Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 7176–7180, August 1993
Evolution

Covariation of mutations in the V3 loop of human immunodeficiency virus type 1 envelope protein: An information theoretic analysis

(convergent evolution/peptide vaccine design)

BETTE T. M. KORBER*†, ROBERT M. FARBER*‡, DAVID H. WOLPERT*, AND ALAN S. LAPEDES*‡

*Santa Fe Institute, 1660 Old Pecos Trail, Santa Fe, NM 87501; and †Theoretical Biology and Biophysics (T10), and ‡Complex Systems Group (T13), Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM 87545

Communicated by Stirling A. Colgate, March 16, 1993

ABSTRACT **The V3 loop of the human immunodeficiency virus type 1 (HIV-1) envelope protein is a highly variable region that is both functionally and immunologically important. Using available amino acid sequences from the V3 region, we have used an information theoretic quantity called mutual information, a measure of covariation, to quantify dependence between mutations in the loop. Certain pairs of sites, including noncontiguous sites along the sequence, do not have independent mutations but display considerable, statistically significant, covarying mutations as measured by mutual information. For the pairs of sites with the highest mutual information, specific amino acids were identified that were highly predictive of amino acids in the linked site. The observed interdependence between variable sites may have implications for structural or functional relationships; separate experimental evidence indicates functional linkage between some of the pairs of sites with high mutual information. Further specific mutational studies of the V3 loop's role in determining viral phenotype are suggested by our analyses. Also, the implications of our results may be important to consider for V3 peptide vaccine design. The methods used here are generally applicable to the study of variable proteins.**

The V3 loop of the human immunodeficiency virus type 1 (HIV-1) envelope protein (env) has been the focus of intense research efforts because it is a potent epitope for neutralizing antibodies (1–3) and T cells (4, 5), and it plays a role in determining cell tropism and viral growth characteristics (6–11). While there is some propensity to conserve amino acid side chain chemistry in the different positions in the loop (12, 13), this conservation often breaks down upon inclusion of phylogenetically distant viruses (13). Such variation presents a difficult challenge for those attempting to design broadly reactive V3 loop-based vaccines (3, 4, 14, 15). Our goal was to quantify the degree of covariation of mutations at different sites by analyzing the available data base (13) of V3 amino acid sequences by using *mutual information*, a concept from information theory (16–19). All pairs of positions in an alignment of 308 distinct V3 loop sequences were compared, and we determined with high confidence that certain pairs are covarying (see Figs. 1 and 2).

The identification of covarying sites is potentially useful for two biological purposes. First, coordinate mutations of these sites may be important for phenotypic changes in the V3 loop, and they could be used as a tentative map for researchers attempting to define functional domains in the V3 region through mutational analysis. Second, they could be used as a guide for reasoned selection of sets of V3 loops for inclusion in a mixture of V3 peptides for vaccine design. Particular pairings with high mutual information values are likely to confer a selective advantage in terms of either structure or function. Therefore, by selecting V3 loop sequences which include pairs of amino acids that are highly predictive of each other, one may be covering important classes of V3 loop sequences that are structurally or functionally related. These relationships may exist across distant phylogenetic groups through parallel or convergent evolution. Thus inclusion of V3 peptides with highly covariant amino acids may be a useful strategy for designing broadly reactive vaccines, in a time when little is known about the phenotypic consequences of the radically divergent V3 loops found throughout the globe (13).

A formal measure of variability (17) at position i is the Shannon entropy, $H(i)$. $H(i)$ is defined in terms of the probabilities, $P(s_i)$, of the different symbols, s, that can appear at sequence position i (e.g., s = A, S, L, . . . for the 20 natural amino acids Ala, Ser, Leu, . . .). $H(i)$ is defined as

$$H(i) = -\sum_{s=A,S,L...} P(s_i) \log P(s_i).$$

Mutual information is defined in terms of entropies involving the joint probability distribution, $P(s_i, s'_j)$, of occurrence of symbol s at position i, and s' at position j. The probability, $P(s_i)$, of a symbol appearing at position i regardless of what symbol appears at position j is defined by $P(s_i) = \Sigma_{s'_j} P(s_i, s'_j)$ and similarly, $P(s'_j) = \Sigma_{s_i} P(s_i, s'_j)$. Given the above probability distributions, one can form the associated entropies

