Hierarchical Binding of DNA Fragments Derived from Scaffold-Attached Regions: Correlation of Properties in Vitro and Function in Vivo[†,‡]

Christian Mielke,[§] Yoshinori Kohwi,[∥] Terumi Kohwi-Shigematsu,[∥] and Jürgen Bode*[,§]

GBF, Gesellschaft für Biotechnologische Forschung, mbH, Genetik von Eukaryonten, Mascheroder Weg 1, D-3300 Braunschweig-Stöckheim, West Germany, and La Jolla Cancer Research Foundation, 10901 North Torrey Pines Road, La Jolla, California 92037

Received February 26, 1990; Revised Manuscript Received April 23, 1990

ABSTRACT: On its upstream side, the human interferon-β gene is flanked by a 7-kb SAR (scaffold-attached region) DNA element. The core of this element is determined and subjected to in vitro reassociations with isolated scaffolds. Binding properties of SAR fragments with decreasing length are quantified and related to consensus sequences like the topoisomerase II box and an ATATTT motif. Characteristics as the stoichiometry, affinity, and cooperativity of the binding process are shown to depend on the length of SAR DNA and suggest a model involving a multiple-site attachment to protein scaffolds. We propose a rational approach for predicting the SAR mediated transcriptional enhancements in vivo from their binding properties in a standardized in vitro assay. The efficiency of this approach is demonstrated for a marker (huIFN-β) and a selector gene (neo^r).

Within eukaryotic nuclei, the genome is organized into topologically constrained domains, which are thought to provide the structural basis for a differential gene expression [reviewed by Gross and Garrard (1987)]. It has been estimated that each nucleus contains between 10000 and 100000 of these domains with sizes between 5 and 120 kb (average, 70 kb).

Recent progress has been made in defining the attachment sequences bordering the domains, as these are associated with the nuclear scaffold or nuclear matrix owing to their intrinsic binding affinity. The term "scaffold" has been reserved for the proteinaceous backbone, which resists extraction with a rather mild detergent, lithium diiodosalicylate (LIS), at physiological salt concentrations (Mirkovitch et al., 1984), whereas the more general term "matrix" was originally introduced for a related structure that is obtained after high-salt extraction procedures. To map scaffold-attached regions (SARs) or matrix-associated regions (MARs), two principal approaches have been developed. One of these, the in vivo assay, uses a nuclear halo mapping protocol that involves trimming down the DNA in LIS extracted nuclei by a combination of restriction enzymes and identifying the DNA segments that remain attached to the scaffold. The other (in vitro) protocol is based on a reconstitution procedure that selects for DNA fragments with an affinity for the extraction-resistant protein backbone from nuclei (Cockerill & Garrard, 1986). It is consistently found that the in vitro procedure yields essentially the same results when matrices or scaffolds have been prepared by the corresponding procedures.

The first SAR sequences were derived for a number of *Drosophila* genes where they were located at the presumptive borders of the corresponding chromatin domains. Here, they sometimes overlapped or coincided with sequences that had previously been identified as enhancers [reviewed by Gasser and Laemmli (1986)]. This group of elements was extended later on to include the chicken lysozyme gene (Phi-Van & Straetling, 1988; Stief et al., 1989), the human interferon-β gene (huIFN-β; Bode et al., 1988) and the β-globin cluster (Jarman & Higgs, 1988). In case of the β-globin cluster, these sequences appear to coincide with DNAse I "superhypersensitive" sites that have been detected within the "dominant control regions" (DCRs) of the domain. For these DCR sequences it was shown that they can provide a number of genes with the property of a tissue-specific and localization-independent expression (van Assendelft et al., 1989; Forrester et al., 1989). SAR sequences next to the lysozyme gene activate the lysozyme enhancer provided that it is localized within the domain (Stief et al., 1989).

The first "MAR" sequences that have been identified due to their affinity for high-salt extracted matrices included an element 5′ to the κ enhancer (Cockerill & Garrard, 1986) and two such elements 5′ and 3′ to the corresponding heavy-chain enhancer (Cockerill et al., 1987). These MARs support the function of the coordinate enhancer (Blasquez et al., 1989; Xu et al., 1989), which is part of an intron. They are, therefore, functionally different from the SAR sequences mentioned above, as their occupation must be regulated during transcription. The regulatory function may be achieved by soluble factors; a negative factor of this type has recently been identified (Scheuermann & Chen, 1989). Other intronic sequences with an affinity for the matrix (which we will continue to call MARs for convenience) have been identified in the dihydrofolate reductase gene (Käs & Chasin, 1987), the interleukin-2 gene (Artelt and Bode, unpublished results), and a plant gene with the quality of a position-independent expression in potato and tobacco plants (Stockhaus et al., 1987; Bode et al., unpublished results).

It must be emphasized that owing to the structural features SAR and MAR elements have in common, they are detected by the same procedures. The present contribution demonstrates that a plant MAR with a high affinity for nuclear scaffolds from animal cells behaves like a SAR element if it is included in gene constructs for transfection into murine cells. Within the animal kingdom, the mentioned MAR sequences are usually shorter than SARs and they have a reduced affinity for scaffold preparations.

[†] This work was supported by a grant from BMFT (0319305A3) and a grant from NIH (RO1 CA39681).

[‡] The nucleic acid sequence in this paper has been submitted to GenBank under Accession Number J02913.

* To whom correspondence should be addressed.

[§] GBF.

[∥] La Jolla Cancer Research Foundation.

0006-2960/90/0429-7475$02.50/0

The human interferon-β gene is flanked by the longest SAR regions that have been characterized to date. This conclusion was first derived from a halo-mapping (in vivo) procedure after transfection into murine cells; this detour avoided a cross-hybridization on Southern blots with human repetitive DNAs which severely restricted a direct inspection of human cells. It was then confirmed by the reconstitution technique, which yielded essentially the same information (Bode & Maass, 1988). Both techniques revealed 7 kb of DNA 5′ to the gene with an uninterrupted affinity for the scaffold. The corresponding element at the 3′-end covers 3 kb or more.

The present contribution delineates some of the factors that are involved in a highly specific SAR–scaffold interaction. The mode of interaction is clearly different from the association of sequence-specific soluble factors and this may be the reason for the fact that it could only be traced by coarse mapping procedures (Bode et al., 1986, Bode & Maass, 1988) or by an exonuclease protection assay (Gasser & Laemmli, 1986) but not by high-resolution footprinting techniques. Our studies further show that SAR elements can be used to enhance the transcriptional activity of a marker (huIFN-β) and a selector gene (neo^r). The enhancement is directly related to the binding strength determined in vitro and to the presence and distribution of certain sequence motifs. It is suggested that variants of the in vitro assay can be applied to screen for SAR elements that can then be used in expression vectors where they support a high-level and position-independent transcription.

