

MOLECULAR STRUCTURE AND VACCINE DESIGN¹

*S. Vajda, R. Kataoka, and C. DeLisi*²

Department of Biomathematical Sciences, Mount Sinai School of Medicine, One Gustave L. Levy Place, New York, New York 10029

H. Margalit and J. A. Berzofsky

Metabolism Branch, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20205

J. L. Cornette

Department of Mathematics, Iowa State University, Ames, Iowa 50011

KEY WORDS: vaccines, AIDS, structural biology, computation, drug design.

CONTENTS

PERSPECTIVES AND OVERVIEW	70
<i>The Molecules: a Brief Review</i>	70
<i>Two Strategies</i>	72
IDENTIFYING IMMUNOGENIC PEPTIDES: RECENT HISTORY	73
<i>Antigenic Determinants Recognized by Antibodies</i>	73
<i>Antigenic Determinants Recognized by T Cells</i>	74
THE ANTIGENIC MOTIF: CONCEPTUAL CONSIDERATIONS.....	74
AMPHIPHATHICITY: A MAJOR ANTIGENIC MOTIF	75
A GENERAL APPROACH TO IDENTIFYING T CELL ANTIGENIC SITES.....	76

¹ The US Government has the right to retain a nonexclusive, royalty-free license in and to any copyright covering this paper.

² Reprint requests should be sent to C. DeLisi, College of Engineering, Boston University, Boston, Massachusetts 02215.

PERSPECTIVES AND OVERVIEW

The development of the rabies vaccine by the French physicist Louis Pasteur introduced the modern era of immunization against pathogenic microbes. Since that time a number of spectacularly successful vaccines have been developed, resulting in the control and, in some cases, virtual eradication of previously devastating diseases such as small pox, poliomyelitis, measles, mumps, and rubella. In addition to countless lives saved and an inestimable reduction in human suffering, the economic impact has also been substantial. In the United States, reduction in the number of cases of polio, measles, mumps, and rubella is producing an estimated annual savings in health care costs of several billion dollars. Recent basic research developments that promise new and more efficient strategies are stimulating the formation of biotechnology companies, which are expected to have an impact on the national economy. But the greatest immediate stimuli for the development of more efficient and reliable vaccine design strategies are the AIDS epidemic and the now almost certain knowledge of the major role of DNA viruses in human cancers (28).

Perhaps coincidental, it is nonetheless fitting, that Louis Pasteur, who knew no disciplinary boundaries, developed the first vaccine. Today some of the most promising strategies draw heavily on a wide range of disciplines, from physics and mathematics to immunology. This article explains the intimate dependence of one such strategy on progress in structural biology. More specifically we discuss the requirements for a predictive understanding of the structures of proteins and peptides that are central to the immune response. These include antibodies, T-cell receptors, products of the major histocompatibility complex (MHC), and antigenic peptides. We begin with a brief review of the relevant aspects of molecular immunology, including the different molecular requirements for initial events in antibody and cellular responses. We briefly discuss immunization strategies and then focus on the biophysical problems underlying the synthesis of peptide antigens.

The Molecules: a Brief Review

The immune response to most antigens consists of two components, humoral and cellular, that act synergistically (17). The humoral response is mediated by antibodies (protein molecules secreted by B-lymphocytes), which bind foreign molecules or cells and thereby target them for destruction by the immune system. The prototypic example of an antibody is the class G immunoglobulin (IgG). IgG consists of two heavy and two light polypeptide chains with a combined molecular weight of about 150 thousand daltons and a true dyad symmetry axis (Figure 1). Each chain consists of distinct structural and functional domains, stabilized by disulfide bonds.