$$H(i) = -\sum_{s_i} P(s_i) \log P(s_i),$$

$$H(j) = -\sum_{s_j} P(s'_j) \log P(s'_j),$$

and

$$H(i, j) = -\sum_{s_i, s_j'} P(s_i, s'_j) \log P(s_i, s'_j).$$

The mutual information, $M(i, j)$, is defined as

$$M(i, j) = H(i) + H(j) - H(i, j).$$

An alternative, mathematically equivalent form of this equation expresses the relation to the log-likelihood ratio of the expected occurrence of pairs (under the assumption of independence) to the observed occurrence:

$$M(i, j) = -\sum_{s_i, s_j'} P(s_i, s'_j) \log P(s_i, s'_j)/P(s_i)P(s'_j).$$

Mutual information is always nonnegative and achieves its maximum value if there is complete covariation. The minimum value of 0 is obtained either when i and j vary completely independently or when there is no variation (17).

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: HIV-1, human immunodeficiency virus type 1.

The above formulae assume that the true probability distributions are known. In practice, however, the true probability distributions are not known, and they must be estimated from a finite data set. This introduces subtle effects in the estimated mutual information. Just as a finite number of tosses of a "fair" coin typically exhibits fluctuations away from the "true" value of 50% heads, so too will the estimated mutual information of an independent distribution exhibit fluctuations away from the "true" value of zero. Since mutual information is always nonnegative, the mutual information of truly independent distributions is consistently overestimated, while the mutual information of covarying distributions can be either overestimated or underestimated, depending on the nature of the fluctuations. It is therefore necessary to assess the statistical significance of any mutual information estimate obtained from limited data. This is complicated by selection effects, as illustrated by another coin-tossing analogy. Consider performing the coin tossing experiment described above on a very, very large number of fair coins. Even if the number of tosses for each coin is large (so one expects close to 50% heads for each coin), there will still be a substantial probability that at least one of the large number of coins will deviate away from the "true" value of 50% heads. This occurs because although the probability of such a fluctuation is small, it is compounded over a large number of coins. If one selected this coin as "interesting" and calculated the statistical significance of the "fair coin hypothesis" for this single *selected* coin, it would appear that the selected coin was not "fair," when in fact all the coins were "fair." A similar effect occurs in estimating mutual information for a large number of pairs of sequence positions. It can be misleading to select a particular high estimate and calculate significance on the basis of a single pair of positions.

An algorithm which employed multiple randomizations of the initial data set was used to determine the statistical significance of the estimated mutual information values, using a very conservative measure that addresses both small sample bias and selection effects (for general reviews of methods of this type, see refs. 20–22). For the final analysis, 750,000 randomizations of the initial data set were done. Highly statistically significant mutual information scores were obtained for several pairs of sites, some on opposite sides of the V3 loop (Figs. 1 and 2). The results of these calculations are displayed in Fig. 2, and a description of the statistical methods is provided in the legend. The data set, and two control data sets, are described in the legend of Fig. 1.

A detailed examination of the highly covarying pairs of sites revealed certain amino acids that were particularly predictive of amino acids in the paired column of the sequence alignment. We quantified this effect by calculating a measure we call "specific information," $I(s_j')$, which is a measure of how much information about site i is gained from knowing a specific symbol, s_j, occurring at j. Defining the conditional entropy, $H(i|s_j')$, as

$$H(i|s_j') = -\sum_{s_i} P(s_i|s_j') \log P(s_i|s_j')$$

allows us to write $I(s_j')$ as

$$I(s_j') = [H(i) - H(i|s_j']P(s_j').$$

Interchanging i and j defines specific information in the other direction. Summing over the symbols s_j' yields the mutual information.

Amino acids with particularly high specific information for a given position are listed in Table 1, together with the amino acids with which they were associated. Table 1 is limited to the pairs of sites that had the highest mutual information scores. These suggested pairings may have functional significance and could be tested through combinatorial mutational analysis. As the pairings occur repeatedly in the data set, involve positions which are known to influence antigenic specificity (3, 27), and may influence the conformation of the loop, it may be prudent to incorporate them into the sets of V3 peptides that will be used in vaccine trials. It is worth noting that the most common pairing of amino acids found in a pair of columns that has a high mutual information value does not necessarily have high specific information or predictive value. For example, the most common pair of amino acids in positions 10 and 12 are Ser-10 and His-12. Ser is found in position 10 in 50% of the sequences and His is found in position 12 in 44% of the sequences. Thus you would expect to see the combination in approximately 22% of the sequences if the association were completely random, and in the actual data it was observed in 20% of the sequences, indicating that there is no particular association between these two amino acids in these positions.

