BIOORGANIC CHEMISTRY 23, 482–498 (1995)

Structural Patterns in Anti-DNA Autoantibodies: A Molecular Modeling Study[1,2]

MELISSA A. EAGAN, JENNIFER M. NORRIS, BRIAN C. COOPER, AND GARY D. GLICK[3]

Department of Chemistry, University of Michigan, Ann Arbor, Michigan 48109-1055

Received May 12, 1995

Anti-DNA autoantibodies are the hallmark of the autoimmune disease systemic lupus erythematosus. Although these antibodies are both diagnostic and pathogenic, little is known about their structures and the manner in which they recognize DNA antigens. To address the first of these points we have predicted the three-dimensional structures of 40 monoclonal anti-DNA F(ab) fragments derived from lupus-prone mice. These antibodies were chosen to encompass several different autoimmune strains along with the known variable region gene families that encode anti-DNA. We find that the structures of the antigen binding regions of these antibodies fall into three main classes, irrespective of both the mouse strain and genetic origins of the antibody. Specifically, high-affinity anti-ssDNA appear to possess a narrow channel that is presumably used for ligand recognition, whereas the binding site on anti-dsDNA is an open surface that is large enough to accommodate a DNA duplex. These findings provide structural data to support the hypothesis that anti-DNA arise by DNA-driven B cell activation and clonal expansion.

INTRODUCTION

The autoimmune disorder systemic lupus erythematosus (SLE) afflicts more than 500,000 Americans (*1*). Women are affected in 90% of cases and African-Americans are several times more likely to develop the disease than Caucasians (*2*). The onset of SLE occurs most commonly between the ages of 17 and 35, and patients who develop lupus nephritis (i.e., kidney damage) experience a 10-year mortality rate of 27% (*3*). Despite SLEs high morbidity and mortality, specific targets for drug intervention have not been identified and current therapies employ nonspecific immunosupressive agents (*4, 5*). These drugs are themselves toxic and in compromised patients they can be lethal.

In human and murine SLE, IgG anti-DNA autoantibodies are both diagnostic and pathogenic (*6*). The involvement of anti-DNA in SLE has been established

[1] Warmly dedicated to Jeremy R. Knowles on the occasion of his 60th birthday.

[2] Supported by NIH Grant 46831. Research Fellowships from General Electric (M.A.E. and J.M.N.) and Eli Lilly (B.C.C.) are gratefully acknowledged. G.D.G. is the recipient of a National Arthritis Foundation Arthritis Investigator Award, an American Cancer Society Junior Faculty Research Award, a National Science Foundation Young Investigator Award, A Camille Dreyfus Teacher-Scholar Award, and a Research Fellowship from the Alfred P. Sloan Foundation.

[3] To whom correspondence should be addressed.

0045-2068/95 $12.00

through (i) correlation of serum antibody levels with disease activity, (ii) isolation of anti-DNA · DNA immune complexes from sites of renal damage, and (iii) the finding that up-regulation of anti-DNA accelerates disease whereas down-regulation of anti-DNA suppresses or prevents disease (*7–11*). Defining the structural and biochemical factors that govern the formation and stability of anti-DNA · DNA complexes is therefore an essential first step toward understanding immune-complex-mediated tissue injury (*12*).

The molecular nature of the DNA epitopes targeted by lupus anti-DNA has only been revealed for a single F(ab) derived from thymidine-specific anti-ssDNA BV04-01 (IgG2b, κ) liganded to $d(pT)_3$ (*13*). In this complex, the trinucleotide is stabilized by hydrophobic interactions where the bases stack between aromatic residues on the protein, and by a single salt bridge to the phosphodiester backbone. The relevance of this crystal structure with respect to binding in solution has been confirmed by Swanson *et al.*, who used DNA footprinting to map the autoreactive epitopes on model DNA ligands (*14*).

In contrast to anti-ssDNA, there are no high-resolution structural data available for anti-DNA · B-DNA complexes so that a detailed molecular understanding of antibody · dsDNA interactions does not yet exist. In one approach to this problem, Stollar and co-workers studied the binding of anti-dsDNA H241 in competition binding experiments using short DNA duplexes and analogs containing modified bases (*15*). These experiments showed that H241 recognizes epitopes in both the major and minor grooves as well as the phosphate backbone of dsDNA. Several groups have used comparative sequence alignments to determine variable (V) region residues that are products of somatic mutation (*16, 17*). Presumably these residues are selected to increase binding to dsDNA. Such predictions have been tested through site-directed mutagenesis by comparing the relative affinity and specificity of the mutants and the wild-type antibodies for DNA ligands (*18–21*).