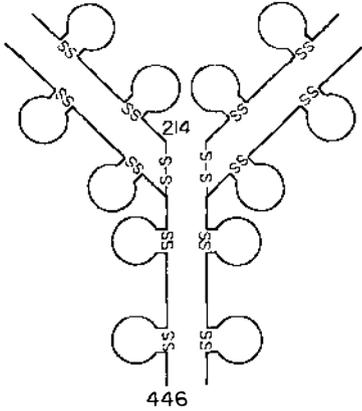


Figure 1 The class G antibody is formed by a symmetric arrangement of two heavy-light chain couples, with independently folded disulfide bonded domains within each chain.

An alignment of the amino terminal domains of a large number of IgG sequences shows that variability from one molecule to the next tends to be confined to three hypervariable regions (21). In the folded molecule these sequences are spatially juxtaposed to form the site that recognizes antigen. Hence, the ability of Igs to recognize the almost limitless number of antigenic shapes arises from the large number of amino acid substitutions, and consequent structural variability, found in hypervariable regions.

The initial step in a humoral immune response is the interaction of a foreign molecule—typically a glycoprotein or a polysaccharide—with Ig-like receptors on the B cell surface. Unfortunately, little is known about the three-dimensional structure of polysaccharides or glycoproteins. Consequently this review is limited to protein antigens.

Because proteins are folded, their contact residues are generally far apart along the sequence; i.e. the antigenic determinants recognized by B cell receptors, and ultimately by the antibodies secreted by B cell progeny, are usually formed by the spatial juxtaposition of sequentially distal residues (5, 6). As indicated below, this is in distinct contrast to antigen recognition by T cells. The difference has important consequences for certain vaccine design strategies.

After the protein antigen is bound, the antibody-antigen complex is internalized, proteolytically digested (40), and antigenic fragments, typically 10–20 residues long, are cycled to the surface where they associate with so-called class II products of the major histocompatibility gene complex (MHC) (34) (Figure 2). This peptide-MHC complex targets the B cell for help by CD4 (helper) T cells. Thus, receptors on helper T cells recognize and respond to a binary complex.

In contrast to class II products, which appear predominantly on cells

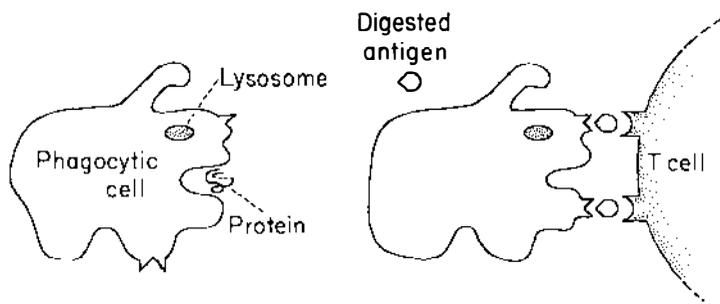


Figure 2 Cells of the immune system endocytose native protein, proteolytically digest it, and recycle the fragments to the surface where they are found associated with class II MHC products.

of the immune system, class I MHC products are found on the plasma membranes of all nucleated cells. Although a given individual has very few class I genes, they are highly polymorphic throughout a species, and they provide a molecular basis for distinguishing different individuals. Class I genes also play an important role in the cellular immune response; they are recognized in conjunction with antigenic peptides by CD8 cells, usually cytotoxic cells that associate with and destroy foreign cells and virally infected host cells. Class I and class II products bear close evolutionary relations to each other and to the immunoglobulin superfamily (19).

Under normal conditions, an encounter with a foreign protein stimulates B and T cells not only to divide and mature but also to produce a stable set of progeny specific for the stimulating antigen. These memory cells assure a rapid, intense reaction to a second exposure to the same antigen.

Two Strategies

The ability to elicit such anamnestic responses is the central concept underlying immunization. The initial or immunizing exposure in a traditional vaccine consists of attenuated or killed organisms, and consequently it carries some small risk of causing the disease it is intended to prevent. The major new approach to vaccine development uses recombinant DNA technology and avoids this problem. The gene coding for a major cell surface antigen of the invading microbe (such as the GP 160 protein in the case of AIDS) is identified, cloned, and produced in large quantities for a massive vaccination program. This procedure is safer because it avoids using the infectious organism.