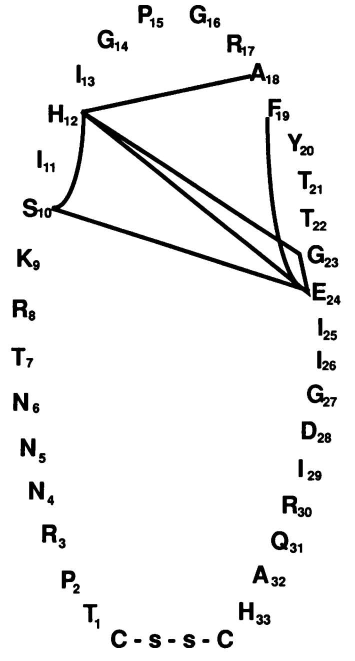

FIG. 1. Sites in the V3 loop that have high mutual information. A consensus sequence is shown (12, 13), with lines between sites indicating positions that have the highest mutual information with the greatest statistical significance. The tip of the loop centers on a relatively conserved motif (GPGR), which can form a type II β-turn (23), and is the focal point of the loop's immunological reactivity. An alignment of 610 unique V3 loop sequences was generated by using the MASE program (24), with particular regard to aligning the tip of the loop. This was reduced to a set of 308 sequences, containing no more than one or two sequences per individual (13); 150 sequences from the LaRosa *et al.* set (12) were used, from 111 individuals (13), excluding probable HIV-1 IIIB contaminations (25). An additional 158 sequences were included from 143 individuals (13), using the most common sequence from an individual when multiple sequences were available. When distinct forms were present in a person, differing in more than 5 amino acids from the most common form, then two sequences were included. When sequences from the blood and the brain of an individual were available, the most common sequence from each set was included. This data set includes some highly divergent sequences from Africa. When it was reduced to 271 non-African sequences, or when all unique sequences were included without regard to limiting the number of sequences from an individual, the same highly correlated sites were observed. The complete alignment can be obtained through the HIV-1 data base (13). Columns in the alignment that were primarily gaps inserted to maintain alignment with sequences that contained insertions were deleted. The apparent "networks" of sites that have high mutual information indicate that there may be higher-order interactions occurring between the sites.

