Advances in
PROTEIN CHEMISTRY

VOLUME 32

ADVANCES IN PROTEIN CHEMISTRY

EDITED BY

C. B. ANFINSEN

National Institute of Arthritis, Metabolism and Digestive Diseases
Bethesda, Maryland

JOHN T. EDSALL

Biological Laboratories
Harvard University
Cambridge, Massachusetts

FREDERIC M. RICHARDS

Department of Molecular Biophysics and Biochemistry
Yale University
New Haven, Connecticut

VOLUME 32

1978

ACADEMIC PRESS • *New York San Francisco London*

A Subsidiary of Harcourt Brace Jovanovich, Publishers

ACADEMIC PRESS, INC.
111 Fifth Avenue, New York, New York 10003

United Kingdom Edition published by
ACADEMIC PRESS, INC. (LONDON) LTD.
24/28 Oval Road, London NW1 7DX

LIBRARY OF CONGRESS CATALOG CARD NUMBER: 44–8853

ISBN 0–12–034232–4

PRINTED IN THE UNITED STATES OF AMERICA

ADVANCES IN PROTEIN CHEMISTRY

Volume 32

CONTRIBUTORS TO THIS VOLUME

Claude Balny

Pierre Douzou

Robert E. Feeney

Elvin A. Kabat

F. MacRitchie

Yin Yeh

CONTRIBUTORS TO VOLUME 32

Numbers in parentheses indicate the pages on which the authors' contributions begin.

CLAUDE BALNY, *INSERM, Unité 128, B.P. 5051, 34033 Montpellier, and Institut de Biologie Physico-Chimique, Foundation Edmond de Rothschild, 75005 Paris, France* (77)

PIERRE DOUZOU, *INSERM, Unité 128, B.P. 5051, 34033 Montpellier, and Institut de Biologie Physico-Chimique, Foundation Edmond de Rothschild, 75005, Paris, France* (77)

ROBERT E. FEENEY, *Department of Food Science and Technology, University of California, Davis, California 95616* (191)

ELVIN A. KABAT, *Departments of Microbiology, Human Genetics and Development, and Neurology, College of Physicians and Surgeons, Columbia University, New York, New York and the National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20014* (1)

F. MACRITCHIE, *CSIRO Wheat Research Unit, North Ryde, New South Wales, 2113, Australia* (283)

YIN YEH, *Department of Applied Science, University of California, Davis, California 95616* (191)

CONTENTS

Antifreeze Protein from Fish Bloods
ROBERT E. FEENEY AND YIN YEH

Proteins at Interfaces
F. MACRITCHIE

THE STRUCTURAL BASIS OF ANTIBODY COMPLEMENTARITY[1]

By ELVIN A. KABAT

Departments of Microbiology, Human Genetics and Development, and Neurology
College of Physicians and Surgeons, Columbia University, New York,
and the National Cancer Institute, National Institutes of Health, Bethesda, Maryland

I. INTRODUCTION

Perhaps the most important problem in structural immunology and immunochemistry today is an understanding of antibody complementarity in terms of three-dimensional structure since this should provide new insights leading to the genetic basis for the generation of diversity and will open new perspectives for research in cellular immunology. More precisely one would like to know the mechanism by

[1] These studies were aided by grants from the National Science Foundation, BMS-72-02219 A04 and PCM76-81029. Work with the PROPHET computer system is sponsored by the National Cancer Institute, National Institute of Allergy and Infectious Diseases, National Institute of Arthritis, Metabolism and Digestive Diseases, National Institute of General Medical Sciences, Division of Research Resources (Contract No. NO1-RR-42147), National Institutes of Health.

which the immune system can produce an enormous repertoire of different receptor sites, estimates of the number of which range from at least 10^4 to as high as 10^7, each distinct site showing a remarkable degree of specificity. Extraordinary progress has been made in the past 10–12 years both in sequencing and from X-ray diffraction studies on immunoglobulins and antibodies.

