Prediction of HIV vaccine

SIR—We read with interest the recent News and Views article[1] on the problems of producing an effective vaccine to HIV, the virus causing AIDS (acquired immune deficiency syndrome).

For some time we have been developing a novel computational approach for analysing protein sequences and predicting regions suitable for use as artificial vaccines. The program appears capable of recognizing not only polar sites, but also epitopic segments of predominantly hydrophobic character. The algorithms used will be published in full elsewhere, but, briefly, the method draws on information from secondary structure predictions, hydrophobicity analyses, scans for patterns of hydrophobic and hydrophilic residues consistent with loop and other structures, and the study of homologous sequences.

The procedure works well when tested on the available database of epitopic sites. Recently, the peptide of a steroid hormone receptor most strongly predicted to be a continuous epitope has been conjugated with a protein carrier and found to raise antibodies that recognize, with extremely high specificity, the nuclei of cells expressing these receptors. Based on this work, and as attachment of a carrier molecule is likely to be necessary for the immunogenicity of peptides and the conjugation procedure used may be important, we recommend that a short C-terminal extension ending in a cysteine residue be made, through which the peptide can be linked to the protein carrier.

We would be most interested to know if the regions of the HIV envelope proteins that we predict to be highly antigenic (Table 1) have been tried as synthetic vaccines, and if so, whether they are effective. This is particularly of interest as most of the strongly predicted epitopes in these proteins contain potential glycosylation sites, especially in the putative extracellular domain of the transmembrane protein, whereas our predicted epitopes are a reduced set devoid of potential glycosylation sites, five of which correspond to predicted surface loops.

We are willing to use this program and other more elaborate methods[2,3] to predict the antigenic regions of other proteins of possible interest in the production of an AIDS vaccine and to make the results freely available, if it is felt that they may be of value.

Table 1 Antigenic sites predicted in the HIV *env* gene product

	Site	Sequence
Transmembrane protein	1	EESQNQQEKNEQE
Exterior protein	2	VPTDPNPQE
	3	PKVSFEPIPIHY
	4	KQSSGGDPE
	5	YAPPISGQIR
	6	RPGGGDMRDN
	7	PTKAKRRVVQREKR

Sequences are given in IUPAC one-letter code. These predicted sites are identical in the three published sequences studied (4–6). It may be worth extending Site 3 to TQACPKVSFEPIPIHYCAPA and closing the cysteines intramolecularly.

BARRY ROBSON
ROBERT V. FISHLEIGH
Department of Biochemistry and Molecular Biology,

CHRISTOPHER A. MORRISON
Immunology Division, Department of Cell and Structural Biology, University of Manchester, Oxford Road, Manchester M13 9PT, UK

1. Newmark, P. *Nature* **324**, 304 (1986).
2. Robson, B. *et al. Int. J. Peptide Protein Res.* **25**, 1 (1985).
3. Robson, B. Platt, E. *J. molec. Biol.* **188**, 259 (1986).
4. Ratner, L. *et al. Nature* **313**, 277 (1985).
5. Meusing, M.A. *et al. Nature* **313**, 450 (1985).
6. Wain-Hobson, S. *et al. Cell* **40**, 9 (1985).

Protein structure and the specific heat of water

SIR—Studying the thermodynamics of the polymerization of collagen into fibrils, we have observed that the concentration of soluble collagen in equilibrium with fibrils is a minimum at 34.3 °C. Puzzled by the possible biological significance of this observation, we read with interest J. Paul's letter (*Nature* **323**, 300; 1986), in which he brought to our attention the fact that the specific heat capacity of water is minimal at 34–35 °C. To account for the apparent correlation between the temperature at which the specific heat capacity of water is minimal and the temperature at which higher organisms have evolved, he suggests that homoiothermic animals having body temperatures around 35 °C have been selected because of their high efficiency at generating or dissipating heat in order to maintain their body temperature. Although the difference in specific heat capacity of water between 20°C (4.1819 J g^{-1} $°C^{-1}$) and 35 °C (4.1782 J g^{-1} $°C^{-1}$) is very small, it remains an attractive proposition that the evolution of warm-blooded animals and the chemical properties of water are related.

Is it feasible to assume that a change in the thermal properties of water is a consequence of a change in the structural organization of water molecules? If so, an alternative interpretation of J. Paul's hypothesis is that the structure rather than the thermal properties of water at 35 °C was the basis for evolutionary development of warm-blooded animals. It is interesting to postulate that the structure of water at 35 °C is the most permissive to the evolution of the wide variety of biological structures that exist today.

Turning to our data on the solubility of collagen, 34.3 °C marks an abrupt transition in the thermodynamic properties of collagen in solution, and we would like to learn about other proteins and biological molecules whose structure/function appears to be responsive to the peculiar structure of water at 35 °C.

KARL KADLER
DARWIN J. PROCKOP
Jefferson Institute of Molecular Medicine, Jefferson Medical College, Thomas Jefferson University, Philadelphia, Pennsylvania 19107, USA

EMBL database offer

SIR—The identification of newly sequenced proteins or parts thereof requires searches of protein sequence databases. May we inform your readers that we have undertaken the translation into protein sequences of the nucleotide sequences in the nucleotide sequence data collection of the European Molecular Biology Laboratory. Our database consists of 5,139 entries with a total of 952,078 amino acids. A tape with a flat file of our translated sequence data collection is available upon request.

The main features of our approach include the inspection of the features table, use of different genetic codes, splicing of exons as far as possible and control of proper reading frames by stop-codon test. In case of conflict, the entry has been omitted, as are sequences shorter than seven amino acids. Splicing was not undertaken if neighbouring exons appear in different entries.

RAINER STULICH
Zentralinstitut für Molekularbiologie, Akademie der Wissenschaft der DDR, Robert-Rössle-Strasse, 1115 Berlin-Buch, DDR

Mendel was no fraud

SIR—As a result of a paper by Fisher[1], and early discussion of the topic, the view has become widespread (for example, ref. 2) among those who have not bothered to read the biological literature, that Mendel falsified some of his data. Several analyses[3–6] do not confirm this view, and results like Mendel's have even been found by others[3]. I wonder if some of the similar claims about other historic figures would also dissolve if suitably scrutinized.

LEIGH M. VAN VALEN
Biology Department (Whitman), University of Chicago, Chicago, Illinois, 60637, USA

1. Fisher, R.A. *Ann. Sci.* **1**, 115–137 (1936).
2. Joyce, C.R.B. *Nature* **324**, 181 (1986).
3. Piegorsch, W.W. *Nature* **256**, 206 (1975).
4. Blixt, S. *Nature* **256**, 206 (1975).
5. Douglas, L. & Novitski, E. *Heredity* **38**, 253–257 (1977).
6. Monaghan, P. & Corcos, A. *J. Hered.* **76**, 307–309 (1985).