

NBP, a protein that specifically binds an enhancer of immunoglobulin gene rearrangement: purification and characterization

Min Li, Ewa Morzycka-Wroblewska¹, and Stephen V. Desiderio

Department of Molecular Biology and Genetics and Howard Hughes Medical Institute Laboratory of Genetics, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205 USA

Immunoglobulin and T-cell receptor (TCR) genes are encoded in discrete germ line DNA segments that are joined by site-specific recombination during lymphocyte development. These DNA rearrangements are mediated by conserved heptamer and nonamer DNA sequence elements that lie near the sites of recombination. In this paper we show that the nonamer element coincides with the recognition site for a specific DNA-binding protein: mutations within the nonamer sequence, but not outside of it, decrease affinity for the binding protein by 300- to 1000-fold. Deletion of the binding site for the protein results in at least a 50-fold decrease in recombination frequency *in vivo*. By a combination of conventional and recognition site affinity chromatography, we have achieved > 20,000-fold purification of the protein from calf thymus, with an overall yield of 22%. The purified protein, which we now call nonamer-binding protein (NBP), has an apparent molecular weight of 63,000 and a frictional ratio of 1.27, suggesting that it exists as a globular monomer in 0.5 M NaCl. Our observations suggest that NBP is a component of the recombinational apparatus.

[Key Words: Immunoglobulin gene rearrangement; lymphocyte development; DNA-binding proteins; specific DNA affinity chromatography]

Received July 10, 1989; revised version accepted August 24, 1989.

The variable regions of immunoglobulin chains are encoded by discrete germ line DNA segments that are brought together by site-specific recombination during lymphoid differentiation (for reviews, see Tonegawa 1983; Alt et al. 1987). For example, the variable regions of immunoglobulin heavy chains are encoded by three germ line elements, V_H , D, and J_H (Early et al. 1980; Sakano et al. 1980); during B-cell differentiation, individual segments from each group are joined to form a complete V_H -D- J_H unit. The genes that encode T-cell antigen receptors (TCR) show similar patterns of segmentation and rearrangement (Davis 1985). As a consequence of these rearrangements, a diverse set of variable regions is generated from a relatively small number of gene segments. Thus, immunoglobulin and TCR gene rearrangement play a central role in establishing the primary immune repertoire.

Unrearranged immunoglobulin and TCR gene segments are accompanied by heptamer and nonamer sequence elements, separated by a spacer region (Early et al. 1980; Sakano et al. 1980). Several lines of evidence indicate that these sequences mediate rearrangement.

First, the heptamer and nonamer sequences are conserved among immunoglobulin and TCR gene families and among vertebrate species (Litman et al. 1985). Second, the heptamer-spacer-nonamer motif is located at the sites of recombination. Joining of coding sequences is accompanied by the formation of a reciprocal product in which the heptamer sequences of the participating gene segments are precisely joined (Lewis et al. 1985; Okazaki et al. 1987; Desiderio and Wolff 1988). Third, the spacer regions fall into two length classes of 12 and 23 bp; recombination normally occurs only between gene segments carrying spacers of different length (Early et al. 1980). Fourth, the heptamer-spacer-nonamer motif is sufficient to support rearrangement of exogenous recombinational substrates in cultured cells (Akira et al. 1987; Hess et al. 1987). Immunoglobulin and TCR gene segments are apparently recombined by a similar mechanism, as suggested by the correct joining of exogenous TCR gene segments in cultured B-progenitor cells (Yancopoulos et al. 1986). On the basis of these observations, it is likely that the conserved heptamer and nonamer elements represent binding sites for proteins involved in rearrangement of immunoglobulin and TCR genes.

We recently identified a protein that binds to DNA

¹Present address: Department of Medicine, School of Medicine, University of California San Diego, La Jolla, California USA.

fragments containing immunoglobulin recombinational signal sequences from diverse sources (Halligan and Desiderio 1987). The specific binding activity was detected in extracts of nuclei from lymphoid cells but not in extracts from nonlymphoid cells. Because the protein was observed to bind to a DNA fragment containing a murine J_K4 nonamer sequence, but not to a fragment containing the heptamer sequence, we inferred that the protein recognizes the nonamer.

In this paper we define the protein's DNA recognition site and examine the function of this site *in vivo*. In addition, we report the purification of this protein. The sequence required for DNA-protein binding was found to coincide precisely with the conserved nonamer recombinational signal: Mutations within the nonamer, but not outside of it, resulted in large decreases in affinity. Deletion of the nonamer sequence was observed to profoundly impair rearrangement of an immunoglobulin gene segment *in vivo*. The protein was purified by a combination of conventional methods and DNA recognition site affinity chromatography. We obtained a purification of >20,000-fold, with an overall yield of 22%. The protein, which we now call nonamer-binding protein (NBP), has an apparent molecular weight of 63,000, as judged by SDS-polyacrylamide gel electrophoresis; it behaves as a globular monomer in 0.5 M NaCl. The specific binding of NBP to an enhancer of immunoglobulin gene rearrangement and its expression in nuclei of lymphoid cells suggest a role for NBP in the assembly of immunoglobulin and TCR genes.

Results

Construction of mutant substrates for binding

Previous experiments identified a protein in extracts of calf thymus nuclei that specifically binds DNA fragments containing recombinational signal sequences

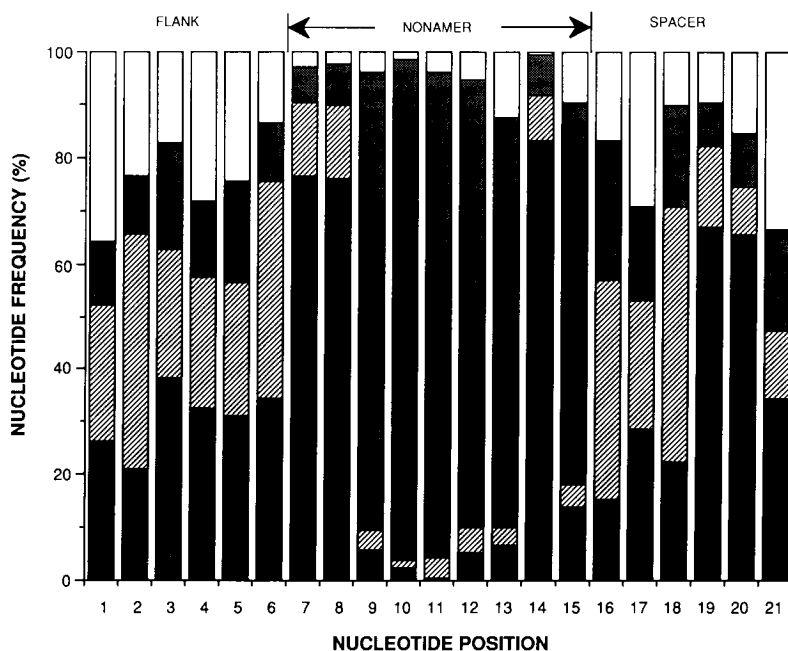
from a variety of immunoglobulin gene segments. Within a fragment from the immunoglobulin J_K4 gene segment, the binding site was localized to a 27-bp interval spanning the nonamer region (Halligan and Desiderio 1987). To define the binding site, we constructed a series of mutant DNA fragments and assayed their ability to compete with a wild-type J_K4 DNA fragment for binding.

Comparison of the nucleotide sequences of 137 immunoglobulin and TCR gene segments revealed that the nonamer motif (consensus GGTTTTGT), is well conserved but exhibits some variability nonetheless (Fig. 1). Because of this fact and our observation that the protein binds DNA fragments containing differing nonamer sequences (Halligan and Desiderio 1987, and B. Halligan and S.V. Desiderio, unpubl.), we elected to construct mutant substrates with multiple nucleotide substitutions. Nine different pairs of complementary, 24-mer oligonucleotides were synthesized, annealed, and ligated into pUC13 to yield a series of plasmids, each containing an 18-bp sequence spanning the J_K4 nonamer (Fig. 2). In one of the plasmids (pNM13/14), the 18-bp core sequence was identical to the wild-type J_K4 sequence. Of the remaining plasmids, six (pNM1/2, pNM3/4, pNM5/6, pNM7/8, pNM9/10, and pNM11/12) contained a 3-bp substitution, whereas two (pNM60/61 and pNM62/63) contained a 2-bp substitution (Fig. 2). Cleavage of each plasmid with *Hind*III and *Eco*RI yielded a 54-bp DNA fragment containing the mutant or wild-type sequence; these fragments served as competitors in protein-binding assays.

Definition of the binding site for the protein

The formation of specific DNA-protein complexes was detected by an electrophoretic mobility shift assay. The source of the protein used in the assay was the active fraction from Bio-Rex-70, the first column used in purifi-

Figure 1. Nucleotide usage near immunoglobulin and TCR nonamer signal sequences. The sequences of 137 immunoglobulin and TCR gene segments within a 21-bp interval spanning the nonamer were aligned by inspection. The nucleotide frequency at each position on the upper strand was computed. Sequences are oriented so that the spacer region lies toward the right. The nonamer region extends from position 7 to 15; positions 16–21 lie in the spacer region. Open bars represent percent C; shaded bars, percent T; hatched bars, percent A; solid bars, percent G. The sequences used in the computation were obtained from the compilation of Kabat et al. (1987), and from the following additional references: Bernard et al. (1978); Tonegawa et al. (1978); Max et al. (1979); Sakano et al. (1979); Early et al. (1980); Max et al. (1980); Bothwell et al. (1981); Cohen and Givol (1983); Chen et al. (1984); Heinrich et al. (1984); Yancopoulos et al. (1984); Hayday et al. (1985); Litman et al. (1985); Concannon et al. (1986); LeFranc et al. (1986); Ichihara et al. (1988).