Although more is known about anti-DNA structure than is known about the nature of DNA epitopes and the interactions that stabilize anti-DNA · DNA complexes, such information is still limited. Crystal structures exist for only two lupus anti-DNA F(ab)'s, BV04-01 (*13*) and HEd10 (*22*), both of which recognize oligo(dT). However, the V-region amino acid sequences for over 250 lupus anti-DNA autoantibodies have been published (*16, 17*). This database has been used to compare anti-DNA primary structure (with emphasis on dissecting patterns of somatic mutations). But extensive use of these data to investigate the three-dimensional (3-D) structure of anti-DNA has not been reported.

Empirical-based molecular modeling procedures, such as those described by Chothia and Lesk, now exist that enable the accurate prediction of complementarity determining region (CDR) conformations (*23*). With the exception of CDR3 of the heavy chain (denoted here as HCDR3) which is highly variable, empirically modeled CDR loop conformations generally show <1.2 Å root mean square deviation (RMSD) for the backbone atoms when compared to the X-ray structure of the corresponding F(ab) (reviewed in *24*). Here we report the predicted 3-D structures for 40 murine antibodies that are representative of the anti-DNA commonly expressed in lupus-prone mice. We find that the antigen recognition sites for antibodies that bind ssDNA are topologically distinct from those that recognize dsDNA.

Specifically, anti-ssDNA CDRs appear to form a long narrow channel whereas the binding sites on anti-dsDNA seem to be flat open surfaces. These conclusions are independent of both the strain of mouse from which the antibody is derived and the V-gene family that encodes it. The results of this study are discussed in terms of both the immune response to DNA associated with SLE and the anti-DNA pathogenicity.

MATERIALS AND METHODS

Molecular modeling calculations were performed on a Silicon Graphics 4D/360 GTX computer running INSIGHT, DISCOVER, HOMOLOGY, and DELPHI software (Biosym Technologies, San Diego, CA). Energy minimization and molecular dynamics simulations were performed *in vacuo* using the AMBER all atom forcefield and a distance-dependent dielectric constant.

F(ab) structure predictions followed the procedures outlined in the Biosym software guide (*25*). Briefly, all of the antibody sequences in the Brookhaven Protein Database (PDB; vide infra) were aligned with the primary sequence of the antibody to be folded. The three antibodies with the highest sequence homology were then selected as reference proteins. Segments of contiguous polypeptide common to the three reference proteins (excluding the hypervariable loops) were defined as structurally conserved regions (SCRs). To generate one structure from the three reference proteins (this structure encompasses the entire F(ab) except for the hypervariable loops), the α-carbon backbone of the SCRs was superimposed and the differences between them were minimized (i.e., best fit superimposition). To fill gaps in the framework region that were not defined by SCRs, the entire PDB was searched for sequences that were homologous to the gaps. These segments were then grafted onto the SCRs to complete the F(ab) framework. Side chains were added to all positions and the model was refined by 1500 iterations of steepest descent energy minimization to remove close contacts.

To complete the model, the canonical designations of the CDRs were assigned using the definitions given by Chothia and Lesk (*23*). The PDB was searched to find those antibodies that possessed the same canonical structure as the loop in question. Candidate loops were grafted onto the model and the one with the best fit (as judged by the highest sequence homology and loop length) was retained. Often there were sequence differences between the antibody under investigation and a particular hypervariable loop selected from the PDB. In those cases we reexamined the loop candidates that were not chosen to see if they had the requisite amino acid at the position(s) in question that could be used to replace the appropriate site on the selected loop. If no homology existed for a particular position, we added the correct side chain manually so that it did not contact other residues on the model. This latter procedure proved most effective for HCDR3 which does not fit into a canonical class. Each "final" loop was then subjected to 500 cycles of conjugate gradient minimization to remove nonbonded contacts, followed by 15 ps of molecular dynamics refinement (1 fs timestep, 300 K). The final RMSD for

each of the loops is typically ≤ 0.3 kcal/Å^2. This entire modeling process generally required 4 h of CPU time and 4 h of interactive time for each antibody.

