Theoretical Evidence that the Ebola Virus Zaire Strain May Be Selenium-Dependent: A Factor in Pathogenesis and Viral Outbreaks?

Ethan Will Taylor[1] and Chandra Sekar Ramanathan

Abstract

A theoretical analysis of the genomic structure of the Ebola virus Zaire strain reveals the existence of several open reading frames (ORFs) containing large numbers of in-frame UGA codons. This clustering of UGA codons is very unlikely to have arisen by chance, and raises the possibility that these ORFs may encode selenoproteins, since, in addition to its usual role as a stop codon, UGA can under certain conditions encode selenocysteine. The other major requirement for selenocysteine insertion at UGA codons appears to be met in this case, due to the presence of selenocysteine insertion sequences (SECIS) in stable stem-loop structures in the appropriate Ebola Zaire mRNAs. Specifically, there is a SECIS in the 3'-untranslated region of the nucleoprotein mRNA, where the largest potential selenoproteins are encoded, one of which may contain up to 16 selenium atoms per molecule. The expression of this hypothetical protein could impose an unprecedented selenium demand upon the host, potentially leading to severe lipid peroxidation and cell membrane destruction. This could also contribute to the characteristic hemorrhaging caused by intravascular blood clotting, due to the thrombotic effect of Se deficiency. The possibility that this gene might contribute to the extreme pathogenicity of the Zaire strain of Ebola virus by this mechanism is also consistent with the observation that this potential selenoprotein gene is not present in the Ebola Reston strain, which was not pathogenic in humans.

It has long been apparent that an increased susceptibility to infectious diseases is common in malnourished human populations. This has traditionally been viewed as simply a consequence of the fact that the immune system must be maintained by adequate nutrition in order to function optimally. Only recently has data begun to accumulate in support of the idea that nutritional factors may sometimes have a direct effect on pathogens, and that passage through nutritionally deficient hosts may facilitate evolutionary changes in infectious agents. In this communication, we present theoretical molecular evidence that the highly pathogenic Zaire strain of the Ebola virus may be dependent on the trace mineral selenium (Se), due to the presence in the Ebola genome of several open reading frames (ORFs) containing clusters of up to 17 in-frame UGA codons, which potentially encode the rare amino acid selenocysteine (SeC). We will argue that, by analogy to other known examples, this raises the possibility that Se deficiency in host populations may actually foster viral replication, possibly triggering outbreaks and perhaps even facilitating the emergence of more virulent viral strains.

This concept is not unprecedented: a classical example of a nutrient effect on viral replication is the well documented induction of endogenous retrovirus expression in cells cultured in arginine-deficient media (recently reviewed by Becker[1]). Note that arginine is an essential component of many viral proteins. Thus, paradoxically, in this case viral replication appears to be triggered by a deficiency of something the virus requires. This would most likely involve some sort of repressor type of mechanism. Based on the data that we will review here, we suggest that, for some viruses, an analogous situation may exist in the case of Se.

Aside from the nonspecific protection that can be achieved *in vivo* by a nutritional boost in immune function, specific antiviral effects have been claimed for various antioxi-

1. Computational Center for Molecular Structure and Design, and Department of Medicinal Chemistry, The University of Georgia, Athens, GA 30602-2352.
Email: wtaylor@rx.uga.edu

dant nutrients, and Se in particular. A surprising range of *in vitro* and *in vivo* antiviral activities has been reported for various simple Se compounds, including inhibition of hepatitis B in humans, influenza virus in culture, and retroviruses like the mouse mammary tumor virus and bovine leukemia virus (reviewed by Schrauzer and Sacher[2] and by Taylor et al.[3]). Recent work has also demonstrated the *in vitro* activity of Se compounds against the human immunodeficiency virus, HIV-1.[4,5]

The most compelling data pertain to Keshan disease, a classical Se-deficiency disease manifested as a non-obstructive cardiomyopathy. Chinese investigators suspected an infectious agent might be a cofactor, and eventually isolated coxsackievirus from the hearts of Keshan disease victims; the combination of the virus and Se deficiency produced cardiomyopathy in mice.[6] Recently, Beck and coworkers have shown that in Se-deficient mice, even a normally non-virulent strain of coxsackievirus B3 can produce myocarditis similar to that seen in Keshan disease[7] (and references therein). Significantly, during passage through Se-deficient mice, the virus mutates into a more virulent strain that is pathogenic even in normal animals on Se-adequate diets.

