Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 6328–6332, July 1991
Medical Sciences

Multiple overlapping homologies between two rheumatoid antigens and immunosuppressive viruses

(systemic lupus erythematosus/scleroderma/70-kDa antigen/CENP-B/cross-reactivity)

ANGELINE DOUVAS* AND STEVEN SOBELMAN

Division of Rheumatology, Department of Medicine, University of Southern California School of Medicine, Los Angeles, CA 90033

Communicated by John Abelson, March 18, 1991

ABSTRACT **Amino acid (aa) sequence homologies between viruses and autoimmune nuclear antigens are suggestive of viral involvement in disorders such as systemic lupus erythematosus (SLE) and scleroderma. We analyzed the frequency of exact homologies of ≥5 aa between 61 viral proteins (19,827 aa), 8 nuclear antigens (3813 aa), and 41 control proteins (11,743 aa). Both pentamer and hexamer homologies between control proteins and viruses are unexpectedly abundant, with hexamer matches occurring in 1 of 3 control proteins (or once every 769 aa). However, 2 nuclear antigens, the SLE-associated 70-kDa antigen and the scleroderma-associated CENP-B protein, are highly unusual in containing multiple homologies to a group of synergizing immunosuppressive viruses. Two viruses, herpes simplex virus 1 (HSV-1) and human immunodeficiency virus 1 (HIV-1), contain sequences exactly duplicated at 15 sites in the 70-kDa antigen and at 10 sites in CENP-B protein. The immediate-early (IE) protein of HSV-1, which activates HIV-1 regulatory functions, contains three homologies to the 70-kDa antigen (two hexamers and a pentamer) and two to CENP-B (a hexamer and pentamer). There are four homologies (including a hexamer) common to the 70-kDa antigen and Epstein–Barr virus, and three homologies (including two hexamers) common to CENP-B and cytomegalovirus. The majority of homologies in both nuclear antigens are clustered in highly charged C-terminal domains containing epitopes for human autoantibodies. Furthermore, most homologies have a contiguous or overlapping distribution, thereby creating a high density of potential epitopes. In addition to the exact homologies tabulated, motifs of matching sequences are repeated frequently in these domains. Our analysis suggests that coexpression of heterologous viruses having common immunosuppressive functions may generate autoantibodies cross-reacting with certain nuclear proteins.**

Disease expression in systemic rheumatic disorders (SRDs) has several features in common with infections caused by immunosuppressive viruses—e.g., human immunodeficiency virus 1 (HIV-1), herpes simplex virus 1 (HSV-1), cytomegalovirus (CMV), and Epstein–Barr virus (EBV). Common immune anomalies include lymphokine dysregulation, polyclonal B-cell activation, autoantibody production, anergy, diminished responses to specific antigens, and altered ratios of $CD4^+$ to $CD8^+$ T lymphocytes (1–4). Clinical similarities include a subacute, exacerbating, and remitting course; inflammation; musculoskeletal complaints; and lymphadenopathy (1, 5).

However, efforts to demonstrate a viral etiology for SRDs have produced inconclusive results (6). Accumulated data on cross-reactivities between SRD antibodies and viruses demonstrate that sera from patients with a single disorder react with viruses of different families (7–10). Two examples are systemic lupus erythematosus (SLE), characterized by high titer autoantibodies to U1, U2, and U4 to -6 small nuclear ribonucleoprotein (RNP) particles (11, 12), and scleroderma, in which 40–50% of patients have antibodies to centromeres, and another 25–30% have antibodies to Scl-70 (scleroderma 70-kDa antigen)/topoisomerase I (12). Antibodies in SLE sera have been found to cross-react with the retroviruses human T-cell lymphotropic virus type I and murine leukemia virus, and the DNA viruses EBV and CMV (13–15). Both scleroderma and SLE sera inhibit replication of the same adenoviral strains *in vitro* (16). Furthermore, antibodies to a single virus—e.g., EBV—occur in several disorders, including SLE, Sjogren syndrome, and rheumatoid arthritis (8, 9, 17).

A recent approach to establishing a viral link has involved the search for amino acid (aa) sequence homologies between major nuclear antigens (na) and viral proteins. This approach is based on bacterial/autoimmune paradigms, in which molecular mimicry is believed to generate anti-self antibodies, which injure cells and tissues (18). An example is the identification of common epitopes between *Klebsiella* nitrogenase reductase and HLA B27.1, which carries an increased risk for ankylosing spondylitis (19–22). Recently viral homologies have been reported in two major nuclear antigens: Scl-70/topoisomerase (23), and the 70-kDa antigen, a component of U1 RNP particles (24, 25). Interestingly, two homologies identified by different investigators in the 70-kDa protein are each associated with a different virus: a type C retrovirus (24) and influenza B virus (25).