The electrostatic potential of the modeled antibodies (F_V region only) was calculated using the DELPHI program (*26*). The solvent ionic strength was set to 0.145 M Na^+ with an ionic radius of 2.00 Å. Formal charges on the protein were employed based on a solvent pH value of 7.4 (Arg, +1; Lys, +1; His, +0.5; Asp, −1; Glu, −1). The dielectric constant was set to 2.0 for the F_V's for all points within the solvent accessible surface area (SASA; based on a 1.8-Å probe) and could increase to 5.0 to account for electric polarizability. The dielectric constant was set to 80 for the solvent.

RESULTS

Selection of anti-DNA and reference mAb's. Four criteria were used to select the monoclonal anti-DNA sequences for folding analysis (Table 1). First, only mAb's whose complete V-region sequences are known were considered. Second, mAb's that represent a range of antigenic specificities were chosen: anti-ssDNA, the largest group of lupus anti-DNA; anti-dsDNA, which are relatively uncommon; and antibodies that bind both single- and double-stranded DNA ("cross-reactive") which compose a moderately sized family of anti-DNA mAb's (*41*). Because each of the mAb's within the three specificity groups are encoded by a variety of genes, our third criterion was to select antibodies that represent each of these gene families and their subfamilies (*16*, *17*; Table 2). We concentrated our analysis on all of the reported V_H gene families as opposed to all of the anti-DNA V_L genes since the heavy chain is believed to have a greater influence on DNA binding. Nevertheless, our panel also spans most of the known anti-DNA V_L gene families. Lastly, all anti-DNA mAb's that are believed to accelerate kidney damage (i.e., glomerulonephritis) in nonautoimmune mice were included in our analysis.

Empirical structure predictions are highly dependent on the reference protein(s). If only a small group of crystal structures were used in modeling our panel of F(ab)'s, all of the resulting structures would be expected to be similar, making it difficult to compare the modeled antibodies (*24*). Twenty-five crystal structures in the PDB were used as reference proteins for the folding calculations. As indicated in Table 3, between 60 and 65% of the 25 reference proteins were used in determining at least one of the CDR conformations in our panel of 40 antibodies, although typically for a given loop (e.g., LCDR1 of the anti-ssDNA) a smaller fraction of the reference proteins (40–50%) was used in the modeling calculations. This outcome is consistent with the observation that CDRs of most anti-DNA have similar canonical structures (Table 3) and that anti-DNA are encoded by a recurrent set of (germline) genes. Overall, it appears that enough diversity exists that the modeled F(ab)'s will not be biased by a small group of reference proteins.

Anti-ssDNA. Before work was initiated on our panel of mAb's, we tested the ability of the modeling protocol to reproduce the crystal structure of anti-ssDNA BV04-01 in its unliganded form. For this exercise LCDR1, LCDR2, LCDR3, and HCDR2 came from 4FAB; HCDR1 was derived from 2FBJ, and HCDR3 from

