gous replacement is most clearly seen in the Pst I digest where the wild-type 4.2-kb fragment (Fig. 4B, lane 1) was replaced with a 4.7-kb fragment in the recombinant (Fig. 4C, lane 4) reflecting the loss of the second genomic Pst I site and the difference in size between the *CAT* plus *SAG1* sequences and the *ROP1* coding sequence they replace.

Most of the transformants analyzed after CAT selection were not *ROP1* knockouts, and it is unclear whether this was due to a phenotypic disadvantage for parasites lacking *ROP1* (and, thus, overgrowth of *ROP1*-positive clones) or whether it reflected the efficiency of homologous integration.

The success of the *ROP1* targeting contrasts with the failure of the *SAG1* construct to homologously integrate. This could be due to the relative importance of *SAG1* and *ROP1*, to factors such as the length and nature of the sequences flanking *CAT* in the relevant plasmid constructs, or both.

Stable transformation of *Toxoplasma*, in conjunction with analysis of mutants and utilization of transmission genetics, should allow the molecular dissection of the biology of intracellular parasitism. In addition, the use of *CAT* as a selectable marker rather than as a simple reporter may be applicable to the study of other systems, especially the related coccidian parasites *Eimeria* and *Plasmodium*.

REFERENCES AND NOTES

1. B. Danneman *et al.*, *Ann. Int. Med.* **116**, 33 (1992).
2. D. Soldati and J. C. Boothroyd, *Science* **260**, 349 (1993).
3. A. Laban, J. F. Tobin, M. A. Curotto de Lafaille, D. F. Wirth, *Nature* **343**, 572 (1990).
4. A. L. M. A. ten Asbroek, M. Ouellette, P. Borst, *ibid.* **348**, 174 (1990).
5. A. Cruz and S. M. Beverley, *ibid.*, p. 171.
6. E. R. Pfefferkorn, R. F. Rothnagel, S. E. Borotz, *Antimicrob. Agents Chemother.* **36**, 1091 (1992).
7. M. J. Gardner *et al.*, *Mol. Biochem. Parasitol.* **48**, 77 (1991).
8. A. A. Divo, T. G. Geary, J. B. Jensen, *Antimicrob. Agents Chemother.* **27**, 21 (1985).
9. K. Kim, L. D. Sibley, J. C. Boothroyd, unpublished data.
10. J. R. Mineo *et al.*, *J. Immunol.* **150**, 3951 (1993).
11. J. L. Burg, C. M. Grover, P. Pouletty, J. C. Boothroyd, *J. Clin. Microbiol.* **27**, 1787 (1989).
12. K. Kim and J. C. Boothroyd, unpublished data.
13. Freshly lysed tachyzoites (2×10^7) from the clonal RH wild-type strain were transfected with plasmid in Cytomix [M. J. van den Hoff, A. M. Moorman, W. H. Lamers, *Nucleic Acids Res.* **20**, 2902 (1992)] as previously described (*2*). After electroporation, each sample was divided and inoculated onto two T25 flasks of confluent HFF growing in Dulbecco's modified Eagle's medium (Gibco/BRL) supplemented with 10% Nuserum (Collaborative Research), 2 mM glutamine, and gentamycin (20 μg/ml). We harvested one flask for CAT activity 1 day after transfection by scraping the infected monolayer and syringing through a 27 gauge needle. The remaining culture lysed after 3 days and was inoculated onto new monolayers with and without 20 μM chloramphenicol (water-soluble chloramphenicol, Sigma). Five to 10% of the culture (2×10^6 to 5×10^6 parasites) was passed as host cells lysed. Cultures that did not lyse host monolayers within 5 to 7 days were scraped, syringed, and passaged with 20 to 25% of the lysate.
14. Transfection was done with 20 μg of ROP1/3 CAT plasmid linearized with Kpn I. Selection was begun 16 hours after transfection with 20 μM chloramphenicol. Parasites were cloned by limiting dilution in drug 10 days after transfection.
15. D. Soldati and J. C. Boothroyd, unpublished data.
16. J. R. Neuman, C. A. Morency, K. O. Russian, *BioTechniques* **5**, 444 (1987).
17. The initial transient expression of CAT seen in the first 1 to 6 days after transfection with 10 μg of DNA and this number of parasites is not detectable by the phase partition CAT assay. It can only be detected the more sensitive thin-layer chromatographic method.
18. L. D. Sibley and J. C. Boothroyd, *Mol. Biochem. Parasitol.* **51**, 291 (1992).