Experimental Procedures

Plasmids. The construction of the 10.1-kb plasmid pCATIFβ from sequences coding for human IFN-β, bacterial CAT, and bacterial neomycin resistance has been described (Klehr & Bode, 1988) and several of its applications have been reported (Wirth et al., 1988). The plasmid contains three *Eco*RI sites at positions 6097, 10100/0, and 3297. Using two rounds of partial *Eco*RI digest/fill-in/ligation reactions, a set of plasmids was obtained in which a single *Eco*RI site was retained (pCATIFβ-a, -b, and -c). The present studies are based on pCATIFβ-a (cf. Figure 7); the application of the other constructs for determining position effects of SAR elements will be reported elsewhere.

Site-Directed Mutagenesis. The preparation of mutants by the Amersham system exploited the method of Eckstein et al.

Polymerase Chain Reaction (PCR). PCRs were performed either on human placenta DNA (400 ng) or on a cloned vector with segment IV as an insert (10 ng). These DNAs were linearized by PstI in 50 μL of PCR buffer [10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM $MgCl_2$, 0.01% w/v gelatin] prior to amplification. Primers contained 20 matching nucleotides adjacent to the 5′aaatATATTT motif on the top strand (primer composition 5′TCTTTTTTAGAAAAACAGGG3′) and overlapping the bottom strand aaaATATTT motif (primer composition 5′GGGG*C*A*TGCAAATATTTTCTGTC3′; nucleotides marked with an asterisk have been mutated to yield a *Sph*I site). Amplification was in 100 μL of PCR reaction buffer using 20 nmol of each dNTP, 10 pmol of each primer, and 2.5 units of Taq-polymerase (Cetus) in a vial with an overlay consisting of 100 μL of mineral oil. Forty cycles (1 min, 94 °C; 1 min, 26 °C; 1.5 min, 72 °C) were performed. The 314-bp PCR fragment was repaired by Klenow polymerase, phosphorylated by polynucleotide kinase, and cloned into the *Sma*I site of the pTZ-18R polylinker. Alternatively, it was cleaved at the terminal and an internal *Sph*I site and cloned (as a 166-bp subfragment) directly into the same site to yield monomeric or oligomeric inserts (Figure 2D). Plasmids were multiplied in *Escherichia coli* (strain DH-5α) and SAR fragments reisolated by using the adjacent *Eco*RI and *Hin*dIII sites of the vector's multiple cloning site. They were end labeled by the Klenow reaction using [α-^{32}P]dATP. Fractions of the labeled fragments were added to stock solutions containing the same fragments at a predetermined concentration.

Cell Culture, Transfection, and Electroporation. The procedures for culturing mouse L cells and for the isolation of nuclei were as described (Bode et al., 1986; Bode & Maass, 1988).

Transfections were performed according to a protocol adapted from Wigler et al. (1977). DNA precipitates were prepared as follows: On a vortex mixer, 1–5 μg of linearized plasmid DNA, dissolved in 250 μL of 125 mM $CaCl_2$, was added dropwise to 250 μL of 50 mM HEPES (pH 7.1), 0.28 M NaCl, and 1.5 mM sodium phosphate. The mixture was left for 1 h at room temperature before use. On day one, 3×10^5 LTK cells in 5 mL of DME/10% FCS were seeded into a 25-cm^2 flask. On day two, the medium was changed and 4 h later 500 μL of the above linearized, carrier-free DNA was added. The cells were left overnight before the medium was changed for 5 mL of DME/10% FCS (day three). On day four, the cells were seeded into a 75-cm^2 flask and provided with 15 mL of selective medium (DME/10% FCS/700 μg/mL of the neomycin analogue G-418). This medium was replaced on days 8 and 10 before clones were counted on day 12.

Electroporation. Logarithmically growing, semiconfluent cells were trypsinized, washed in PBS (137 mM NaCl, 2.7 mM KCl, 8 mM sodium phosphate, 1.5 mM potassium phosphate, pH 6.8), and collected by centrifugation (400*g*, 5 min). The cellular pellet was distributed to 500 μL of electroporation buffer (20 mM HEPES, pH 7.05. 137 mM NaCl, 5 mM KCl, 7 mM sodium phosphate, 6 mM dextrose) containing 10 μg of the linearized plasmid and treated in a BTX transfector, typically at 1100 μF and 200 V/cm. After treatment, there remained about 35% of surviving cells as determined by trypan blue exclusion. The cellular pellet was left on ice for 10 min, dissolved in 15 mL of DME/10% FCS, and seeded into a 75-cm^2 flask (day one). The medium was changed on day two. The further selection procedure was as for transfection.

Gene Expression Studies. The methods used for determining the inducible expression of huIFN-β have been described by Dinter and Hauser (1987). Expression of the neo^r gene was determined via mRNA levels in total RNA that had been extracted from cells close to confluence by the acid guanidinium thiocyanate–phenol–chloroform extraction method of Chomczynski and Sacchi (1987). After they were dried and denatured, samples containing 5 μg of total RNA were applied to an 1.2% agarose gel containing 20 mM morpholinopropanesulfonic acid (pH 7.0)/5 mM NaOAc/1 mM EDTA/6.6% formaldehyde and separated during a 5-h run at 100 V. After washing in 10×SSC, RNA was blotted with the same buffer to a nylon membrane (Gene Screen plus); the membrane was baked for 2 h at 80 °C and probed with a nick-translated DNA fragment by using standard autoradiography. The film (Kodak X-Omat AR) was scanned with a laser densitometer (Pharmacia Ultroscan) and the signals were referenced to an internal control.

Nuclear Scaffolds. A typical experiment started with two plates of 150 cm^2 (5×10^7 cells). Nuclei were suspended in 100 μL of nuclear buffer (5 mM Tris-HCl, pH 7.4, 0.005 mM spermine, 0.125 mM spermidine, 20 mM KCl, 1% thiodiglycol, 0.1% digitonin from a frozen aliquot), which was freshly brought to 0.2 mM $PhMeSO_2F$ and 1% aprotinin. They were stabilized in this medium by incubation at 37 °C for 30 min.