J_K4 ...GCAGGTTTTTGTAAAGGG...

pNM13/14 GGATCCGCAGGTTTTTGTAAAGGGGAATTC

pNM 1/2 GGATCC**TAC**GGTTTTTGTAAAGGGGAATTC

pNM 3/4 GGATCCGCAT**TG**TTTTTGTAAAGGGGAATTC

pNM 5/6 GGATCCGCAGGT**GGG**TGTAAAGGGGAATTC

pNM 7/8 GGATCCGCAGGTTTTT**CAC**AAAGGGGAATTC

pNM 9/10 GGATCCGCAGGTTTTTGT**CCC**GGGAATTC

pNM 11/12 GGATCCGCAGGTTTTTGTAAAT**TT**GAATTC

pNM 60/61 GGATCCGCAT**TT**TTTTTGTAAAGGGGAATTC

pNM 62/63 GGATCCGCAGGTTTTT**AC**AAAGGGGAATTC

Figure 2. Sequences of wild-type and mutant DNA fragments used in competition experiments. Complementary, 24-base-long oligonucleotide pairs (NM1/NM2, NM3/NM4, NM5/NM6, NM7/NM8, NM9/NM10, NM11/NM12, NM13/NM14, NM60/NM61, and NM62/NM63) were synthesized and annealed; the resulting duplex DNA, flanked by *Eco*RI and *Bam*HI restriction sites, was ligated into the plasmid pUC13. The resulting plasmids were designated pNM1/2, pNM3/4, pNM5/6, pNM7/8, pNM9/10, pNM11/12, pNM13/14, pNM60/61, and pNM62/63. (Top line) Sequence of the murine J_K4 segment in an 18-bp interval spanning the nonamer. The nonamer sequence is underlined. (Remaining lines) Sequences of synthetic DNA fragments from the *Bam*HI site to the *Eco*RI site. Only the upper strand is shown. Mutations within the 18-bp interval corresponding to the J_K4 sequence are shown in bold-face type.

cation of the protein (see below); the labeled DNA fragment was the 27-bp *Dde*I-*Hin*PI fragment of pJ_K4.copy96 (Halligan and Desiderio 1987), which spans the J_K4 nonamer sequence. Figure 3A shows the results of an assay in which the 54-bp *Hind*III-*Eco*RI fragment from pNM13/14 was used as a specific competitor. In the absence of the pNM13/14 fragment, we observed a species of altered mobility that was not detected in the absence of added protein (Fig. 3A, lanes b and a). In the presence of increasing amounts of the specific competitor, the yield of this species decreased approximately linearly, demonstrating that binding is directed to the 18-bp sequence shared by the pNM13/14 fragment and the J_K4 fragment and that the affinity of the protein for each of these fragments is similar (Fig. 3A, lanes c-k).

When the mutant fragments were used as specific competitors in the binding assay, we found that fragments from pNM1/2, pNM9/10, and pNM11/12 competed with the J_K4 fragment for binding, whereas fragments from pNM3/4, pNM5/6, and pNM7/8 competed very poorly or not at all (Fig. 3B). The amount of bound, labeled DNA recovered from each reaction was quantitated by densitometry and normalized to the amount of bound, labeled DNA recovered in the absence of competitor; this value (B/B₀) is expressed in Figure 4A as a function of the mass of added competitor. Assuming that the protein binds DNA with a stoichiometric ratio

of 1, we estimate that the fragments from pNM3/4, pNM5/6, and pNM7/8 have at least a 1000-fold lower affinity for the binding protein than does the wild-type fragment from pNM13/14. In contrast, the fragments from pNM1/2, pNM9/10, and pNM11/12 bound the protein with an affinity similar to that of the wild-type fragment. To exclude the possibility that the 5 T/A base pairs in the center of the nonamer are sufficient for binding, we measured the relative affinities of fragments from pNM60/61 and pNM62/63, in which the initial GG or terminal GT dinucleotide of the nonamer was mutated (Fig. 2). These mutations also resulted in large (300- and 1000-fold) decreases in affinity, indicating that these dinucleotides are crucial for binding (Fig. 4B). The mutants pNM3/4, pNM5/6, pNM7/8, pNM60/61, and pNM62/63 therefore define a site, from 7 to 9 bp in length, that directs specific protein binding. This site coincides with the conserved nonamer recombinational sequence (Figs. 2, and 4A,B). We shall henceforth refer to the protein as NBP.

Deletion of the binding site for NBP impairs recombination in vivo

If the nonamer element plays a role in recombination, deletion of the nonamer would be expected to impair rearrangement. To test this, we assayed the rearrangement of wild-type and mutant V_H segments that were stably integrated into the genome of a B-lymphoid progenitor cell line by retroviral transduction (Desiderio and Wolff 1988; Morzycka-Wroblewska et al. 1988). The wild-type substrate for recombination, pLJHCR-2 (Fig. 5), has been described (Morzycka-Wroblewska et al. 1988) and contains three murine immunoglobulin gene segments: V_H, DJ_H, and J_H. Between the V_H and DJ_H segments lies the *gpt* gene of *Escherichia coli*. The immunoglobulin gene segments are arranged so that joining of V_H to DJ_H or to J_H results in an inversion of the intervening DNA. The mutant substrate pLJHCR-2ΔN, which was derived from pLJHCR-2, carries a deletion that removes the nonamer element and 2 bp of spacer DNA from the recombinational signal sequence of V_H (Fig. 5).

The pLJHCR-2 and pLJHCR-2ΔN constructs were packaged and transmitted to the B-progenitor cell line HAFTL-1 (Alessandrini et al. 1987). Derivatives of HAFTL-1 that contained integrated provirus were selected in G418. Rearrangement of the substrates in individual clones of HAFTL-1 cells was assayed by digestion of genomic DNA with *Kpn*I and hybridization to probes specific for *gpt* or *neo* sequences (Morzycka-Wroblewska et al. 1988). Digestion of unrearranged proviral DNA creates a 2.4-kb fragment that hybridizes to a *gpt*-specific probe and a 4.0-kb fragment that hybridizes to a *neo*-specific probe. Proviral DNA that has undergone V_H-to-DJ_H joining yields a 4.8-kb fragment that hybridizes to both probes. Of 183 HAFTL-1 clones containing the wild-type substrate, 80 yielded a 4.8-kb *neo*-containing *Kpn*I fragment that also hybridized to the *gpt* probe, consistent with V_H-to-DJ_H joining within the substrate (Fig. 5). Of 139 clones containing the mutant sub-

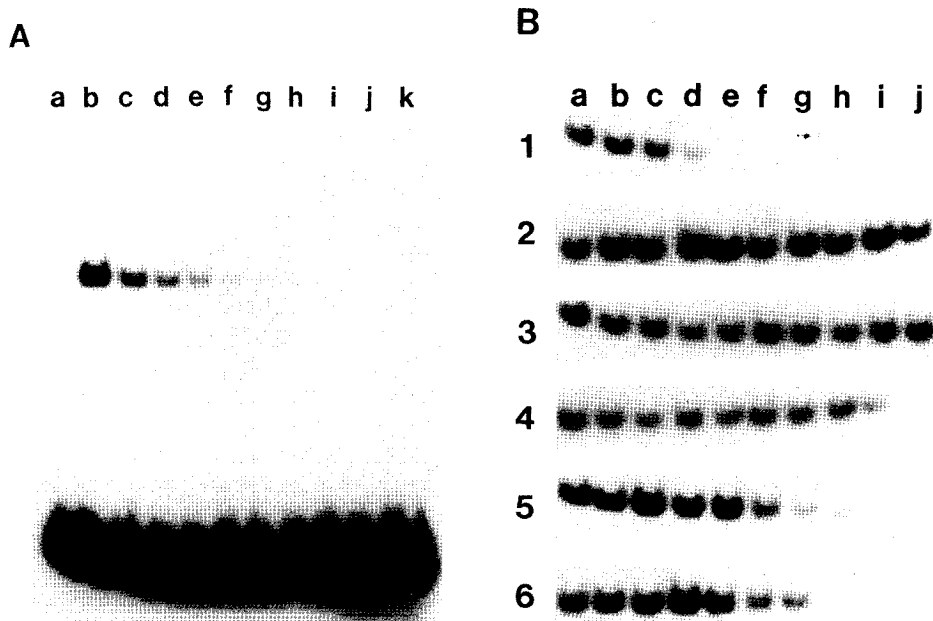


Figure 3. Assays for competition by wild-type and mutant DNA fragments (A) Assay for competition by a DNA fragment containing the consensus nonamer sequence. Labeled, 27-bp *DdeI-HinPI* fragment (0.01 ng) from pJ_R4.copy96 was assayed for formation of a specific DNA-protein complex in the presence of 2 μ l (2 μ g) of partially purified protein (Bio-Rex-70 pool), as described in Methods. Reactions contained variable amounts of the 54-bp *HindIII-EcoRI* fragment from pNM13/14 (specific competitor), which was quantitated by fluorimetry. (Lane a) No extract, no specific competitor; (lane b) no specific competitor; (lanes c-k) reactions containing the pNM13/14 fragment in twofold increments from 10.2 to 2600 pg. (B) Assays for competition by mutant fragments. Assays were performed as described in A and Methods. (Lane a) contains no specific competitor, whereas lanes b-j contain twofold increments of a 54-bp, specific competitor fragment. The sources of competitor fragments and the amounts added are pNM1/2, 7.8–2000 pg (1); pNM3/4, 78–20,000 pg (2); pNM5/6, 74–19,000 pg (3); pNM7/8, 90–23,000 pg (4); pNM9/10, 4.7–1,200 pg (5); pNM11/12, 6.3–1600 pg (6).