The recombinant DNA strategy has drawbacks as well as advantages. One potential difficulty is that different portions of a protein tend to have different effects on the immune response. In particular, certain fragments can preferentially elicit suppressor T cells that diminish the response to

other portions of the same molecule (37). Other difficulties include the failure of bacterial systems to modify eucaryotic proteins post-translationally, and the tendency of these proteins to form insoluble inclusion bodies (the signature of improperly folded proteins) in the cytoplasm. A number of possible solutions to these problems are being explored, including the development of eucaryotic cloning systems. A better understanding of the determinants of protein folding would also be desirable. In fact, as we indicate below, progress on the other approach to vaccine development—identification and synthesis of antigenic peptides—is also tied to a deeper understanding of folding.

This second strategy identifies the relatively limited number of amino acids—typically 10–15—on the foreign target that are recognized by the immune system. Since the number of residues is limited, the peptide sequences can readily be synthesized in large quantities. Underlying this approach is the hope that some general sequence correlates of antigenicity will be discovered. Such a discovery would open the possibility for a rational approach to introducing relatively minor modifications into the native peptide sequence, which would increase its effectiveness as a vaccine. The use of peptide vaccines would also offer an optimal approach to safety by minimizing the possibility of stimulating undesirable components of the immune response.

This article focuses on the peptide synthesis strategy. Our goal is to delineate some of the problems associated with the development of efficient, cost effective vaccines and to assess prospects for progress. A review of recombinant DNA approaches can be found in (30).

IDENTIFYING IMMUNOGENIC PEPTIDES: RECENT HISTORY

Antigenic Determinants Recognized by Antibodies

A number of guiding principles for identifying sequences antigenic for B cells (2, 3, 6, 26) have been developed, all of which can be subsumed under the idea that the antibody binding site will generally have access only to residues on the protein surface. Methods for defining the surface of a protein are reasonably well developed (25). An immediate implication of the surface requirement is that in a polar solvent, surface residues tend to be polar, so a sequence of hydrophilic residues is a candidate for an antigenic determinant (20). This is a potentially useful rule, because only the sequence is required to make an identification.

A second indicator of antigenicity is side chain mobility (26). Estimates of mobility can be obtained from data on crystallographic temperature coefficients. Since mobility of surface residues is somewhat greater than for interior residues, this correlate is not independent of hydrophilicity.

It is also limited to those sequences for which higher order structural information exists or can be readily obtained. Its use is therefore more restricted than the use of indicators based on sequence properties.

The strategy of identifying and synthesizing sequences recognized by antibodies is beset with a major problem: The antigenic sites are not generally sequentially contiguous; they may be far apart along the sequence and brought into physical proximity by the fold of the protein (5, 6). Therefore, synthesizing a short contiguous sequence of residues that mimics a local antigenic configuration formed by several sequentially separated segments is a difficult problem even if the spatial arrangement of the antigenic protein is known—and in most cases it is not known.

Antigenic Determinants Recognized by T Cells

In contrast to B cell antigenic sites that are formed by the fold of a protein, T cells recognize 10–15-residue-long peptide fragments. Peptides of this length can be readily synthesized in the large quantities required for a vaccination program. However, for large viral coat proteins, often with thousands of residues, the task of synthesizing all possible contiguous peptides, even in quantities necessary for experimental studies, is formidable. Hence the search for sequence or structural correlates of antigenicity is the basis for a rational search strategy.

The search for T cell antigenic sites is far simpler than for B cell sites because short length implies that only sequentially localized residues are recognized. Length, however, is only part of the problem; one must also identify sequence motifs that are strong correlates of antigenicity and that can therefore be used to identify antigenic sites. The search for such motifs can be guided by current structure-function concepts (12, 13, 15).