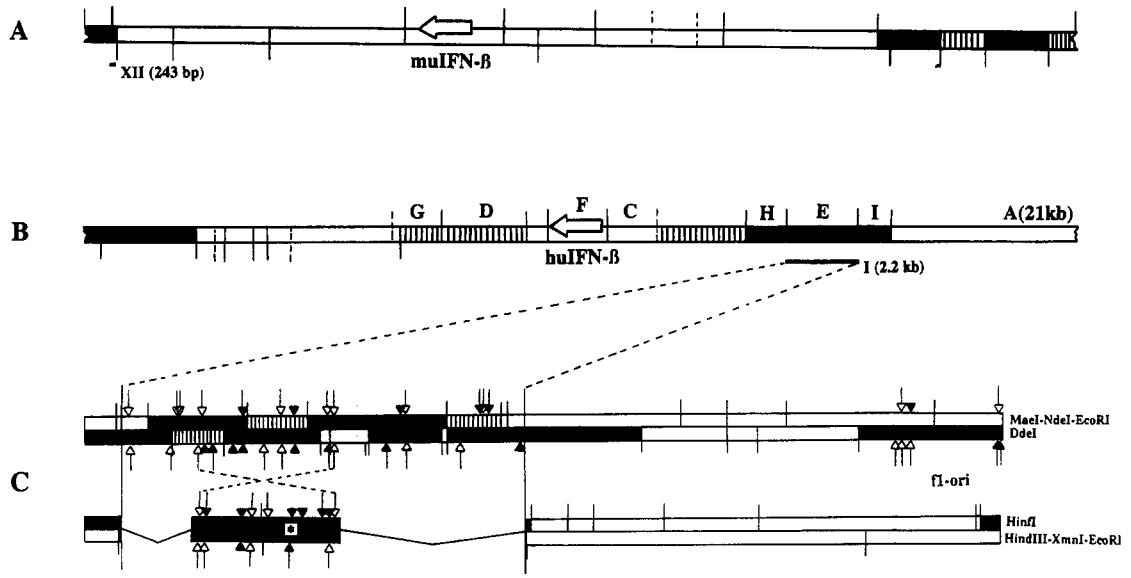


FIGURE 1: Domain and SAR structures of interferon-β genes. The figure shows the 20-kb domain of the murine interferon gene (A), the 14-kb domain of the human interferon gene (B), and the distribution of typical scaffold-binding motifs in one of the major SAR fragments (C). The murine domain has been restricted by *Eco*RI (lines pointing up) and *Hin*dIII (lines pointing down) and the human domain by *Eco*RI (lines pointing up) and *Bgl*II (lines pointing down). Other restrictions are marked by broken lines. (C) The 2.2-kb *Eco*RI fragment (I) and a 0.83-kb subfragment (IV, see Figure 2 for numbering) are shown in the context of plasmid pTZ18R sequences. Arrows mark topoisomerase II consensus sequences and solid arrowheads presence of a motif ATATTT. Where both motifs coincide they have been marked by a solid arrow. Scaffold association under standard conditions (see text) is characterized as follows: Hatched regions 50–80% and black regions 81–100% association.

Extraction of nonscaffold proteins was then achieved by adding 4 mL of 15 mM lithium 3,5-diiodosalicylate (LIS) medium. LIS medium is a stock solution of 10 mL of 20 mM *N*-(2-hydroxyethyl)piperazine-*N'*-2-ethanesulfonic acid (Hepes)–NaOH, pH 7.4, 0.1 M lithium acetate, 1 mM EDTA, 9 mL of which were freshly provided with 1 mL of 1% digitonin (solubilized by shortly heating in a microwave oven), and 55 mg of LIS powder. The mixture was carefully homogenized by two strokes in a loosely fitting Dounce homogenizer and centrifuged (2400 g, 5 min, 4 °C). Resulting nuclear halos were transferred into 20 mL of sterile digest buffer (20 mM Tris-HCl, pH 7.4, 0.05 mM spermine, 0.125 mM spermidine, 20 mM KCl, 70 mM NaCl, 10 mM $MgCl_2$). They were then gently rocked in a capped tube for 30 min and centrifuged as before; this procedure was repeated four times. Nuclear halos were degraded by 500 units of *Eco*RI in 600 μL for 1 h. After a careful homogenization by pipeting, 50 units each of *Bgl*II and *Hin*dIII were added and the degradation continued by two more hours. The resulting pellets were provided with 600 μg of sonified *E. coli* genomic DNA, incubated (45 min, 37 °C), and washed with digestion buffer in order to remove >95% of mouse DNA and to clear the endogenous scaffold attachment sites.

Binding Experiments. The resulting scaffold was filled up to 700 μL with digestion buffer containing 980 μg of *E. coli* genomic DNA. It was then partitioned into seven aliquots (7 $\times$ 10^6 cell equivalents of nuclear scaffolds each) and provided with restriction fragments that had been trace labeled by Klenow polymerase using [α-^{32}P]dATP. All samples were agitated on a rocking platform at 37 °C overnight and separated into a pellet (P) and supernate (S) fraction by centrifugation (4 °C, 14000*g*, 45 min) and two washing steps using 200 μL of digestion buffer. They were then subjected to DNA preparations as described (Bode & Maass, 1988). After dissolving the purified DNA pellets in 200 μL of Tris–EDTA buffer, DNA contents were determined and corrected such that within a given series of supernatant- or pellet-derived samples the same values were obtained. DNA, 10–40 μL (typically 1000–3000 cpm), was then applied per slot of an agarose gel (1–3.5% depending on the fractionation range). After electrophoresis, the gel was dried onto Whatman 3MM paper on a gel dryer and autoradiographed overnight by exposure to Agfa Curix RP1 film.

For determining binding curves, similar procedures were followed with the exception that labeled fragments in the S and P fractions were quantified by scintillation counting. The free ligand concentrations (L) obtained in an equilibrium binding experiment were fitted by nonlinear regression analysis to obtain estimates for two dissociation constants (K_1 and K_2) and the number of binding sites (n). For this purpose, variants of the programs "Enzfitter" (Elsevier Biosoft) and DNRP53 (Duggleby, 1984, 1988) were adapted to fit a modified Adair equation in the form

$$Y = n(A + 2B)/(2(1 + A + B))$$

with $A = (1/K_1)L$ and $B = (1/K_1)(1/K_2)L^2$ (see Figure 3C,D).

Results

Domain Structures of the Human and Murine IFN-β Genes. We have shown recently that the human IFN-β gene is flanked by two DNA regions that exhibit strong scaffold-binding characteristics by the criteria of both an in vivo and an in vitro assay. These SARs are about 14-kb apart and apparently they define the huIFN-β gene domain. Within that region there was a stretch of DNA with an intermediate affinity for the scaffold, immediately adjacent to the 3'-end of the gene (Figure 1B). With the present knowledge about SAR function at hand, it was difficult to decide whether this DNA could be classified as a true scaffold-attached region and whether such a property is important for the inducible expression of the interferon-β genes. We therefore decided to investigate the SAR structure around the analogous murine IFN-β gene that was available as a cosmid clone (Dirks et al., 1989). The analysis scheme followed the same protocol as used before for the in vitro determination of SARs (Bode & Maass, 1988). The results are briefly summarized in Figure 1A together with the restriction fragments on which this analysis is based. It is seen that the domain is larger here (20 kb), that