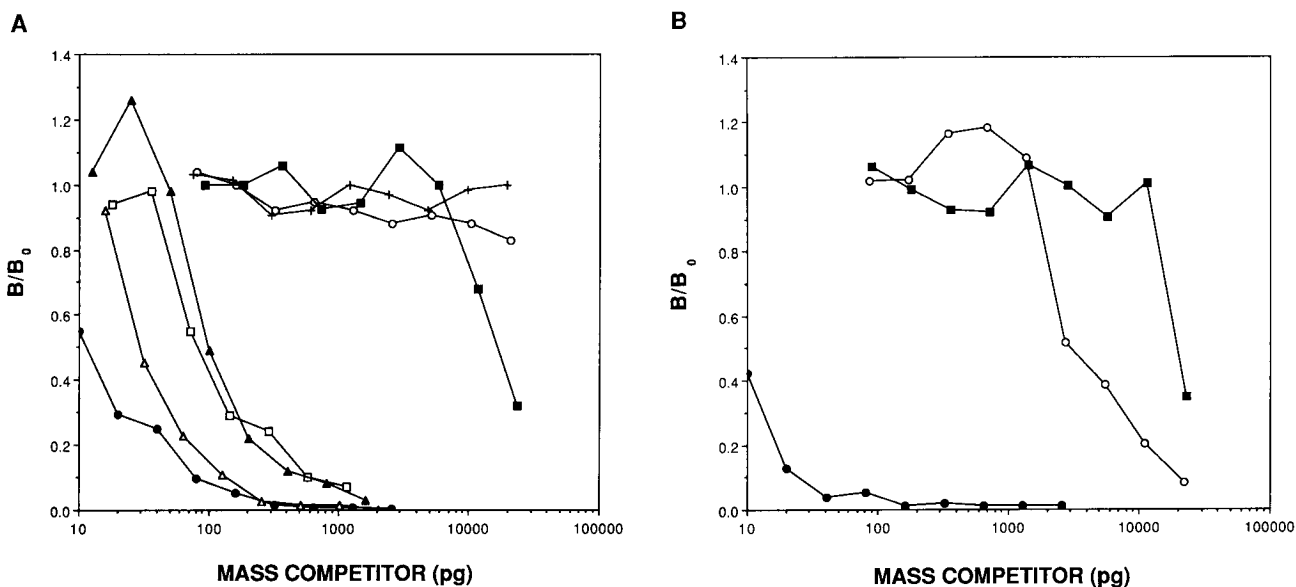


Figure 4. Effects of mutation on the relative affinity of DNA-protein binding. (A) DNA fragments containing trinucleotide substitutions. The amount of labeled DNA bound in the presence of competitor was quantitated by densitometry and normalized to the amount bound in the absence of competitor. This value (B/B₀) is displayed as a function of the mass of competitor added. (Δ) Competitor fragment from pNM1/2; (\circ) pNM3/4; (+) pNM5/6; (\blacksquare) pNM7/8; (\square) pNM9/10; (\blacktriangle) pNM11/12; (\bullet) pNM13/14. (B) DNA fragments containing dinucleotide substitutions. Data are displayed as described in A. (\circ) Competitor fragment from pNM60/61; (\blacksquare) pNM62/63; (\bullet) pNM13/14.

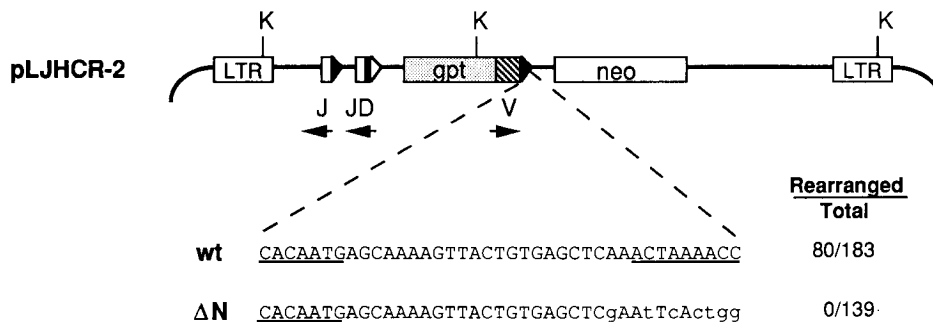


Figure 5. Effect of nonamer deletion on rearrangement *in vivo*. (Top line) The retroviral substrate for recombination, pLJHCR-2. The Moloney murine leukemia virus long terminal repeats (LTRs), the *gpt* gene, and the *neo* gene are indicated. The D sequence is indicated by a solid box, the J_H sequences, by open boxes, and the V_H sequence, by a hatched box. The transcriptional orientations of the immunoglobulin gene segments are indicated by arrows. Recombinational signal sequences carrying 23- and 12-bp spacers are indicated by solid and open triangles, respectively. Sequences surrounding the retroviral LTR are not included. (K) *KpnI* restriction sites. (Center and bottom lines) Sequences of pLJHCR-2 and pLJHCR-2ΔN in the 39-bp interval 3' to the V_H coding region. Residues in the mutant sequence that differ from wild-type are written in lowercase letters. The number of cell clones containing rearranged provirus and the number of cell clones assayed are shown (right).

strate, none showed evidence of V_H -to-D J_H or V_H -to- J_H joining by this assay (Fig. 5). Thus, removal of 11 bp from the recombinational signal sequence of the V_H gene segment, including the nonamer and 2 bp of the adjacent spacer sequence, results in at least a 50-fold decrease in the frequency of rearrangement.

The deletion in the substrate pLJHCR-2ΔN was found to abolish specific binding of NBP to the 3' flank of the V_H gene segment; conversely, a DNA fragment containing the deleted sequence was able to bind NBP specifically (data not shown). We conclude that the deletion in pLJHCR-2ΔN spans a recognition site for NBP.

Recovery of active NBP from SDS-polyacrylamide gels

To determine the approximate size of the polypeptide(s) that binds to the nonamer sequence, protein from the Bio-Rex-70 pool (see below) was denatured in SDS and 2-mercaptoethanol and fractionated by electrophoresis through an SDS-polyacrylamide gel. The lane containing the protein was cut into 10-mm slices. Protein was eluted from each slice, renatured in guanidine hydrochloride, and assayed for DNA binding activity by the electrophoretic mobility shift assay (Fig. 6A). Activity was recovered from a single gel slice, which contained protein ranging in molecular weight from 55,000 to 65,000 (Fig. 6A, lane f). The mobility of the DNA-protein complex formed in the presence of renatured protein was identical to the mobility of the complex formed with native protein (Fig. 6A, lanes b and f). To determine whether the renatured activity had the same specificity as the native protein, reactions were performed in the presence of 1.5 ng of specific competitor DNA (Fig. 6B). The wild-type DNA fragment from pNM13/14 (Fig. 6B, lane c) and mutant fragments from pNM1/2 (Fig. 6B, lane d), pNM9/10 (Fig. 6B, lane h), and pNM11/12 (Fig. 6B, lane i) competed with the J_k4 fragment for binding to the renatured protein, whereas mutant fragments from pNM3/4, pNM5/6, and pNM7/8 did not (Fig. 6B, lanes e, -g). Thus, the renatured binding ac-

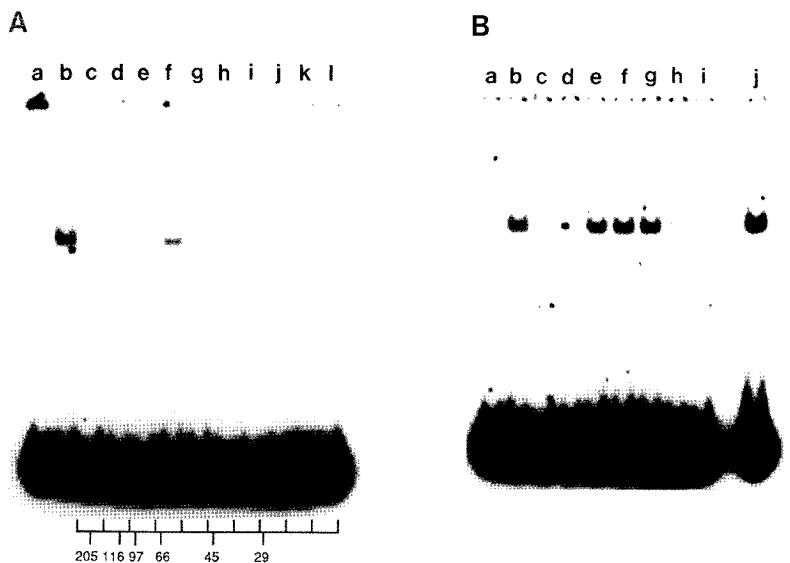
tivity, like the native protein, recognizes a site that coincides with the conserved nonamer sequence.

Purification of NBP

As a prerequisite to understanding the function of NBP, we purified the protein on the basis of its specific DNA binding activity. The purification is summarized in Table 1. Calf thymus proved to be an excellent source of activity. An extract of calf thymus nuclei was prepared, and specific binding activity was quantitatively precipitated in $(\text{NH}_4)_2\text{SO}_4$. The precipitate was chromatographed on the cationic exchange resin Bio-Rex-70 (Fig. 7A; Table 1). The DNA binding activity eluting at 280 mM NaCl was shown to be specific for the nonamer sequence by the competition assays presented above. Additional DNA binding activities eluted at higher NaCl, but these were found to be nonspecific by the mobility shift assay (data not shown). The active pool from Bio-Rex-70 was chromatographed on heparin agarose (Fig. 7B; Table 1). In this step, the nonamer binding activity was separated from the major protein peak. The active pool from heparin agarose was chromatographed on nonspecific DNA-Sepharose (Fig. 7C, Table 1); in this step, 28-fold purification was achieved with a yield of 70%.

The nonamer binding activity pool from nonspecific DNA-Sepharose was still inhomogeneous. Recognition site affinity chromatography, developed by Rosenfeld and Kelly (1986) and Kadonaga and Tjian (1986), has proved useful in the purification of a number of specific DNA-binding proteins and seemed likely to be applicable here. We therefore constructed a specific nonamer affinity matrix. The active fraction from nonspecific DNA-Sepharose was chromatographed on the nonamer affinity column in the presence of nonspecific competitor DNA poly[d(I-C)]; Fig. 8A; Table 1). The column was washed with binding buffer containing 205 mM NaCl. The column was then developed with the same buffer containing 580 mM NaCl. Nonamer binding activity was

Figure 6. Recovery of nonamer-binding activity after electrophoresis through SDS-polyacrylamide (A) Protein (600 μ g) from the active Bio-Rex-70 fraction was boiled in SDS and 2-mercaptoethanol and fractionated by electrophoresis through SDS-polyacrylamide. The lane containing the protein was cut into 10-mm wide slices; protein was eluted and renatured in guanidinium chloride as described in Methods. Renatured protein was assayed for specific binding to the 27-bp J_{K4} probe by the mobility shift assay. (Lane a) No protein; (lane b) 2 μ l (2 μ g) of the Bio-Rex-70 pool; (lanes c-l) 5 μ l of protein from each gel slice. The positions of molecular weight standards in relation to the gel slices assayed are indicated at bottom. (B) Renatured binding activity retains specificity for the nonamer. The active gel fraction was assayed for binding to the 27-bp J_{K4} probe (0.01 ng per reaction) in the presence of specific competitor fragments. (Lane a) No protein or specific competitor added; (lane b) no specific competitor; (lanes c-i) assays performed in the presence of 1500 pg of the 54-bp *Hind*III-*Eco*RI fragment from pNM13/14, pNM1/2, pNM3/4, pNM5/6, pNM7/8, pNM9/10, or pNM11/12; (lane j) protein (2 μ g) from the Bio-Rex-70 pool, assayed in the absence of specific competitor.



nearly quantitatively (92%) recovered in the 580 mM NaCl eluate. By the Bradford assay, which has a sensitivity of ~ 1 μ g of protein per ml [Bradford 1976], no protein was detectable in the active pool from the nonamer affinity column. On this basis we estimate that at least a 10-fold purification was obtained in this step. Examination of the column fractions by SDS-polyacrylamide gel electrophoresis and silver staining indicates that the purification obtained by affinity chromatography was, in fact, far greater (Fig. 8B, see below).