In a sense these have only opened to our view the magnitude of the problem and the need for a somewhat different approach if it is to be solved within a reasonable time. X-Ray diffraction studies on a particular class of protein, such as the cytochromes *c*, hemoglobins, snake venoms in a wide range of species, have shown that there is a remarkable preservation of secondary and tertiary structure of the molecule and indeed of sequence in the active-site region. Indeed, if one knows the three-dimensional structure of cytochrome *c* from one species one can approximate both the folding of the polypeptide chain and the shape of the active-site region of another cytochrome *c*; a stereo model of cytochrome *c* has been constructed using tuna cytochrome *c* for one chain and horse for the other (Takano *et al.*, 1973). This preservation of site structure over the evolutionary time span is the result of natural selection; individuals with mutations such that the polypeptide chain could not fold up to form a functional molecule would not survive. Some mutations in proteins may not be lethal but may alter the molecule functionally, producing disease; many such hemoglobin mutants have been studied (for sequence differences, see Dayhoff, 1972, 1973, 1976).

Since mutations are occurring constantly, what one observes in examining the various species are those mutations which permit the proper three-dimensional structure to form. The resulting differences in amino acid sequence are recognized as species differences. Since the sequence for each species remains constant, there may well be some specific advantages of any given amino acid residue in one species, but from an overall point of view the differences are merely *mutational noise* compatible with satisfactory three-dimensional folding to give a functional molecule. Thus, if one has elucidated from X-ray diffraction studies at sufficiently high resolution the three-dimensional structure of one or two members of a class of proteins, such as the hemoglobins or the cytochromes *c*, one can predict, except for relatively minor differences, the three-dimensional structures of all members of the class. One can also predict with substantial confidence the effects of mutations at a given position and validate the predictions crystallographically. This has been done for human hemoglobin mutants (Greer, 1971). One can

thus envisage acquiring a reasonably complete understanding of the three-dimensional structure of each of these classes of proteins within a finite time.

Antibody complementarity is a far more complex phenomenon. Heterogeneity and diversity are the hallmarks of the antibody response (Kabat, 1961, 1966a,b, 1976a). A protein antigen generally contains many antigenic determinants, and the antibody response even to one determinant may result in a population of antibodies of different specificities, each recognizing different aspects of the same determinant. This has been shown most clearly with polysaccharide antigens containing a single sugar such as dextran, a dextran with 96% $\alpha 1 \rightarrow 6$ linkages and 4% $\alpha 1 \rightarrow 3$ like linkages giving rise to a population of antibody molecules with combining sites complementary to different lengths of the $\alpha 1 \rightarrow 6$ polysaccharide chain, e.g., to different numbers of $\alpha 1 \rightarrow 6$ glucose residues or, in other terms, having antibody combining sites of different sizes (Kabat, 1956, 1957, 1960, 1976a).

Antibodies to polysaccharide antigens such as dextrans and levans produced in humans, while heterogeneous, often show restricted heterogeneity as shown by acrylamide gel electrophoresis (Yount *et al.*, 1968), isoelectric focusing patterns (Cisar *et al.*, 1975), starch gel electrophoresis of the separated polypeptide chains (Edelman and Kabat, 1964) and in possessing fewer genetic markers than found on the total immunoglobulin of the same individual (Allen *et al.*, 1964; Yount *et al.*, 1968).

Antibodies to the streptococcal groups A and C carbohydrates and to the pneumococcal polysaccharides are also heterogeneous, but it has been possible in individual rabbits to obtain homogeneous antibodies or antibodies of substantially restricted heterogeneity (Haber, 1970; Krause, 1970) as demonstrated by the above criteria and by microzone electrophoresis (see Kochwa and Kunkel, 1971) in amounts suitable for sequencing; such studies have shown substantial differences in sequence among antibodies of relatively similar specificity (Jaton, 1974, 1975, 1976; Haber *et al.*, 1975; Thunberg and Kindt, 1975; Braun *et al.*, 1976a,b; for additional references, see Kabat *et al.*, 1976b).