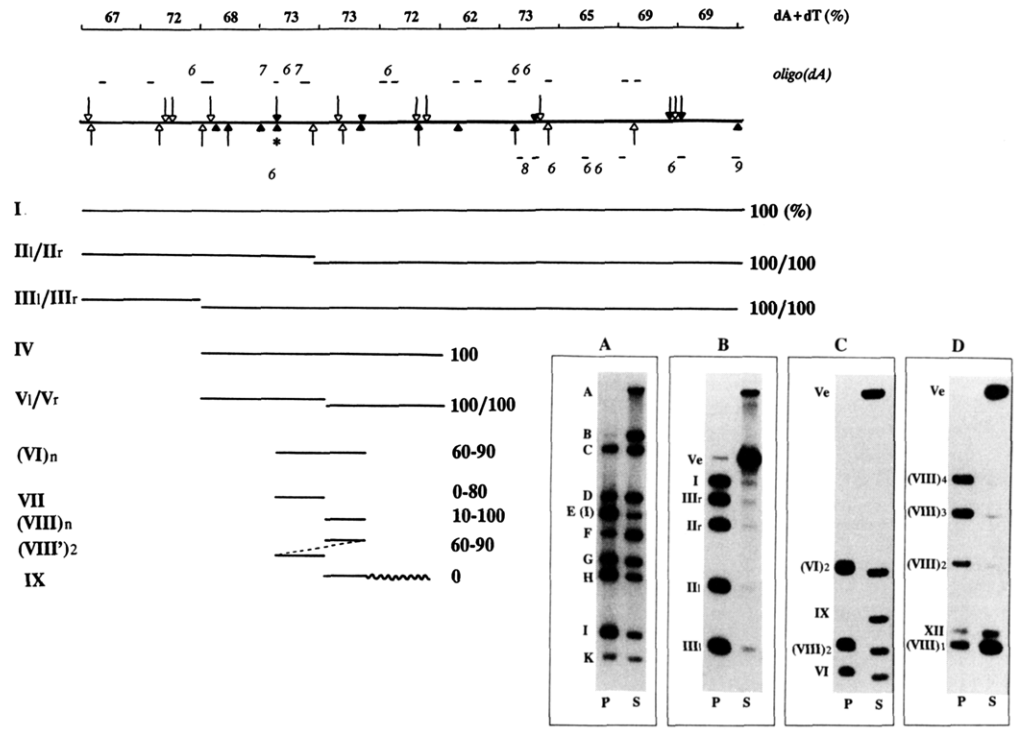


FIGURE 2: Dissecting the huIFN-β upstream SAR. SAR structure and scaffold-binding potential have been compiled and are shown in the context of structural features that have been implicated in SAR function. The top line shows the distribution of A + T contents in 200-bp sections of the 2.2-kb fragment I. oligo(dA) tracts (i.e., tracts of A_n with $n > 5$) have been characterized by the value of n (cursive numbers). Short horizontal bars next to these numbers mark sections that fulfill a major criterion for DNA bending (i.e., the condition $AAAN_{7-10}AAA$). Binding potentials (i.e., the percentage of a given fragment that is recovered from the pellet of a standard assay) are separated by a slash if they refer to the left/right parts of a fragment. In cases where binding is affected by the presence of competing fragments or where oligomeric forms were assayed, a range is indicated. Inserts A–D display several analyses of this type. Ve, vector fragment (usually linearized pTZ18R).

the gene has a more central position with respect to the attachment sequences, and that there are only strongly binding flanking elements. A comparison between both structures (A and B) seems to indicate that for the function of this inducible gene only the flanking, high-affinity elements are essential. With these results in mind we selected one of these elements for further characterization.

An *Eco*RI digest of pCosIFN-β, a cosmid harboring 36 kb of human DNA around the huIFN-β gene, reveals a balanced distribution of fragments between the pellet (SAR) and supernate fractions; i.e., four of the fragments (labeled A, F, G, and K in Figure 2A) are essentially nonbinding and five (C, D, E, H, and I) are attached to a considerable extent. These findings were derived under our standard conditions, i.e., if the concentration of restriction elements is chosen such that equal counts are recovered from the scaffold (pellet P) and supernate (S) fractions (Bode & Maass, 1988). More specific information about the relative binding affinities is possible by overloading the scaffold. Figure 2 shows a distribution of fragments under conditions that direct 90% of the counts to the supernatant (insert A). In this situation, an internal competition of scaffold-binding sequences occurs, which enriches in the pellet fraction the fragments with the highest affinity. It is clearly seen in Figure 2A that this quality can be ascribed to the *Eco*RI fragments E (2.2 kb) and I (0.8 kb) which are adjacent in the restriction map forming two parts of apparently the same upstream SAR element.

The majority of the studies to follow will be based on the elements introduced here, i.e., the center of the huIFN upstream SAR region and a small SAR subfragment found downstream from the murine gene. Sequences from other sources will be added where correlations with the current literature seem relevant or where sequence–function correlations are to be made.

Equilibrium Binding Studies Using SAR Fragments of Different Length. The in vitro binding assays upon which these studies are based comprise a number of variables that had to be appreciated before comparative data could be generated. First, they require competitor DNA to generate specificity. Unless stated otherwise, this competitor was sonicated *E. coli* DNA, applied at concentrations of 140 μg per 100 μL of assay, which approximates a 10000-fold excess over the radioactively labeled SAR fragments added to the same mixture. Second, the input concentration of scaffold-binding fragments affects the specificity of the assay in an unexpected manner: it has been repeatedly observed that at low input levels the enrichment in the pellet fraction is poor but improves as saturation is approached. Such a property could only be explained by a cooperative binding process during which affinities increase with increasing loads of DNA. This idea was verified under the standard conditions of our assay (i.e., in the presence of *E. coli* competitor, Figure 3A,B) and under conditions that permit the derivation of an equilibrium binding curve (Figure 3C,D).

Figure 3A shows a titration of scaffolds with increasing amounts of the 2.2-kb *Eco*RI fragment E (which will carry the roman number I henceforth, cf. Figure 2 for assignments). Plotting of DNA_{bound} versus DNA_{added} was preferred here over DNA_{bound} versus DNA_{free} since after the initial lag nearly quantitative binding was observed up to an apparent point of saturation where 3000–3500 molecules of I were bound per individual scaffold (i.e., the equivalent derived from a single cell). These values seem small if they are compared with published estimates [>10000, cf. Cockerill and Garrard (1986)], but this could arise from the fact that we have to deal with a large fragment with many potential subsites. This is supported by the data obtained with a 0.3-kb fragment from I (VI, cf. Figure 2B), binding of which levels off around 14000