19. We thank members of the Boothroyd laboratory for helpful suggestions, E. Pfefferkorn for sharing data before publication, and S. Tomavo and F. Seeber for critical review of the manuscript. Supported by grants (to J.C.B.) from the NIH, the MacArthur Foundation, and the Burroughs-Wellcome Fund. K.K. was supported by fellowships from the NIH and the University of California Universitywide AIDS Research Program. D.S. was supported by fellowships from the European Molecular Biology Organization and the Swiss National Science Foundation.

1 June 1993; accepted 5 August 1993

Genetic Identification of a Hantavirus Associated with an Outbreak of Acute Respiratory Illness

Stuart T. Nichol,* Christina F. Spiropoulou, Sergey Morzunov, Pierre E. Rollin, Thomas G. Ksiazek, Heinz Feldmann, Anthony Sanchez, James Childs, Sherif Zaki, Clarence J. Peters

A mysterious respiratory illness with high mortality was recently reported in the southwestern United States. Serologic studies implicated the hantaviruses, rodent-borne RNA viruses usually associated elsewhere in the world with hemorrhagic fever with renal syndrome. A genetic detection assay amplified hantavirus-specific DNA fragments from RNA extracted from the tissues of patients and deer mice (*Peromyscus maniculatus*) caught at or near patient residences. Nucleotide sequence analysis revealed the associated virus to be a new hantavirus and provided a direct genetic link between infection in patients and rodents.

An outbreak of a respiratory illness with high mortality was recently reported in the shared border region (Four Corners) of New Mexico, Arizona, and Colorado in the southwestern United States (*1*). Patients were defined as having unexplained adult respiratory distress syndrome (ARDS) or acute bilateral pulmonary interstitial infiltrates in the presence or absence of prodromal symptoms (*2*). Mortality in confirmed patients has been in excess of 75%, frequently in previously healthy adults between 20 and 40 years of age. Serologic surveys of patients failed to detect evidence of agents normally associated with severe respiratory illness but did detect immune cross-reactivity with previously characterized hantavirus antigens (*1*). This finding was unexpected because hantaviruses had not previously been associated with outbreaks of acute human disease in North America nor had hantaviruses found elsewhere in the world been associated with a severe, predominantly respiratory illness.

Hantaviruses are rodent-borne viruses belonging to the family Bunyaviridae. They possess a negative sense, single-stranded RNA genome consisting of three segments, designated large (L), medium (M), and small (S), which encode the virus polymerase protein (L), the glycoproteins G1 and G2, and the nucleocapsid (N) protein, respectively (*3–7*). At least four distinct virus serotypes have been clearly defined that differ in their overall geographic distribution, rodent host, and degree of pathogenicity for humans (*8*). The Hantaan (HTN) serotype viruses, associated with field mice (*Apodemus agrarius*) and found predominantly throughout Korea, China, and far eastern Russia, cause severe hemorrhagic fever with renal syndrome (HFRS). The Seoul (SEO) serotype viruses are probably found worldwide, which reflects the range of their primary host, *Rattus norvegicus*. The SEO viruses have been associated with a generally more moderate form of HFRS, particularly in Korea and China. Recently, three cases of mild HFRS disease associated with SEO-related virus were described in the United States, although the SEO virus was probably introduced into the United States by colonization of its Eurasian rat host (*9*). The Puumala (PUU) serotype viruses, found throughout Scandinavia and Europe west of the Ural Mountains, are associated with a relatively mild form of HFRS. The primary

Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, GA 30333.

*To whom correspondence should be addressed.