THE ANTIGENIC MOTIF: CONCEPTUAL CONSIDERATIONS

One may ask, is there reason to believe in the existence of sequence or other motifs that might provide clues to function or activity? The answer is yes for both general and specific reasons. With respect to general reasons, an abundance of protein sequence and structural data indicate that conserved sequence motifs are widespread over functionally unrelated proteins. One example is the epidermal growth factor (EGF) domain, a sequence about 40 residues long and defined by a particular disulfide bond topology. The EGF motif is found not only in the EGF precursor but also in the low density lipoprotein receptor, in urokinase, and in several other proteins. Such highly homologous fragments seem to be found only in proteins of eucaryotes (16).

Structure-function relations can be observed in the classification of beta structures into distinct architectural categories [each with a similar function (32)] or into characteristic secondary structural arrangements such as the helix-turn-helix motif for certain classes of repressors. At another level, the database of known sequences can be divided into some 30 functional categories that are distinguishable on the basis of three or four sequence characteristics (24). Collectively, these findings point to the presence of several interrelated levels of structure-function motifs. Although unrelated to the possible existence of antigenic motifs, these findings make such a hypothesis natural and plausible.

For T cell antigenic sites a more direct reason for expecting structural motifs can be found. While the number of MHC molecules in a particular individual is very small, these few MHCs in a normal immune system are sufficient to recognize peptides from virtually any protein. One explanation for this observation hypothesizes that immunodominant peptides can assume only one or, at most, a few structural motifs. This idea is developed in the last section. Here we note that if this reasoning is accepted, it may provide an intuitive basis for expecting motifs, but it gives no clues about what their distinguishing characteristics might be. We discuss this subject below.

AMPHIPATHICITY: A MAJOR ANTIGENIC MOTIF

In the early 1980s, motivated by the rapid increase in protein sequence data, we embarked upon a project to find sequence properties that correlated with function and activity. Within a year we found that over 50% of the known protein database could be clustered into 30 functional categories, with each category distinguishable on the basis of three or four sequence characteristics (23). A relevant example is the relation between sequence and structure in globins. The surface of a globin is formed by five (cylindrical) alpha helical segments arranged in a globular structure. The helices form the boundary between two different environments: the solvent, which is polar, and the interior of the protein, which tends to be apolar. At such a boundary the free energy of a peptide is minimized: polar residues point out into the solvent, and apolar residues point inward. One might therefore expect to find such an arrangement in globins. In fact, amphipathic alpha helicity is the dominant characteristic distinguishing globins from other proteins in the database.

Similar considerations apply to membrane-binding peptides (22), which are also at a polar-apolar boundary. One might therefore guess that the ability of a peptide to form an amphipathic cylinder would facilitate recognition by a T cell receptor. Such a possibility is consistent with the

observation that peptides binding T cell receptors when they are presented in conjunction with class II do not bind T cell receptors when they are in solution.

The presentation requirement for T cell recognition can be explained by peptide flexibility. If immunodominant peptides have many conformations in solution, rather than a single dominant structure, the conformation suitable for binding T cell receptors is relatively dilute. On the other hand, binding to a surface such as a membrane embedded MHC molecule locks the peptide into a single stable structure. Because the function of T-cells is to interact with other cells, the requirement for recognition of antigen with MHC may be to direct T-cell recognition of antigens to appropriate cell surfaces and prevent competition by antigen in solution. Since an individual's MHC repertoire is very small, the association between peptide and MHC is expected to be relatively nonspecific, and hydrophobicity is a good candidate for a stabilizing force. On the other hand, interaction with the T cell receptor is relatively specific, suggesting the precise alignments associated with polarity. DeLisi & Berzofsky (14) therefore suggested a structure in which polar and apolar residues tend to be spatially partitioned. The question is, do known T cell antigens display amphipathicity? Or, are the sequences of immunodominant peptides consistent with amphipathic structures?