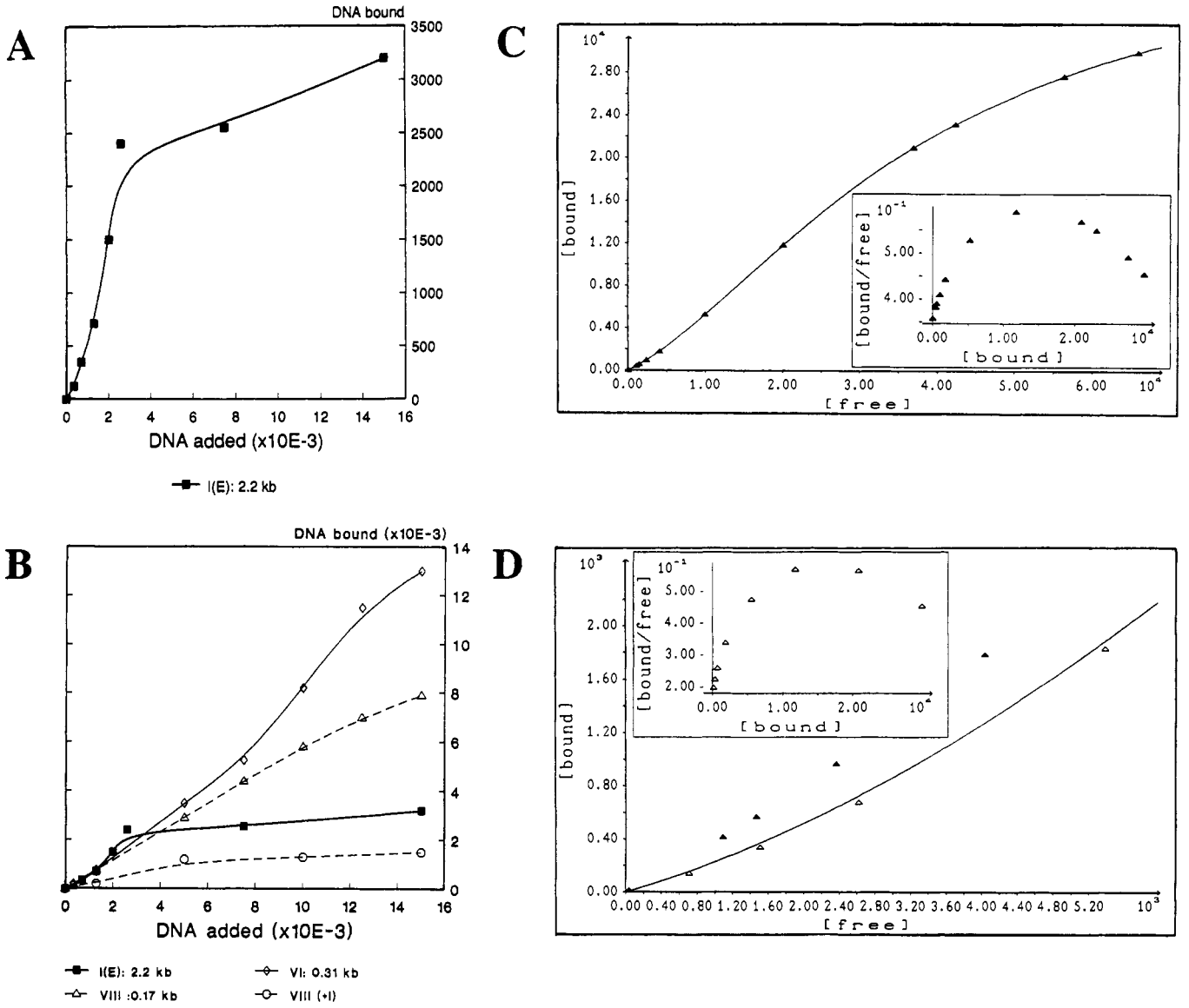


FIGURE 3: Scaffold association of SAR elements depending on the kind of competitor DNA. (A, B) Strong binding in the presence of *E. coli* DNA (140 μg/100 μL). Scaffolds were loaded with a given amount of fragments I (2.2 kb), VI (0.31 kb), or VIII (0.17 kb; dashed lines). Another experiment shows the association of labeled VIII in the presence of saturating amounts of unlabeled I. Data were referenced to the cellular scaffold equivalent. (A) Binding of fragment I alone. (B) The same data are superimposed by the corresponding figures for the other fragments. (C, D) Reversible association under conditions favoring equilibrium binding. In the presence of 50 μg/100 μL of mouse DNA, scaffolds were loaded with a given amount of labeled fragment I. Parallel experiments were performed in the absence (C) or in the presence (D) of an additional 5 ng/100 μL of an unlabeled MAR fragment from a potato gene (X); this quantity would correspond to 4×10^3 cpm of I. Solid triangles mark the data for I alone, while open triangles refer to the corresponding data obtained in the presence of IX. Scatchard transformations of the binding curves are shown in the inserts. For the structure of X, see Figure 4. In terms of the arbitrary scaling (counts per minute bound versus counts per minute free) used here, the curves can formally be fitted by two dissociation constants each [$6.0/2.4 \times 10^4$ for (C) and $10.7/0.9 \times 10^4$ for (D)].

molecules per cellular scaffold equivalent. Again, there is a sigmoid response in the respective binding curve (Figure 3B). At variance are the data for a short 0.17-kb subfragment (VIII), which has been prepared by PCR techniques. From the data in Figure 3B it can be concluded that this species shows low-affinity binding and this is essentially true for any other SAR subfragment of a corresponding size (see below). Its binding sites seem to mostly overlap with those of the parental fragment (I) since a scaffold preloaded with saturating amounts of unlabeled I has lost the majority of the corresponding sites (cf. the dashed lines in Figure 3B).

True equilibrium conditions for the binding of I could be established by changing the competitor DNA. If we use mouse DNA at about one-third the concentration applied for *E. coli* competitor DNA, a plot of DNA bound versus DNA free has the shape shown in Figure 3C. This curve can formally be fitted by two dissociation constants differing by factors of 3–10, supporting its cooperative nature. Furthermore, the fitting gave a value $n = 4.3 \times 10^4$ (cpm) corresponding to 3000 sites per scaffold, which is in fair agreement with the above data.

The same setup has been extended (Figure 3D) by adding a specific competitor, i.e., an unlabeled 1.3-kb SAR fragment (X) derived from a plant gene (ST-LS1, cf. Figure 4 and Table I). Clearly, this fragment equilibrates with the same sites on the scaffold that are otherwise taken by fragment I, demonstrating that our assay selects SAR fragments that interact in a similar or identical way.

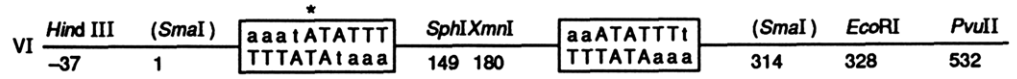
We conclude that a SAR–scaffold association is an essentially reversible, cooperative process. Competitor DNA has to be added to prevent nonspecific binding and this competitor is derived best from bacterial sources that are apparently free from the features that mediate binding. DNA from eukaryotes

Table I: Structural Features Found in SAR (MAR) Elements (1–3) and in an A + T Rich Control[a]

sequence	A + T (%)	Topo II cons (+ATATTT)	ATATTT + AAATAT	$(A)_{n>4}$ + $(T)_{n>4}$	$AA(N)_{9-10}AA$ + $TT(N)_{9-10}TT$	$AAA(N)_{7-10}AAA$ + $TTT(N)_{7-10}TTT$
1. huIFN-β-SAR (I)	69	10 (3)	6	15	23	10
2. IgH-MAR (XVI)	65	6 (2)	6	6	18	11
3. ST-LS1-MAR (X)	67	7 (2)	9	5	12	9
4. control (XI)	66	3 (0)	1	7	19	7

[a] Data refer to the average number of motifs per kilobase of DNA.