Physical characterization of NBP

By SDS-polyacrylamide gel electrophoresis and silver staining, a predominant polypeptide of 63 kD was detected in the 580 mM NaCl eluate from the nonamer affinity column (Fig. 8B). The 63-kD polypeptide comigrated with the nonamer binding activity (Fig. 8A,B), in agreement with our earlier assessment of the apparent molecular weight of NBP (Fig. 6A). Furthermore, when protein from the active Bio-Rex-70 pool was fractionated

by SDS-polyacrylamide gel electrophoresis and transferred to nitrocellulose, a 63-kD polypeptide was observed to bind specifically to a radiolabeled, 27-bp DNA probe containing the nonamer (data not shown). To prove that the 63-kD polypeptide recovered from the affinity column represented NBP, protein from the activity peak was fractionated by SDS-polyacrylamide gel electrophoresis and the 63-kD polypeptide was recovered from the gel. The eluted protein was renatured and shown to possess nonamer-binding activity (Fig. 8C). We conclude that the 63-kD polypeptide represents NBP.

In previous assays for nonamer-binding activity in protein eluted from SDS gels, we had inconsistently observed, in addition to the 63-kD species, a nonamer-binding species of ~ 45 kD (data not shown). It is possible that the 45-kD polypeptide is a proteolytic fragment of the 63-kD protein, although alternative explanations, including heterogeneous post-translational modification and the existence of a distinct nonamer-binding species, have not been eliminated.

The subunit composition and shape of NBP in solu-

Table 1. Affinity purification of NBP

Fraction	Protein (mg)	Activity (units)	Specific activity (units/mg)	Total yield (%)	Purification (fold)
I Nuclear extract	1,030	45,000	43.7	100	
II $(\text{NH}_4)_2\text{SO}_4$ ppt	747	45,000	60.2	100	1.4
III Bio-Rex-70	72	19,070	265	42.4	6.1
IV Heparin-agarose	5.1	15,225	2,990	33.8	68
V DNA-Sepharose	0.13	10,655	83,898	23.7	1,920
VI Nonamer affinity	<0.011	9,864	>874,000	21.9	>20,000

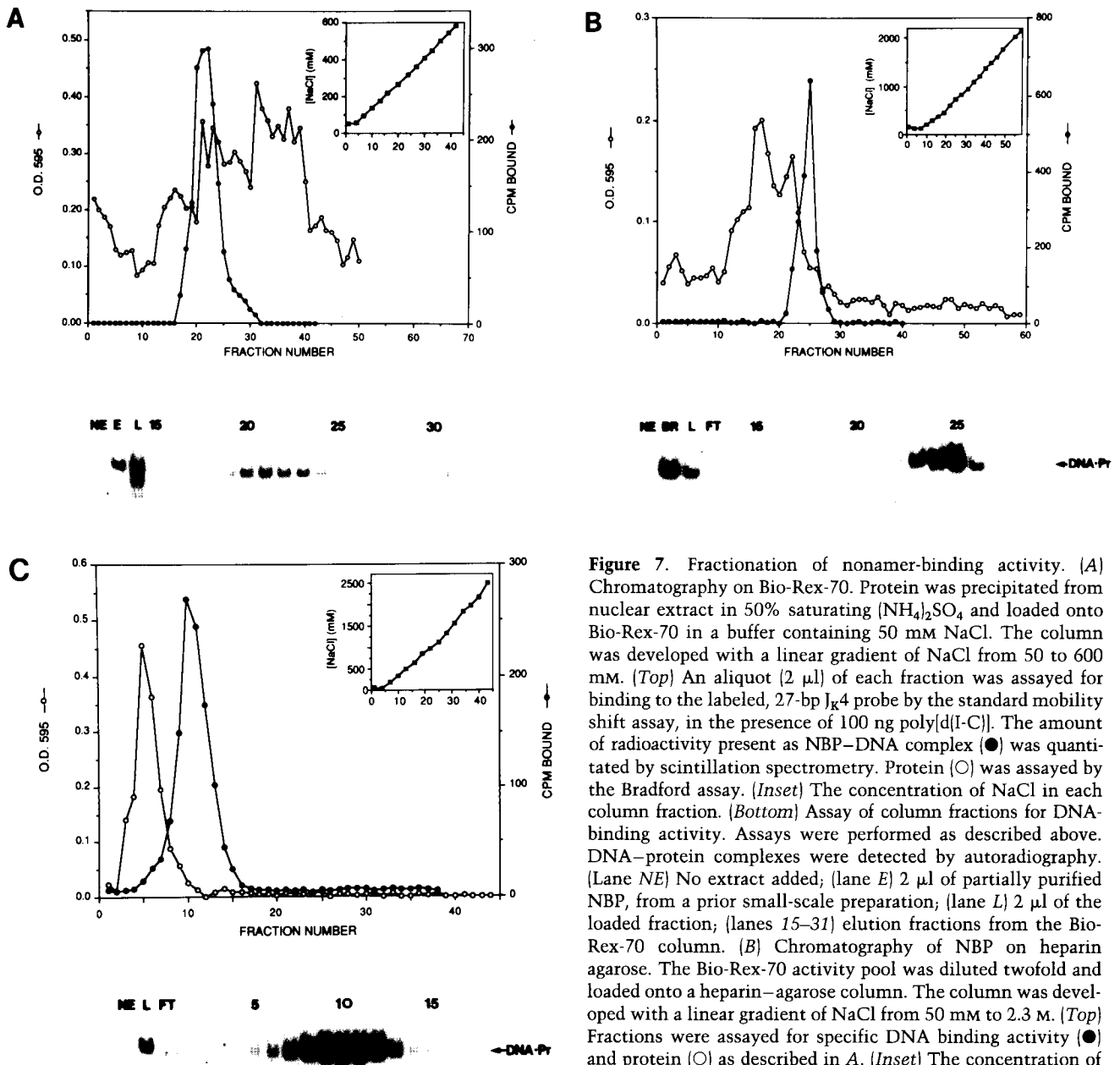


Figure 7. Fractionation of nonamer-binding activity. (A) Chromatography on Bio-Rex-70. Protein was precipitated from nuclear extract in 50% saturating $(\text{NH}_4)_2\text{SO}_4$ and loaded onto Bio-Rex-70 in a buffer containing 50 mM NaCl. The column was developed with a linear gradient of NaCl from 50 to 600 mM. (Top) An aliquot (2 μl) of each fraction was assayed for binding to the labeled, 27-bp $J_{\kappa}4$ probe by the standard mobility shift assay, in the presence of 100 ng poly[d(I-C)]. The amount of radioactivity present as NBP-DNA complex (\bullet) was quantitated by scintillation spectrometry. Protein (\circ) was assayed by the Bradford assay. (Inset) The concentration of NaCl in each column fraction. (Bottom) Assay of column fractions for DNA-binding activity. Assays were performed as described above. DNA-protein complexes were detected by autoradiography. (Lane NE) No extract added; (lane E) 2 μl of partially purified NBP, from a prior small-scale preparation; (lane L) 2 μl of the loaded fraction; (lanes 15-31) elution fractions from the Bio-Rex-70 column. (B) Chromatography of NBP on heparin agarose. The Bio-Rex-70 activity pool was diluted twofold and loaded onto a heparin-agarose column. The column was developed with a linear gradient of NaCl from 50 mM to 2.3 M. (Top) Fractions were assayed for specific DNA binding activity (\bullet) and protein (\circ) as described in A. (Inset) The concentration of NaCl in each column fraction. (Bottom) Assay of column

fractions for DNA binding activity. Assays were performed as described in A). (Lane NE) No extract added; (lane BR) 2 μl active pool from Bio-Rex-70; (lane L) 2 μl of the loaded material; (lane FT) 2 μl of the flowthrough; (lanes 14-29) elution fractions from the heparin-agarose column. The position of the NBP-DNA complex is indicated. (C) Chromatography of NBP on nonspecific DNA-Sephadex. The heparin agarose activity pool was dialyzed against a buffer containing 50 mM NaCl and loaded onto a nonspecific DNA-Sephadex column. The column was developed with a linear gradient of NaCl from 50 mM to 2.5 M. (Top) Fractions were assayed for specific DNA binding activity (\bullet) and protein (\circ), as described in A. (Inset) The concentration of NaCl in each column fraction. (Bottom) Assay of column fractions for DNA binding activity. Assays were performed as described in A. (Lane NE) No extract added; (lane L) 2 μl of the loaded material; (lane FT) 2 μl of the flowthrough; (lanes 1-17) elution fractions from the nonspecific DNA-Sephadex column. The position of the NBP-DNA complex is indicated.

tion were determined by a calculation of the native molecular weight and frictional ratio, according to the method of Siegel and Monty (1966). This calculation is based on two experimentally derived parameters: sedimentation coefficient ($S_{20,w}$) and Stokes' radius (r_s). Ve-

locity sedimentation analysis of NBP in 0.5 M NaCl yielded a sedimentation coefficient of 4.0S (Fig. 9A). By analytic gel filtration on Superose 12, we determined the Stokes' radius of NBP to be 31.5×10^{-8} cm (Fig. 9B). Assuming a partial specific volume of $0.725 \text{ cm}^3/\text{g}$, we