TABLE 1
Anti-DNA Used for Homology Modeling

Antibody	Strain	Age (months)	Isotype	Specificity[a] ssDNA	Specificity[a] dsDNA	Ref.
				ssDNA specific		
111-55	(NZB × NZW)F_1	6–8	IgG2a	+++	-	27
111-34	(NZB × NZW)F_1	6–8	IgG2a	+++	....	27
10-43	(NZB × NZW)F_1	6–8	IgG	++	—	27
165-14	(NZB × NZW)F_1	6–8	IgG1	++	....	27
165-3	(NZB × NZW)F_1	6–8	IgG1	++	—	27
DNA2	(NZB × NZW)F_1	10	IgG2a	++	—	28
DNA12	(NZB × NZW)F_1	10	IgG2a	++	—	28
33-2	MRL/lpr/lpr	4	IgG2b	++	—	29
4H8	MRL/lpr/lpr	6	IgG3	++	—	30
H161	MRL/lpr/lpr	6	IgG3	++	—	31
17-s130	(NZB × NZW)F_1	6–8	IgG1	++	....	27
bfd03	(NZB × NZW)F_1	NA	IgG2a	++	—	32
A52	(NZB × NZW)F_1	9	IgG2b	++	—	33
17-s83	(NZB × NZW)F_1	6–8	IgG1	++	—	27
1A11	MRL/lpr/lpr	6	IgG3	+	—	30
CAL27	MRL/lpr/lpr	NA	IgG2a	+	....	34
				dsDNA specific		
43-2	MRL/lpr/lpr	4	IgG2a	—	+++	29
163-47	(NZB × NZW)F_1	6–8	IgG2b	—	+++	27
BV16-13	(NZB × NZW)F_1	NA	IgG2a	—	+++	35
163-72	(NZB × NZW)F_1	<6	IgM	—	++	27
163-42	(NZB × NZW)F_1	<6	IgM	—	++	27
BV17-31	(NZB × NZW)F_1	NA	IgG2b	—	++	35
N4-1	(NZB × NZW)F_1	7	IgG2a	—	+	36
N4-2	(NZB × NZW)F_1	7	IgG2a	—	+	36
N4-16	(NZB × NZW)F_1	4	IgG2a	—	+	36
N4-18	(NZB × NZW)F_1	4	IgG2a	—	+	36
N4-27	(NZB × NZW)F_1	4	IgG2a	—	+	36
N4-36	(NZB × NZW)F_1	4	IgG2a	—	+	36
N4-10	(NZB × NZW)F_1	7	IgG2a	—	+	37
				Cross-reactive		
185-33	(NZB × NZW)F_1	6	IgG	+++	+++	27
DP17	MRL/lpr/lpr	6	IgG1	+++	+++	38
172-s125	(NZB × NZW)F_1	6–8	IgG2b	+++	+++	27
H8	MRL/lpr/lpr	3	IgG2a	+++	+++	31
A6.1	(NZB × NZW)F_1	6	IgG2a	++	++	39
3H9	MRL/lpr/lpr	6	IgG3	+++	+	30
N14-4	(NZB × NZW)F_1	4	IgG2a	+	++	36
7B6.8	MRL/lpr/lpr	5	IgG3	+	+	40
CAL9	MRL/lpr/lpr	NA	IgG2a	+	+	34

[a] Binding specificity is based on the literature values. Most of these binding measurements were conducted by solution-phase competition enzyme-linked immunosorbant assay (ELISA). Because competition ELISA does not measure equilibrium association (or dissociation) constants, it is difficult to compare the results from different studies. In an effort to relate the predicted structural features of these mAb's with their antigen binding properties, we have nevertheless attempted to assess the relative affinity of the different antibodies in our panel using the following criteria: strong binding, ≤ 0.5 μg/ml DNA competitor (denoted +++); medium binding, 0.5–1.5 μg/ml DNA competitor (denoted ++); and weak binding, ≥ 1.5 μg/ml DNA competitor (denoted +). A dash (—) denotes no binding.

TABLE 2
Gene Families of Anti-DNA Used for Homology Modeling

Antibody	V_H[a]	D_H[b]	J_H[c]	V_K[d]	J_K[e]	Ref.
		ssDNA specific				
111-55	7183	DSP16.2	3	1	2	27
111-34	J558	NA	2	2	4	27
10-43	J558	NA	2	1	1	27
165-14	J558	DFL16.1	2	10	4	27
165-3	J558	DSP2.3,4	2	8	5	27
DNA2	7183	DFL16.2/DSP2.2	3	9	2	28
DNA12	7183	DFL16.2/DSP2.2	3	9	2	28
33-2	7183	DFL16.2/DSP2.2	2	9	5	29
4H8	J558	NA	2	4	3	30
H161	J558	DFL16	2	1	5	31
17-s130	7183	DSP2.6	1	19	2	27
bfd03	J558	DFL16.1	3	9B	1	32
A52	J558	SP2.3	2	8	1	33
17-s83	J558	DSP2.3,4	3	1	2	27
1A11	J558	NA	2	4	3	30
CAL27	J558	DFL16.1	2	23	2	34
		dsDNA specific				
43-2	J558	DSP2	4	23	4	29
163-47	J558	NA	1	9	2	27
BV16-13	Q52	DFL16.2	3	1	1	35
163-72	J558	DFL16.1	3	5	1	27
163-42	7183	DFL16.1	2	OX 1	1	27
BV17-31	Q52	DFL16.2	3	1C	4	35
N4-1	S107	Q52	1	4,5	5	36
N4-2	S107	Q52	1	4,5	5	36
N4-16	S107	Q52	1	4,5	5	36
N4-18	S107	Q52	1	4,5	5	36
N4-27	S107	Q52	1	4,5	5	36
N4-36	S107	Q52	1	4,5	5	36
N4-10	S107	Q52	1	4,5	5	37
		Cross-reactive				
185-33	Q52	Q52	2	12	1	27
DP17	J558	DFL16.2	3	19	1	38
17-s125	J558	DFL16.1	2	5	5	27
H8	J558	DSP2	4	8	2	31
A6.1	J558	DFL16.1	4	9	2	39
3H9	J558	NA	2	4	3	30
N14-4	S107	FL16.1	3	8	2	36
7B6.8	J558	SP2.6/2.7	2	4	1	40
CAL9	J558	DSP2.5,7	2	5	4	34

