Available online at www.sciencedirect.com

Journal of Theoretical Biology 228 (2004) 459–465

Journal of Theoretical Biology

www.elsevier.com/locate/jtbi

Proteomic analyser with applications to diagnostics and vaccines

Geoffrey W. Hoffmann*

Department of Physics and Astronomy, University of British Columbia, 6224 Agricultural Road, Vancouver, BC, V6T1Z1 Canada

Received 29 July 2003; received in revised form 19 February 2004; accepted 20 February 2004

Abstract

This paper describes a method for proteomic analysis with applications to diagnostics and vaccines. A panel of N ($\geqslant 1$) reagents called $X(j)$, with $j = 1$ to N, is used. The binding strength of each of the $X(j)$ reagents to each other is measured, for example by an ELISA assay, giving an $N \times N$ matrix $\boldsymbol{K}$. The matrix $\boldsymbol{K}$ is used to define another set of N reagents called $Y(j)$, with $j = 1$ to N, each of which is a linear combination of the $X(j)$ reagents and each of which is tailored to be complementary to one of the $X(j)$ reagents. Each of the N pairs of reagents $X(j)$ and $Y(j)$ defines an axis in an N-dimensional shape space. The definition of these axes facilitates proteomic analysis of diverse biological samples, for example, mixtures of proteins such as serum samples or T cell extracts. A method for defining and measuring similarity between pairs of biological samples and between sets of biological samples in the context of the set of N reagent pairs is described. This leads to methods for using the N reagent pairs in the diagnosis of diseases and in the formulation of preventive and therapeutic vaccines. The relationship of this work to previous research on shape space is discussed.

Keywords: Antibodies; Complementary pairs of reagents; Shape space; N dimensions

1. Introduction

Immune system V region proteomics is important because the immune system V region repertoire is changed or "skewed" in many diseases, including cancer, autoimmune diseases and graft versus host disease (Pilch et al., 2002; Wucherpfennig et al., 1992; Imberti et al., 1991; Smith et al., 1995; Rebai et al., 1994; Ebling et al., 1988). This skewing opens possibilities for innovations in diagnostic testing. It is also possible that some diseases can be prevented and/or treated if the skewing is sufficiently characterized and counteracted by a suitable perturbation, namely an immunization precisely tailored to reverse the particular skewing.

A full proteomic description of the specific (V region) components of a particular immune system would constitute a list of the concentrations of each of millions of lymphocytes, antibodies and specific T cell factors, together with the isotypes, amino acid sequences and three-dimensional structures of the corresponding V regions. Even with the spectacular advances that are currently being made in proteomics, such a description is not a realistic goal, and even if it were, achieving it may not be particularly useful. Each individual has his or her own set of V regions, due to different V region genes, different MHC (major histocompatability complex) genes that affect the expressed repertoire of T cells, and different histories of exposure to a wide range of antigens. Furthermore, different somatic mutations in each individual contribute significantly to the generation of the V region repertoire.

One recent approach to diagnostic proteomics is the SELDI-MS technology (Surface-Enhanced Laser Desorption/Ionization-Mass Spectrometry) coupled to pattern recognition software. This is not suited for V region proteomics because it is based on mass differences between proteins, and while (for example) IgG antibodies with different V regions can have slightly different masses, each person has a unique spectrum of antibodies. On the other hand, ELISA-based protein array technologies are becoming available that are suitable for V region proteomics as described in this paper.

I here describe a method for proteomic analysis that builds on our previously defined concept of serological distance coefficients (Hoffmann and Tufaro, 1989). In the earlier work, experimentally measurable similarity coefficients $S[A,B|C]$ specify the extent to which a pair

*Fax: +1-604-822-5324.
E-mail address: hoffmann@physics.ubc.ca (G.W. Hoffmann).

0022-5193/$ - see front matter © 2004 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2004.02.011

of substances, A and B, are similar in the context of a diverse reagent, C. The definition of $S[A,B|C]$ is the fraction of C that binds both A and B divided by the sum of (i) the fraction that bind A but not B, (ii) the fraction that binds B but not A and (iii) the fraction that binds both A and B. The value of $S[A,B|C]$ is then necessarily a number between zero and one. This definition was applied (conceptually) also to similarities between complex mixtures of substances, such as the antibodies of two serum samples, A and B. A "distance coefficient" $D[A,B|C]$ between two sera, A and B, in the context of C, was defined as one minus the similarity coefficient in the same context. Methods for the experimental measurement of these coefficients and their possible use in the diagnosis and prognosis of disease conditions were described.