In order to answer these questions, a structure needs to be specified. The only systematic way to proceed is to consider the regular structures—alpha, beta, etc.—and then to use Fourier analysis of the sequence to search for periodicity in polarity that is compatible with the assumed structure. For example, if the structure is an alpha helix we would look for a dominant sinusoidal variation in polarity along the sequence with a period near 100 degrees. The initial analysis was performed when only 12 immunodominant peptides had been studied. Nine of them were consistent with an amphipathic alpha helix; only one was consistent with amphipathic beta structures (180 degree period). More recently, 39 out of 52 T helper antigenic sites were found to be consistent with alpha amphipathicity, with a level of statistical significance that was better than 99% (11, 27).

A GENERAL APPROACH TO IDENTIFYING T CELL ANTIGENIC SITES

The results obtained by applying the amphipathic hypothesis and other pattern recognition algorithms (33, 38) offer the promise of a structure-activity relation. The concept is also of some practical importance, because algorithms based on it seem to correctly identify at least one of the known antigenic sites in each protein for which multiple sites are known. In

addition, when several sites are predicted for proteins that had not been previously studied, at least one is invariably found to be antigenic. On the other hand, when new sites are correctly predicted we do not know whether they are among the most antigenic for the protein. Moreover, it is not likely that a monovalent (single peptide sequence) vaccine will be effective for an entire population of outbred individuals, such as the human population. A polyvalent vaccine using several peptides will probably be needed, and if several peptides must be identified, a truly rational strategy (optimal in time and cost effectiveness) will require much greater than 70% reliability.

The problem to be solved is the following: For a given MHC product sequence, and a protein against which T cell immunization is to be achieved, which peptide segments from the protein will bind the MHC best? Several segments are sought rather than a single best segment because the method is still an approximation and therefore a screening procedure, albeit one that is expected to be substantially more accurate than a screen based on amphipathicity. The question in this form also assumes that the T cell receptor repertoire is large enough so that for any MHC-peptide complex a T cell receptor will be available to bind it with stability sufficient to assure a response (29).

In fact the problem might be somewhat more subtle: The most stable MHC-peptide complex need not lead to the most stable ternary complex; in extreme cases the receptor repertoire might have gaps for particular binary complexes (34). The most general approach to the problem therefore requires consideration of ternary complexes. The number of such complexes is so large, however, and knowledge of the receptor repertoire so limited that an ability to predict the stability of the ternary complex for any specified set of sequences is at present of little more than theoretical interest.

As with all complex problems, the first approach is to try to understand the simplest relevant problem. In this case the simplest problem is to calculate the structure of isolated peptides 10–20 residues long. A variety of indirect evidence (35) and some direct evidence indicates that structures in this size range can be calculated with reasonable accuracy. The direct evidence is limited by the availability of crystal structures and nuclear Overhauser data required for rigorously testing predictions.

Mellitin is a small molecule whose crystal structure is available. Energy minimization of the 26-residue-long peptide finds a structure with the same general features as the observed structure (31), and with a root mean square deviation between the observed and calculated backbone atoms of 3.8 Å. The result is actually better than this figure indicates because it reflects an improper proline geometry (S. Vajda et al, in preparation). This

is easily demonstrated by forcing the proline dihedral angles into their observed positions and leaving the remainder of the molecule unchanged. The root mean square error between calculated and observed structure is then 2 Å, which is also the resolution of the observed structure (Figure 3).

Of equal importance is the speed with which structures in this size range can be calculated. New algorithms based on dynamic programming decrease 10–20 fold the number of function evaluations required to find a global minimum (41). This means, for example, that the CPU time typically required to calculate the structure of a fifteen residue sequence on an IBM 3090 processor is 5–7 hours. The structures of immunodominant peptides can therefore be determined routinely, and the question of motifs can be addressed in terms of detailed structural analysis.