[a] V_H, heavy chain variable region gene.
[b] D_H, heavy chain diversity gene.
[c] J_H, heavy chain joining gene.
[d] V_L, light chain variable region gene.
[e] J_K, light chain joining gene.

TABLE 3A
Reference Antibodies Used for Modeling Anti-dsDNA (CDR Loops Only)

Antibody	Reference antibodies[a]					
	LCDR1	LCDR2	LCDR3	HCDR1	HCDR2	HCDR3
43-2	3HFM	3HFM	3HFM	6FAB	2HFL	2HFL
163-47	6FAB	6FAB	3HFM	MAM	6FAB	FDL
BV16-13	1NBV	1NBV	MCG	FDL	IFDL	IMN
163-72	MAM	2HFL	FAI	MAM	6FAB	2FBJ
163-42	2HFL	2FBJ	3HFM	IMN	IGF	IGF
BV17-31	1NBV	1NBV	IGF	FDL	FDL	IMN
N4-1	2HFL	2HFL	MAM	MAM	IMN	IGF
N4-2	2HFL	2HFL	2FBJ	MAM	MAM	IGF
N4-16	2HFL	2HFL	IGF	MAM	IMN	IMN
N4-18	2HFL	2HFL	IMN	2FBJ	MAM	IMN
N4-27	2HFL	2HFL	IMN	MAM	IMN	IGF
N4-36	2HFL	2HFL	IMN	MAM	MAM	IMN
N4-10	2HFL	2FBJ	IMN	MAM	IMN	IMN

[a] Antibodies are denoted according to PDB accession codes: BAF (*42*), BIL (*43*), FAI (*44*), FDL (*45*), IGF (*46*), IMM (*47*), IMN (*47*), MAM (*48*), MCP (*49*), MCW (*50*), PFC (*51*), REI (*52*), 1NBV (*13*), 2FBJ (*53*), 2FVW (*54*), 2HFL (*55*), 2IGF (*46*), 2IG2 (*46*), 2IG2 (*56*), 2MCG (*57*), 2MCP (*49*), 3FAB (*58*), 3HFM (*59*), 3MCG (*57*), 4FAB (*60*), 6FAB (*61*).

2IGF. The calculations were performed as described under Materials and Methods and the RMSD value for each of the CDRs after refinement (backbone only) relative to the X-ray structure of BV04-01 is ≤0.7 Å, whereas the RMSD for the entire F_V is 0.95 Å. The greatest deviation between the X-ray structure and the model occurs for HCDR3 where the RMSD = 2.4 Å; however, this result is not unexpected since HCDR3 cannot be defined by a single canonical designation (*23*, *24*). These results suggest that our modeling protocol can reproduce X-ray crystal structures to a relatively high degree of accuracy.

All of the anti-ssDNA are moderately homologous with respect to primary amino acid sequence, ~60–65% identity for each sequence compared to any given sequence in the group (Table 4; the reader is referred to Refs. (*16*) and (*17*) for comprehensive reviews of anti-DNA primary sequence alignments and comparisons). However, when the anti-ssDNA are divided into groups based on V_H and V_L gene families (e.g., all of the mAb's encoded by V_HJ558) and then aligned, the sequence identity rises to ≥90%. Inspection of the modeled F(ab)'s shows that the homology observed in the primary sequence alignments is also present in the 3-D structures of these antibodies (Fig. 1).