In this report, we propose that multiple viruses may interact to produce immune anomalies and generate cross-reacting epitopes in SRDs. This proposal is based on extensive comparison of the structure of 8 na to 61 proteins representing 20 viruses. A key issue addressed in this analysis is whether single short homologies to individual viruses are likely to generate the multiple epitopes contained in SRD-associated antigens (25–28). In this report, we analyze the frequency of occurrence of such homologies (of ≥5 aa and ≥6 aa) in a control group of 41 proteins.

MATERIALS AND METHODS

A VAX/VMS version V4.7 computer and the Wisconsin University Genetics Computer Group (GCG) software were used to retrieve protein or DNA sequences from the National Biomedical Research Foundation (NBRF), GenBank, and EMBL data banks by the STRINGS program. DNA sequences were translated into amino acid sequences in the proper frame. Retroviral sequences were located through STRINGS,

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: SLE, systemic lupus erythematosus; SRD, systemic rheumatic disorder; A.I., antigenicity index; RNP, ribonucleoprotein; HIV, human immunodeficiency virus; HSV, herpes simplex virus; CMV, cytomegalovirus; EBV, Epstein–Barr virus; aa, amino acid(s); na, nuclear antigen(s); IE, immediate-early.

*To whom reprint requests should be addressed.

followed by elimination of entries representing repeat or partial sequences, or different isolates of the same strain. Endogenous retroviral and oncogene sequences were not included in the current analysis. Banks were then searched for viral families containing infectious human pathogens, including Retroviridae, Herpesviridae, Adenoviridae, Paramyxoviridae, and Togaviridae. Viruses with only invertebrate hosts were eliminated.

Control cellular proteins were selected arbitrarily based on data availability. Ovalbumin was the only nonprimate protein included. Efforts were made to include unusually long proteins (e.g., dystrophin and fibronectin) as well as short sequences (Table 1). To increase the representation of unique control proteins, single representatives of highly conserved families (i.e., the actins) were selected if homologies were common to all members of the same family. na sequences were obtained by STRINGS or from the original publications. The individual na, with references, appear in Table 1.

The GCG programs BESTFIT, WORDSEARCH, and COMPARE were used to identify matches of ≥5 aa between proteins within a file or proteins in different files. All matches were verified against the published sequences. Hydrophilicity, predictions of secondary structure, and antigenic index (A.I.) were based on the algorithms of Kyte–Doolittle (34), Chou–Fasman (35), Garnier–Osguthorpe–Robson (36), and Jameson–Wolf (37), respectively. Average, maximum, and minimum A.I. values for each protein were calculated arithmetically. A.I. values for stretches of length n were calculated by summating the A.I. values of the individual amino acids and dividing by n.

RESULTS

Table 1 reports all exact homologies of ≥5 aa identified by comparing 8 na and 41 control proteins to 61 viral proteins representing 20 viruses (Table 1). Homologies of ≥6 aa appear separately in the last column. Controls represent normal proteins with no pathogenic association. The rate of homologies of ≥5 aa to viruses is similar for the na and control groups: Homologies to viral proteins (a total of 19,827 aa) occur on the average once every 37 aa in the na group (3813 aa), and once every 38 aa in the controls (11,743 aa). The rate for the na is a composite of CENP-B, 70-kDa antigen and C, which have higher individual rates (once every 22, 22, and 20 aa, respectively), and topoisomerase I, Ro, LA, A, and B″, having lower rates of 45, 56, 67, 70, and 98, respectively. The Ro, LA, A, and B″ antigens are primarily associated with SLE, and the latter two are components of U1 and U2 RNP particles, respectively. C is also a U1 RNP component along with the 70-kDa antigen. The average frequency for the total of antigens included in Table 1 is ≈8-fold higher than expected by random chance, given the size of the comparison pool and the probability of occurrence of any specific pentamer (20^5).

Table 1. Occurrence of exact homologies of ≥5 aa between 8 autoimmune na, 41 control proteins, and 61 viral proteins

Protein	No. of aa	Homologies ≥5 aa	Homologies ≥6 aa
na			
CENP-B	594	27	6
70K	614	28	4
Topo I	765	17	1
Ro	616	11	0
LA	405	6	0
A	281	4	0
B″	392	4	0
C	159	8	0
Controls			
41 proteins	11,743	308	15