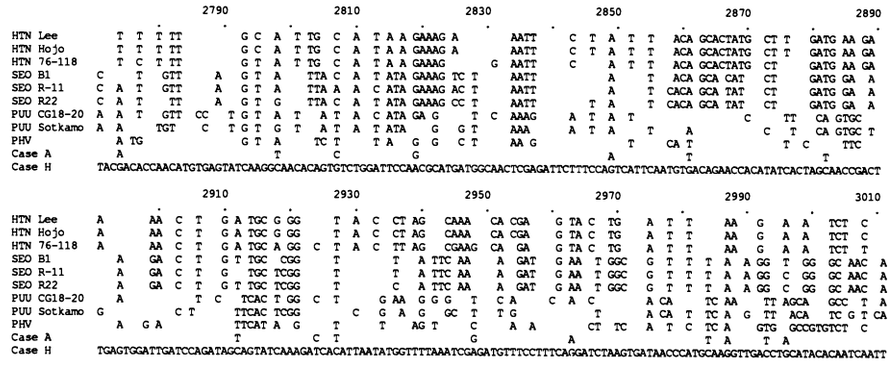

Fig. 1. Comparison of nucleotide sequences of a hantavirus associated with ARDS in the Four Corners region. The nucleotide sequence of 241 bp of the PCR product derived from the two representative hantaviruses detected in New Mexico (case H) and reported in Arizona ARDS cases was compared with the same genome region of previously characterized hantaviruses (*11*). Sequence differences are shown relative to the case H sequence. All sequences are in the viral complementary DNA (+) sense and numbered relative to the overall hantavirus sequence alignment (nucleotides 1 to 3722, including gaps).

rodent host is the bank vole (*Clethrionomys glareolus*). The fourth serotype, Prospect Hill (PH), is known from only two virus isolates, both of which are from indigenous North American rodent species. The original isolate was from meadow voles (*Microtus pennsylvanicus*) in Frederick, Maryland, and the other from Minnesota. PH cross-reactive antibodies have been found in other North American microtine and cricetid rodents and in humans involved in trapping mammals (*10*), but PH has not been associated with any human disease. Other, less well characterized hantaviruses, such as Thottapalayam from a shrew in India, may represent additional serotypes.

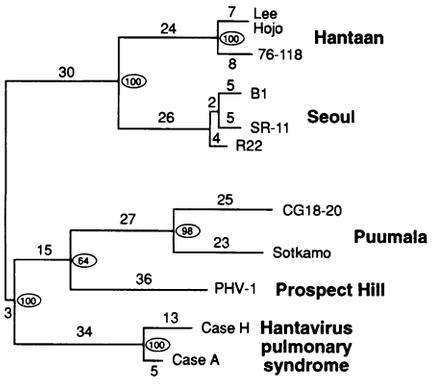

Fig. 2. Phylogenetic relation of the ARDS-associated hantavirus from the Four Corners region (bottom) to previously characterized hantaviruses. Phylogenetic analysis of the virus sequence differences within 241 bp of the amplified fragment was performed by the maximum parsimony method (*12*). The horizontal distances represent the number of nucleotide step differences (indicated adjacent to the lines) present between branch nodes and taxa (that is, viruses). Bootstrap confidence limits exceeding 50% are indicated in ovals next to each branch node.

The two pairs of hantavirus serotypes, HTN and SEO and PUU and PH, each share considerable nucleotide sequence similarity (approximately 70%). Precise regions of sequence conservation within the G2 protein coding region of the M segment of the virus genome were identified, and deoxyoligonucleotide primers were designed for detection of small amounts of hantavirus of either pair of serotypes by a nested reverse transcriptase–polymerase chain reaction (RT-PCR) assay (*11*). The first-round primers were expected to amplify a DNA fragment that contained predicted conserved sequence targets that could be used in a nested PCR by a second pair of PCR primers to give additional specificity and sensitivity to the detection assay. Specific nested PCR DNA products of the correct size [278 base pairs (bp)] were obtained with only the PUU-PH primer set

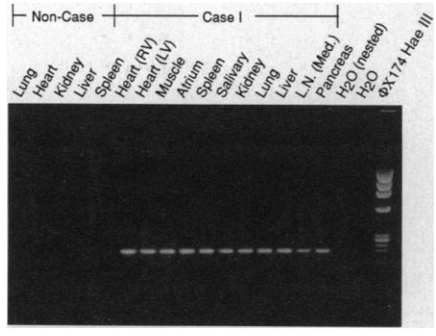

Fig. 3. Agarose gel electrophoresis analysis of hantavirus PCR DNA products from nested RT-PCR assays of total RNA extracted from ARDS patient autopsy specimens. Shown is a hantavirus-specific DNA band, 172 bp in length, amplified from total RNA from various tissues from case I. RV and LV, heart right and left ventricles, respectively; L.N. (Med.), medial lymph nodes.

when used with total RNA extracted from tissues from several of the outbreak patients (that is, nested HTN-SEO primer reactions were negative, as were control autopsy materials). Amplified DNA from the tissues of two representative ARDS patients (case H in New Mexico and case A in Arizona) was extracted from the agarose gel and analyzed with an automated thermocycle sequencing technique (Applied Biosystems, Foster City, California) with the same primers used for PCR product synthesis.