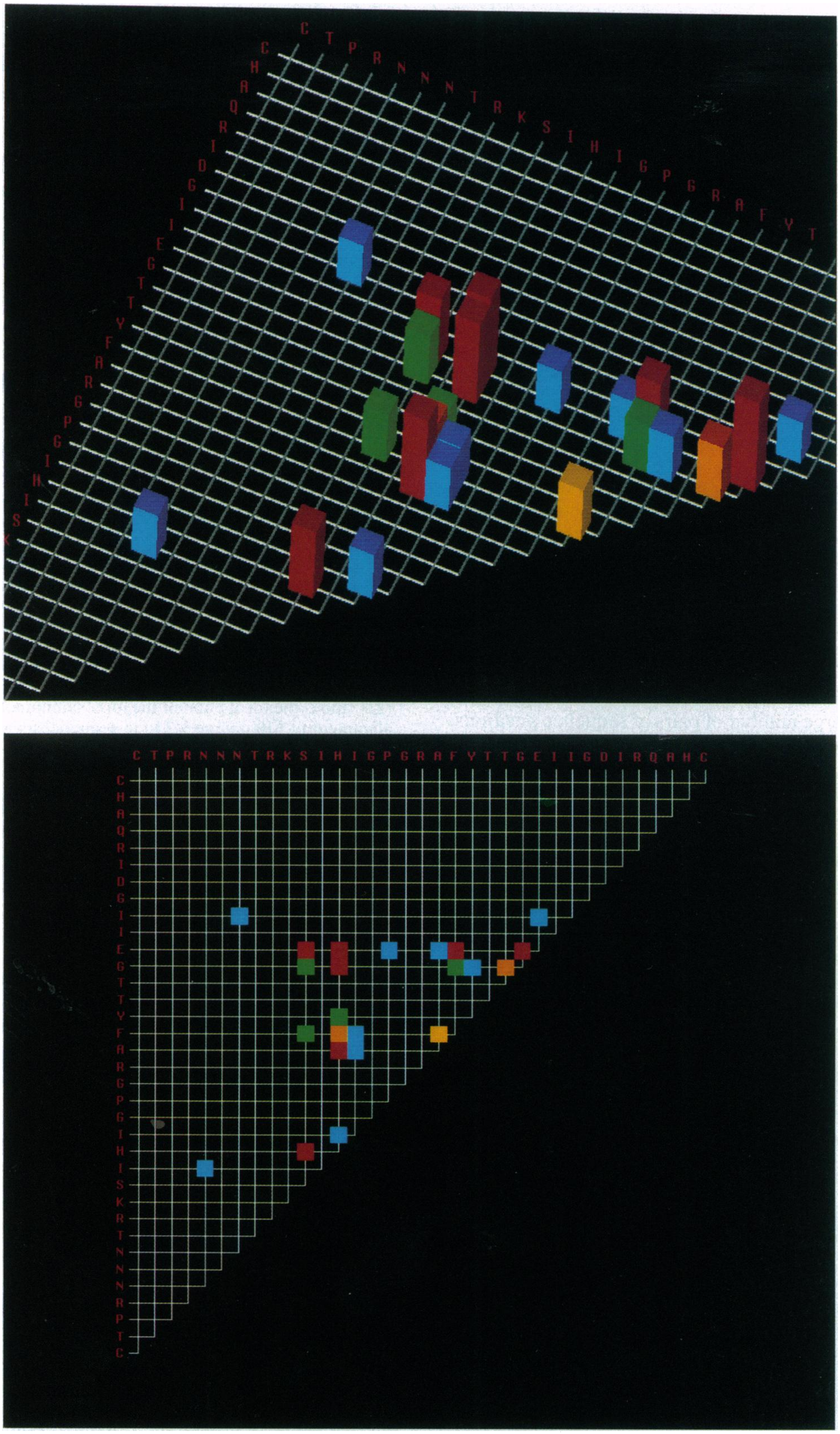

FIG. 2. Mutual information (height) and statistical significance (color) of pairs of sites in the V3 loop. (*Upper*) The V3 consensus sequence is drawn on each axis: each interaction represents the mutual information and statistical confidence between a pair of positions. The height scale is proportional to the value of the mutual information. Columns with greater than 95% conservation were excluded; these are apparent as rows of zero height. To calculate statistical significance, sequence data within each position were permuted at random. This breaks covariation between pairs of positions and generates a data set the same size as the original set, with identical single site probabilities. This procedure was followed 750,000 times, yielding 750,000 randomized data sets. Positive values of mutual information estimates in the randomized data sets are due solely to finite size bias effects. Selection effects were handled the following way. For each pair of positions, a mutual information estimate was calculated from the original data. For each such estimate the number of randomized data sets that had at least one pair of positions anywhere in the set with a mutual information value at least as high as the original pair was calculated. The probability, *P*, that a randomized data set

Table 1. Predictive amino acids in sites with high mutual information

Most significant associations	Most common pair (% in data set)	Predictive amino acids (% in data set)	Expected percentage
23–24	G–E (23%)	**G** ↔ **E** or **D** (42%)	30%
12–24	H–E (14%)	**H** ← E or D (24%)	20%
		R → K or R (6%)	1%
12–18	H–A (39%)	**H**, N, or P ↔ **A** (60%)	54%
		T ← **V** (8%)	1%
12–23	H–G (32%)	**H**, **T**, or N ↔ **G** (51%)	42%
19–24	F–E (19%)	**F** ↔ **E** or **D** (38%)	30%
		V → K or R (7%)	1%
10–24	S–D (17%)	**S** ↔ **E** or **D** (33%)	23%
10–12	S–H (20%)	G → H or R (22%)	13%
		S ← N or P (19%)	11%