The amino acid sequences of the 8 na, in the order listed, were obtained from refs. 26 and 27–33. All exact homologies of ≥5 aa between 61 viral proteins and either the 8 na or 41 control proteins. Topo I, topoisomerase I. The 61 viral proteins correspond to the following 20 viruses: HSV-1, HSV-2, measles, rubella, Rous sarcoma virus, HTLV-1, HIV-1, SRV-1 (simian AIDS virus), MC29 (avian leukemia virus), mink cell focus-forming virus, mouse mammary tumor virus, murine leukemia virus, simian virus 40, simian sarcoma virus, feline leukemia virus, squirrel monkey retrovirus, avian sarcoma virus, CMV, EBV, and hepatitis B virus. Control proteins included immunoglobulins; globins; histones; ClqA; ClqB; renin; angiotensin; ovalbumin; serum albumins; RNase A; DNase I; heat shock proteins; pepsinogen; myelin basic protein; alkaline phosphatase; T-cell receptor subunits α, β, and δ; carboxypeptidase A and B; gastrin; α_1-antitrypsin; trypsin inhibitor; tropomyosin α chain; C-reactive protein; cytochrome P-450; fibronectin; and dystrophin. Multiple members of some protein families are represented.

We also compared the na to the total control group and determined that the average frequency of ≥5 aa, once in 93 aa, is in the same range as the majority of the na/viral matches (see above). The highest frequency, 54, was seen in CENP-B. It appears, on this basis, that pentamer homologies appear with similar frequency in all comparison groups. We therefore segregated and compared homologies of ≥6 aa (Table 1).

Homologies of ≥6 aa between viruses and controls average 1 per 2.7 control proteins (once every 769 aa; Table 1). The 15 homologies include a heptamer match between HSP-70 and EBV (GGSGSGP) and an octamer match between ovalbumin and the simian retroviral protein $p27^{gag}$ (GSAEAGVD). The frequency of occurrence is 5.5-fold higher than by random chance. The average rate of ≥6 aa homologies is higher for the na group (once every 347 aa), due to multiple matches in CENP-B and 70-kDa antigen. It was of interest, therefore, to further analyze these homologies.

Table 2 reports all exact homologies of ≥5 aa identified in comparing the 70-kDa protein to the file of 61 viral proteins. The table excludes the ETPEEREERRR consensus sequence and the ERKRR motif described previously (24, 25). There is some debate as to whether the entire sequence representing 614 aa is actually translated (38). Our rationale for considering the longer sequence is 2-fold. (*i*) Differences in reported lengths may result from alternative splicing, as proposed by Spritz *et al.* (39). This is in itself a possible mechanism generating "foreign" epitopes (see *Discussion*). (*ii*) A significant degree of homology between viral proteins and sequences immediately upstream of the 70-kDa protein provides a mechanism for direct viral intervention in the expression of these sequences (see *Discussion*).

The upper section of Table 2 reports homologies to the N terminus of the 70-kDa protein. Half of the matches (8 of 16) are to a single virus, HSV-1. Two hexamers (SGGGGS and VEAEAG) and one pentamer (AASSA) match a single HSV-1 protein, the immediate-early (IE) protein. A possible pathogenetic significance of these matches, related to the role of IE in the activation of HIV-1 is suggested below (see *Discussion*). A ninth HSV-1 matching site is located in the C terminus of the 70-kDa protein.

The C terminus of the 70-kDa protein (lower section of Table 2) has recurring stretches of the alternating basic and acidic amino acids arginine and aspartic acid (RDRDR. . .). Shorter forms of these sequences (e.g., RDRD, DRDR) are interspersed among the longer stretches. Half of the matches to the C terminus of the 70-kDa protein (6 of 12) are to HIV-1, and an additional homology is seen to the simian AIDS virus SRV-1. The RDRDR. . . stretches in particular match the HIV-1 transmembrane glycoprotein gp41, a proteolytic product of gp160 (40). Following the longest RDRDR. . . stretch in the autoantigen is a gap of 9 aa followed by a second match (ERGRD) to gp41.

Table 2. Sequence homologies between viral proteins and the N and C termini of the 70-kDa protein