Our results on the structures of known T helper cell immunodominant peptides are incomplete but suggestive. The global energy minimum for each of the 18 fragments that we have studied to date (Table 1) is attained by α -helical structures, generally with slight distortions at some of the residues (R. Kataoka et al, in preparation). In some instances the peptides must change dramatically relative to their native conformation in order to adopt these structures. A well-studied example is residues 52 through 61 of the hen egg-white lysozyme. In the native protein the sequence is part of a β -sheet, with a bend separating the two short strands. The global energy minimum of the isolated peptide is, however, a helix with a potential energy of -111.9 kcal/mol. For comparison, when we minimize the energy of this fragment in the neighborhood of its native conformation (i.e. β -

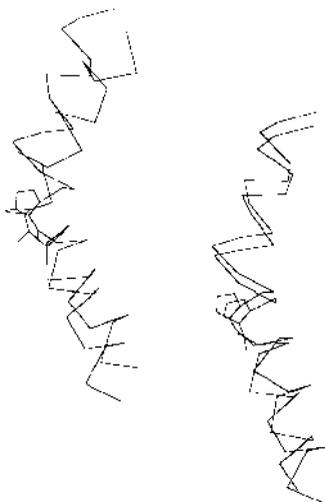


Figure 3 The root mean square deviation between the backbone atoms in the energy-minimized structure with no constraints imposed and the observed structure is 3.8 Å (*left*). When proline is constrained to be in its observed conformation (*right*), the calculated and observed backbone structures agree to within experimental error (~ 2 Å).

Table 1 Amphipathicity and predicted conformation of some antigenic sites

Restriction element	Protein	Antigenic site	Amphipathic ^a		Calculated helical region
			Site	Score	
E(k)	Sperm whale	69–78	64–78(64–78)	14.2(20.6)	69–78
A(d)	myoglobin	102–118	99–117(100–111)	20.1(13.5)	102–111
A(d)	Influenza haemagglutinin	129–140	126–139(126–139)	5.3(11.2)	129–140
A(d)	Beef cytochrome <i>c</i>	11–25	9–29(10–23)	22.7(17.5)	11–20
A(d)	Hen ovalbumin	323–339	329–346(322–332)	18.0(9.0)	323–339
A(d)	Repressor protein Cl	12–26	8–25(8–25)	19.5(24.1)	12–26
E(k)	Pigeon cytochrome <i>c</i>	93–104	92–103(92–103)	4.3(10.9)	93–104
		45–58	—(43–48)	—(9.9)	45–58
? ^b	Pork insulin	5–16	4–16(7–16)	5.5(9.6)	5–16
? ^b	Staphylococcal nuclease	61–80	58–75	17.9	61–75
E(d)	Sperm whale myoglobin	132–145	128–145(126–141)	15.3(18.0)	132–145
A(b)	Pig lactate dehydrogenase	211–223	—	—	213–223
A(k)	Hen egg-white lysozyme	46–61	—	—	46–61
A(k)		74–86	72–86(—)	9.9(—)	79–86
E(k)	Staphylococcal nuclease	81–100	73–94	26.9	81–100

^a For fragments of 11 (in parenthesis 7) residues long (14).

^b Restriction element is uncertain.

structure), the minimum energy is -77.9 kcal/mol.¹ The higher stability of the helical conformation for this isolated peptide is in agreement with the results of stimulation and inhibition experiments (1).

Although alpha helices tend to be dominant, they are just barely dominant: All of the fragments can assume a large number of different conformations with energies only slightly higher than the energy of the alpha helix. The immunodominant peptides are flexible in solution, and able to adopt numerous conformations, but each peptide has the alpha helix as a common conformation. It is suggested that the alpha-helical-like structures bind best to the MHC and therefore are stabilized for recognition by the T cell receptor in conjunction with MHC.

On the other hand, the class I crystal structure (7) suggests that the

¹ Evidentially, in the native protein, the sequence interacts more favorably with the rest of the structure when it is beta rather than alpha.

MHC combining site is large and has at least 15–20 potential contact residues. Based on what we know about bond energies, probably only 5–8 of these residues are sufficient to stabilize interaction with a peptide. Consequently, peptides, in principle, can form contacts with the site in many ways by small variations on a dominant structural motif. These two features, degeneracy of potential contacts and peptide flexibility, explain why a given peptide is able to bind more than one member of a polymorphic population while, at the same time, in a given individual only a few MHC molecules are required to bind a large number of different peptides.