It is doubtful whether studies on antibodies would have advanced as rapidly had not a substantial body of data on the specificity, sequence, and three-dimensional structure been accumulated on the immunoglobulins and related molecules associated with certain neoplastic diseases. A large proportion of patients with a disease of plasma cells called multiple myeloma had been known for over a century to excrete in their urine large amounts of a class of protein named

Bence Jones proteins. These proteins because of their unusual behavior of precipitating when heated to 60°–70°C at pH 4.5–5, again going into solution on boiling, and reprecipitating on cooling, have proved to be of great value in the diagnosis of multiple myeloma. Bence Jones proteins were shown by Edelman and Gally (1962) to be the light chains of immunoglobulins. It was recognized quite early that Bence Jones proteins from different individuals varied in chemical (Putnam and Miyake, 1954), physical (Gutman *et al.*, 1941; Moore *et al.*, 1943), and immunological properties (Bayne-Jones and Wilson, 1922; Moore *et al.*, 1943; Korngold and Lipari, 1956; Migita and Putnam, 1963; Osserman *et al.*, 1957); indeed among humans no two individuals have been shown to have identical Bence Jones proteins, and estimates of as many as 4000 different human Bence Jones proteins have been made (Quattrocchi *et al.*, 1969). This diversity is intimately related to antibody specificity.

The serum of patients with multiple myeloma often had high levels of serum globulin. When the Tiselius electrophoretic method was used and later with paper or gel electrophoresis, these sera were shown to contain high levels of proteins with a very narrow range of electrophoretic mobility unlike the broad heterogeneous immunoglobulin peaks. The mobilities of the sharp peaks in the sera of these patients varied from individual to individual. These serum immunoglobulins also differed in molecular weight. A macroglobulin now known as IgM was first isolated by Waldenström (1944, 1948), and the neoplastic disease is termed Waldenström macroglobulinemia.

Another crucial development was the finding that the BALB/c (Potter, 1972, 1977a) and later that the NZB strains of mice (Warner, 1975) when injected with paraffin oil develop a disease like multiple myeloma and also often excrete Bence Jones proteins. This not only provided an experimental model, but also permitted detailed comparison of mouse Bence Jones proteins and immunoglobulins with their human counterparts, an indispensable prerequisite for the study of antibody specificity. Relatively enormous quantities (kilograms in some instances) of Bence Jones proteins were obtainable from the urine of patients; plasmapheresis yielded substantial quantities of myeloma proteins. Large amounts of the corresponding mouse proteins were also obtainable; proteins from such neoplasms were in almost all instances monoclonal and homogeneous.

About 12 years ago a number of myeloma immunoglobulins from both humans and mice were found to have specific receptor sites for various substances, and these are now considered to be monoclonal antibodies. These include human antibodies reacting with ϵDNP-

(see list of abbreviations below) or TNP-lysine, streptolysin O anti-IgG, anti-blood group I and i, a hydroxy derivative of vitamin K, lipoproteins, etc., and mouse myeloma proteins with antibody sites reacting with DNP, phosphocholine, $\alpha 1 \rightarrow 6$ and $\alpha 1 \rightarrow 3$ linked dextrans, $\beta 2 \rightarrow 1$ and $\beta 2 \rightarrow 6$ linked fructosans, galactans, nonreducing terminal DGlcNAc, etc. (for a more complete listing, see Kabat, 1976a; Potter, 1977a,b). These, however, represent but a very small selection of the diversity of specific sites which the antibody-forming mechanism can generate.

It is manifestly impossible to envisage, within a reasonable period of time, a solution to the problem of antibody complementarity by elucidation of the structure of all kinds of antibody combining sites. One must add at least another dimension to such attempts. This review will consider how one may gain more insight into the problem by attempting to make predictions as to the shape, structure, contacting and structural elements of complementarity regions of antibody combining sites which, if verified crystallographically, would provide insight into the structures of many kinds of antibody combining sites.

Abbreviations and Definitions

DNP—dinitrophenyl as in ϵDNP-Lys

CDR—complementarity-determining regions or segments, the three portions of the light and of the heavy chain showing hypervariability and which were predicted (Wu and Kabat, 1970) and subsequently found to form the antibody combining site.