Sequence	Location in the 70-kDa protein	No. of aa*	Virus (protein)	No. of aa*
N terminus				
SGGGGS	5–10	6	HSV-1 (IE)	6
SGGGG	5–9	5	HSV-1 (pol)	5
GERLD	64–68	5	HSV-1 (TK)	5
PAARP	94–98	5	HSV-1 (DNA binding)	5
AASSA	101–105	5	HSV-1 (IE)	5
VEAEAG	143–148	6	HSV-1 (IE)	6
AEAGV	145–149	5	SRV-1 ($p27^{gag}$)	5
APRDP	190–194	5	HSV-1 (DNA binding)	5
RRQQE	253–257	5	HSV-1 (TK)	5
GERLD	64–68	5	HSV-2 (TK)	5
GRAAS	99–103	5	HSV-2 (TK)	5
AEAGV	145–149	5	SRV-1 ($p27^{gag}$)	5
VAEGL	151–155	5	EBV (coat)	5
PQPPRA	156–161	6	Rubella	6
HNQPY	210–214	5	SV40 (large T)	5
PSPLP	401–405	5	CMV (early)	5
C terminus				
RDRDRDR	407–413	7	HIV-1 (gp41)	5
GGGDM	488–492	5	HIV-1 (gp120)	5
RDRDR	524–528	5	HIV-1 (gp41)	5
RDRDRDRDRDR	542–552	11	HIV-1 (gp41)	5
ERGRD	562–566	5	HIV-1 (gp41)	5
GLEGL	578–582	5	HIV-1 (3'orf)	5
RSSRS	467–471	5	SRV-1 (coat)	5
SRERAR	471–476	6	EBV (na)	6
DSRDM	585–589	5	EBV (93K)	5
DSRDM	585–589	5	EBV (140K reduc.)	5
GYLAP	598–602	5	HSV-1 (exo)	5
RERRE	415–419	5	$p30^{gag}$	5

Sequences were retrieved from NBRF, GenBank, or EMBL banks, and homology comparisons were made. The N and C termini are defined as aa 1–406 and 407–631, respectively. The complete list of 20 viruses is reported in Table 1. gp, Glycoprotein; pol, polymerase; TK, thymidine kinase; SV40, simian virus 40; orf, open reading frame; reduc., reductase; exo, exonnuclease.

*Number of amino acids matching in the nuclear antigen and virus, respectively. All matches were verified against the published sequences.

Other hydrophilic stretches in close proximity to RDRDR. . . also contain viral homologies (Table 2). The sequence RERRE (aa 415) contains homology to $p30^{gag}$ and also occurs in similar form aa 428 (RERRR), 449 (ERRR), 475 (RRERE), 527 (RERRR), and 537 (RERRR). These stretches are similar to the consensus sequence ETPEEREERRR and the motif ERKRR described elsewhere (24, 25), both of which react with human anti-U1 antibodies. Other correlations between known epitopes and these sequences are presented in the *Discussion*. To further predict the antigenic probability of the hydrophilic stretches, the A.I. was calculated (as described in *Materials and Methods*). The maximum, average, and minimum A.I. values for the 70-kDa antigen are 1.7, 0.85, and −0.6, respectively. The 11 aa of the longest RDRDR. . . stretch have a summated A.I. of 1.63, with an A.I. of 1.7 for the central 8 aa. The other two stretches also have relatively high A.I. values of 1.47 and 1.26 for sites 1 and 3, respectively. Several viruses contained no matches to the 70-kDa antigen, including HTLV-1, measles, and hepatitis B.

Table 3 reports matches between CENP-B and the same 20 viruses. At the N terminus of the autoantigen there are two matches to each of 4 immunosuppressive viruses, including 2 to the simian AIDS virus SRV-1. The C-terminal one-third

Table 3. Sequence homologies between viral proteins and the N and C termini of CENP-B

Protein sequence	Location in CENP-B	No. of aa	Virus (protein)	No. of aa
N terminus				
DQAAG	201–205	5	HSV-1 (DNA binding)	5
QAGLP	249–253	5	HSV-1 (gp-D)	5
LPVKG	88–92	5	SRV-1 (gag p27)	5
ETSLW	191–195	5	SRV-1 (protease)	5
ASQDV	182–186	5	HIV-1 (gag)	5
RTPAA	144–148	5	FeLV ($12p^{gag}$)	5
LLLAG	288–292	5	FeLV ($30p^{gag}$)	5
EGSGGS	158–163	6	EB-V (na)	6
LAGRL	290–294	5	EB-V (93k)	5
C terminus				
EEEGE	412–416	5	HSV-1 (pol)	5
EEEGE	421–425	5	HSV-1 (pol)	5
QGVVE	473–477	5	HSV-1 (IE)	5
DEDDDD	521–526	6	HSV-1 (IE)	6
EDGDE	528–532	5	HSV-2 (pol)	5
EEEEE. . .	401–414	14	MC29 (v-myc)	5
EEEEEE	418–423	6	MC29 (v-myc)	5
EEEEEE	425–430	6	MC29 (v-myc)	5
EEEEE	453–457	5	MC29 (v-myc)	5
EEDEE	456–460	5	MC29 (v-myc)	5
SDSEEE	507–512	6	MC29 (v-myc)	6
DSDEEE	450–455	6	CMV (gp-B)	6
DSDEE	450–454	5	CMV (LM-P)	5
DEDDDD	521–526	6	CMV (30K)	6
EEEGGE	428–433	6	HIV-1 (gp41)	6
EEEEV	439–443	5	HIV-1 (3'orf)	5
FAMVK	546–550	5	SRV-1 (pol)	5
DDDDE	524–528	5	SV40 (large T)	5