The first column lists the pairs of sites that had the highest, most significant, mutual information scores; these are the site pairs for which, with confidence $1 - P > 99.9999\%$, the null hypothesis of statistical independence between the sites can be rejected. (Comparable scores were never observed in 750,000 randomizations of the entire data set.) The site pairs are listed according to the ranking of their mutual information scores, numbered as in Fig. 1. The second column shows the most common amino acid combination observed and the percentage of sequences which included that pair. The third column shows amino acids with high specific information, I, as discussed in the text. The arrows indicate which amino acids predicted the other, and the percentage of sequences which included the combinations is given. For example, in pair 12–18, a V (Val) in position 18 has a high specific information, I, and is associated with a T (Thr) in position 12, as determined by conditional probabilities; however, T does not have a particularly high specific information, hence the left-pointing arrow. The expected percentage of sequences to have a T–V pairing in positions 12 and 18, if their association were random, based on the actual frequencies of T and V in their respective positions, is 1%, shown in the fourth column. In situations where more than one amino acid was predicted, for example the E (Glu) or D (Asp) in position 24 of the position 23 and 24 association, the 42% in the third column represents the frequency that G is found in association with either E or D, and the 30% in the fourth column represents the product of the frequencies of (E plus D) in position 24 and the frequency of G in position 23. Boldface amino acid symbols in the third column indicate that the pair was found widely distributed in the sequence set and was likely to have arisen independently on multiple occasions. Such pairings met the following four criteria: (*i*) the pair was found in viruses that were from at least three different continents and so were from geographically distinct sources; (*ii*) there were a minimum of 20 people whose viral sequences included the pair; (*iii*) the pair was found in at least three distant branches of env fragment gp120 trees generated by the method of maximum parsimony (13, 26); and (*iv*) a minimum of four people from whom multiple sequences were available contained a mixture of viral sequences which included some cases of the pair in question and some cases which did not have the pair. The pairs shown here are not the only combinations that may be of interest, but they were the outstanding pairs considering both "predictiveness" and frequency.

A further consideration for peptide vaccine design would be to extend peptides to encompass the covariant sites (between positions 10 and 24) as a minimal boundary, which may allow peptides to more readily fold into a shape which mimics the intact protein. In a recent study by Wang *et al.* (28), features were defined that contributed to the immunogenicity of V3 peptides. The longest peptide they tested served as the best immunogen. It was the only peptide among their set that spanned the full region between positions 10 and 24; however, it also extended beyond position 10, slightly past the N-terminal cysteine of the V3 loop (28).

High mutual information between certain sites suggests that functional studies of the V3 loop using site-directed mutagenesis may depend upon simultaneously altering amino acids on both sides of the loop. Indeed, this has been shown to be the case for some of the positions linked through mutual information analysis—de Jong *et al.* (8) showed that simultaneous mutations were required at site 10 in conjunction with sites 21 through 24, located across the loop, to get a complete conversion in viral phenotype from non-syncytium-inducing, low-replicating to syncytium-inducing, high-replicating. Our analysis indicated sites 10, 23, and 24 were covariant. Also, Chesebro *et al.* (9) have recently further defined critical regions of the V3 loop for imparting macrophage tropism. Blocks of amino acids from macrophage-tropic V3 loops were inserted into the background of T-cell-tropic viruses. In one example from their paper, a single amino acid change at position 12 in Fig. 1, from Ser to His, created noninfectious virus. Altering position 12 in conjunction with positions 20–29 caused a phenotype switch from T-cell- to macrophage-tropic. Thus virus viability as well as such phenotypic "switches" may require simultaneous mutations in covarying sites. Chesebro *et al.* go on to point out that, in the critical sites which have been demonstrated to be important for macrophage tropism, the amino acids observed to be conserved in macrophage-tropic strains can also be found in some T-cell-tropic strains; therefore, the amino acids in these positions acting alone do not appear to account for macrophage tropism and are likely to be acting coordinately with other sites. We propose that rather than exchanging blocks of amino acid sequences, mutational analysis could be done which incorporates only the sites we have observed to be strongly mutually interactive (Fig. 1), with an emphasis on the paired amino acid combinations shown in Table 1.

When sites related by high mutual specific information were compared with alignments of V3 regions of viruses with distinct tropism and cytopathicity, several of the positions that appear to be significant in terms of phenotype (7–11) were also seen to covary (Figs. 1 and 2). This correlation supports the hypothesis that mutual information can identify functionally interactive sites. Some of the sites which are thought to be critical for viral tropism for which we have calculated high mutual information are 12 and 24 (Westervelt *et al.*, ref. 11), 10 and 24 (Fouchier *et al.*, ref. 10), and the sites detected by de Jong *et al.* (ref. 8). While several of the sites we predicted to be mutually interactive were substantiated by experimental evidence, additional linkages were observed that also may be relevant for the generation of viable V3 loops with specific phenotypes. These positions may have been missed in experiments to date, due to their relative conservation among cloned samples used for experiments in tissue culture (specifically, positions 18 and 23).