The DNA bands were found to contain hantavirus-like sequences (Fig. 1), but their nucleotide sequence differed from that of any of the known hantaviruses by at least 30%. A 7% nucleotide difference was also observed between PCR bands from case H and those from case A. Phylogenetic analysis of the sequence differences by the maximum parsimony method (*12*) indicated that the hantavirus associated with ARDS in the Four Corners region was novel, representing a distinct lineage, and was most closely related to PH, the only other hantavirus recovered from rodent species indigenous to North America (Fig. 2). This relation is supported by the serologic cross-reactivity obtained with convalescent-phase sera from patients (*13*).

On the basis of sequences derived from the two representative hantaviral ARDS cases, we designed a new second-round (nested) pair of PCR primers [+2816 AAGGTAACACAGT(G/C)TCTGGATTC and −2955 GGTTATCACTTAGATC(C/T)TGAAAGG]. A broader screening of autopsy specimens was then carried out by an RT-PCR assay with the original first-round primers followed by the new second-round primers, which generate a specific, 172-bp DNA band. Frozen-tissue autopsy specimens were available from 10 patients who met the case definition (*2*). When tested with the new primer set, all 10 patients were found to be PCR positive (Fig. 3). Lung, heart, liver, kidney, and spleen tissues were frequently found to be positive. At the time of death, the virus was apparently distributed extensively throughout the body. Currently, the site of virus replication is unknown, and no precise quantitation of virus in different tissues has been performed. However, immunohistochemical analysis of tissues with a monoclonal antibody that cross-reacts with conserved hantavirus nucleoprotein epitopes revealed widespread endothelial involvement, with deposition of antigen in lung, kidney, heart, pancreas, adrenals, and skeletal muscle (*14*). The pathophysiologic role of the endothelial involvement in the increased vascular permeability, which is a hallmark of this syndrome, remains to be elucidated.

A systematic rodent trapping effort initiated this year in the Four Corners region

allowed a calculation of the relative density of the rodent population. The deer mouse (*Peromyscus maniculatus*) was the dominant rodent species, accounting for 63% of all captures. Initial serologic testing (by immunoglobulin G enzyme-linked immunosorbent assay) of all small rodents captured at or near case and control houses indicated an extremely high seropositivity to PUU-PH antigens (approximately 33%) among *P. maniculatus* (*15*). Of 54 seropositive *P. maniculatus* tested, 44 (82%) gave a positive PCR band, which indicates a high persistence of the virus in this rodent reservoir even in the presence of a substantial humoral immune response. Similar results were recently obtained with PCR analysis of seropositive, SEO virus–infected rats (*16*). Indeed, the presence of virus with persistent shedding in excreta, concurrent with the presence of circulating antibodies, is a hallmark of hantaviral infection in other rodent species (*17*). In addition, of 45 seronegative *P. maniculatus* tested, 14 (31%) were PCR positive. These likely represent recently infected individuals that have not yet mounted an immune response or, alternatively, may indicate a greater sensitivity of detection of virus infection by PCR. These data suggest that the virus can be maintained and actively transmitted in this rodent species and implicate *P. maniculatus* as an important virus reservoir. Hantaviral infection of other rodent species (*P. truei, P. boylei, Tamias dorsalis*) in this region could also be detected but to a much lesser extent (*15*).

Nucleotide sequence analysis was performed on PCR products obtained from each of the PCR-positive patients and from PCR-positive rodents captured at or in the vicinity of the residences of the patients. Despite the fact that the region of the genome (within the G2 coding region) targeted for amplification is generally highly conserved among hantaviruses and that virus-infected samples were collected over a period of less than 2 months from a limited geographic area, a surprisingly high level of genetic diversity was observed among the samples (Fig. 4). However, identical sequences were consistently obtained from independent RNA preparations from the same tissue or PCR products from different tissues within an individual.

Comparison of the human and rodent sequences demonstrated a direct genetic link between the virus in infected *Peromyscus* rodents and the virus in the human hantaviral ARDS cases. Viral sequences from two of the cases (G and H, from Pinehill/Ramah, New Mexico) were genetically closely clustered. The viral sequence of case G was identical to that obtained from three *P. maniculatus* captured at or near the patient's residence. Similarly, the viral sequence of case H was identical to those of two *P. maniculatus* trapped at or near the patient's residence. These results suggest that these individuals may have acquired infection in or near their residences. A third genetically closely related patient (case J) was from 35 miles northeast, in the Crownpoint area. However, an additional patient (case I), from Black Mesa, Arizona, was also contained within this genetic cluster.