Retrieval of viral sequences from protein and gene banks and matching to CENP-B was performed as described in Table 1. The N and C termini are defined as aa 1–400 and 401–594, respectively. Abbreviations are the same as in Table 2.

of CENP-B is unusual both in terms of its composition and the viral matches identified. The complete sequence reveals two extended domains of acidic amino acids in the C-terminal one-third of the protein, known to contain epitopes reacting with anti-centromere antibodies from scleroderma patients (41). Sequences in these domains, composed primarily of glutamic acid, aspartic acid, or combinations of the two, also occur in 6 viruses. A total of 16 of 18 viral matches to the C terminus of CENP-B map to these domains. The human pathogens HSV-1 and CMV contain 4 and 3 matches, respectively, including a total of 3 hexamers. The sequence DEDDD is common to both viruses. Two sites in the acidic domains are homologous to HIV-1 proteins, including a hexamer match to gp41. The leukemogenic virus MC29, which has structural similarities to adenoviruses (see *Discussion*), matches to 6 acidic sites. The acidic homologies lie in regions of CENP-B having maximum surface probability. These regions have a uniform A.I. of 0.9, compared to an average of 0.52 for the entire protein.

Fig. 1 reveals that most of the homologies in CENP-B and the 70-kDa antigen overlap or are contiguous with viral sequences. A stretch of 33 aa in the first acidic domain of CENP-B (aa 401–433) is covered by six alternating matches to two retroviruses (MC29 and HIV-1) and the DNA virus HSV-1. Two additional stretches of 11 and 13 aa, also in the acidic domains, are covered by 4 and 3 overlapping matches, respectively. A total of 40% of the C-terminal one-third of CENP-B, including 73% of the first acidic domain, is covered by viral matches. Most of the matches to the N terminus of the 70-kDa protein also overlap or are contiguous. Two stretches of 12 and 19 aa contain multiple overlaps.

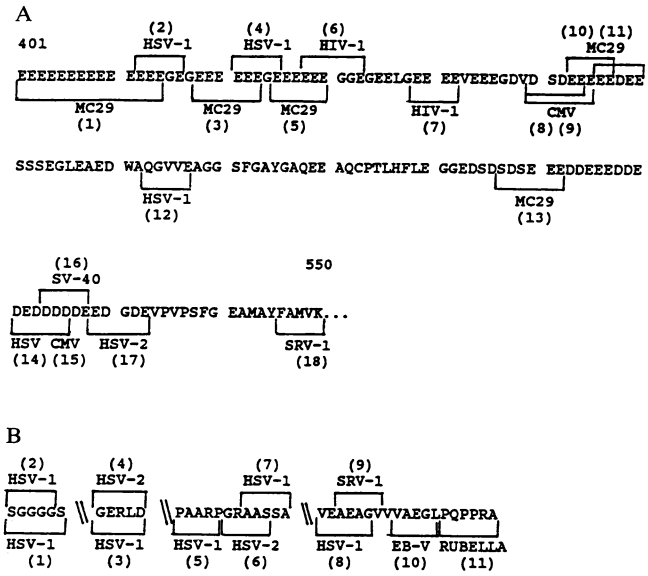

FIG. 1. Overlapping and contiguous homologies between immunosuppressive viruses and CENP-B and 70-kDa protein. (*A*) Amino acids 401–550 of CENP-B. All matching sites derived from Table 3 are shown. The sites are numbered according to their position on CENP-B and therefore do not correspond to the numbering in Table 3. (*B*) Overlapping and contiguous homologies derived from Table 2 are mapped along discontinuous but sequential segments of 70-kDa protein. Discontinuities are indicated by \\. The discontinuous segments of 70-kDa protein containing matching sites numbered (1) to (11) correspond to amino acids 5–10, 64–68, 94–105, and 151–169.

Although not mapped in Fig. 1, a number of the exact matches to the C terminus of the 70-kDa protein occur within a short segment of the protein. For example, 7 matches (totaling 41 aa) occur within a 100-aa stretch of 70-kDa protein, between positions 488 and 589 (Table 2). As noted above, partial homologies are also interspersed in this stretch. There are four instances in which a single viral protein contains two matches that lie in close proximity on either 70-kDa antigen or CENP-B. Two homologies between HSV-1 polymerase and CENP-B map to positions 412–416 and 421–425 in CENP-B. Two MC29 v-myc homologies overlap at positions 453–459 and 456–460 in CENP-B. Two homologies between the 70-kDa and the IE protein of HSV-1 occur at positions 101–105 and 143–148. There is a gap of 9 aa between 2 gp41 homologies to the 70-kDa protein, at positions 542–552 and 562–566. In addition, as noted above, matches to a single virus tend to cluster in a relatively short region of either antigen. Three of the HIV-1 matches to the 70-kDa protein occur within 40 amino acids (positions 542–582). Two HSV-1 protein matches to the 70-kDa protein (DNA-binding and IE) map at positions 94–98 and 101–105. Thus, there is a high density of matches both to multiple viruses, and to individual viruses, including, in some instances, to a single protein of the same virus. This type of clustering also occurs in other pathogenetic models (see *Discussion*).