Chart I



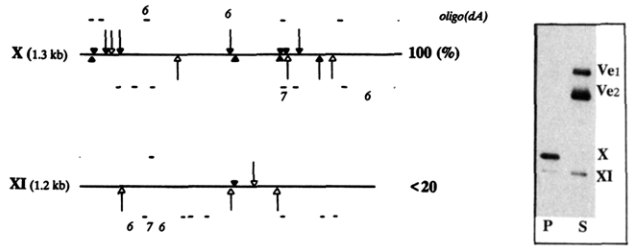


FIGURE 4: Comparison of a MAR fragment with a comparably A + T rich control. X is a 1.3-kb fragment isolated from the first intron of the ST-LS1 gene (potato) and XI a 1.2-kb fragment from mouse mitochondria. Symbols are as in Figures 1 and 2. The insert depicts the relative association in an assay where both fragments were present together with the respective vector fragments (Ve_1, pBR322; Ve_2, pUC19, both linearized via the respective *Eco*RI sites). For a full evaluation, see Table I.

affects binding substantially as it decreases the apparent affinities.

For the studies to follow we have used reassociation mixtures that comprise comparable amounts of binding and nonbinding DNA fragments in the presence of 140-μg *E. coli* genomic DNA per 100 μL of sample. Ideally, the scaffolds were loaded to a point of saturation, which leads to equal counts in the supernate and pellet fractions ("standard conditions").

Sequence Motifs and Structural Features Recurring in SAR Elements. All SAR elements have several features in common. They are relatively A + T rich (60–70%, sometimes more), and they contain recurring motifs corresponding to the topoisomerase II consensus as defined for *Drosophila* (Gasser & Laemmli, 1987) and a sequence ATATTT first assigned by Cockerill and Garrard (1986), which may or may not coincide with the central bases of this consensus. Further features that have been ascribed to SARs is their propensity for bending (Anderson, 1986; Homberger, 1989) and the presence of several homopolymeric oligo(dA)·oligo(dT) tracts (Adachi et al., 1989, Izaurralde et al., 1989, Käs et al., 1989). The experiments described here were aimed at obtaining a useful control that matches a given SAR or MAR fragment in its size and A + T content but behaves radically different in its in vitro binding properties. This approach should allow the distinction of features that are due to the general A + T contents or the specific properties of a SAR DNA, respectively. A comparison of these fragments with regard to their efficiencies in vitro and in vivo should then yield information about SAR functions that is only obtained with difficulty by other means.

A couple with these properties were found with the main intronic 1.3-kb MAR element detected in a light-inducible plant gene, ST-LS1 (X; Bode et al., unpublished results) and an 1.2-kb A + T rich fragment (XI) from mouse mitochondria DNA. The latter fragment contains a mitochondrial origin of replication and promoted a head-to-tail multimeric insertion after transfection into recipient cells (Lutfalla et al., 1985); this property was lost in our transfection protocol, which was optimized to yield low copy numbers. The results are shown in Figure 4 and evaluated together with other SAR/MAR sequences in Table I. From these data it seems that the sequences are not dramatically different in their contents of homopolymeric oligo(dA)·oligo(dT) tracts and $AA(N)_{9-10}AA$ or $AAA(N)_{7-10}AAA$ motifs, which are indicators for the propensity of bending (Anderson, 1986). On the other hand, there were increasing differences with regard to the content of Topo II boxes of the type GTN(A/T)A(T/C)ATTNATNN(G/A) [this sequence had to occur with an overall precision of 80% but had to comprise the "core-motif" A(T/C)ATTN (7 versus 3)]; a motif ATATTT (9 versus 1); and Topo II boxes in which the core motif is ATATTT (2 versus 0). In a comparative binding experiment, the SAR sequence (X) was found scaffold-associated close to 100% whereas the propensity of XI was low (<20%) but nevertheless different from that of vector DNA (Figure 4, insert).

Dissecting the Upstream SAR of the Human IFN-β Gene. The relevance of the mentioned motifs for binding was investigated next by fractionating the 2.2-kb SAR element I, which was cloned into the single *Eco*RI site of vector pTZ18R. The results have been compiled in Figures 1C and 2 and lead to the following conclusions:

The predominant motifs derived above [Topo II box (light arrow), ATATTT (solid arrowhead), and the consensus motif (solid arrow)] are distributed evenly over the 2.2-kb sequence of SAR "I". Binding is observed for several SAR I fragments, the shortest of which comprises 380 bp and containes five simple Topo II boxes (Figure 2, insert B, species I–III).

In the prokaryotic part these motifs are rare except for a short stretch around the F1 origin, which is an integral part of the vector (Figure 1C).

SAR fragments containing multiple motifs of this type can direct considerable stretches of linked non-SAR DNA to the scaffolds, cf., the bordering fragments that cover SAR and vector sequences (Figure 1C).

In the plasmids degraded by *Mae*I/*Nde*I/*Eco*RI and *Dde*I there appears to be a decay in the bindings potential below about 300 bp, i.e., in sequences that are reduced to the extent of leaving a single dominant motif or less (cf. Figure 1C, white and hatched segments within the SAR sequence).

These observations prompted efforts to define a minimum-length SAR element that is still functional by a variety of criteria. For this purpose we have chosen the segment VI, which contains overlapping tracts of ATATTT at both ends of both strands. Moreover, at its 5′-end it reproduces a sequence (aatATATTT), which is also part of other SAR or MAR elements. With a set of two matching primers, this fragment was obtained by PCR techniques either from the plasmid or from human placenta DNA. The fragment was cloned into the *Sma*I blunt site of the pTZ18R vector (Chart I). VI is the basic construct of a series (VI–VIII) composed