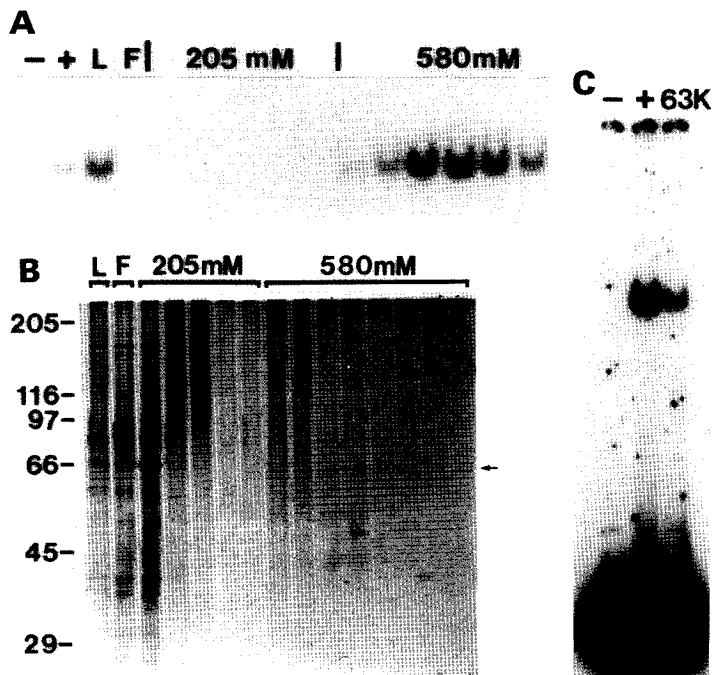


Figure 8. Purification of NBP by nonamer affinity chromatography. (A) Assay of column fractions for DNA binding activity. The activity pool from nonspecific DNA–Sephacryl was concentrated fivefold, dialyzed against binding buffer, and loaded onto a nonamer DNA affinity column in the presence of poly[d(I-C)], as described in Methods. The column was washed with 15 ml of binding buffer containing 205 mM NaCl, and activity was eluted by passage of 10 ml of binding buffer containing 580 mM NaCl. The volume of each fraction was 1 ml. Fractions were assayed for DNA binding activity as described in Fig. 7. (Lane –) No extract added; (lane +) 2 μ l of Bio-Rex-70 activity pool; (lane F) 2 μ l of the loaded material; (lane F) 2 μ l flowthrough; (lanes 205 mM) 2 μ l from fractions 1, 4, 7, 10, and 13, which were eluted with 205 mM NaCl; (lanes 580 mM) 2 μ l from fractions 16–11, which were eluted with 580 mM NaCl. (B) SDS–polyacrylamide gel electrophoresis of fractions from the nonamer DNA affinity column. An aliquot of each column fraction was precipitated by addition of trichloroacetic acid (TCA) to 17% and deoxycholate (DOC) to 0.7 mg/ml, followed by incubation on ice for 15 min.

Protein was collected by centrifugation at 12,000g for 20 min at 4°C, dissolved in 15 μ l of sample buffer, neutralized by the addition of Tris base (pH 9.0), heated at 95°C for 10 min, and fractionated by SDS–polyacrylamide gel electrophoresis. Protein was detected by silver staining. (Lane L) Twenty microliters of the material loaded onto the affinity column; (lane F) 500 μ l of the flowthrough fraction; (lanes 205 mM) 500 μ l of fractions 1, 4, 7, 10, and 13; (lanes 580 mM) 500 μ l of fractions 16–23. The arrow indicates the position of the 63-kD species. (Left) The positions and apparent molecular weights of standards (in kD). (C) The 63-kD polypeptide has nonamer binding activity. Aliquots (100 μ l) of fractions 16–20 from the affinity column were pooled and precipitated with 17% TCA in the presence of 0.7 mg/ml DOC, as described above. The protein pellet was dissolved in 15 μ l of sample buffer, neutralized, heated to 95°C, and fractionated by electrophoresis through SDS–polyacrylamide alongside a sample of each elution fraction. The portion of the gel containing the pooled protein was removed, and the remainder of the gel was stained with silver. The gel lane containing the pooled protein was aligned with the stained gel, and a 2-mm slice, corresponding to the 63-kD species, was excised. Protein was eluted and renatured as described in Methods. The renatured protein was assayed for binding to the 35-bp *EcoRI*–*HinPI* probe from *J κ 4* (Lane –) No protein added; (lane +) 2 μ l Bio-Rex-70 activity pool; (lane 63K) 5 μ l renatured 63-kD protein (1% of the renatured sample).

estimate the native molecular mass of NBP to be 53 kD, with a frictional ratio (f/f_0) of 1.27. These results suggest that NBP exists as a globular monomer in 0.5 M NaCl.

Discussion

The structural features of immunoglobulin gene rearrangement suggest a mechanism in which participating DNA segments are cleaved at the junction between the heptamer and the coding sequence, to yield an intermediate in which four DNA ends—two coding ends and two flanking ends—are held in proximity (Morzycka-Wroblewska et al. 1988). From a consideration of the structures of aberrant rearrangement products, it is apparent that a heptamer element, in the absence of an intact nonamer, can mediate joining at a very low frequency (Kleinfeld et al. 1986; Reth et al. 1986). It is nonetheless clear that the nonamer element plays an intimate role in rearrangement. Experiments presented here demonstrate that deletion of the nonamer is associated with at least a 50-fold decrease in the frequency of rearrangement *in vivo*. In addition, rearrangement is

constrained by the spatial relationship between the heptamer and nonamer elements, as evidenced by the observation that recombination generally occurs between gene segments that carry nonamer elements at specific distances (12 and 23 bp) from the heptamer.

One interpretation of these observations is that the heptamer and nonamer elements are recognized by separate components of the recombinational apparatus and that efficient rearrangement requires interaction between these components. For example, we might imagine that the binding of a protein(s) to the nonamer stabilizes an interaction between a specific endonuclease and the heptamer, via protein–protein contacts. Another possibility, not exclusive of the first, is that interactions between nonamer binding proteins are involved in the appropriate pairing of immunoglobulin or TCR gene segments. If the nonamer element represents the binding site for a specific component of the recombinational apparatus, why do individual nonamer sequences exhibit such apparent variability? One answer may be that these sequence differences lead to differences in affinity and consequently in the frequency of

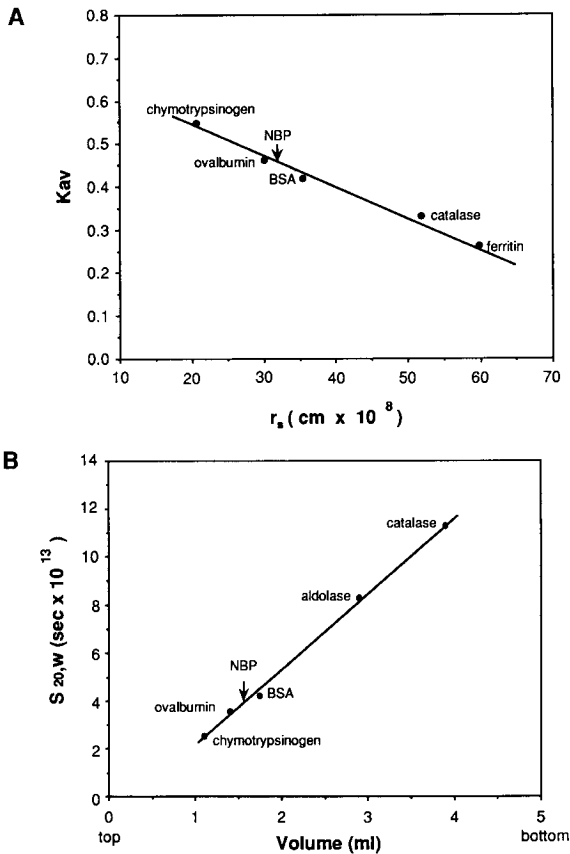


Figure 9. Hydrodynamic properties of NBP (A) Determination of Stokes' radius. Protein (4 μ g) from the DNA-Sepharose activity pool was fractionated by gel filtration on a Superose-12 FPLC HR 10/30 column, as described in Methods. Fractions (250 μ l) were assayed for DNA binding activity. Standards were detected by absorbance at 280 nm and by Bradford assay. The partition coefficient, K_{av} , was calculated for each species from its elution volume, the void volume of the column (V_0), and the total volume of the column (V_t), according to the equation

$$K_{av} = \frac{V_e - V_0}{V_t - V_0}$$

where $V_0 = 7.1$ ml, $V_t = 20.3$ ml, and V_e is the elution volume. V_0 was determined from the elution volume of dextran blue. The partition coefficients of standards are plotted as a function of Stokes' radius (r_s). The K_{av} for NBP, which corresponds to a Stokes' radius of 31.5×10^{-8} cm, is indicated on the standard curve. (B) Determination of Svedberg constant. Protein (1.5 μ g) from the nonspecific DNA-Sepharose activity pool was combined with standards and sedimented through a 4.9-ml, linear 15-30% glycerol gradient, as described in Methods. Fractions (100 μ l) were assayed for DNA binding activity, and standards were localized by SDS-polyacrylamide gel electrophoresis and staining with Coomassie blue. The sedimentation coefficients of standards ($S_{20,w}$) are plotted as a function of gradient volume. The position of NBP, which corresponds to a sedimentation coefficient of 4×10^{-13} sec, is indicated on the standard curve.

rearrangement. Such a mechanism would allow evolution to adjust the immunological repertoire by altering the sequence of a specific, *cis*-acting DNA element.

The DNA-binding protein that we have purified from

calf thymus recognizes a site, from 7 to 9 bp long, that coincides with the conserved nonamer. In addition to its presence in thymus, previous experiments identified a similar binding activity in extracts of two immature B-lymphoid cell lines but not in extracts of two fibroblastoid cell lines nor in an erythroleukemia cell line (Halligan and Desiderio, 1987 and B. Halligan and S.V. Desiderio, unpubl.). Thus, nonamer binding activity is not ubiquitous and appears to be preferentially expressed in lymphoid cells. On the basis of its sequence specificity, its preferential expression in lymphoid cells, and the profound impairment of rearrangement upon deletion of its binding site, it seems likely that NBP is a component of the recombinational apparatus that assembles immunoglobulin and TCR genes.

The mutations we placed within the nonamer element reduced the affinity of DNA fragments for NBP by a factor of 300- to at least 1000-fold, relative to the affinity of the wild-type fragment. These decreases in affinity correspond to decreases in binding free energy, ranging from 3.4 to >4.1 kcal/mole, and are comparable to the energy of a hydrogen bond ($\sim 3-6$ kcal/mole) (Saenger 1984). DNA fragments carrying 3-bp substitutions in the spacer region had ~ 10 -fold lower affinity for NBP than the wild-type substrate. Although the effects of these mutations are far smaller than the effects of mutations within the nonamer, this observation suggests that NBP may also interact, albeit weakly, with nucleotides within the spacer region. Interestingly, the spacer mutations that we tested overlap a 3-bp motif, PuGG, that lies 2 bp from the nonamer in most immunoglobulin and TCR spacer regions (see Fig. 1). Whether mutation of this motif affects rearrangement has not yet been determined.