With two exceptions, na/viral homologies contain different sequences from control/viral homologies. One exception is the 70-kDa protein sequence matching simian $p27^{gag}$ (Table 2), contained also in the ovalbumin octamer GSAEAGVD. The CENP-B sequence EEEEE. . . (Table 3) also occurs as a pentamer in gastrin. The overall rate of homology between na and controls was lower than that between controls and viruses.

DISCUSSION

This analysis of the primary structures of eight SRD-associated antigens reveals that two polypeptides, CENP-B and the 70-kDa protein, contain multiple homologies to human viral pathogens (Table 1). The frequency of hexamer homologies was particularly unusual, occurring once every 99 and 153 aa in CENP-B and the 70-kDa protein, respectively, compared to once every 769 aa in the control group, and an equivalent or lower frequency in each of six other SDR antigens (Table 1). In addition to their frequency, viral homologies to the 70-kDa protein and CENP-B are notable in four other respects: (*i*) the majority lie in regions known to contain epitopes for human autoantibodies; (*ii*) most homologies involve very hydrophilic sequences; (*iii*) they cluster in short regions of both proteins rather than having a random distribution (as is typical of controls); (*iv*) the majority of homologies are limited to a select group of viruses that are immunoinfective and immunosuppressive. These properties are discussed in greater detail below.

Correlation Between Viral Homologies and Epitopic Sites. At least six epitope regions have been identified in the 70-kDa protein, and several have been identified in CENP-B (25, 26, 41–43). Although all sources do not agree exactly on the location of 70-kDa protein epitopes, we estimate that at least 14/27 homologies identified in Table 2 lie within previously identified epitope sites (43, 44). Another 10, most matching to HSV-1, lie within the extreme N terminus encoded by the 3.9-kilobase mRNA fragment (27, 39). This area has not been analyzed for the presence of epitopes. As suggested by Spritz *et al.* (39), shorter and longer versions of the 70-kDa protein, the shorter lacking the N terminus, may arise from alternative splicing. It will be of interest to determine whether alternative splicing plays a role in the generation of new epitopes in SRDs. Viruses are known to have a role in aberrant splicing of cellular proteins (45).

All of the homologies identified to CENP-B (Table 3) are potentially within epitopic areas, as determined from existing studies (26, 41). In particular, the acidic C domains appear to be important in immunoreactivity, although epitopes also exist in the N-terminal two-thirds of the protein (41). Thus, both the antigenic probabilities (see *Results*) and the correlation with known epitopic sites predict that many of the homologies identified will be antigenic. This likelihood is further supported by the hydrophilicity and high surface probability of most of the sequences, as discussed below.

The Hydrophilic Nature of Many Homologies. It is apparent (Tables 2 and 3) that most of the viral homologies in the 70-kDa protein and CENP-B are hydrophilic. Only 5% of those in the control group are distinctly hydrophilic (i.e., ≥50% hydrophilic amino acids). Homologies identified here, therefore, have a high probability of being exposed to the aqueous environment and particularly to the immune system. Moreover, previously identified homologies in the 70-kDa protein known to have immunoreactivity are also hydrophilic, including the ETPEEREERRR consensus sequence and the ERKRR motif (24, 25). The acidic homologies in CENP-B are interesting from the perspective of their evolutionary relation to both viruses and cellular oncogenes. The consensus sequence SDSEEE (Table 3) has also been identified in adenoviral strains and in both v-myc and c-myc (Table 3; ref. 46). Other acidic consensus sequences also occur in both DNA and RNA viruses (46). The function of these sequences is not known (46). However, their occurrence in extended stretches of CENP-B creates the potential for interaction of virally encoded functions with either the gene encoding CENP-B or directly with centromeres. It is interesting that monoclonal antibodies raised against an acidic HSV-1 IE epitope give a punctate metaphase chromosome staining pattern in HSV-1-infected cells (47).