Along similar lines, it is of considerable interest that Ziegler has pointed out a correlation between high rates of endemic Kaposi's sarcoma (KS) in African subsistence farmers and geographic regions in Africa where the soils are of volcanic origin.[8] These include regions surrounding the entire East African Rift Valley and the Nigeria-Cameroon border. It is widely documented that low Se levels in plants and Se deficiency syndromes of livestock are common in areas with soils of volcanic origin: the Rift Valley is a typical example. Furthermore, Se deficiency in humans has been specifically documented in northern Zaire (e.g.[9]). Since recent evidence strongly suggests that KS involves a novel herpesvirus, this association of KS in Africa with low Se areas suggests a possible analogy to Keshan disease and coxsackievirus. Significantly, we have also found very large ORFs with start codons and up to 11 in-frame UGA codons in herpesviruses like cytomegalovirus and Epstein Barr virus (reported at 8th ICAR, Taylor et al.[5]), suggesting that some herpesviruses may also be "Se-dependent".

Rather than being indirect (e.g. involving a nonspecific antioxidant effect), the possibility that some antiviral effects of Se might involve virally-encoded selenoproteins has apparently not been considered until very recently.[3,5] Even though first demonstrated about ten years ago, it has still has not become widely appreciated that SeC can be encoded by the UGA codon, which usually serves as a stop codon in the genetic code. Conventional analyses of potential protein coding regions in genes still do not usually discriminate UGA from the other two stop codons, and thus they fail to reveal that proteins might be encoded in regions containing UGA codons. Such regions are routinely assumed to be inactive due to the presence of stop codons, which is probably true in the vast majority of cases, because efficient SeC incorporation is only possible when the mRNA contains a *cis*-acting signal known as a SeC insertion sequence.[10] However, as shown in Figure 2, such consensus SeC insertion sequences capable of forming the required characteristic stem-loop RNA structures *are* present in several Ebola mRNAs that also encode UGA-rich ORFs.

We recently reported a similar potential for selenoproteins to be encoded in HIV-1[3,5,11] and in coxsackievirus B3,[11] in regions overlapping known genes. In both cases, the link between Se deficiency and the associated viral diseases (AIDS and viral myocarditis, respectively) is strongly supported by an extensive body of literature (reviewed in[2,3,7,12]).