Several groups have hypothesized that overall charge of the V3 loop may affect phenotype (8, 10). It is interesting to note that 7 of the 11 informative amino acids listed in Table 1 predict alternative amino acids with conserved charge, either negatively charged Glu (E) and Asp (D), or positively charged Arg (R) and Lys (K) or His (H). Such charge conservation may reflect structural constraints.

While the pairs of sites with high mutual information are likely to be structurally or functionally related, the apparent

has at least one pair of positions anywhere in the set with a value as high as the one of interest in the real data was estimated by dividing this number by the total number of random data sets (750,000). Color represents statistical significance with the corresponding estimated probabilities: red, $P < 1.3 \times 10^{-6}$ (didn't occur in 750,000 randomized data sets); orange, $5 \times 10^{-6} < P < 5 \times 10^{-4}$; yellow, $5 \times 10^{-4} < P < 2 \times 10^{-3}$; green, $2 \times 10^{-3} < P < 10^{-2}$; blue, $10^{-2} < P < 10^{-1}$. Combinations of sites that had mutual information values with P values $> 10^{-1}$ are not shown. (*Lower*) Two-dimensional plot of *Upper*, with the same color coding indicating statistical significance.

interdependence may, alternatively, result from an evolutionary heritage from distinct founder viruses. A qualitative attempt was made to rule out the latter possibility among the pairs in V3 found to have the highest mutual information. Each pair in Table 1 was examined independently to determine if representative pairs were found in phylogenetically diverse sequences and among sequences with distant geographic origins, and most pairs met these criteria (Table 1). These observations suggest that genetic linkage is not a dominating influence in the present study. A systematic means of addressing this issue would be a useful development in the application of mutual information algorithms to protein sequences. An additional area of interest is the application of this tool to proteins with known structure, to determine if known structural elements can be detected, in contrast to the functional elements related here. In contrast to the V3 loop sequences considered here, there is presently not enough data to identify statistically significant covariation in intact genes from the HIV-1 genome; such analysis was attempted for HIV-1 full env and reverse transcriptase. Sequence data for HIV are rapidly accumulating, and it is reasonable to expect that within a few years it will be possible to use mutual information to look for potentially interactive sites in distant regions of the linear sequences of HIV-1 genes.

The techniques described here can be of general use for identification of functional relationships in variable proteins, assuming that involved sites are constrained to covary and that the covariation reflects functionality. The technique would not, however, be able to differentiate between functional constraints, which may be a consequence of interactions with host molecules in the case of the HIV-1 V3 loop, and constraints imposed by the inherent structure of the protein. In a sense, natural selection does a first round of mutational experiments, and the resulting ensemble of variable sequences can yield valuable clues to important interrelationships between potentially distant domains in a protein. These putatively identified domains can then become targets for mutational and functional analyses.

We thank Drs. Gerald Myers, James Theiler, Carla Kuiken, and Joseph Bryngelson for thoughtful comments and suggestions. D.H.W. was supported in part by National Library of Medicine Grant F37 LM00011.