2. Measurement of similarity using many reagents

The improved method utilizes a number N ($>>1$) of reagents, rather than a single diverse reagent. Each reagent can be an individual substance, for example a protein, possibly an antibody, or a mixture of substances. This produces a much larger data set than using a single diverse reagent, but it is still a very small set compared with the complete listing of V regions and their concentrations mentioned in the second paragraph above. The result is a measure of similarity based on an N-dimensional shape space, that is a more powerful tool for applications to diagnostics and vaccines.

The concept of an N-dimensional shape space has been discussed by Perelson and Oster (1979), and a formulation that permits an experimental determination of the dimensionality of a shape space has been described by Lapedes and Farber (2001). It will become clear that the N-dimensional shape space of this paper is different from both of these; I compare the different approaches near the end of the paper.

We denote the N reagents by $X(j)$ (with $j = 1$ to N), and use them most simply at a uniform concentration C_0. We measure the binding (relative affinity) of each of these reagents to each other using, for example, an ELISA assay. This produces a matrix $\boldsymbol{K}$ with elements K_{jk} ($j = 1$ to N, $k = 1$ to N). Such $\boldsymbol{K}$ matrices for IgM antibodies have been described by Holmberg et al. (1989) and Kearney et al. (1987).

We next define N additional reagents, that we denote as $Y(j)$, ($j = 1, N$). Each of the $Y(j)$ reagents is made up of a linear combination of the $X(j)$ reagents, with the amount of the kth component being proportional to K_{jk}. Those components that have strong binding to $X(j)$ are present in $Y(j)$ at a high concentration, while those with little or no binding are included at a low or zero concentration. For a given value of j, $X(j)$ and $Y(j)$ are complementary to each other, and together the pair defines an axis in the N-dimensional shape space. There are N such pairs, that together define the N axes of an N-dimensional shape space.

There are two possible ways of normalizing the concentrations of the $Y(j)$ reagents to establish a symmetry between the $X(j)$ reagents and the $Y(j)$ reagents. One is to make the total concentration of the components of $Y(j)$ such that the binding signal obtained for $Y(j)$ binding to $X(j)$ (in the case of an ELISA assay, with $Y(j)$ binding to $X(j)$ on the plate), in the linear range of the assay, is equal to the converse binding signal (binding of $X(j)$ to $Y(j)$, also in the linear range of the assay). The other method is to simply set the total concentration of each $Y(j)$ equal to C_0. The former method leads to the definition of a convenient virtual N-dimensional origin for the shape space, namely a hypothetical sample to which $X(j)$ and $Y(j)$ bind equally in the assay, for all values of j.

We measure the binding of each $X(j)$ reagent ($j = 1, N$) to each $Y(k)$ ($k = 1, N$) reagent. This produces the $N \times N$ matrix $\boldsymbol{J}$ with elements J_{jk}. On the basis of mass-action, and subject to linearity of the assay, the expected relative values of the elements of $\boldsymbol{J}$ are

$$J_{jk} = \sum_{l=1}^{N} K_{jl} K_{lk}. \tag{1}$$

The diagonal elements of this matrix ($j = k$) specify the level of binding between the reagents $X(j)$ and $Y(j)$, that have been specifically tailored to be complementary to each other. Hence their mutual binding will produce a strong signal, while there will be relatively weak signals for off–diagonal terms. Thus $\boldsymbol{J}$ is an approximately diagonal matrix.

We now consider a set of biological samples obtained from M individuals. These samples may be, for example but not exclusively, serum, T-lymphocyte extracts, saliva or urine. We use the index i for the samples, so $i = 1$ to M. We measure the binding of each of the reagents $X(j)$ ($j = 1$ to N) to each of the samples, again using for example an ELISA assay. For each sample i we thus obtain N absorbance values $A_{iX(j)}$. Together all the elements $A_{iX(j)}$ constitute an $M \times N$ matrix that we call $\boldsymbol{A_X}$.