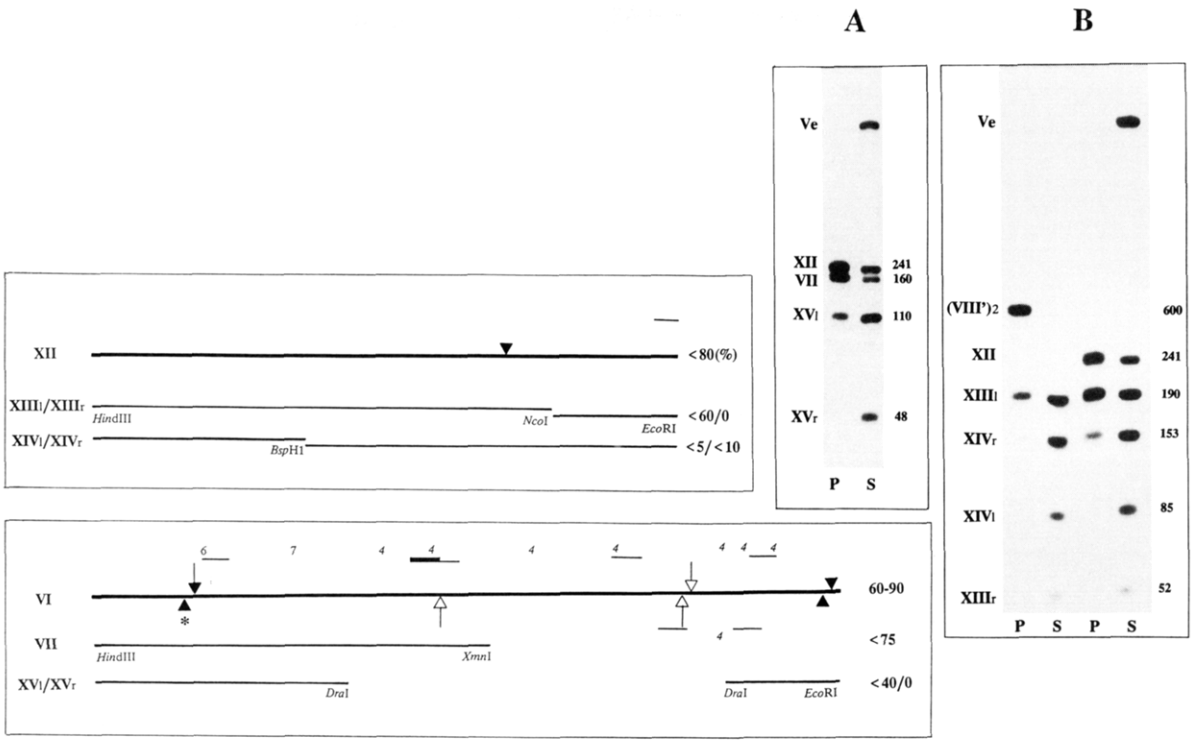


FIGURE 5: Successive loss of scaffold-binding potential for short SAR subfragments. Fragments XII (from the muIFN-β downstream SAR) and VI (from the ulIFN-β upstream SAR) have been subdivided by internal restriction sites. Potential SAR-binding motifs are designated as for Figure 2 with the exception that $(A)_4$ tracts and $AAN_{9-10}AA$ motifs (thin horizontal bars) are shown in addition to $(A)_{>4}$ tracts and $AAAN_{7-10}AAA$ motifs (thick horizontal bar). The weak binding mode for XII-derived subfragments is shown in insert A and of VI-derived fragments in insert B. Fragment lengths (base pairs) are shown in the right lanes of inserts A and B.

from the same elements. This series comprised oligomers of the 314-bp insert as well as oligomers derived from the *Sph*I–*Sma*I subfragment, each cloned into pTZ18R. Unless stated otherwise, these species were excised via the flanking *Hin*dIII and *Eco*RI sites and labeled by Klenow polymerase. In one of these experiments, the remote *Pvu*II site of the vector was chosen in place of *Eco*RI to trace the effect of a certain amount of non-SAR DNA (IX).

Figure 2 (insert C) comprises binding of element VI, i.e., the 314-bp SAR sequence within the 366-bp *Hin*dIII–*Eco*RI fragment in a standard SAR binding assay. Here it is compared with its own dimer (VI_2) and a cloned 381-bp fragment containing a dimeric insert composed of 166 bp to the right of the intrinsic *Sph*I site, $(VIII)_2$. Furthermore, the same 166-bp fragment, excised from a vector as a 420-bp *Hin*dIII–*Pvu*II fragment (IX), has lost its binding capacity due to a significant stretch of plasmid DNA. Several conclusions can be drawn from this experiment. First, dimerization of VI does not grossly improve its already existing binding capacity. Second, dimerizing its right-hand part creates a fragment, $(VIII)_2$, that in its size and binding behavior resembles VI, probably because it maintains the same number of plain ATATTT boxes.

When the right-hand half of VI (VIII) is oligomerized to a series $(VIII)_n$ and the resulting mixture is tested, the monomer unit is no longer able to cope; i.e., in this context it is mostly displaced into the supernate (Figure 2D). Related properties have been demonstrated for a 241-bp fragment derived from the proximal end of the downstream mu IFN-β-SAR (XII, cf. Figures 1A and 5). These findings seem to indicate that fragments around 200 bp are below a critical length that is necessary for accumulating sufficient sequence information for strong binding. The example of IX (Figure 2C) has taught that extensions of this basic element have to contain additional sequence information for binding in order to yield an efficient SAR element.

The "Weak-Association" Mode. If compared with plasmid DNA, 150–200-bp SAR-derived fragments may still show a considerable tendency to associate with the scaffold, at least in the absence of major amounts of longer SAR elements. In Figure 5 we have looked into the behavior of a number of sub-SAR fragments in the absence of these strong competitors. The labeled species were derived from the above construct VI, which could be subdivided by its internal *Dra*I and *Xmn*I sites and from the murine SAR fragment XII. Within XII, binding of the 153- and 85-bp subfragments (XIV_l and XIV_r) becomes indistinguishable from the properties of plasmid DNA (fragment "Ve" in Figure 5B) and this is even true in a two-component sub-SAR/plasmid mixture (not shown). Within VI, the 160-bp *Hin*dIII–*Xmn*I fragment (VII) is comparable to the full length 241-bp fragment XII, while the association is successively lost with the 110-bp *Hin*dIII–*Dra*I and the 48-bp *Eco*RI–*Dra*I fragments (XV_l/XV_r, respectively) (Figure 5A). This occurs despite the presence of considerable oligo(A)·oligo(T) tracts and even ATATTT motifs.

Mutagenesis of ATATTT Boxes. The suggested contribution of ATATTT runs to the strong binding mode has been tested by mutagenesis. For this purpose we have chosen the segment aatATATTT, which is present both in the huIFN-β upstream SAR and at the 3′-end of the murine IgH enhancer (marked with an asterisk in Figure 6 and elsewhere). In both SAR elements the motif forms a nucleation center for the formation of non-B DNA structures under superhelical tension (see below). On the other hand, it is not part of any $AAAN_{7-10}AAA/AAN_{9-10}AA$ and oligo(dA) stretches so that potential alternative recognition sites remain unaffected.