1. Goudsmit, J., Debouck, C., Meloen, R. H., Smit, L., Bakker, M., Asher, D. M., Wolff, A. V., Gibbs, C. J., Jr., & Gajdusek, D. C. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 4478–4482.
2. Palker, T. J., Clark, M. E., Langlois, A. J., Matthews, T. J., Weinhold, K. J., Randall, R. R., Bolognesi, D. P. & Haynes, B. F. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1932–1936.
3. Javaherian, K., Langlois, A. J., LaRosa, G. J., Profy, A. T., Bolognesi, D. P., Herlihy, W. C., Putney, S. D. & Matthews, T. J. (1991) *Science* **250**, 1590–1593.
4. Takahashi, H., Nakagawa, Y., Pendleton, C. D., Houghton, R. A., Yokomuro, K., Germain, R. N. & Berzofsky, J. A. (1992) *Science* **255**, 333–336.
5. Hart, M. K., Palker, T. J., Matthews, T. J., Langlois, A. J., Lerche, N. W., Martin, M. E., Scearce, R. M., McDanal, C., Bolognesi, D. P. & Haynes, B. F. (1990) *J. Immunol.* **145**, 2677–2685.
6. Westervelt, P., Gendelman, H. E. & Ratner, L. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 3097–3101.
7. Hwang, S. S., Boyle, T. J., Lyerly, H. K. & Cullen, B. R. (1991) *Science* **253**, 71–74.
8. de Jong, J.-J., Goudsmit, J., Keulen, W., Klaver, B., Krone, W., Tersmette, M. & de Ronde, A. (1992) *J. Virol.* **66**, 757–765.
9. Chesebro, B., Wehrly, K., Nishio, J. & Perryman, S. (1992) *J. Virol.* **66**, 6547–6554.
10. Fouchier, R. A. M., Groenink, M., Kootstra, N. A., Tersmette, M., Huisman, H. G., Miedema, F. & Schuitemaker, H. (1992) *J. Virol.* **66**, 3183–3187.
11. Westervelt, P., Trowbridge, D. B., Epstein, L. G., Blumberg, B. M., Li, Y., Hahn, B. H., Shaw, G. M., Price, R. W. & Ratner, L. (1992) *J. Virol.* **66**, 2577–2582.
12. LaRosa, G. J., Davide, J. P., Weinhold, K., Waterbury, J. A., Profy, A. T., Lewis, J. A., Langlois, A. J., Dreesman, G. R., Boswell, R. N., Shadduck, P., Holley, L. H., Karplus, M., Bolognesi, D. P., Matthews, T. J., Emini, E. A. & Putney, S. D. (1990) *Science* **249**, 932–935.
13. Myers, G., Korber, B. T. M., Berzofsky, J. A., Smith, R. F. & Pavlakis, G. F., eds. (1991) *Human Retroviruses and AIDS 1991* (Theoret. Biol. Biophys. Group, Los Alamos Natl. Lab., Los Alamos, NM), Sect. 3.
14. Holley, L. H., Goudsmit, J. & Karplus, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 6800–6804.
15. Korber, B., Wolinsky, S., Haynes, B., Kunstman, K., Levy, R., Furtado, M., Otto, P. & Myers, G. (1992) *AIDS Res. Hum. Retroviruses* **8**, 1461–1465.
16. Kullback, S. (1959) *Information Theory and Statistics* (Wiley, New York).
17. Blahut, R. E. (1987) *Information Theory and Statistics* (Addison-Wesley, Reading, MA).
18. Lapedes, A., Barnes, C., Burks, C., Farber, R. & Sirotkin, K. (1989) in *Computers and DNA: SFI Studies in the Sciences of Complexity*, eds. Bell, G. & Marr, T. (Santa Fe Inst., Santa Fe, NM), Vol. 7, pp. 157–182.
19. Farber, R. M. & Lapedes, A. S. (1992) *J. Mol. Biol.* **226**, 471–479.
20. Efron, B. (1979) *SIAM Rev.* **21**, 460–480.
21. Efron, B. (1983) *J. Am. Stat. Assoc.* **78**, 316–331.
22. Efron, B. & Tibshirani, R. (1991) *Science* **253**, 390–395.
23. Chandrasekhar, K., Profy, A. T. & Dyson, H. J. (1991) *Biochemistry* **30**, 9187–9194.
24. Faulkner, D. V. & Jurka, A. (1988) *Trends Biol. Sci.* **13**, 321–322.
25. LaRosa, G. J., Weinhold, K., Profy, A. T., Langlois, A. J., Dreesman, G. R., Boswell, R. N., Shadduck, P., Bolognesi, D. P., Matthews, T. J., Emini, E. A. & Putney, S. D. (1991) *Science* **253**, 1146.
26. Myers, G., Berzofsky, J. A., Rabson, A. B., Smith, T. F. & Wong-Staal, F., eds. (1990) *Human Retroviruses and AIDS 1990* (Theoret. Biol. Biophys. Group, Los Alamos Natl. Lab., Los Alamos, NM), Sect. 3.
27. Wolfs, T. F. W., Zwart, G., Bakker, M., Valk, M., Kuiken, C. L. & Goudsmit, J. (1991) *Virology* **185**, 195–205.
28. Wang, C. Y., Looney, D. J., Li, M. L., Walfield, A. M., Ye, J., Hosein, B., Tam, J. P. & Wong-Staal, F. (1991) *Science* **254**, 285–288.