We repeat this process using the set of N complementary reagents, $Y(j)$. We measure the binding of each $Y(j)$ reagent to each sample i, to obtain the matrix $\boldsymbol{A_Y}$ consisting of the elements $A_{iY(j)}$. Subject to the assay being linear, we can however also compute expected relative values of $A_{iY(j)}$ using the product of the matrix $\boldsymbol{A_X}$ and the matrix $\boldsymbol{K}$:

$$A_{iY(j)}(\text{expected}) \propto \sum_{k=1}^{N} A_{iX(k)} K_{kj}. \tag{2}$$

The results of these summations are then normalized such that the average of the computed $A_{iY(j)}$ matrix elements is the same as the average of the $A_{iX(j)}$ matrix elements. Hence, remarkably, we can have the benefit of an analysis in terms of the N $X(j)/Y(j)$ axes in shape space without needing to prepare the $Y(j)$ reagents, and without making measurements on all our samples using them! This is because the $\boldsymbol{A_X}$ and $\boldsymbol{K}$ matrices already contain all the physical information. On the other hand, by including the actual measurement of $A_{iY(j)}$ using the $Y(j)$ reagents we have a technology that is more robust, because the individual measurements are then automatically screened for self-consistency. This is analogous to sequencing both strands of DNA, in which case any sequencing errors are immediately revealed, since one sequence predicts the other.

The difference $A_{iX(j)} - A_{iY(j)}$ is a coordinate for the sample i on the $X(j)- Y(j)$ axis, that can be either positive or negative. It specifies whether the sample i is more $X(j)$-like $\left(A_{iX(j)} - A_{iY(j)} < 0\right)$ or more $Y(j)$-like $\left(A_{iX(j)} - A_{iY(j)} > 0\right)$. There are N such coordinates for each sample. Fig. 1 illustrates this for just two of the N coordinates.

It is expected that the N-dimensional coordinates for young, healthy individuals form one cluster (Hoffmann submitted) while the points for individuals with various diseases cluster around other, disease-specific points. Let a subset of the M samples be derived, for example, from people who have been classified to have a given disease (the "D set", consisting of, say, M_D samples) and let another subset be from healthy individuals (the "H set", consisting of M_H samples). We obtain $M_H N$ ELISA absorbance results $A_{H(i)X(j)}$ for the healthy group, where i is an index for the sample that goes from 1 to M_H, and j is the index for the reagents $X(j)$ that goes from 1 to N. We likewise obtain $M_D N$ results $A_{D(i)X(j)}$ from the disease group, where i goes from 1 to M_D.

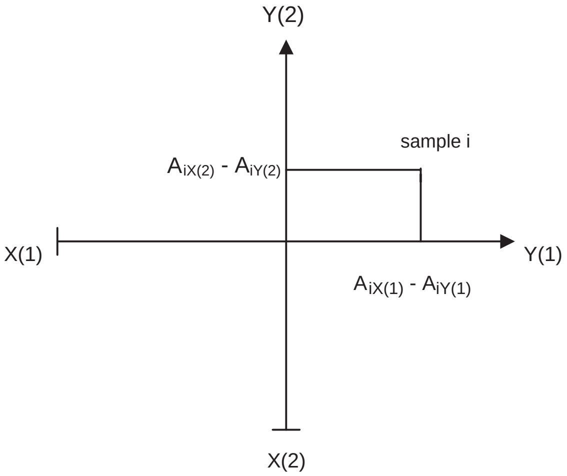

Fig. 1. The reagents $X(1)$ and $Y(1)$ are complementary to each other and define an axis in shape space, and the reagents $X(2)$ and $Y(2)$ define a second axis. The coordinates of sample i are determined by measuring the amount of binding of the reagents $X(1)$, $Y(1)$, $X(2)$ and $Y(2)$ to the sample. Here sample i binds more to $X(1)$ than $Y(1)$ and more to $X(2)$ than $Y(2)$. Hence it is more similar to $Y(1)$ than to $X(1)$ and more similar to $Y(2)$ than to $X(2)$.