Given the ability to calculate peptides, one can calculate the stability of every overlapping peptide fragment, e.g. ten residues long, with a given MHC product. Alternatively, one can calculate the energy of interaction of every cleavage fragment with the MHC product. The latter approach would eliminate arbitrary choice of fragment length and at the same time would substantially reduce the number of calculations. Enzymes involved in processing have recently been identified (39), and the information on digestion sites required for such a procedure should become available within the next 4–7 years. The present lack of detailed knowledge of the digestion pathway is another reason for identifying more than just the best binding fragment.

Identifying fragments that interact stably with MHC requires knowledge of the MHC structure, or at least the structure of the domains that interact with antigen, and the structure of the peptide fragment in its bound conformation. At present the most important information available is a human class I crystal structure (7). At least two major problems, however, must be addressed—one theoretical and the other experimental—before maximum use can be made of this information.

Since all class I products are expected to have the same general fold, the crystal structure provides a good first approximation for the structure of any class I product. An approximate structure, however, is unlikely to be sufficiently accurate to identify the fragments that bind best to a particular class I sequence. One must be able to accurately generate the structure of any class I from the generic approximation. Although much has been written on this so-called homologous extension (18) problem, a systematic evaluation of its reliability is not available. Even if structures could be accurately generated, the class I system at present is not ideal from an experimental point of view, because peptide binding to class I has not yet been measured definitively, though some progress is being made (10).

With respect to class II, a number of thermodynamic and kinetic studies of peptide class II product interactions are available (4, 9). However, experimental data on the structure of class II is not available. Although class I and class II peptide-binding domains bear weak sequence homology,

and some modeling of class II based on class I has been attempted, the class I as a generic fold for class II must be considered speculative (8).

With well-supported and well-planned research efforts, a truly rational vaccine design strategy should be effective within the next 5–10 years. The low-energy conformations of short peptides can now be determined accurately, and during the next few years computation of such structures should become routine. A high priority is elucidation of the digestion pathway within macrophages, B cells, and other antigen-presenting cells, and determination of generic structures for class II MHC products and T cell receptors. This information, coupled with improvements in homologous extension algorithms, will permit rapid determination of any of the millions of different immunoglobulin family molecules and will have a major impact on fundamental immunology.