In the Ebola virus genome (Zaire strain), there are several ORFs with highly significant clusters of in-frame UGA codons. Overlapping the major nucleoprotein (NP) gene, there are two ORFs in the -1 reading frame, containing 17 and 11 UGA codons, respectively (Figure 1). The first ORF has excellent potential to be expressed by a ribosomal frameshift from the NP coding region, due to the presence of an "ideal" heptameric shift sequence and an RNA pseudoknot (PK) 8 bases downstream (Figure 3A). This frameshift site comes very near the beginning of the ORF, and could permit the formation of a fusion protein consisting of the N-terminal 314 residues of NP fused to a 181 residue C-terminal module potentially containing 16 SeC residues, encoded in the -1
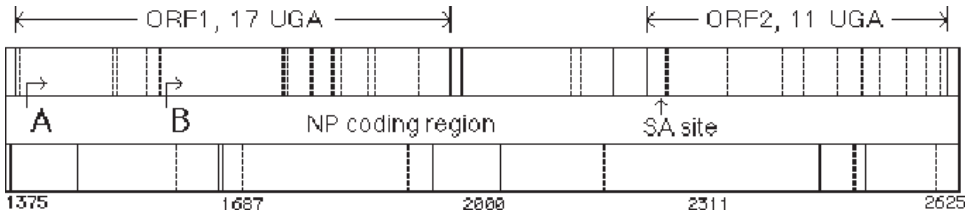
Figure 1. UGA-rich open reading frames (ORFs) overlapping the Ebola Zaire nucleoprotein (NP) coding region. The figure shows a schematic of the three reading frames for a portion of the NP gene, with stop codons shown as vertical lines. The dotted lines are UGA stop codons, which can potentially encode selenocysteine. There are two UGA-rich ORFs -1 to the main NP reading frame, ORF1 with 17 and ORF2 with 11 in-frame UGA codons. Neither ORF1 or ORF2 has a start codon. ORF1 could be expressed as an NP fusion protein containing a selenoprotein module, by means of a frameshift at either one of two potential -1 frameshift sites, shown with an arrow symbol as **A** and **B**. These frameshift sites are shown in detail in Figure 3. ORF2 is less likely to be a functional gene, because it could only be expressed from an edited or spliced NP mRNA, and splicing has never been demonstrated in filoviruses (see text). However, there is a complete splice acceptor (SA) site at the very beginning of the ORF, and several potential upstream splice donor sites (not shown). The only other requirement for selenoprotein synthesis, a selenocysteine insertion sequence (SECIS) in the required stem-loop structure, is present in the 3'-untranslated region of the NP mRNA (Figure 2A). (Figure produced using the beta version of the gene finder program, developed in collaboration with Dr. Dan Everett, University of Georgia).
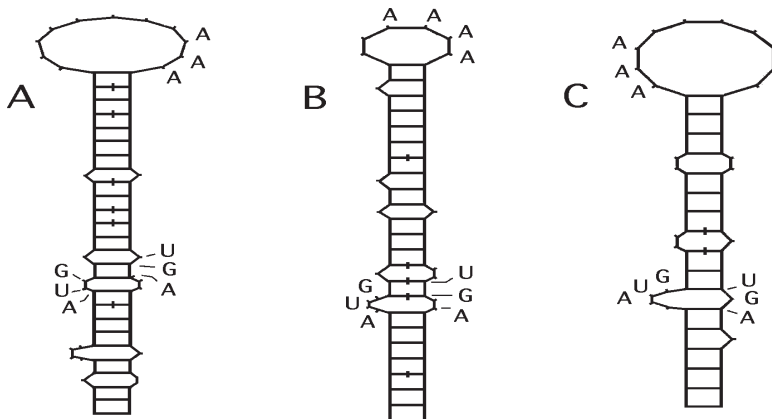

Figure 2. Schematic RNA secondary structures predicted for selenocysteine insertion sequences (AUG...AAA...UGA) in the Ebola virus RNA with potential to form the required stem-loop structures.[10] **A**: In the 3'-untranslated region of the nucleoprotein mRNA, bases 2758-2836 in GenBank #L11365; E = - 10.1 kcal/mole. **B**: At the 3' end of the vp35 mRNA, bases 4094-4160; E = - 13.4 kcal/mole. **C**: At the 3' end of the vp30 mRNA, bases 9029-9087; E = - 9.4 kcal/mole. Note that the computed stability of these structures is comparable to that determined by Sanchez et al.[19] for the 5'-end stem-loop structures in the Ebola mRNAs (average E = -13.3 kcal/mole, range = -7.6 to -20). All base pairs (shown as ladder rungs) are Watson-Crick except those marked by a slash, which are GU base pairs. These structures were predicted using the Zuker FOLD program[20] as implemented in the GCG software package (Program Manual for the Wisconsin Package, Ver. 8, September 1994, Genetics Computer Group, 575 Science Drive, Madison, WI 53711).
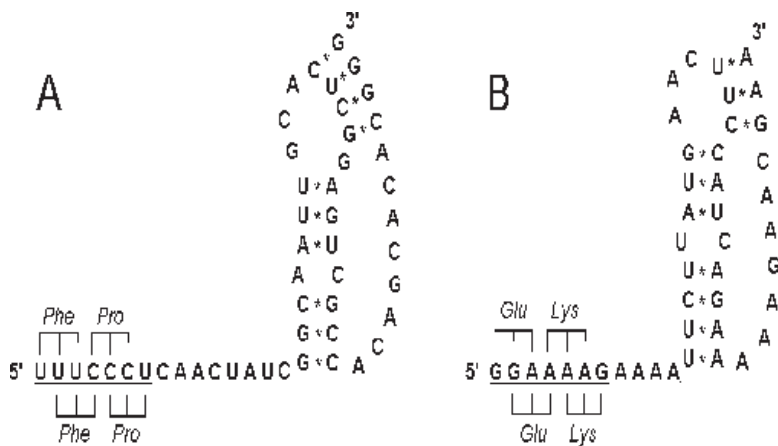
Figure 3. Potential -1 frameshift sites near the beginning of the major UGA-rich ORF in the Ebola Zaire nucleoprotein (NP) coding region, consisting of slippery sequences (underlined) and potential RNA pseudoknots. The location of these sites are indicated by **A** and **B** in Figure 1. Codon-anticodon interactions of the P- and A-site tRNAs are shown schematically both before (below sequence) and after slippage (above sequence). **A**: An "ideal" (XXXYYYZ) heptameric -1 frameshift sequence beginning at position 1405 in GenBank #L11365, located 8 nucleotides upstream from a potential pseudoknot. The A-C bulge shown in the major stem probably forms a hydrogen bonded purine-pyrimidine A:C base pair, as these have been observed in some experimental RNA stem structures. The ORF in the -1 reading frame has a total of 17 in-frame UGA codons, 16 of which are downstream from this frameshift site. **B**: A second near-ideal frameshift site and potential pseudoknot in the NP coding region beginning at position 1582, 178 bases downstream from that shown in **A**. This could produce a shift into the same ORF with 17 UGA codons, but the potential selenoprotein module would contain only 11 SeC residues (Figure 1).