For each value of j we average the values of $A_{H(i)X(j)}$ for $i = 1$ to M_H:

$$A_{HavX(j)} = \frac{1}{M_H}\sum_{i=1}^{M_H} A_{H(i)X(j)}, \; j = 1, N. \tag{3}$$

We likewise average the values of $A_{D(i)X(j)}$:

$$A_{D_{av}X(j)} = \frac{1}{M_D}\sum_{i=1}^{M_D} A_{D(i)X(j)}, \; j = 1, N. \tag{4}$$

Similarly, using the $Y(j)$ reagents we obtain the average values

$$A_{H_{av}Y(j)} = \frac{1}{M_H}\sum_{i=1}^{M_H} A_{H(i)Y(j)}, \; j = 1, N. \tag{5}$$

and

$$A_{D_{av}Y(j)} = \frac{1}{M_D}\sum_{i=1}^{M_D} A_{D(i)Y(j)}, \; j = 1, N. \tag{6}$$

Now we consider a set of M_U samples that are unknown in that they are from individuals that may or may not have the disease. We measure the binding of each of the N reagents $X(j)$ to each of the $U(i)$ samples ($i = 1$ to M_U, $j = 1$ to N), giving the results $A_{U(i)X(j)}$. We also measure and/or compute the binding of each of the N reagents $Y(j)$ to each sample, giving the values $A_{U(i)Y(j)}$. One measure of the similarity of sample $U(i)$ to the average of the healthy samples ("H_{av}"), in the context of just one $X(j)/Y(j)$ pair of reagents, is then

$$S\left[U(i), H_{av}|X(j)/Y(j)\right] = \left(A_{U(i)X(j)} - A_{U(i)Y(i)}\right) \tag{7}$$

The corresponding similarity of $U(i)$ to H_{av} in the context of the complete set of the N reagent pairs $X(j)/Y(j)$ is obtained by summing over j:

$$S[U(i), H_{av}|N_{X(j)/Y(j)}] = \sum_{j=1}^{N}\left(A_{U(i)X(j)} - A_{U(i)Y(j)}\right) \times \left(A_{H_{av}X(j)} - A_{H_{av}Y(j)}\right). \tag{8}$$

The similarity of sample $U(i)$ to the average of the disease set of samples ("D_{av}") would then be likewise

$$S[U(i), D_{av}|N_{X(j)/Y(j)}] = \sum_{j=1}^{N}\left(A_{U(i)X(j)} - A_{U(i)Y(j)}\right) \times \left(A_{D_{av}X(j)} - A_{D_{av}Y(j)}\right). \tag{9}$$

These measures of similarity or other measures of clustering in the N-dimensional space can then be used as the basis for a diagnosis.

The same set of reagents $X(j)$ and $Y(j)$, $j = 1$ to N, can be used for diagnosis of multiple diseases. All that is additionally needed is a set of samples for each disease, from which the values of $A_{D_{av}X(j)}$ and $A_{D_{av}Y(j)}$ ($j = 1$ to N) for each disease are determined.

So far we have included all of the N reagents in the analysis. We do not need to do this. For the diagnosis of a particular disease or condition we can instead include only those reagents that optimize specificity, sensitivity and simplicity, either individually or jointly.

An advantage of this diagnostic method is that it is based on N-dimensional shape space, with $N>>1$, in contrast to the two-dimensional map of the previously published serological distance coefficient diagnostic method (Hoffmann and Tufuro, 1989). N-dimensional vectors with $N>>1$ contain much more precise information than two-dimensional vectors. The method consequently is expected to provide more specific diagnoses.

Another advantage of this method over the precursor method (Hoffmann and Tufaro, 1989) is that it eliminates the need to do absorptions, which is the most labour-intensive part of that earlier method.

3. An example: diagnosis of SARS

We are currently faced with an important new disease, namely SARS. A corona virus has been identified as the culprit,[1] but in Canada only about 50% of confirmed SARS patients were found to be positive for direct detection of the virus, namely polymerase chain reaction or virus culture (Frank Plummer, personal communication). Ultimately, about 95% of confirmed cases developed antibody to SARS coronavirus at 4 weeks. This raises the question of whether SARS can be caused by a proteomic stimulus similar to that caused by the virus.