Literature Cited

- Allen, P. M., Matsueda, G. R., Evans, R. J., Dunbar, J. B., Marshall, G. R., Unanue, E. R. 1987. *Nature* 327: 713
- Atassi, M. Z. 1978. *Immunochemistry* 15: 909
- Atassi, M. Z., Lee, C. L. 1978. *Biochem. J.* 171: 429
- Babbitt, B. P., Allen, P. M., Matsueda, G., Haber, E., Unanue, E. R. 1985. *Nature* 317: 359
- Benjamin, D. C., Berzofsky, J. A., East, I. J., Gurd, F. R. N., Hannum, C., et al. 1984. *Annu. Rev. Immunol.* 2: 67
- Berzofsky, J. A. 1985. *Science* 229: 932
- Bjorkman, P. J., Saper, M. A., Samraoui, B., Bennett, W. S., Strominger, J. L., Wiley, D. C. 1987. *Nature* 329: 506
- Brown, J. H., Jardetzky, T., Saper, M. A., Samraoui, B., Bjorkman, P. J., Wiley, D. C. 1988. *Nature* 332: 845
- Buus, S., Sette, A., Colon, S. M., Jenis, D. M., Grey, H. M. 1986. *Cell* 47: 1071
- Chen, B. P., Parham, P. 1989. *Nature* 337: 743
- Cornette, J. L., DeLisi, C., Margalit, H., Berzofsky, J. A. 1989. *Methods Enzymol.* 178: In press
- DeLisi, C. 1988. *Science* 240: 47
- DeLisi, C. 1989. Computation and the human genome project: an historical perspective. In *Computers and DNA*, ed. G. I. Bell, T. Marr, I: 13. Reading, Mass: Addison-Wesley
- DeLisi, C., Berzofsky, J. A. 1985. *Proc. Natl. Acad. Sci. USA* 82: 7048
- DeLisi, C., Vajda, S. 1989. The human genome project and structural biology. In *Proceedings of the 6th Conversation in Biomolecular Stereodynamics*, ed. R. H. Sarma, M. H. Sarma. New York: Adenine. In press
- Doolittle, R. F., Feng, D. F., Johnson, M. S., McClure, M. A. 1986. Relationships of human protein sequences to those of other organisms. In *Cold Spring Harbor Symp. Quant. Biol.* 51: 447
- Eisen, H. 1980. *Immunology*. New York: Harper & Rowe. 547 pp. 2nd ed.
- Feldman, R. J., Bing, D. H., Potter, M., Mainhart, C., Furie, B., et al. 1984. *Ann. NY Acad. Sci.* 439: 12
- Hood, L., Steinmetz, M., Malissen, B. 1983. *Annu. Rev. Immunol.* 1: 529
- Hopp, T. P., Woods, K. R. 1981. *Proc. Natl. Acad. Sci. USA* 78: 3824
- Kabat, E. 1968. *Structural Concepts in Immunology and Immunochemistry*. New York: Holt, Rinehart & Winston. 310 pp.
- Kaiser, E. T., Kezdy, F. J. 1984. *Science* 223: 249
- Klein, P., Jacquez, J., DeLisi, C. 1986. *Math. Biosci.* 81: 177
- Klein, P., Kanehisa, M., DeLisi, C. 1984. *Biochim. Biophys. Acta* 787: 221
- Lee, B. K., Richards, F. 1971. *J. Mol. Biol.* 55: 379
- Lerner, R. A. 1984. *Adv. Immunol.* 36: 1
- Margalit, H., Spouge, J. L., Cornette, J. L., Cease, K. B., DeLisi, C., Berzofsky, J. A. 1987. *J. Immunol.* 138: 2213
- Marx, J. L. 1989. *Science* 243: 1012
- Ogasawara, K., Maloy, W. L., Schwartz, R. H. 1987. *Nature* 325: 450
- Patzer, E. J., Obijeski, J. F. 1985. Vac-

- cines prepared from translation products of cloned viral genes. In *Immunology of Viruses*, ed. M. H. V. Van Regenmortel, A. R. Neurath, 1: 153. Amsterdam: Elsevier. 493 pp.
31. Pincus, M. K., Klausner, R. D., Scheraga, H. A. 1982. *Proc. Natl. Acad. Sci. USA* 79: 5107
 32. Richardson, J. 1977. *Nature* 268: 495
 33. Rothbard, J. B., Taylor, W. R. 1988. *EMBO J.* 7: 93
 34. Schaeffer, E. B., Sette, A., Johnson, D. L., Bekoff, M. C., Smith, J. A., et al. 1989. *Proc. Natl. Acad. Sci. USA* 86: 4649
 35. Scheraga, H. A. 1984. *Carlsberg Res. Commun.* 49: 1
 36. Schwartz, R. H. 1985. *Annu. Rev. Immunol.* 3: 237
 37. Shastri, N., Oki, A., Miller, A., Sercarz, E. E. 1985. *J. Exp. Med.* 162: 332
 38. Stille, C. J., Thomas, L. J., Reyes, V. E., Humphreys, R. E. 1987. *Mol. Immunol.* 10: 1021
 39. Takahashi, H., Cease, K. B., Berzofsky, J. A. 1989. *J. Immunol.* 142: 2221
 40. Unanue, E. R., Allen, P. M. 1987. *Science* 236: 551
 41. Vajda, S., DeLisi, C. 1989. *Biopolymers*. In press