The Clustered, Overlapping Distribution of Homologies. In contrast to homologies identified in control proteins, the majority of the 70-kDa protein and CENP-B homologies are contiguous or overlapping (Fig. 1). This is also typical of the

cross-reacting homologies between *Klebsiella* nitrogenase reductase and HLA B27.1, in which epitopes cluster and overlap within a short region (22). As in the *Klebsiella* homologies, we identified four instances in which the 70-kDa protein and CENP-B contained two closely spaced homologies to a single viral protein (see *Results*). If, as we propose, simultaneous activation of immunosuppressive viruses occurs, the distribution of homologies seen in Fig. 1 creates a high density of potential epitopes for cross-reactivity with the resulting anti-viral antibodies.

Homologies to Immunosuppressive Viruses. In addition to containing conserved sequences, certain viruses of different families synergize in infecting cells of the immune system. Both HSV-1 and EBV activate the expression of HIV-1 long terminal repeat sequences (48, 49). A crucial role in this activation is played by the IE protein (50). Coinfection of T cells with CMV and HIV-1 enhances the expression of both viruses and involves an IE CMV protein (51). An analysis of homologies in 70-kDa protein (Table 2) and CENP-B (Table 3) reveals that these four viruses—HSV-1, HIV-1, CMV, and EBV—account for 71% of the matches to the 70-kDa protein (20 sites) and 52% of the matches to CENP-B (14 sites) (Tables 2 and 3). Interestingly, 6 homologies to the two IE proteins of these viruses were identified in 70-kDa protein and CENP-B, including 3 hexamers (Tables 2 and 3). Our analysis suggests that synergistic viruses of different families that are prevalent in the environment may play a role in autoreactivity in SRDs (and predict that a virus similar to HIV-1, but less lethal, exists). This model is not only congruent with current infectious disease paradigms but also accounts for the multiple epitopes known to occur in na.

We thank Dr. Suraiya Rasheed for discussions on viral interactions and Dr. David Sadava for review of the manuscript. This work was supported by U.S. Public Health Service Grant AR39554 and by an Arthritis Foundation Investigator award.