Because immunoglobulin gene rearrangement is apparently initiated by specific cleavage of DNA at the junction of heptamer and coding sequences, we assayed NBP for specific (heptamer-directed) and nonspecific endonucleolytic activity and found NBP to possess neither (M. Li and S.V. Desiderio, unpubl.). If separate components of the recombinational apparatus act at the nonamer and the heptamer, as we have argued above, then the lack of associated endonucleolytic activity is not surprising. In light of these considerations, it will be of interest to determine whether the binding of NBP renders immunoglobulin recombinational signal sequences susceptible to site-specific endonucleolytic or strand-exchange activities.

NBP is apparently a rare component of thymic nuclei. On the basis of an upper limit of 11 μ g of protein recovered in the affinity chromatography step and a 22% overall yield of nonamer binding activity, we estimate that NBP represents $<0.005\%$ of protein in the crude nuclear extract. The scarcity of NBP is consistent with action at a relatively small number of sites, as would be expected were NBP to function in immunoglobulin and TCR gene rearrangement. Experiments described here further demonstrate the efficacy of recognition site affinity chromatography, in conjunction with conventional chromatographic methods, in purification of rare

DNA-binding proteins. The availability of affinity-purified NBP will facilitate functional studies, not least by providing peptide sequence information that should be useful in identifying the gene for NBP.

Methods

DNA constructions

Mutant and wild-type substrates for protein binding. Complementary, 24-nucleotide-long oligonucleotide pairs (NM1/NM2, NM3/NM4, NM5/NM6, NM7/NM8, NM9/NM10, NM11/NM12, NM13/NM14, NM60/NM61, and NM62/NM63; Fig. 2) were synthesized by the phosphoramidite method [Beaucage and Caruthers 1981], using the Applied Biosystems model 380B automated synthesizer. Tritylated oligonucleotides were isolated by reverse-phase chromatography on an octadecylsulfate (C18) column (Keystone Scientific). Complementary strands were annealed as described [Kadonaga and Tjian 1986], and the resulting duplex DNA, flanked by *EcoRI* and *BamHI* restriction sites, was introduced into the plasmid pUC13. The resulting plasmids were designated pNM1/2, pNM3/4, pNM5/6, pNM7/8, pNM9/10, pNM11/12, pNM13/14, pNM60/61, and pNM62/63. The DNA sequence of each of the cloned fragments was verified by the dideoxynucleotide chain-termination method [Sanger et al. 1977]. The construction of plasmid p λ 4.copy96, which contains multiple tandem copies of the wild-type λ 4 heptamer-spacer-nonamer sequence, has been described [Halligan and Desiderio 1987].

Substrates for immunoglobulin gene rearrangement. Construction of the plasmid pLJHCR-2 (Fig. 5), which was used as a substrate for rearrangement of wild-type immunoglobulin gene segments *in vivo*, has been described [Morzycka-Wroblewska et al. 1988]. The V_H segment of the plasmid pLJHCR-2 Δ N (Fig. 5) is identical to that of pLJHCR-2, except for an 18-bp deletion that encompasses 2 bp of spacer sequence, the entire nonamer, and 7 bp of 3'-flanking sequence.

All plasmids were propagated as monomers in the *E. coli* strain DH1 [Hanahan 1983]. Plasmid DNA was prepared by a detergent lysis procedure [Bothwell et al. 1981] and purified over two CsCl equilibrium gradients.

Biochemical assays

Specific binding of protein to DNA. The formation of specific DNA-protein complexes was detected by a modification of the electrophoretic mobility shift assay of Strauss and Varshavsky, as described [Halligan and Desiderio 1987]. Wild-type DNA probes used in these assays were either a 27-bp *DdeI*-*HinPI* or a 35-bp *EcoRI*-*HinPI* DNA fragment from p λ 4.copy96. To prepare the probes, p λ 4.copy96 was digested with the appropriate restriction endonucleases; the products were dephosphorylated by treatment with calf intestinal alkaline phosphatase and radiolabeled with [γ -³²P]ATP by T4 polynucleotide kinase. The labeled DNA probes, which carry the nonamer sequence of λ 4, were purified by polyacrylamide gel electrophoresis. The specific activity of the probes was typically ~2000 Ci/mmol.

Binding reactions (20 μ l) contained labeled DNA fragment, binding buffer [80 mM NaCl, 10 mM HEPES-NaOH (pH 7.5), 40 mM (NH₄)₂SO₄, 10 mM 2-mercaptoethanol, 1 mM EDTA, 4% (wt/vol) glycerol, 0.1% Triton X-100], 30 ng poly[d(I-C)] and 2 μ l of protein. In some experiments, a second competitor DNA was included at varying concentrations. Standard binding assays contained 0.1 ng of labeled DNA fragment; competition assays contained 0.01 ng of labeled fragment. After 30 min incubation

at room temperature, 5 μ l of assay buffer, supplemented with 50% glycerol, 0.1% bromphenol blue, and 0.1% xylene cyanol, was added. Reaction products were fractionated by electrophoresis through a 6% polyacrylamide slab gel (16 cm \times 1.5 mm) in 33.5 mM Tris-HCl (pH 7.5), 16.5 mM sodium acetate, and 5 mM EDTA. Electrophoresis was carried out at 110 V for 2 hr at room temperature. Radioactivity was detected by autoradiography. The amount of DNA present in specific DNA-protein complexes was quantitated indirectly, by densitometric analysis of autoradiographs, and directly, by scintillation spectrometry of the species excised from polyacrylamide gels; the values obtained by these two methods were similar. One unit of binding activity is defined as the amount required to bind 1 fmole of the 27-bp λ 4 probe in the standard electrophoretic assay.

SDS-polyacrylamide gel electrophoresis. SDS-polyacrylamide gel electrophoresis was performed as described [Laemmli 1970]. Sample buffer contained 100 mM 2-mercaptoethanol, 60 mM Tris-HCl (pH 6.8), 10% glycerol (wt/vol), 1% SDS, and 0.0005% bromphenol blue. Molecular weight standards (Sigma) were myosin (205 kD), β -galactosidase (116 kD), phosphorylase B (97.4 kD), BSA (66 kD), ovalbumin (45 kD), and carbonic anhydrase (29 kD). Protein was detected by silver stain [Rapid-Ag-Stain, ICN].

Recovery and renaturation of protein from SDS-polyacrylamide gels. After electrophoresis, lanes containing protein were cut into 10-mm-wide slices and crushed into a paste. Protein was eluted by incubating the gel paste with a buffer containing 150 mM NaCl, 20 mM HEPES-NaOH (pH 7.5), 5 mM DTT, 0.1 mM EDTA, 0.1% SDS, and 0.1 mg/ml BSA for 3 hr at room temperature. Protein was renatured by the method of Hager and Burgess [1980], with modifications. After elution from SDS-polyacrylamide, protein was precipitated by addition of 4 volumes of cold (-20°C) acetone and incubation for 45 min in a dry-ice/ethanol bath. Protein was collected by centrifugation at 12,000g for 15 min. The pellet was washed with a solution containing 80% acetone, 20% dilution buffer [150 mM NaCl, 20 mM HEPES-NaOH (pH 7.5), 5 mM DTT, 0.1 mM EDTA, and 0.1 mg/ml BSA], dried, dissolved in 5 μ l of dilution buffer supplemented with 6 M guanidine-HCl, and incubated at room temperature for 20 min. The solution was then diluted 50-fold with dilution buffer and incubated for an additional 12 hr at room temperature.

Glycerol gradient sedimentation. The active fraction from nonspecific DNA-Sepharose (0.75 μ g protein) was combined with sedimentation standards [40 μ g of each catalase (11.3S), aldolase (8.3S), BSA (4.2S), ovalbumin (3.6S), and chymotrypsinogen (2.5S)] in 100 μ l buffer G [0.5 M NaCl, 10 mM HEPES-NaOH (pH 7.5), 10 mM 2-mercaptoethanol, 5 mM EDTA, 0.5 mM PMSF] and layered onto a 4.9-ml linear 15–30% glycerol gradient that was prepared in buffer G. The gradient was spun in a Beckman SD 50.1 rotor at 45,000 rpm for 26 hr at 4°C. After centrifugation, 100 μ l fractions were collected. A 2- μ l aliquot of each fraction was assayed for specific DNA-binding activity. To determine positions of the standards, a 15- μ l aliquot of each fraction was analyzed by SDS-polyacrylamide gel electrophoresis and staining with Coomassie blue.

Superose-12 gel filtration chromatography. An aliquot of the active fraction from nonspecific DNA-Sepharose (4 μ g of protein in 100 μ l) was loaded onto a Superose-12 FPLC HR 10/30 column (Pharmacia), which had been equilibrated in buffer B (see below), supplemented with 0.5 M NaCl. Protein was eluted in the same buffer at a flow rate of 12 ml/hr, 5 μ l of each 250 μ l

fraction was assayed for binding activity. Standards used were ferritin (r_s 61.0×10^{-8} cm), catalase (r_s 52.2×10^{-8} cm), BSA (r_s 35.5×10^{-8} cm), ovalbumin (r_s 30.5×10^{-8} cm), and chymotrypsinogen (r_s 20.9×10^{-8} cm).

Protein assay. Protein concentrations were determined by the method of Bradford (1976). BSA was used as a standard.

DNA assay. DNA fragments used as probes and competitors in binding experiments were quantitated fluorimetrically after staining with the dye Hoechst 33258.

Purification of NBP

Materials. Calf thymus was obtained from a local slaughterhouse. T4 polynucleotide kinase and T4 DNA ligase were obtained from New England Biolabs. Sepharose CL-2B was purchased from Pharmacia, Bio-Rex-70 (200–400 mesh) from Bio-Rad, and heparin agarose from BRL. CNBr was obtained from Boehringer-Mannheim. Radioisotopes were supplied by Dupont-NEN.

Measurement of NaCl concentration. The concentration of NaCl in column fractions was determined by measuring conductivity in comparison to a standard curve.