Two other patients (cases B and C) were from southwestern Colorado. Their viral sequences were closely related to one another and to those from rodents captured nearby but were distinct from the New Mexico clusters. Case A was officially recorded as an Arizona case. This individual had resided in the vicinity of Springerville, Arizona, 11 days before the onset of illness in nearby Snowflake, Arizona (and death 3 days later), but had resided in Hesperus, in southwestern Colorado, before this trip. The viral sequence from this patient was found to be closely related to that obtained from other human and rodent samples from Colorado and identical to that of a *P. maniculatus* captured near the patient's residence in Hesperus, which suggests that this individual acquired the infection in Colorado before traveling to Arizona.

On the basis of the limited sequence region analyzed, cases D, E, and F and additional rodents do not appear to form a geographic cluster. A clearer picture may emerge as these studies are extended to other regions and segments of the virus genome. However, these initial results of extensive virus genetic diversity obtained over a small sampling period, together with examples of geographic clustering of virus variants, suggest that this virus may have been independently maintained for some time in rodent populations in multiple areas of the southwestern United States. More extensive studies of virus and rodent populations will be needed to confirm this hypothesis and to gain a better understanding of the epidemiology and public health impact of this disease. As *P. maniculatus* represents one of the most abundant small mammals in North America, it is of particular importance to determine the potential role it may play as a reservoir for hantaviruses associated with severe ARDS.

	Case	Rodent	Location		
1	A		Snowflake	AZ*	g t t c
2		1(A)	Hesperus	CO	g t t c
3		2	Grand Junction	CO	g t t c
4	B		Cortez	CO	g t t t g
5	C		Grand Junction	CO	g c t t g g
6		3	Grand Junction	CO	t t g c
7	D		Sedan	NM	g a
8	E		Farmington	NM	g a
9		4(H)	Pinehill/Ramah	NM	t c
10		5	Springerville	AZ	t c
11		6	Springerville	AZ	t c
12	F		Alpine Ranchero	AZ	g t c
13		7	Black Mesa	AZ	g c
14	G		Pinehill/Ramah	NM	g t
15		8(G)	Pinehill/Ramah	NM	g t
16		9(G)	Pinehill/Ramah	NM	g t
17		10(G)	Pinehill/Ramah	NM	g t
18	H		Pinehill/Ramah	NM	g c c a
19		11(H)	Pinehill/Ramah	NM	g c c a
20		12(H)	Pinehill/Ramah	NM	g c c a
21		13(G)	Pinehill/Ramah	NM	g c c
22	I		Black Mesa	AZ	g t c c
23	J		Crownpoint area	NM	t g c c
24		14	Black Mesa	AZ	g a t a c t g
				Consensus:	A C T C A C G T C C A A T A T T T G T
					2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2 2
					8 8 8 8 8 8 8 8 8 8 8 9 9 9 9 9 9 9 9
					1 3 4 4 4 5 6 7 7 8 8 0 1 2 2 2 4 4 5
					9 5 3 6 9 5 1 3 6 2 8 6 2 1 4 7 5 8 4

Fig. 4. Nucleotide sequence differences among the PCR products derived from hantaviral ARDS patients from the Four Corners and from rodents trapped in the region. Shown is a genetic cluster analysis (*18*) of 139 bp of virus sequences relative to the consensus sequence. Nucleotide difference positions are numbered as in Fig. 1. The asterisk indicates that the infection was acquired in Colorado (see text).