Several years ago there was a similar situation with AIDS and HIV, but then cases of the syndrome that were negative for HIV were defined as "idiopathic CD4+ T-lymphocytopenia", rather than AIDS (Smith et al., 1993; Ho et al., 1993; Spirat et al., 1993; Duncan et al., 1993). The definition of AIDS was narrowed to include only those people who are positive for HIV (Morbidity and Mortality Weekly Report, 1999).

The method described here may be useful for identifying any additional causes of SARS. The SARS corona virus may produce one form of repertoire skewing, while other agents may induce a similar but distinct skewing. The method described may thus enable a diagnosis for SARS that is independent of the detection of a corona virus or any other virus.

4. Application to vaccine formulation

In addition to its diagnostic role, the formalism and method developed here is useful for designing and evaluating highly specific multi-component proteomic perturbations to the immune system, that function as preventive and/or therapeutic vaccines.

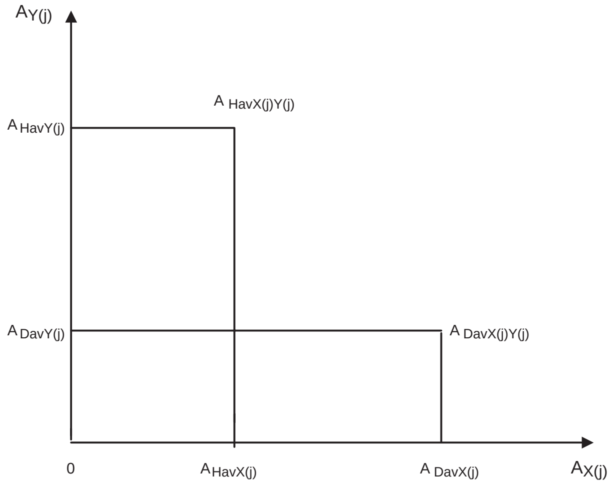

Fig. 2. Average absorbances $A_{H_{av}X(j)}, A_{D_{av}X(j)}, A_{D_{av}Y(j)}$, and $A_{D_{av}Y(j)}$ plotted on the $A_{X(j)}$ and $A_{Y(j)}$ axes. The average disease state, $A_{D_{av}X(j)Y(j)}$, and the average healthy state, $A_{H_{av}X(j)Y(j)}$, from the perspective of the $X(j)$ and $Y(j)$ pair of reagents is shown. (Note that this is a different perspective on the N-dimensional shape space from that of Fig. 1.)

For a single pair of reagents $X(j)$ and $Y(j)$ and a given disease D, we can plot the values $A_{D_{av}X(j)}, A_{H_{av}X(j)}$, $A_{D_{av}Y(j)}$ and $A_{H_{av}Y(j)}$ on the axes $A_{X(j)}$ and $A_{Y(j)}$ as shown in Fig. 2. Hence the points labelled $A_{D_{av}X(j)Y(j)}$ and $A_{H_{av}X(j)Y(j)}$ are defined for the average disease and average healthy states, respectively. We need a stimulus that (firstly for this pair of reagents), moves the system from $A_{D_{av}X(j)Y(j)}$ towards $A_{H_{av}X(j)Y(j)}$. An appropriate stimulus consists of two components, one for motion from right to left (for example, Fig. 2) and one for motion in the vertical direction. The reagent $Y(j)$ stimulates the complementary $X(j)$ cells, and hence moves the system along the $X(j)$ axis (the horizontal axis). The reagent $X(j)$ stimulates $Y(j)$ cells, and moves the system in the vertical direction. We next need to determine the appropriate concentrations of the reagents.

At first sight, we might choose a concentration of $Y(j)$ proportional to $A_{H_{av}X(j)} - A_{D_{av}X(j)}$ and a concentration of $X(j)$ proportional to $A_{H_{av}Y(j)} - A_{D_{av}Y(j)}$. A problem with this is however that some such tentative relative concentrations are negative, and we cannot include a negative amount of a reagent in the formulation of a vaccine. This problem can be resolved by substituting a positive amount of the reagent $X(j)$ for any computed negative amount of reagent $Y(j)$ [since $X(j)$ is complementary to $Y(j)$], and likewise a positive amount of $Y(j)$ for any negative amount of $X(j)$. The relative amount of $X(j)$ needed in the vaccine, from the perspective of the $X(j)/Y(j)$ pair of reagents, will be denoted by $R[X(j)]$ and