DGlc—Dglucose

DFru—Dfructose; in oligosaccharides the furanose or five-membered ring is indicated by *f*, as DFru*f* is Dfructofuranosyl.

IM—the isomaltose or $\alpha 1 \rightarrow 6$ linked series of oligosaccharides of DGlc followed by a number which indicates the size of the oligosaccharide; thus IM3 is the trisaccharide DGlc$\alpha 1 \rightarrow$ 6DGlc$\alpha 1 \rightarrow$ 6DGlc. All glycosidic sugar units are in the six-membered pyranose form unless the furanose form is indicated.

NMR—nuclear magnetic resonance

ESR—electron spin resonance

Idiotypic specificity—Antibodies and myeloma globulins of a given specificity show differences in antigenic specificity ascribable to differences in sequence in the variable regions of their light and heavy chains. Some of these idiotypic determinants are inhibitable by the antigenic determinant or ligand for which the antibody or myeloma protein is specific, whereas others are not. Those idiotypic determinants which are inhibitable are hypothesized to in-

volve residues of the CDR whose side chains project away from the interior of the combining site and whose three-dimensional structures are altered by conformational changes occurring when the combining site interacts with the specific ligand so that they no longer react with the antiidiotypic antibody (for a detailed discussion, see Kabat *et al.*, 1976a). Noninhibitable idiotypes involve the V-region away from the site. Idiotypic specificity is thus another important parameter of antibody heterogeneity and diversity. No amino acid residues have as yet been implicated in the specificity of either inhibitable or noninhibitable idiotypes. Cross-reactions occur among idiotypes of a given specificity, and a given idiotype may be present in small amounts in the serum of normal animals. In an immunized animal, the antibody and nonantibody immunoglobulins may show substantial amounts of an idiotype which was not detectable in the serum prior to immunization (Oudin, 1974; Hopper and Nisonoff, 1971; Natvig and Kunkel, 1973; Childs and Feizi, 1975; Casenave *et al.*, 1974).

Immunoglobulin G (IgG) subclasses—There are four subclasses, called IgG1, IgG2, IgG3, and IgG4, which show sequence differences in the C-domains (see Fig. 4); they also show differences in their antigenic determinants.

II. Immunochemical Studies of the Size and Shape of Antibody Combining Sites

Since Landsteiner showed that a low molecular weight hapten like arsanilic acid would compete for the antibody combining site with an antigen, prepared by coupling arsanilic acid to a protein, and inhibit precipitation, a tool was available to study the structure of antibody combining sites. Using synthetic dipeptides (Landsteiner and van der Scheer, 1932, 1934; see Landsteiner, 1945) and disaccharides (Goebel *et al.*, 1934) coupled to proteins, it could be demonstrated that the C-terminal amino acid or the nonreducing end of the disaccharide in reacting in the antibody combining site contributed the major portion of the binding energy. In addition, with such azoproteins the antigenic determinant often involved not only the grouping introduced, but also a portion of the protein onto which it was attached (Hooker and Boyd, 1933). Thus arsanilic acid azotyrosine was a better inhibitor than arsanilic acid itself. Since the grouping diazotized to the protein could attach to any accessible tyrosine or histidine in the protein, and since the sequence around each tyrosine varied, the antigen was very heterogeneous and the antibody response was

accordingly extremely complex and one could obtain no data on the size of antibody combining sites.