Buffers used in the purification. Buffer B is 50 mM NaCl, 20 mM HEPES-NaOH (pH 7.5), 10 mM 2-mercaptoethanol, 2 mM EDTA, 0.2 mM PMSF, 10% (wt/vol) glycerol. Buffer E is 250 mM sucrose, 100 mM NaCl, 50 mM HEPES-NaOH (pH 7.5), 10 mM 2-mercaptoethanol, 0.2 mM PMSF. Buffer H is 250 mM sucrose, 50 mM HEPES-NaOH (pH 7.5), 25 mM KCl, 10 mM 2-mercaptoethanol, 5 mM $MgCl_2$, 0.2 mM PMSF. Buffer N is 80 mM NaCl, 40 mM $(NH_4)_2SO_4$, 10 mM HEPES-NaOH (pH 7.5), 10 mM 2-mercaptoethanol, 5 mM EDTA, 0.2 mM PMSF, 4% (wt/vol) glycerol. Buffer P is 50 mM Tris-HCl (pH 8.0), 10 mM $MgCl_2$, 5 mM DTT, 1 mM spermidine, 0.1 mM EDTA. Ligation buffer is 50 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 10 mM DTT, 1 mM spermidine, 2 mM ATP, 0.01% (wt/vol) BSA.

Preparation of nonspecific DNA–Sepharose. DNA was conjugated to Sepharose as described by Alberts and Herrick (1971), with modifications. Sepharose CL-2B (wet volume 25 ml) was washed with water and resuspended in water to give a 50-ml slurry. Activation was performed in an ice-water bath. Cyanogen bromide (4 gm) was dissolved in 3 ml dimethylformamide and added dropwise to the slurry. The pH value of the reaction was maintained between 11.0 and 11.5 by addition of 5 N NaOH. When the pH value became steady (after ~5 min on ice), the reaction was stopped with 400 ml of ice-cold water. The resulting CNBr-derived Sepharose was washed with 500 ml of water followed by 500 ml of 10 mM KH_2PO_4/K_2HPO_4 (pH 8.0) buffer. The coupling reaction was immediately carried out by resuspending the activated Sepharose in 45 ml 10 mM KH_2PO_4/K_2HPO_4 buffer (pH 8.0) containing 6 mg of sheared salmon sperm DNA and incubating at room temperature for 12 hr on a rotary shaker. Coupling efficiency was estimated at 76 μ g DNA/ml matrix. DNA–Sepharose was washed with 500 ml water, followed by 500 ml 0.1 M ethanolamine-HCl (pH 8.0). Any activated sites remaining were blocked by incubation with 40 ml of 0.1 M ethanolamine-HCl (pH 8.0) at room temperature for 5 hr on a rotary shaker.

Preparation of the nonamer affinity column. The 27-mer oligonucleotides SD7 (5'-AATTCAGGCAGGTTTTTGTAAGGGGG-3') and SD8 (5'-GATCCCCCTTTACAAAACCTGCTTG-3') (300 μ g of each) were annealed in 100 μ l P buffer by

incubation at 100°C for 2 min, 65°C for 10 min, 37°C for 10 min, and room temperature for 10 min. The 5' termini were phosphorylated by addition of 10 μ l 30 mM $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (0.1 μ Ci/mmol) and 2.5 μ l T4 polynucleotide kinase (400 U/ μ l), followed by incubation at 37°C for 2 hr. The reaction was stopped by addition of 50 μ l 50 mM EDTA, 2% SDS. After extraction with phenol and precipitation in ethanol, the DNA pellet was dried in vacuo and resuspended in 100 μ l ligation buffer. T4 DNA ligase (4 μ l at 10 Weiss U/ μ l) was added, and the reaction was incubated at 14°C for 10 hr. Ligation products were extracted with phenol, precipitated in ethanol, dried, and dissolved in 200 μ l 10 mM KH_2PO_4/K_2HPO_4 (pH 8.0), 2 mM EDTA. Concatemered DNA was coupled to 8 ml of CNBr-activated Sepharose, as described above. The efficiency of coupling, based on the specific activity of the DNA and the radioactivity retained on the resin, was 83% (61 μ g of DNA per milliliter of resin).

Preparation of nuclear extract. All purification steps were performed at 4°C. Nuclei were prepared from 200 grams calf thymus, as described (Halligan and Desiderio 1987), except that cells were lysed in a Waring blender at medium speed in 500 ml homogenization buffer (buffer H). Nuclei were collected by centrifugation at 1800g for 7 min, incubated with 100 ml of buffer E at 0°C for 20 min, and pelleted by centrifugation at 2000g for 10 min. The supernatant was designated nuclear extract. The concentration of protein in the nuclear extract was generally 10–15 mg/ml.

Ammonium sulfate precipitation. To the nuclear extract, 0.3 grams of ammonium sulfate per ml of extract was slowly added with constant stirring. The precipitate was collected by centrifugation at 18,000g for 30 min. The pellet was resuspended in 50 ml buffer B and dialyzed for 10 hr against buffer B. A precipitate that formed during dialysis was removed by centrifugation at 15,000g for 30 min. The supernatant was designated crude extract. The concentration of protein in crude extract was typically 15–20 mg/ml.

Bio-Rex-70 chromatography. The crude extract was diluted three-fold with buffer B and loaded onto a Bio-Rex-70 column (2.5 \times 15 cm, 74 ml), which had been washed with 600 ml 2.5 M NaCl, 100 mM HEPES-NaOH (pH 7.5) and equilibrated with 740 ml buffer B. After loading, the column was washed with 150 ml buffer B. The DNA binding activity was eluted with a linear gradient of NaCl from 50 to 600 mM (850 ml). The specific DNA binding activity eluted as a single peak at 280 mM NaCl. Fractions eluting between 260 and 300 mM NaCl were pooled and designated the Bio-Rex-70 activity pool (protein concentration, 1.5–2.5 mg/ml; volume, 36 ml).

Heparin agarose chromatography. The Bio-Rex-70 activity pool was diluted twofold in buffer B without NaCl. This diluted pool was then loaded onto a heparin–agarose column (1.5 \times 20 cm, 35 ml), which had been washed with 200 ml of 2.5 M NaCl in buffer B and equilibrated with 350 ml of buffer B. After loading, the matrix was washed with 200 ml of buffer B. DNA binding activity was eluted with a 650-ml linear gradient of NaCl from 50 mM to 2.3 M. Specific binding activity eluted as a single peak at 780 mM NaCl. Fractions containing activity were pooled (protein concentration, 100–200 μ g/ml; volume, 45 ml) and dialyzed for 3 hr against 500 ml buffer B containing 5 mM EDTA.

Nonspecific DNA–Sepharose chromatography. The dialyzed heparin agarose activity pool was loaded onto a nonspecific DNA–Sepharose column (1.5 \times 8 cm, 14 ml), which was equilibrated with 300 ml buffer B, supplemented with 5 mM EDTA (B-EDTA). After loading, the matrix was washed with 100 ml

B-EDTA and DNA binding activity was eluted with a linear gradient of NaCl from 50 mM to 2.5 M. The specific binding activity eluted as a single peak at 300 mM NaCl. Fractions containing DNA binding activity were pooled (protein concentration, 1.5–2.5 µg/ml; volume, 7 ml) and dialyzed against 500 ml of binding buffer.

Recognition site affinity chromatography. The nonamer DNA affinity column (5 × 0.5 cm, 1 ml) was washed with 20 ml 2 M NaCl and equilibrated with 30 ml buffer N. The activity pool from nonspecific DNA–Sephacrose was diluted threefold in buffer B without NaCl. The activity was concentrated by loading onto Bio-Rex-70 and eluting in the reverse direction with buffer B containing 400 mM NaCl. Overall, the activity from nonspecific DNA–Sephacrose was concentrated five fold in this step, without detectable loss of activity. The concentrated material was dialyzed against binding buffer, and poly[d(I-C)] was added to a final concentration of 1.5 µg/ml. The resulting solution was loaded under gravity onto the nonamer DNA affinity column, and the flowthrough was reloaded three times. After loading, the column matrix was washed with 15 ml of binding buffer containing 205 mM NaCl. The activity was eluted with 10 ml of binding buffer containing 580 mM NaCl.

Assay for immunoglobulin gene rearrangement

Cell lines. The Harvey murine sarcoma virus-transformed cell line HAFTL-1 (Alessandrini et al. 1987) was cultured in RPMI 1640 medium supplemented with 10% fetal calf serum and 50 µM 2-mercaptoethanol (RPMI-10). The cell line Ψ-2 (Mann et al. 1983) was propagated in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (DEME-10). Lymphoid cells were cloned by limiting dilution.

Packaging of recombinant retrovirus and transmission to lymphoid cells. Packaging of the pLJHCR-2 and pLJHCR-2ΔN constructs in the Ψ-2 cell line and transmission to the lymphoid cell line HAFTL-1 were performed as described (Desiderio and Wolff 1988), except that G418 was used at a concentration of 1.5 mg/ml to select for lymphoid cells carrying provirus.

Assay for rearrangement of the LJHCR-2 substrate in the HAFTL-1 cell line. The assay has been described in detail elsewhere (Desiderio and Wolff 1988; Morzycka-Wroblewska et al. 1988) and is summarized here. HAFTL-1 cells were infected in duplicate with either the LJHCR-2 or the LJHCR-2ΔN virus. After 10–14 days of selection in RPMI-10 supplemented with G418 (1.5 mg/ml), G418-resistant [G418^r] derivatives were apparent. G418^r cell populations were expanded to 2 × 10⁷ cells and cloned by limiting dilution in 96-well microtiter plates in RPMI-10 supplemented with 1.5 mg/ml G418 (0.1 ml per well). After 10–14 days, clones were transferred to 1 ml of the same medium, when cells had achieved a density of ~1 × 10⁶ cells/ml, cultures were split 1 : 2 and permitted to expand to a density of ~1 × 10⁶ to 2 × 10⁶ cells/ml.

Genomic DNA was isolated from individual cell clones and assayed for proviral rearrangement. Cells (1 ml at 1 × 10⁶ to 2 × 10⁶/ml) were harvested by centrifugation at 3000 rpm for 5 min in a Dupont Microspin 24S centrifuge. After discarding the supernatant, cell pellets were lysed by addition of 200 µl 100 mM NaCl, 10 mM Tris-Cl (pH 8.0), 1 mM EDTA, 1% SDS, and 0.4 mg/ml proteinase K and incubated for 16 hr at 37°C. Samples were extracted twice with phenol and twice with ethyl ether. DNA was precipitated by addition of 2 volumes absolute ethanol. Pellets were washed with 1 ml of 70% ethanol, dried, and dissolved in 200 µl *KpnI* restriction endonuclease buffer [10 mM NaCl, 10 mM Tris-Cl (pH 7.5), 10 mM MgCl₂, 1 mM DTT,

100 µg/ml BSA]. After the addition of 3 µl of *KpnI* (30 units), samples were incubated for 16 hr at 37°C. Reactions were supplemented with 1 µl (10 units) additional *KpnI*, and incubation was continued for 3 hr at 37°C. Reactions were stopped by addition of 50 µl of 2% SDS, 50 mM EDTA. After extraction with phenol and precipitation in ethanol, samples were dissolved in 50 µl of 10% glycerol, 5 mM EDTA, 0.1% SDS, 0.01% bromophenol blue, and 0.01% xylene cyanol. One-half of each sample (~20 µl) was analyzed for rearrangement of proviral DNA by agarose gel electrophoresis, transfer to nitrocellulose, and hybridization to *gpt*- and *neo*-specific DNA probes as described (Desiderio and Wolff 1988).