The predicted anti-ssDNA structures appear to be quite similar; however, they can be divided into three groups based on overall CDR topology (Fig. 2). Specifically, SASA calculations (using a water probe with a radius of 1.8 Å) reveal significant differences between the mAb's. The binding site on one group of mAb's (designated group 1) is a flat open surface whereas the CDRs in the other two

TABLE 3B
Reference Antibodies Used for Modeling Anti-ssDNA (CDR Loops Only)

	Reference antibodies[a]					
Antibody	LCDR1	LCDR2	LCDR3	HCDR1	HCDR2	HCDR3
111-55	4FAB	4FAB	4FAB	4FAB	IGF	2FBJ
111-34	1NBV	IMN	1NBV	IGF	6FAB	6FAB
10-43	1NBV	1NBV	1NBV	6FAB	2HFL	2FBJ
165-14	6FAB	1NBV	REI	6FAB	6FAB	6FAB
165-3	IMN	IMN	2HFL	FAI	FAI	6FAB
DNA2	6FAB	MAM	2HFL	MAM	2IG2	2IG2
DNA12	MAM	MAM	2HFL	MAM	2IG2	6FAB
33-2	FAI	REI	IGF	IGF	IGF	6FAB
4H8	3HFM	IMN	3HFM	MAM	6FAB	IMN
H161	1NBV	1NBV	1NBV	1NBV	6FAB	FDL
17-s130	6FAB	IMN	MAM	4FAB	2IG2	3FAB
bfd03	FDL	6FAB	REI	MAM	2HFL	6FAB
A52	IMN	IMN	REI	MAM	6FAB	IMN
17-s83	1NBV	1NBV	1NBV	2FBJ	2HFL	FDL
1A11	3HFM	2HFL	MAM	2FBJ	2HFL	IMN
CAL27	3HFM	3HFM	3HFM	2FBJ	FAI	IMN

[a] Antibodies are denoted according to PDB accession codes: BAF (*42*), BJL (*43*), FAI (*44*), FDL (*45*), IGF (*46*), IMM (*47*), IMN (*47*), MAM (*48*), MCP (*49*), MCW (*50*), PFC (*51*), REI (*52*), 1NBV (*13*), 2FBJ (*53*), 2FVW (*54*), 2HFL (*55*), 2IGF (*46*), 2IG2 (*46*), 2IG2 (*56*), 2MCG (*57*), 2MCP (*49*), 3FAB (*58*), 3HFM (*59*), 3MCG (*57*), 4FAB (*60*), 6FAB (*61*).

TABLE 3C
Reference Antibodies Used for Modeling Cross-Reactive Anti-DNA (CDR Loops Only)

	Reference antibodies[a]					
Antibody	LCDR1	LCDR2	LCDR3	HCDR1	HCDR2	HCDR3
185-33	FDL	FDL	FDL	FDL	FDL	6FAB
DP-17	MAM	IGF	3HFM	MAM	6FAB	IMN
172-s125	3HFM	2HFL	IMN	IGF	6FAB	INBV
H8	IMN	IMN	IMN	6FAB	6FAB	IMN
A6.1	6FAB	IGF	6FAB	6FAB	6FAB	2IG2
3H9	2IG2	2HFL	BJL	2FBJ	6FAB	IMN
N14-4	IMN	IMN	2HFL	MAM	MAM	2FBJ
7B6.8	MAM	2HFL	MAM	2FBJ	6FAB	6FAB
CAL9	MAM	2HFL	MCG	FAI	FAI	IMN

[a] Antibodies are denoted according to PDB accession codes: BAF (*42*), BJL (*43*), FAI (*44*), FDL (*45*), IGF (*46*), IMM (*47*), IMN (*47*), MAM (*48*), MCP (*49*), MCW (*50*), PFC (*51*), REI (*52*), 1NBV (*13*), 2FBJ (*53*), 2FVW (*54*), 2HFL (*55*), 2IGF (*46*), 2IG2 (*46*), 2IG2 (*56*), 2MCG (*57*), 2MCP (*49*), 3FAB (*58*), 3HFM (*59*), 3MCG (*57*), 4FAB (*60*), 6FAB (*61*).