REFERENCES AND NOTES

1. *Morb. Mortal. Wkly. Rep.* 42, 421 (1993).
2. *ibid.*, p. 441.
3. F. Gonzalez-Scarano and N. Nathanson, in *Virology*, B. N. Fields *et al.*, Eds. (Raven, New York, ed. 2, 1990).
4. C. S. Schmaljohn, A. L. Schmaljohn, J. M. Dalrymple, *Virology* 157, 31 (1987).
5. L. B. Giebel, R. Stohwasser, L. Zoeller, E. K. Bautz, G. Darai, *ibid.* 172, 498 (1989).
6. J. Arikawa, H. F. Lapenotiere, L. Iacono-Connors, M. Wang, C. S. Schmaljohn, *ibid.* 176, 114 (1990).
7. M. A. Parrington, P. W. Lee, C. Y. Kang, *J. Gen. Virol.* 72, 1845 (1991).
8. K. T. McKee Jr, J. W. LeDuc, C. J. Peters, in *Textbook of Human Virology*, R. B. Belshe, Ed. (Mosby–Year Book, St. Louis, ed. 2, 1991), pp. 615–632.
9. G. E. Glass, A. J. Watson, J. W. LeDuc, J. E. Childs, *Nephron*, in press.
10. R. Yanagihara, *Rev. Infect. Dis.* 12, 449 (1990).
11. The complete M genome segment nucleotide se-

quences of hantaviruses Hantaan strains 76118 (GenBank numbers M14627 and Y00386), Lee (D00377), and Hojo (D00376); Seoul strains B-1 (X53861), SR-11 (M34882), and R22 (S68035); Puumala strains CG18-20 (M29979) and Sotkamo (X61034); and Prospect Hill strain PHV-1 (X55129) were aligned with the GAP, PILEUP, and LINEUP programs of the GCG software package (Genetics Computer Group, Madison, WI) run on a VAX computer. Predicted conserved positions for the synthesis of nested RT-PCR oligonucleotide primers for HTN-SEO viruses or PUU-PH viruses were as follows. HTN-SEO first-round primers: +2548 GATATGAATGATTG(T/C)TTTGT and −2859 CCATCAGGGTCT(T/C)TCCA; second-round: +2590 TGTATAATTGGGAC(T/A)GTATCTAA and −2751 GCAAAGTTACATTT(T/C)TTCCT (position numbering of the oligonucleotide 3′ terminus was relative to the total aligned and gapped sequence length of 3722 nucleotides); PUU-PH first-round primers: +2671 TTTAAGCAATGGTG(C/T)ACTAC(T/A)AC and −3108 CCATAACACAT(A/T)GCAGC; second-round: +2770 AGAAAGAAATGTGCATTTGC and −3012 CCTGAACCCCATGC(A/T/C)CCATC. Because of the hazardous nature of the agent, all steps of the homogenization of autopsy tissue samples and the total RNA extraction and purification were performed under biosafety level 3 conditions. RNA extraction, first-round RT-PCR reactions, and subsequent product DNA gel electrophoresis analysis were performed essentially as described [L. L. Rodriguez, G. J. Letchworth, C. F. Spiropoulou, S. T. Nichol, *J. Clin. Microbiol.* **31**, 2016 (1993)], except the following cycle profile run on a Perkin-Elmer 9600 thermocycler was used: 41°C for 1 hour, followed by 40 cycles at 94°C for 40 s, at 38°C for 45 s, and at 72°C for 60 s. Second-round reaction conditions used 3% of the first-round reaction product, no RT step; the following profile was used: 35 cycles at 94°C for 40 s, at 43°C for 45 s, and at 72°C for 60 s.

12. PAUP: Phylogenetic Analysis Using Parsimony, version 3.0 s; D. L. Swofford (Illinois Natural History Survey, Champaign, 1991). The software was run on a SUN SPARC 10 workstation. Two equally parsimonious trees were obtained (with minor branch order differences among the SEO viruses) with the use of the BANDB option. Bootstrap confidence limits were obtained with 10,000 repetitions, and the ALLTREES option was used to examine the tree-length frequency distribution.
13. T. G. Ksiazek and P. E. Rollin, unpublished data.
14. S. Zaki, unpublished data.
15. *Morb. Mortal. Wkly. Rep.* **42**, 477 (1993).
16. R. R. Arthur, R. S. Lofts, G. E. Glass, J. W. LeDuc, J. E. Childs, *Am. J. Trop. Med. Hyg.* **47**, 210 (1992).
17. J. W. LeDuc, *Lab. Anim. Sci.* **37**, 413 (1987).
18. D. G. Higgins and P. M. Sharp, CABIOS **5**, 151 (1989).
19. We thank the reporting physicians, the Navajo Nation, and the staff of the Indian Health Service, the New Mexico, Arizona, and Colorado state health laboratories, and the Centers for Disease Control and Prevention epidemiologic team and environmental assessment teams for provision of samples and case information. We are also grateful to M. Monroe and S. Trappier for excellent technical assistance and to B. Mahy for support and encouragement. C.F.S. and S.M. are supported by a U.S. Department of Agriculture Animal Molecular Biology National Research Initiative grant 90-37266-5473 through the University of Nevada, Reno. H.F. is supported by a National Research Council fellowship.