frame (bases 1411 to 1953 in GenBank #L11365; subsequent numbering refers to the same sequence). Slightly downstream there is a second near-ideal frameshift site and potential PK, also in the NP coding region beginning at position 1582 (Figure 3B). This could provide a "second chance" to express the selenoprotein module, whenever the first frameshift failed. This second site follows the sixth UGA codon in the ORF, so a frameshift here would yield a potential selenoprotein module with only 11 SeC residues. These redundant frameshift sites could provide for either an increased probability of translating the selenoprotein module, or for two alternate forms of the NP fusion protein.

Because there are three different stop codons (UGA, UAA and UAG), if this is not a real gene and the potential shift sites and PKs are mere "artifacts", the chance of the next 16 stop codons following shift site **A** in the "blocked" -1 reading frame (Figure 1) all being UGA could be estimated as $(1/3)^{16}$, or less than one in 43 million. Thus, the high significance of this clustering of UGA codons, combined with the presence of the frameshift signals and the distinctive SeC insertion sequence in the 3'-untranslated region of the Ebola NP mRNA, argues overwhelmingly that this must be the gene of an actual selenoprotein. However, one cannot completely rule out the possibility that it could be a vestigial gene that may have only recently become inactive.

A second UGA-rich ORF overlapping the Ebola NP gene, encoded between bases 2212 and 2598, contains 11 UGA codons over 129 residues (ORF2 in Figure 1). This ORF lacks a start codon, but could be expressed

from an edited or spliced RNA. There is a definite splice acceptor site very near the beginning of the ORF, a CAGA sequence preceded by a pyrimidine rich sequence and an upstream "CURAY" sequence (CUGAC). There are various potential splice donors in the large NP mRNA that could bring this region in-frame to the main NP ORF or the upstream selenoprotein ORF with 16 UGAs. Since there are no reports of Ebola replication and transcription in the nuclei of infected cells, this ORF and the associated potential splice sites may be mere artifacts. However, Borna disease virus provides a precedent for nuclear replication/transcription and RNA splicing of a negative nonsegmented single stranded RNA virus.[13] Thus, we cannot rule out the possibility that splicing of this Ebola mRNA could occur, e.g. in the unknown "reservoir" species that is the natural host for Ebola virus. Furthermore, RNA editing can also bring such an "out of frame" ORF into frame, and RNA editing is known to occur in a number of viruses, including Ebola Zaire.