A powerful approach to estimating the size of antibody combining sites came from the use of homopolymers as antigens. Dextrans were shown (Kabat and Berg, 1953) to be antigenic in humans, two injections of 0.5 mg a day apart giving rise to substantial amounts of precipitating antibodies and to the development of wheal and erythema skin sensitivity; these findings were confirmed by Maurer (1953). The antibodies were shown to be specific for chains of $\alpha 1 \rightarrow 6$ linked glucoses, and when oligosaccharides of the isomaltose series became available (Jeanes *et al.*, 1953; Jones *et al.*, 1956; Turvey and Whelan, 1957) they could be used as a probe or molecular ruler to explore the sizes of human antidextran sites. It was observed that inhibiting power on a molar basis increased with increasing chain length but that the increment in inhibiting power per added glucose decreased and an upper limit was reached, generally with six sugars (Kabat, 1956, 1957, 1960). In its most extended form, the hexasaccharide (IM6), isomaltohexaose, measures $34 \times 12 \times 7$ Å. It was also observed that each person did not produce a homogeneous population of antibody combining sites, since the ratio of inhibiting power of the tri- (IM3), tetra- (IM4), penta- (IM5), and hexasaccharides to one another differed for each individual. Moreover, absorption of the antidextran on Sephadex and successive elution with IM3 followed by IM6 fractionated the antibody into two populations (Schlossman and Kabat, 1962; Gelzer and Kabat, 1964) which varied in their relative inhibitability by small, as compared with larger, oligosaccharides, confirming that they were mixtures with combining sites of different sizes. Other studies (see Kabat, 1966a, 1976a; Goodman, 1969, 1975; Schlossman *et al.*, 1968) showed that antibody combining sites for various kinds of antigenic determinants—carbohydrate, synthetic polypeptides, protein, nucleic acid, etc.—produced in various species varied from a lower limit of between one and two sugars, 4 to 6 Å, to an upper limit of about 6 or 7 sugars, or about 34 Å, in most extended length.

From the standpoint of evaluating structural differences in antibody combining sites, these purified antidextrans gave relatively little insight because of their heterogeneity, and only when various myeloma antidextrans (Leon *et al.*, 1970; Lundblad *et al.*, 1972; Cisar *et al.*, 1974, 1975) were recognized could progress be made. Myeloma globulins have the disadvantage as compared with antibodies that one does not know precisely the antigen to which they are complementary. Nevertheless in a system such as the $\alpha 1 \rightarrow 6$ dextran myeloma

antidextran, when their combining site sizes are at or close to the upper limit found for $\alpha 1 \rightarrow 6$ antidextrans and when they are tested with a variety of oligosaccharides of different structures and found to be $\alpha 1 \rightarrow 6$ specific, the possibility is minimal that the myeloma globulin is more specific for other unrecognized determinants of an entirely different structure (Cisar *et al.*, 1974, 1975). Indeed the high specificity of antigen–antibody interactions makes it possible to obtain information about the size and shape of antibody combining sites as well as to elucidate the structures of antigenic determinants by identifying the structure which competes most effectively with antigen for the combining site (Kabat, 1961, 1976a).

Figure 1 shows the findings with four IgA myeloma antidextrans. Three of these (W3434, W3129, QUPC 52) were specific for $\alpha 1 \rightarrow 6$ linked, and the fourth for $\alpha 1 \rightarrow 3$ linked, glucose (UPC 102). The specificity of the inhibition and the reaching of an upper limit are seen. In both W3434 and W3129, IM5, IM6, and IM7 are equal as inhibitors on a molar basis and better then IM4 and IM3 (Fig. 1A and B), indicating a site complementary to a chain of five glucoses. With QUPC 52, however, IM6 and IM7 are clearly better than IM5, which in turn is better than IM4, etc., indicating a site complementary to six glucoses. Figure 1A and B also shows the $\alpha 1 \rightarrow 6$ glucosyl specificity, since disaccharides of glucose linked other than $\alpha 1 \rightarrow 6$ react much less strongly than IM2; DGlc$\alpha 1 \rightarrow 6$ DGlcNAc also is not as good IM2, again indicating the need for a second glucose. The fourth myeloma antidextran is specific for $\alpha 1 \rightarrow 3$ linked glucose, nigerotriose, -tetraose, and -pentaose being equally potent and better than nigerose (Fig. 1E), which in turn was better than diglucoses linked $\alpha 1 \rightarrow 4$, $\alpha 1 \rightarrow 2$ and $\alpha 1 \rightarrow 6$. Two other myeloma antidextrans with $\alpha 1 \rightarrow 3$ specificity had been studied. One, an IgM myeloma, MOPC 104E, had a site complementary to the trisaccharide (Leon *et al.*, 1970), whereas with the other, J558 belonging to the IgG3 subclass, nigeropentaose was the best inhibitor (Lundblad *et al.*, 1972).