[1] http://www.who.int/mediacentre/releases/2003/pr31/en/

is given by

$$R[X(j)] = \left[A_{H_{av}Y(j)} - A_{D_{av}Y(j)}\right]\left[\frac{1+\operatorname{sign}\left(A_{H_{av}Y(j)} - A_{D_{av}Y(j)}\right)}{2}\right] + \left[A_{H_{av}X(j)} - A_{D_{av}X(j)}\right]\left[\frac{1-\operatorname{sign}(A_{H_{av}X(j)} - A_{D_{av}X(j)})}{2}\right], \tag{12}$$

where sign $x = 1$ for $x > 0$, and sign $x = -1$ for $x < 0$. Similarly, the relative amount of $Y(j)$ in the vaccine, denoted by $R[Y(j)]$, is given by

$$R[Y(j)] = \left[A_{H_{av}X(j)} - A_{D_{av}X(j)}\right]\left[\frac{1+\operatorname{sign}\left(A_{H_{av}X(j)} - A_{D_{av}X(j)}\right)}{2}\right] + \left[A_{H_{av}Y(j)} - A_{D_{av}Y(j)}\right]\left[\frac{1-\operatorname{sign}\left(A_{H_{av}Y(j)} - A_{D_{av}Y(j)}\right)}{2}\right]. \tag{13}$$

In the example of Fig. 2, both components in the expression for $R[X(j)]$ are positive, and both components in the expression for $R[Y(j)]$ are zero.

The total specific component of the vaccine is then obtained by summing over j. This is thus a method for formulating an immunogenic (vaccine) stimulus using the base set of N reagents. We then still have a single undetermined parameter, namely the ratio of the actual total concentration needed in the vaccine to the numerical values as computed. This parameter can be determined empirically by titration.

5. Application to personally customised vaccines

The preceding description is in terms of vaccines suitable for a particular disease and for many people. Such vaccines are applicable especially as preventive immunisations. An individual patient may however have skewing that is unique to that patient. In such cases a personally tailored approach may be beneficial. One method is to replace the average absorbance values $A_{D_{av}X(j)}$ and $A_{D_{av}Y(j)}$ with the patient's absorbance values $A_{D(i)X(j)}$ and $A_{D(i)Y(j)}$, respectively, in Eq. (12) and (13). Another step in the direction of personally tailored vaccines is to replace $A_{H_{av}X(j)}$ with $A_{H(i)X(j)}$ and $A_{H_{av}Y(j)}$ with $A_{H(i)Y(j)}$, in Eq. (12) and (13), where $A_{H(i)X(j)}$ and $A_{H(i)Y(j)}$ are obtained using historical samples from when the individual i was healthy. Hence N-dimensional perturbations can be tailored to inhibit and/or reverse pathological skewing of V region repertoires at the levels of both populations and individuals.

6. Other Applications

While the concept of using $X(j)/Y(j)$ axis coordinates emerged in the context of the V region network of interactions of the immune system, this method can also be used more generally to characterise and monitor broader proteomic changes in an individual.

Similarity coefficients as defined here can be expected to be a powerful tool for gaining an improved understanding of the idiotypic network. The idiotypic network is the network of V regions that recognise each other (in addition to foreign substances) and is believed to play a central role in the regulation of the immune system (Hoffmann et al., 1988).

7. Criteria for the selection of the N reagents

The N reagents $X(j)$ need to be substances with reproducible, stable, diverse three-dimensional shapes. They may include for example monoclonal antibodies and/or other proteins from one or more species. One possibility is that all of the $X(j)$ reagents are monoclonal antibodies, for example all of the IgG class. This would create a symmetry in the system that allows for essentially unlimited diversity in shapes, while ensuring that all the reagents have a similar intrinsic ability to cross-link complementary receptors. (IgG antibodies have two V regions, and thus a single IgG molecule is able to cross-link complementary receptors.) This is relevant for applications to vaccine formulation, since cross-linking of receptors is believed to be the mechanism for the specific stimulation of lymphocytes. This would be preferable to using proteins with varying degrees of polymerization, some of which would be much stronger immunogenic stimuli than others.

