977–2982.
55. Peterson, P. K., Gekker, G., Shapiro, R., Freiberg, M. & Keane, W. F. (1984) *Infect. Immun.* **43,** 561–566.
56. Tabata, Y. & Ikada, Y. (1988) *Biomaterials* **9,** 356–362.
57. Savill, J. S., Henson, P. M. & Haslett, C. (1989) *J. Clin. Invest.* **84,** 1518–1527.