Acknowledgments

We thank Brian Halligan for advice in the early stages of this work, Clark Riley and Brian Byers for oligonucleotide synthesis, our colleagues in the Department of Molecular Biology and Genetics for stimulating discussions, and Thomas Kelly for his review of the manuscript. The work was supported by grant CA-16519 from the National Institutes of Health and the Howard Hughes Medical Institute.

References

- Akira, S., K. Okazaki, and H. Sakano. 1987. Two pairs of recombination signals are sufficient to cause immunoglobulin V-D-J joining. *Science* **238**: 1134–1138.
- Alberts, B. and G. Herrick. 1971. DNA-cellulose chromatography. *Methods Enzymol.* **21**: 198–217.
- Alessandrini, A., J.H. Pierce, D. Baltimore, and S.V. Desiderio. 1987. Continuing rearrangement of immunoglobulin and T-cell receptor genes in Ha-*ras*-transformed lymphoid progenitor cell line. *Proc. Natl. Acad. Sci.* **84**: 1799–1803.
- Alt, F.W., T.K. Blackwell, and G.D. Yancopoulos. 1987. Development of the primary antibody repertoire. *Science* **238**: 1079–1087.
- Beaucage, S.L. and M.H. Caruthers. 1981. Deoxynucleoside phosphoramidites—A new class of key intermediates for deoxypolynucleotide synthesis. *Tetrahed. Lett.* **22**: 1859–1862.
- Bernard, O., N. Hozumi, and S. Tonegawa. 1978. Sequences of mouse immunoglobulin light chain genes before and after somatic changes. *Cell* **15**: 1133–1144.
- Bothwell, A.L.M., M. Paskind, M. Reth, T. Imanishi-Kari, K. Rajewsky, and D. Baltimore. 1981. Heavy chain variable region contribution to the NP^b family of antibodies: Somatic mutation evident in a q2a variable region. *Cell* **24**: 625–637.
- Bradford, M. 1976. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem* **73**: 248–254.
- Chen, Y.-H., N. Gascoigne, J. Kavaler, D.I. Cohen, and M.M. Davis. 1984. Somatic recombination in a murine T-cell receptor gene. *Nature* **309**: 322–326.
- Cohen, J.B. and D. Givol. 1983. Allelic immunoglobulin V_H genes in two mouse strains: Possible germline gene recombination. *EMBO J.* **2**: 2013–2018.
- Concannon, P., E. Lai, M. Klein, S. Siu, E. Strauss, L. Pickering, P. Kung, R. Gatti, and L. Hood. 1986. Human T-cell receptor genes: Organization, diversity, and polyphorphism. *Cold Spring Harbor Symp. Quant. Biol.* **51**: 785–795.
- Davis, M. 1985. Molecular genetics of the T-cell receptor beta chain. *Annu. Rev. Immunol.* **3**: 537–560.
- Desiderio, S.V. and K.R. Wolff. 1988. Rearrangement of exogenous immunoglobulin V_H and DJ_H gene segments after retroviral transduction into immature lymphoid cell lines. *J. Exp. Med.* **167**: 372–389.

- Early, P., H. Huang, M. Davis, K. Calame, and L. Hood. 1980. An immunoglobulin heavy chain variable region is generated from three segments of DNA: V_H , D , and J_H . *Cell* **19**: 981–992.
- Hager, D.A. and R.R. Burgess. 1980. Elution of proteins from sodium dodecyl sulfate-polyacrylamide gels, removal of sodium dodecyl sulfate, and renaturation of enzymatic activity: Results with sigma subunit of *Escherichia coli* RNA polymerase, wheat germ DNA topoisomerase, and other enzymes. *Anal. Biochem* **109**: 76–86.
- Halligan, B.D. and S.V. Desiderio. 1987. Identification of a DNA-binding protein that recognizes the nonamer recombinational signal sequence of immunoglobulin genes. *Proc. Natl. Acad. Sci.* **84**: 7019–7023.
- Hanahan, D. 1983. Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Bio.* **166**: 557–580.
- Hayday, A., et al. 1985. Structure, organization, and somatic rearrangement of T-cell gamma genes. *Cell* **40**: 259–269.
- Heinrich, G., A. Traunecker, and S. Tonegawa. 1984. Somatic mutation creates diversity in the major group of mouse immunoglobulin kappa light chains. *J. Exp. Med.* **159**: 417–435.
- Hesse, J., M. Lieber, M. Gellert, and K. Mizuuchi. 1987. Extrachromosomal DNA substrates in pre-B cells undergo inversion or deletion at immunoglobulin V-D-J joining signals. *Cell* **49**: 775–783.
- Ichihara, Y., H. Matsuoki, and Y. Kurosawa. 1988. Organization of human immunoglobulin heavy chain diversity gene loci. *EMBO J.* **7**: 4141–4150.
- Kabat, E.A., T.T. Wu, M. Reid-Miller, H.M. Perry, and K.S. Gottesman. 1987. *Sequences of proteins of immunological interest*, 4th ed. U.S. Dept. of Health and Human Services, Washington, D.C.
- Kadonaga, J.T. and R. Tjian. 1986. Affinity purification of sequence-specific DNA binding proteins. *Proc. Natl. Acad. Sci.* **83**: 5889–5893.
- Kleinfield, R., R.R. Hardy, D. Tarlinton, J. Dangel, L.A. Herzenberg, and M. Weigert. 1986. Recombination between an expressed immunoglobulin heavy-chain gene and a germline variable gene segment in a $Ly1 + B$ -cell lymphoma. *Nature* **322**: 843–846.
- Laemmli, U.K. 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**: 680–685.
- LeFranc, M., A. Forster, R. Baer, M.A. Stinson, and T.H. Rabbits. 1986. Diversity and rearrangement of the human T-cell rearranging gamma genes: Nine germ-line variable genes belonging to two subgroups. *Cell* **45**: 237–246.
- Lewis, S., A. Gifford, and D. Baltimore. 1985. DNA elements are asymmetrically joined during site-specific recombination of kappa immunoglobulin genes. *Science* **228**: 677–685.
- Litman, G.W., L. Berger, K. Murphy, R. Litman, K. Hinds, and B.W. Erickson. 1985. Immunoglobulin V_H gene structure and diversity in *Heterodontus*, a phylogenetically primitive shark. *Proc. Natl. Acad. Sci.* **82**: 2082–2086.
- Mann, R., R.C. Mulligan, and D. Baltimore. 1983. Construction of a retrovirus packaging mutant and its use to produce helper-free defective retroviruses. *Cell* **33**: 153–159.
- Max, E.E., J.G. Seidman, and P. Leder. 1979. Sequences of five potential recombination sites encoded close to an immunoglobulin kappa constant region gene. *Proc. Natl. Acad. Sci.* **76**: 3450–3454.
- Max, E.E., J.G. Seidman, H. Miller, and P. Leder. 1980. Variation in the crossover point of kappa immunoglobulin gene V-J recombination: Evidence from cryptic genes. *Cell* **21**: 793–799.
- Morzycka-Wroblewska, E., F.E.H. Lee, and S.V. Desiderio. 1988. Unusual immunoglobulin gene rearrangement leads to replacement of recombinational signal sequences. *Science* **242**: 261–263.
- Okazaki, K., D.D. Davis, and H. Sakano. 1987. T-cell receptor beta gene sequences in the circular DNA of thymocyte nuclei: Direct evidence for intermolecular DNA deletion in V-D-J joining. *Cell* **49**: 477–485.
- Reth, M., P. Gehrman, E. Petrac, and P. Wiese. 1986. A novel V_H -to- DJ_H mechanism in heavy-chain-negative (null) pre-B cells results in heavy chain production. *Nature* **322**: 840–842.
- Rosenfeld, P.J. and T.J. Kelly. 1986. Purification of nuclear factor I by DNA recognition site affinity chromatography. *J. Biol. Chem.* **261**: 1398–1408.
- Saenger, W. 1984. *Principles of nucleic acid structure*, Springer-Verlag, New York, New York.
- Sakano, H., K. Huppi, G. Heinrich, and S. Tonegawa. 1979. Sequences at the somatic recombination sites of immunoglobulin light-chain genes. *Nature* **280**: 288–294.
- Sakano, H., R. Maki, Y. Kurosawa, W. Roeder, and S. Tonegawa. 1980. Two types of somatic recombination are necessary for the generation of a complete immunoglobulin heavy chain gene. *Nature* **286**: 676–683.
- Sanger, F., S. Nicklen, A.R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci.* **74**: 5463–5467.
- Siegel, L.M. and K.J. Monty. 1966. Determination of molecular weights and frictional ratios of proteins in impure systems by use of gel filtration and density gradient centrifugation. Application to crude preparations of sulfites and hydroxylamine reductases. *Biochim. Biophys. Acta* **112**: 346–362.
- Tonegawa, S. 1983. Somatic generation of antibody diversity. *Nature* **302**: 575–581.
- Tonegawa, S., A.M. Maxam, R. Tizard, O. Bernard, and W. Gilbert. 1978. Sequence of a mouse germ-line gene for a variable region of an immunoglobulin light chain. *Proc. Natl. Acad. Sci.* **75**: 1485–1489.
- Yancopoulos, G.D., S.V. Desiderio, M. Paskind, J.F. Kearney, D. Baltimore, and F.W. Alt. 1984. Preferential utilization of the most J_H -proximal V_H gene segments in pre-B cell lines. *Nature* **311**: 727–733.
- Yancopoulos, G.D., K. Blackwell, H. Suh, L. Hood, and F.W. Alt. 1986. Introduced T-cell receptor variable region gene segments recombine in pre-B cells: Evidence that B- and T-cells use a common recombinase. *Cell* **44**: 251–259.