Traditionally immunologists have focussed on high affinity interactions, such that an antibody is "specific for" (has a high affinity for) only a very small number of substances. If we include low affinity interactions, each antibody interacts with a much larger fraction of substances, including other antibodies. ELISA technology provides the option of measuring relatively low-affinity interactions, and in order to define directions in shape space precisely, we would prefer that the matrices K and A be not too sparse. This can be achieved by adjusting the conditions of the ELISA such that low-affinity interactions fall within the dynamic range of the assay.

Another possibility for the choice of the $X(j)$ reagents is to use exclusively soluble proteins of a size comparable to each other and without any repeating determinants, again ensuring that they are of similar immunogenicity. The focus of the method is on three-dimensional shapes, rather than on sequences (as in RNA or DNA nucleotide sequences). The method does not require any of the $X(j)$ reagents to be proteins, but proteins do constitute a convenient library of diverse shapes. We would again be interested in including low affinity interactions.

8. The specificity of the method and the value of N

The specificity of the method depends on the value of N and the accuracy of the assay method. If the values of $A_{iX(j)} - A_{iY(j)}$ are obtained simply as Boolean numbers, when $N = 20$ the shape space would have 2^{20} distinguishable points. With an ELISA assay the results are however analogue rather than Boolean, and each coordinate might have 10 distinguishable values. Then already with $N = 5$ the shape space would have 10^5 distinguishable points, and with $N = 20$ there would be 10^{20} distinguishable points. This theoretical remarkable resolution is expected to be important for applications to diagnostics and vaccines. It can be tested in experiments in which known mixtures of the X(j) reagents themselves are analysed using the method, and the experimentally determined coordinates are compared with the theoretical predictions.

9. Relationship to some other work on shape space

In their work on shape space Perelson and Oster estimated limits on the size of the repertoire that is needed to reliably respond to antigen, and were also concerned with the necessity not to make antibodies to self. The focus of the theory is the relationship between the volume of shape space covered by the reactivity of a single antibody and the total volume of shape space, and hence the number of different antibodies needed to reliably cover shape space. The main parameters in the theory are the dimension of their shape space N, the size of the repertoire N_{Ab}, and the distance in shape space within which an antibody can bind all antigens, ε. These parameters are interdependent, and the theory did not include a method for measuring N or ε. On the basis of literature values of the frequencies of antigen specific cells, they estimated that N could not be more than 5 or 10.

Lapedes and Farber described a shape space for which a dimensionality can be determined using experimental data. They used MN experimental data points, namely the binding of M antigens to N antisera, to map the shapes of each of the antigens and sera to points in a D–dimensional shape space (Lapedes and Farber, 2001). The method involves minimizing a function of the experimental data points and the space shape coordinates. The relationship of this shape space to that of Perelson and Oster is not clear to me, since it does not have ε as a parameter. They found D to have a value of 4 to 5.

The earlier papers are based on the premise that there is an intrinsic dimensionality for shape space relevant to immunological recognition. This premise plays no role in our theory, which is a distinct formalism.

Our theory is an extension of and improvement on our earlier paper on serological distance coefficients, in which similarity was defined in the context of a single diverse reagent (Hoffmann and Tufaro, 1989). Here we define similarity in the context of an approximately orthogonal set of N axes in shape space. In immunology context is of over-riding importance, since antibodies are made in the context of a set of self antigens, T cells and other antibodies. The dimension N of the space is something we are free to choose, and the choice determines the level of specificity. The larger the value of N, the higher the specificity of the method. The theory leads to new methods for diagnostics and vaccines.

Acknowledgements

I thank Robert Forsyth and Graham Clowes for helpful comments on the manuscript.

References

Duncan, R.A., von Reyn, C.F., Alliegro, G.M., Toossi, Z., Sugar, A.M., Levitz, S.M., 1993. Idiopathic CD4+ T-lymphocytopenia—four patients with opportunistic infections and no evidence of HIV infection. N. Engl. J. Med. 328, 393–398.

Ebling, F.M., Ando, D.G., Panosian-Sahakian, N., Kalunian, K.C., Hahn, B.H., 1988. Idiotypic spreading promotes the production of pathogenic autoantibodies. J. Autoimmun. 1 (1), 47–61.