There are several additional potential selenoprotein ORFs overlapping the first 6 genes of Ebola virus, including the vp24, vp30, vp35 and vp40 regions, all of which have potential SeC insertion sequences in their mRNAs (shown for vp30 and vp35 in Figure 2). Because the sequence has not yet been released, we are unable to report on the polymerase coding region, where we have consistently found potential overlapping selenoprotein genes in a number of other viruses. On the Ebola minus strand genomic RNA there are also potential SeC insertion sequences and several UGA-rich ORFs (up to 9 UGAs), some with start codons, and some potentially expressed from spliced genomic RNAs. On both plus and minus strands, some of these potential genes have start codons in the context of Kozak-like sequences, suggesting they may be programmed to bind ribosomes and initiate protein synthesis. All these data will be presented in detail in a subsequent publication.

If viruses like HIV-1, coxsackievirus B3 and Ebola do encode selenoproteins, why does all the evidence suggest that dietary Se inhibits viral replication, whereas Se deficiency triggers replication? Why would Se not "feed" the virus? The answer must lie in how viruses use Se.

As discussed previously,[3] due to the inefficiency of frameshifting and SeC insertion mechanisms, these hypothetical viral selenoproteins could only be formed in very small amounts. Thus, in most cases they are not likely to be major structural proteins; some might have regulatory roles, acting in the mid-phase of the life cycle, and might not even be packaged in virions. If even one such selenoprotein were involved in negative feedback on replication (a repressor type function), decreased levels of that protein would provide the virus a way to respond to low Se levels by leaving the cell in search of a new host. By such a mechanism, the virus could satisfy a basal dependence on Se by escaping from a cell where Se levels had become dangerously low.

Since Se is an essential antioxidant, critical as a component of glutathione peroxidase in blood cells and liver cells (the very cell types that Ebola and many other viruses prefer to infect), very low Se levels are potentially associated with oxidative stress, lipid peroxidation and cell death. Thus, viral survival might be enhanced by the stimulation of replication under low Se conditions.

At the same time, host/viral competition for a limited amount of Se - particularly in a malnourished host - could significantly contribute to pathogenesis. This could be particularly acute with Ebola virus, due to the unprecedented high Se requirement implicit in the ORF with 16 UGA codons.

Dietary Se is also known to have immunopotentiating effects (reviewed in[12]). Thus, in addition to any direct effects exerted via (hypothetical) viral selenoproteins, Se deficiency can also weaken the immune system's ability to fight viral infection, permitting increased replication, rapid mutation, and facilitating the emergence of more virulent strains, as Beck et al. suggest in the case of coxsackievirus.[7]

Given the unique dependence of selenoprotein genes upon a trace nutrient whose availability varies widely in geographical areas and host populations, the presence and activity of such genes would very likely be strain specific, as we suggested for coxsackievirus.[11] This could help explain why some viral strains can be very virulent,

even when significant subsets of indigenous populations have antibodies to a similar, presumably much less virulent viral strain, which appears to be true even for filoviruses.[14,15]

Certainly, prolonged Se deficiency in a host population could eventually lead to the inactivation and loss of any viral selenoprotein genes. Whether that loss would lead to more virulent strains, or whether those strains might undergo a compensatory attenuation by passage through the host population, would be difficult to predict.

However, in the case of Ebola virus, there is some reason to think that the presence of a gene with an exceptionally high Se demand could be a factor in the pathogenicity of specific viral strains. This is supported by the striking observation that in the Ebola Reston strain, which was devoid of pathogenicity in the 3 humans that were infected, there is no equivalent to the major potential selenoprotein gene overlapping the NP gene in Ebola Zaire (ORF1 in Figure 1). In the Ebola Reston NP mRNA, the UGA-rich ORFs are disrupted by non-UGA stop codons, there are fewer UGA codons, no analogous frameshift sites or PKs, and no SECIS element in the 3'-UTR. Thus there is no way that this potential selenoprotein gene could be expressed in Ebola Reston. This is a definite major difference at the gene level between these strains, which have previously been considered to be very close genetically. This potential NP-associated selenoprotein gene is also absent in Marburg virus, which also has a lower mortality rate than Ebola Zaire.