The shapes of the inhibition curves in Fig. 1A, B, and E are atypical since IgA myeloma proteins are mixtures of monomers and polymers. Using a lower molecular weight dextran N-150N, rather than the higher molecular weight native B512, typical inhibition curves as shown in Fig. 1C were obtained. If the myeloma antidextran was separated into monomer and polymer portions, the usual inhibition curves could be obtained (Cisar *et al.*, 1974).

When two of these myeloma antidextrans were studied by equilibrium dialysis (Fig. 2A), a surprising result was obtained. QUPC 52, with the larger size site, had an association constant, K_a, of 8.4×10^3

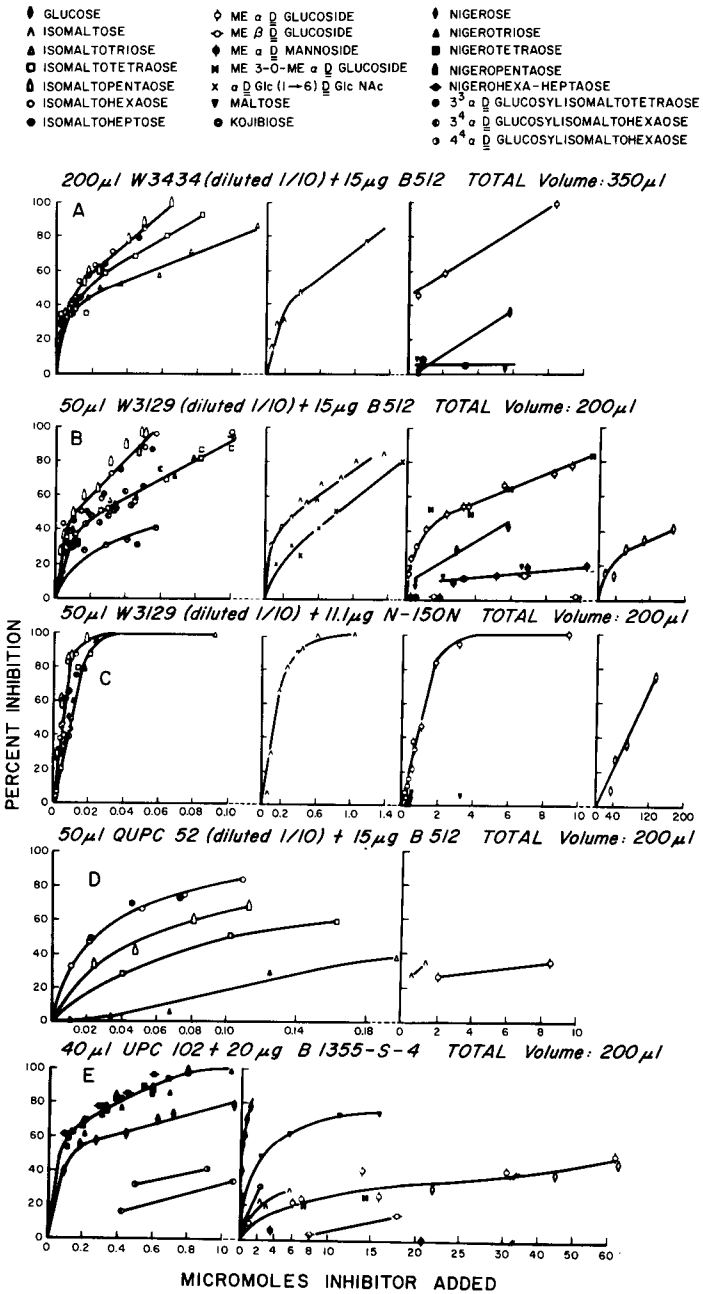

FIG. 1. Inhibition by various oligosaccharides of precipitation of mouse myeloma proteins by dextran. From Cisar *et al.* (1974), with permission.