9 August 1993; accepted 4 October 1993

TECHNICAL COMMENTS

Alpha Helix Propensity of Amino Acids

Michael Blaber *et al.* state that they have found a correlation between changes in stability for substitution mutants of T4 lysozyme at two positions in α helices and the amount of nonpolar surface buried by the substituted side chain when it folds to the native structure (*1*). If this correlation is real, their work represents a significant advance in our understanding of the structural basis for the different α helix propensities of the amino acids. However, before the validity of the correlation [figure 2 of (*1*)] can be properly evaluated, several technical issues should be addressed.

For a statistical correlation involving a small data set to have scientific significance, it must be generally true. A provisional test of the proposed correlation for a subset of the amino acid residue (*2*) substitutions (L, V, I, S, T, K, E, and N) at position 44 can be made from examination of the substitutions at a second site, position 131. Although data are presented for only five substitutions at this position (L, V, I, S, and T), there is no significant correlation between free energy differences (ΔΔG) and buried surface area.

The correlation between the free energy of folding and buried surface area for substitutions at position 44 depends on the exclusion of three data points in addition to those for A, G, and P. The exclusion of R because of a crystal contact seems justifiable (though K should then be excluded for the same reason). However, the logic for discarding the critical F and W points [described in note 22 of (*1*)] is unclear. Blaber *et al.* appear to argue that, because these two mutants crystallize in space groups that are different from that of the other mutants, the observed trans χ_1 angle (which distinguishes these side chains from all others except the wild-type S residue) may be an artifact of the different crystal environments. The fact that the conformation of tryptophan is similar in all four asymmetric environments (and is similar to that of phenyalanine in each of its two asymmetric environments) suggests that this rotamer is *not* an artifact of crystallization.

Finally, we would caution against the deceptive ease with which statistical correlations can be made. Most of the data presented by Blaber *et al.* falls into a fairly narrow range of values for both ΔΔG and for nonpolar surface buried. Nine of 17 residues (excluding G, A, and P) could have, within the experimental error of ±0.1 kcal/mol, the same ΔΔG value of +0.63 kcal/mol. Even more striking is the narrow range into which most surface area values fall. Particularly noteworthy are the similarities of buried area for A, V, E, T, and S and of calculated areas for extended side chains of Q, R, K, Y, H, and F. This might be expected because, in many cases, the β carbon atom often makes the greatest contribution to buried nonpolar surface. As a consequence of this tight clustering of values within a narrow range, it only takes one or two outlying data points to establish an apparent correlation. This reduction in the number of data points that significantly contribute to a correlation increases the ever-present danger that an apparent correlation might be found by omitting data points from an otherwise random-looking scatterplot. Only when a rigorous and compelling argument is at hand can one, in the search for a better correlation, safely exclude a subset of the data.

While burial of side chain hydrophobic surface may play a role in determining the rank order and magnitude of helix propensity, the correlation reported by Blaber *et al.* does not convincingly establish that it is the structural basis of amino acid α helix propensity.

David Shortle
Department of Biological Chemistry, Johns Hopkins University School of Medicine, Baltimore, MD 21205–2185
Neil Clarke
Department of Biophysics and Biophysical Chemistry, Johns Hopkins University School of Medicine

REFERENCES AND NOTES

1. M. Blaber, X. Zhang, B. W. Matthews, *Science* **260**, 1637 (1993).
2. Abbreviations for the amino acid residues are as follows: A, Ala; C, Cys; D, Asp; E, Glu; F, Phe; G, Gly; H, His; I, Ile; K, Lys; L, Leu; M, Met; N, Asn; P, Pro; Q, Gln; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; and Y, Tyr.

9 July 1993; accepted 14 September 1993

Response: Shortle and Clarke correctly state that if a statistical correlation involving a small data set is to have scientific significance, it must be generally true. We have tested (*1*) the correlation in question by comparing the area buried on helix formation [as calculated by Richards and Richmond (*2*)] with experimentally determined scales of helix propensity that are based on substitutions in proteins (*1*, *3*, *4*), model peptides (*5*), host-guest experiments (*6*), and frequencies observed in known protein structures (*7*). In every case the correlation was positive, with an average value of 0.49