1. Lane, H. C. & Fauci, A. S. (1985) *Annu. Rev. Immunol.* **3,** 477–500.
2. Rosenthal, C. J. & Franklin, D. C. (1975) *Arthritis Rheum.* **18,** 207–217.
3. Steinberg, A. D., Klassen, L. W., Budman, D. K. & Williams, G. W. (1970) *Arthritis Rheum.* **22,** 114–122.
4. Sakane, T., Steinberg, A. D., Arnett, F. C., Reinertsen, J. L. & Green, I. (1979) *Arthritis Rheum.* **22,** 770–776.
5. Kelley, W. N., Harris, E. D., Jr., Ruddy, S. & Sledge, C. B. (1989) *Textbook of Rheumatology* (Saunders, Philadelphia).
6. Fox, R. (1988) *J. Virol. Methods* **21,** 19–27.
7. Sany, J. (1987) *Presse Med.* **16,** 725–728.
8. Fox, R. I., Chilton, T., Scott, S., Benton, L., Howell, F. V. & Vaughan, J. H. (1987) *Rheum. Dis. Clin. North Am.* **13,** 275–292.
9. Fox, R. (1988) *J. Virol. Methods* **21,** 19–27.
10. Krieg, A. M., Khan, A. S. & Steinberg, A. D. (1989) *Arthritis Rheum.* **32,** 322–329.
11. Lerner, M. R. & Steitz, J. A. (1979) *Proc. Natl. Acad. Sci. USA* **76,** 5494–5499.
12. Schumacher, E. H., Klippel, J. H. & Robinson, D. R., eds. (1988) *Primer on the Rheumatic Diseases* (Arthritis Found., Atlanta), 9th Ed.
13. Kurata, A., Katamine, S., Fukuda, T., Mine, M., Ikari, N., Kanazawa, H., Matsunaga, M., Eguchi, K. & Nagataki, S. (1985) *Clin. Exp. Immunol.* **62,** 65–74.
14. Rucheton, M., Graafland, H., Fanton, H., Ursule, L., Ferier, P. & Larsen, C. J. (1985) *Virology* **144,** 468–480.
15. Okatamo, T., Tamura, T. & Takano, T. (1983) *Clin. Exp. Immunol.* **54,** 747–755.
16. Nakabayashi, K., Saito, M., Nagasawa, T. & Takada, M. (1985) *Rheumatol. Int.* **5,** 61–67.
17. Pruijn, G. J., Kusters, H. G., Gmelig Meyling, F. H. & van der Vliet, P. C. (1986) *Eur. J. Biochem.* **154,** 363–370.
18. Oldstone, M. B. A. & Notkins, A. L. (1984) *Concepts in Viral Pathogenesis* (Springer, New York), Vol. 2, pp. 195–202.
19. Calin, A., Marder, A. & Becks, E. (1983) *Arthritis Rheum.* **26,** 1460–1464.
20. Schwimmbeck, P. L., Wu, D. T. Y. & Oldstone, M. B. A. (1987) *J. Exp. Med.* **166,** 173–181.
21. Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennet, W. S., Strominger, J. L. & Wiley, D. C. (1987) *Nature (London)* **329,** 512–518.
22. Ewing, C., Ebringer, R., Tribbick, G. & Geysen, H. M. (1990) *J. Exp. Med.* **171,** 1635–1647.
23. Maul, G. G., Jimenez, S. A., Riggs, E. & Ziemricka-Kotula, D. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 8492–8496.
24. Query, C. C. & Keene, J. D. (1987) *Cell* **51,** 211–220.
25. Guildner, H. H., Netter, H. J., Szostecki, C., Jaeger, E. & Will, W. (1990) *J. Exp. Med.* **171,** 819–829.
26. 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.
27. Theissen, H., Etzerodt, M., Reuter, R., Schneider, C., Lottspeich, F., Argos, P., Luhrmann, R. & Philipson, L. (1986) *EMBO J.* **5,** 3209–3217.
28. D'Arpa, P., Machlin, P. S., Ratrie, H., Rothfield, N. F., Cleveland, D. W. & Earnshaw, W. C. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 2543–2547.
29. Ben-Chetrit, E., Gandy, B. J., Tan, E. M. & Sullivan, K. F. (1989) *J. Clin. Invest.* **83,** 1284–1292.
30. Chambers, J. C., Kenan, D., Martin, B. J. & Keene, D. (1988) *J. Biol. Chem.* **263,** 18043–18051.
31. Bringman, P. & Lurhman, R. (1986) *EMBO J.* **5,** 3509–3516.
32. Habets, W. J., Sillekens, P. T. G., Hoet, M. H., Schalken, J. A., Roebroek, A. J. M. & van Venrooij, W. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 2421–2425.
33. Sillekens, P. T. G., Beijir, R. P., Habets, W. J. & van Venrooij, W. (1988) *Nucleic Acids Res.* **16,** 8307–8321.
34. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
35. Chou, P. Y. & Fasman, G. D. (1978) *Adv. Enzymol.* **47,** 45–147.
36. Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* **120,** 97–120.
37. Wolf, H., Modrow, S., Motz, M., Jameson, B. A., Herman, G. & Foertsch, B. (1988) *Comput. Appl. Biosci.* **4,** 187–192.
38. Query, C. C., Bently, R. C. & Keene, J. C. (1989) *Cell* **57,** 89–101.
39. Spritz, R. A., Strunk, K., Surowy, C. S., Hoch, S. O., Barton, D. E. & Francke, U. (1987) *Nucleic Acids Res.* **15,** 10373–10391.
40. Desai, S. M., Kalyanaraman, V. S., Casey, J. M., Srinivasan, A., Andersen, P. R. & Devare, S. G. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 8380–8384.
41. Earnshaw, W. C., Machlin, P. S., Bordwell, B. J., Rothfield, N. F. & Cleveland, D. W. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 4979–4983.
42. Francoeur, A. M. (1989) *J. Clin. Immunol.* **9,** 256–263.
43. Cram, D. S., Fisicaro, N., Coppel, R. L., Whittingham, S. & Harrison, L. C. (1990) *J. Immunol.* **145,** 630–635.
44. Guldner, H. H., Netter, H. J., Szostecki, C., Lakomek, H. J. & Will, H. (1988) *J. Immunol.* **141,** 469–475.
45. Rosson, D., Dugan, D. & Reddy, P. E. (1987) *Proc. Natl. Acad. Sci. USA* **84,** 3171–3175.
46. Ralston, R. & Bishop, J. M. (1983) *Nature (London)* **306,** 803–806.
47. Gimble, J. M., Duh, E., Ostrove, J. M., Gendelman, H. E., Max, E. E. & Rabson, A. B. (1988) *J. Immunol.* **141,** 469–475.
48. Gimble, J. M., Duh, E., Ostrove, J. M., Gendelman, H. E., Max, E. E. & Rabson, A. B. (1988) *J. Virol.* **62,** 4104–4112.
49. Kenney, S., Kamine, J., Markowitz, D., Fenrick, R. & Pagano, J. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 1652–1656.
50. Albrecht, M. A., DeLuca, N. A., Byrn, R. H., Schaffer, P. A. & Hammer, S. M. (1989) *J. Virol.* **63,** 1861–1868.
51. Skolnik, P. R., Kosloff, B. R. & Hirsch, M. S. (1988) *J. Infect. Dis.* **122,** 508–514.