Since the hypothetical selenoprotein overlapping the Ebola Zaire NP gene could only be expressed as an NP fusion protein, it is possible that it could be formed as an NP variant comprising as much as a few percent of the total NP present in virions (more likely a fraction of a percent), in which case it might be possible to detect selenium in Ebola Zaire virions. This percentage would be expected to decrease in late infections if cellular stores of SeC became depleted. In regard to the possible function of such a viral selenoprotein, it is tempting to speculate that it might provide some type of antioxidant protection to the Ebola virions in a rapidly degenerating cellular environment.

Ebola is classified as a "hemorrhagic fever" virus, and produces the characteristic hemorrhaging due to the formation of blood clots ("disseminated intravascular coagulation"), leading to the obstruction and rupture of small blood capillaries. For this reason, counterintuitively, the anticoagulant drug heparin has been used to reduce the bleeding in Ebola patients.

It is very well documented that Se plays a significant role in the regulation of blood clotting via its effects on the thromboxane/prostacyclin ratio. Se has an anti-clotting effect, whereas Se deficiency has a pro-clotting or thrombotic effect.[16] Se deficiency has been associated with thrombosis and even hemorrhaging, which has been documented in a number of animals with severe Se deficiency (often artificially induced), but is almost never seen in humans, probably because such an extreme Se deficiency is rarely attained due to the diversity of human diets.

Thus, the possibility that a rapid depletion of Se due to the formation of viral selenoproteins could be a factor contributing to the severity of the hemorrhagic symptoms is mechanistically very feasible. Our analysis suggests that severe Ebola infections could produce an artificial and extreme Se depletion, resulting in extensive cellular damage due to lipid peroxidation, combined with enhanced thrombosis. This could also contribute to the associated immune deficiency that has been observed in Ebola infections.

To our knowledge, indicators of Se status and lipid peroxidation have never been examined in Ebola patients. However, Se has apparently been used with great success by the Chinese in the palliative treatment of an infectious hemorrhagic fever.[17] Although this did not involve Ebola virus, there are a number of different hemorrhagic fever viruses, and they may share common mechanisms. This example provides yet another reason to expect that pharmacological doses of Se may also have some benefit in Ebola infections.

In the light of the extensive data on the antiviral effects of Se, the association between coxsackievirus and Keshan disease, and the geographic correlation for KS proposed by Ziegler, it is certainly intriguing

that a number of emerging viruses have emanated from these same regions of Africa, that are potentially low in Se. By providing compelling theoretical evidence for the existence of selenoprotein genes in a number of viruses, now including Ebola virus, we have attempted to provide a unifying theoretical model to explain some of these disparate observations.

Taken as a whole, these observations and theoretical findings suggest the basis for a new paradigm in antiviral chemotherapy: the use of nutritional factors to alter the dynamics of the virus-host interaction so as to re-establish a balance in which the natural host defenses can be more effective. In essence, this is the fundamental concept of orthomolecular medicine, so perhaps this is not such a "new" paradigm after all. What *is* new and exciting is that this simple concept may be more widely applicable, to more virulent viral diseases, and a broader range of vitamins and minerals - in this case Se - than previously thought possible.

Finally, because SO_2 reacts with Se compounds in soil, making it more difficult for plants to absorb, it has long been suspected that fossil fuel burning and acid rain may be contributing to a gradual decrease of Se in the food chain.[18] Thus, like deforestation in jungles and rain forests, the resulting alterations in global Se cycling and distribution may be yet another example of how human activity possibly contributes to the emergence of new viral diseases. Ultimately, it is only a deeper understanding of the impact of these human activities on both microbes and their hosts that will empower us to rectify the resulting imbalances in our shared ecosystem.