Hoffmann, G.W. Is the immune system a self-symmetrizing system? Manuscript submitted for publication.

Hoffmann, G.W., Tufaro, F., 1989. Serological distance coefficients. Immunol. Lett. 22, 83–90.

Holmberg, D., Andersson, Å., Carlsson, L., Forsgren, S., 1989. Establishment and functional implications of B-cell connectivity. Immunol. Rev. 110, 89–103.

Ho, D.D., Cao, Y., Zhu, T., Farthing, C., Wang, N., Gu, G., Schooley, R.T., Daar, E.S., 1993. Idiopathic CD4+ T-lymphocytopenia—immunodeficiency without evidence of HIV infection. N. Engl. J. Med. 328, 380–385.

Hoffmann, G.W., Kion, T.A., Forsyth, R.B., Soga, K.G., Cooper-Willis, A., 1988. The N–Dimensional Network. In: Perelson, A.S. (Ed.), Theoretical Immunology Part Two, Vol. III. of Santa Fe Institute Series Studies in the Science of Complexity, Addison-Wesley Publishing Company, Reading, MA, pp. 291–319.

Imberti, L., Sottini, A., Bettinardi, A., Puoti, M., Primi, D., 1991. Selective depletion in HIV infection of T cells that bear specific T cell receptor Vβ sequences. Science 254, 860–862.

Kearney, J.F., Vakil, M., Nicholson, N., 1987. Non-random VH gene expression and idiotype-antiidiotype expression on early B cells. In: Kelsoe, G., Schulze, D. (Eds.), Evolution and Vertebrate Immunity: The Antigen Receptor and MHC Gene Families. Texas University Press, Austin, pp. 175–190.

Lapedes, A., Farber, R., 2001. The geometry of shape space: application to influenza. J. Theor. Biol. 212, 57–69.

Morbidity and Mortality Weekly Report, 1999. CDC Atlanta, USA, 48(RR13), 1–31.

Perelson, A., Oster, G., 1979. Theoretical studies of clonal selection: minimal antibody repertoire size and reliability of self-nonself discrimination. J. theor. Biol. 81, 645–667.

Pilch, H., Hohn, H., Neukirch, C., Freitag, K., Knapstein, P.G., Tanner, B., Maeurer, M.J., 2002. Antigen-driven T-cell selection in Patients with cervical cancer as evidenced by T-cell receptor analysis and recognition of autologous tumor. Clin. Diagn. Lab. Immunol. 2002;9(2):267–278.

Rebai, N., Pantaleo, G., Demarest, J.F., Ciurli, C., Soudeyns, H., Adelsberger, J.W., Vaccarezza, M., Walker, R.E., Sekaly, R.P., Fauci, A.S., 1994. Analysis of the T-cell receptor β–chain variable-region (Vβ) repertoire in monozygotic twins discordant for human immunodeficiency virus: evidence for perturbations of specific Vβ segments in CD4+ T cells of the virus-positive twins. Proc. Natl. Acad. Sci. (USA) 91, 1529–1533.

Smith, F.S., Rencher, S.D., Heslop, H.E., Hurwitz, J.L., 1995. T cell receptor repertoire of CD4+ and CD8+ T cell subsets in the allogeneic bone marrow transplant recipient. Cancer Immunol. Immunotherapy. 41 (2), 104–110.

Smith, D.K., Neal, J.J., Holmberg, S.D., 1993. Unexplained opportunistic infections and CD4+ T-lymphocytopenia without HIV infection. An investigation of cases in the United States. N. Engl. J. Med. 328, 373–379.

Spira, T.J., Jones, B.M., Nicholson, J.K., Lal, R.B., Rowe, T., Mawle, A.C., Lauter, C.B., Shulman, J.A., Monson, R.A., 1993. Idiopathic CD4+ T-lymphocytopenia—an analysis of five patients with unexplained opportunistic infections. N. Engl. J. Med. 328, 386–392.

Wucherpfennig, K.W., Newcombe, J., Li, H., Keddy, C., Cuzner, M.L., Hafler, D.A., 1992. T cell receptor Vα-Vβ repertoire and cytokine gene expression in active multiple sclerosis lesions. J. Exp. Med. 175, 993–1002.