Acknowledgments

References

1. Becker Y. (1995) Endogenous retroviruses in the human genome - a point of view. *Virus Genes* 9:211-218.
2. Schrauzer G.N., Sacher J. (1994) Selenium in the maintenance and therapy of HIV-infected patients. *Chem-Biol Interact* 91:199-205.
3. Taylor E.W., Ramanathan C.S., Jalluri R.K., Nadimpalli R.G. (1994) A basis for new approaches to the chemotherapy of AIDS: novel genes in HIV-1 potentially encode selenoproteins expressed by ribosomal frameshifting and termination suppression. *J Med Chem* 37:2637-2654.
4. Sappey C., Legrand-Poels S., Best-Belpomme M., Favier A., Rentier B., Piette J. (1994) Stimulation of glutathione peroxidase activity decreases HIV type 1 activation after oxidative stress. *AIDS Res Human Retrovir* 10:1451-1461.
5. Taylor E.W., Ramanathan C.S., Nadimpalli R.G., Schinazi R.F. (1995) Do some viruses encode selenoproteins? Evaluation of the theory in the light of current theoretical, experimental and clinical data. *Antiviral Res* 26:A271-86.
6. Bai J., Wu S., Ge K., Deng X., Su C. (1980) The combined effect of selenium deficiency and viral infection on the myocardium of mice. *Acta Acad Med Sin* 2:29-31.
7. Beck M.A., Shi Q., Morris V.C., Levander O.A. (1995) Rapid genomic evolution of a non-virulent Coxsackievirus B3 in selenium-deficient mice results in selection of identical virulent isolates. *Nature Med* 1:433-436.
8. Ziegler J.L. (1993) Endemic Kaposi's sarcoma in Africa and local volcanic soils. *Lancet* 342:1348-1351.
9. Vanderpas J.B., Contempre B., Duale N.L., Goossens W., Bebe N., Thorpe R., Ntambue K., Dumont J., Thilly C.H., Diplock A.T. (1990) Iodine and selenium deficiency associated with cretinism in northern Zaire. *Am J Clin Nutr* 52:1083-1086.
10. Berry M.J., Larsen P.R. (1993) Recognition of UGA as a selenocysteine codon in eukaryotes: a review of recent progress. *Biochem Soc Trans* 21:827-32.
11. Taylor E.W., Ramanathan C.S., Nadimpalli R.G. (1995) A general approach to predicting potential new genes in nucleic acid sequences: application to the human immunodeficiency virus. In: *Proceedings of the First World Congress on Computational Biomedicine, Public Health and Biotechnology*. Austin, TX: World Scientific, Tokyo, in press.
12. Taylor E.W. (1995) Selenium and cellular immunity: evidence that selenoproteins may be encoded in the +1 reading frame over-

lapping the human CD4, CD8 and HLA-DR genes. *Biol Trace Elem Res* 49:85-95.

13. Cubitt B., Oldstone C., Valcarcel J., de la Torre J.C. (1994) RNA splicing contributes to the generation of mature mRNAs of Borna disease virus, a non-segmented negative strand RNA virus. *Virus Res* 34:69-79.

14. Johnson E.D., Gonzales J.P., Georges A. (1993) Filovirus activity among selected ethnic groups inhabiting the tropical forest of equatorial Africa. *Trans R Soc Trop Med Hyg* 87:536-538.

15. Becker S., Feldmann H., Will C., Slenczka W. (1992) Evidence for occurrence of filovirus antibodies in humans and imported monkies: do subclinical filovirus infections occur worldwide? *Med Microbiol Immunol Berl* 181:43-55.

16. Meydani M. (1992) Modulation of the platelet thromboxane A2 and aortic prostacyclin synthesis by dietary selenium and vitamin E. *Biol Trace Elem Res* 33:79-86.

17. Hou J.C., Jang Z.F., He, Z.F. (1993) Inhibitory effect of selenite on complement activation and its clinical significance. *Chung Hua I Hsueh Tsa Chih* 73:645-646.

18. Frost D.V. (1987) Why the level of selenium in the food chain appears to be decreasing. In: *Selenium in Biology and Medicine,* G.F. Combs, Jr., J.E. Spallholz, O.A. Levander, J.E. Oldfield, Eds. AVI Van Nostrand, New York. Part A, pp. 534-547.

19. Sanchez A., Kiley M.P., Holloway B.P., Aupern D.D. (1993) Sequence analysis of the Ebola virus genome: organization, genetic elements, and comparison with the genome of Marburg virus. *Virus Res* 29:215-240.

20. Zuker M., Steigler P. (1981) Optimal computer folding of large RNA sequences using thermodynamics and auxillary information. *Nucl Acids Res* 9:133-148.