Details of the various mutageneses are shown in Figure 6; assays have been performed for the 310-bp IgH-MAR fragment (XVI_r) and the 448-bp huIFN-β-SAR fragment (V_l). Binding has been quantified by referencing to other intrinsic

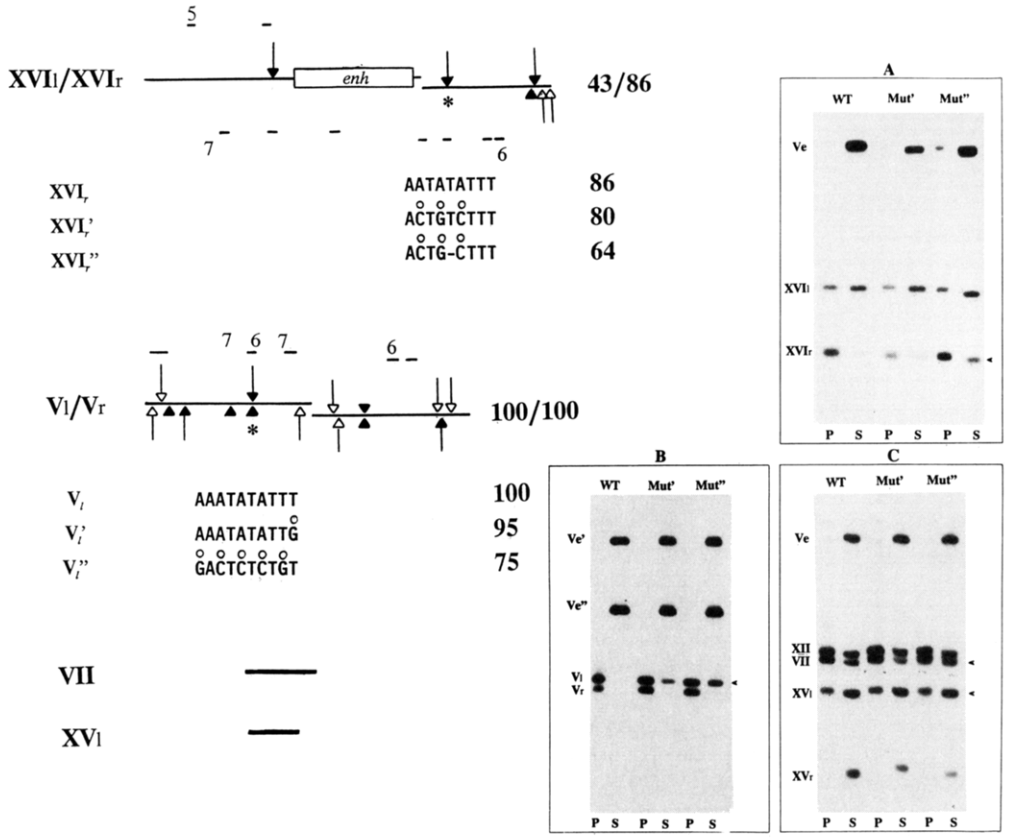


FIGURE 6: Mutation of ATATTT boxes reduces the strong binding mode of SAR fragments. aatATATTT sequences have been mutagenized as shown in the figure. The effects of these mutations have been traced for sequence XVI_r (A) and for V_l (B). Fragment XVI_r is a MAR element downstream from the IgH enhancer; fragments V_l is from the IFN-β upstream SAR. Ve' and Ve'' are the 1.8- and 0.9-kb *Eco*RI–*Xmn*I and *Xmn*I–*Hin*dIII fragments originating from the pTZ vector.

fragments that have not undergone mutagenesis (XVI_l and Ve'', respectively); see Figure 6A,B. It is seen that in both examples the decrease in binding follows the extent of mutagenesis, although it never nearly approaches zero. This is not surprising, however, since only one out of several related boxes is affected.

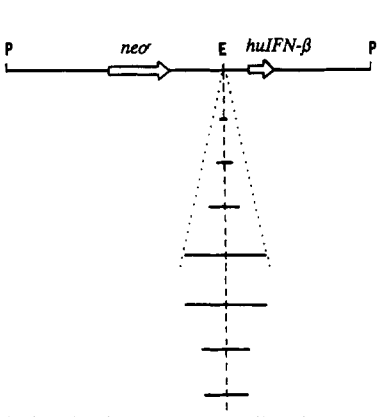
We have tried to amplify these effects by reducing the segments of DNA to the immediate site of mutagenesis by studying the shorter fragments VII and XV_l in an analogous way; both fragments belong to a size class that exhibits highly variable binding, depending on the presence of SAR length fragments. Figure 6C demonstrates that on this level the effect of mutagenesis is negligible. These data suggest that weak and strong binding modes are different as they respond differentially to the same mutations.

Strongly Binding SAR Sequences Promote Transcription. So far, our studies suggest the existence of different binding modes. A base level of binding is attained at a certain A + T content but strong binding requires a critical length of DNA and recurring motifs of the nature mentioned before. Around the interferon-β genes, fragments with this architecture extend over many kilobases of DNA (Figure 1). In the experiments to follow we asked for the biological effects of fragments that we have characterized by their in vitro properties. As a first and simple approach we report here the influence of a single element of this type that is present between the interferon and aminoglycoside-3'-phosphotransferase genes, i.e., upstream from huIFN-β and downstream from neo^r (Figure 7). Moreover, we chose conditions that strongly favored the integration of single copies. These conditions could be adjusted by applying low concentrations of plasmids that had been linearized at their *Pvu*I site and by the omission of carrier DNA. Since in our constructs the gene used for selection (neo^r) is physically linked to the marker gene (huIFN-β), gene transfer could also be achieved by electroporation, i.e., by a protocol that, in principle, eliminates multiple-copy integrations. The results in Figure 7 demonstrate that a SAR sequence of 2.2 kb enhances the induced transcription of the interferon gene at least 10-fold, independent of the gene-transfer mechanism (experiments 5A and 5B), whereas a 448-bp fragment (V_l) still yields a factor of 4 (experiment 3). Experiment 2 applies a sub-SAR fragment that is without an effect. All results are independent of SAR orientation. Two controls have been added to this series: a 1.3-kb MAR fragment derived from a light-inducible gene of potato (X) that in its structural features and binding strength resembles the huIFN-β-SAR fragment I and a similarly sized control fragment that (apart from a similar A + T content) lacks most of these properties (experiments 6 and 7, respectively). It is seen that the in vivo effects are as expected if expectations are founded on the results in Figure 4.

Although the enhancements are less pronounced, the same trends can be derived for the expression level of the neo^r gene that is transcribed under the 5'-transcriptional control of the Moloney murine sarcoma virus (MoMSV) LTR. In many cases, the transcriptional activity of this gene was reflected by the number of clones (Figure 7), showing that expression occurred in a range that does not induce complete resistance to the antibiotic.

Discussion

Scaffold- or matrix-attached regions (SARs or MARs) are usually defined as several hundred base pairs long, about 70% A + T rich DNA sequences that have a number of structural features in common. Early attempts to define a consensus sequence in these elements led to the discovery of "Topo II



	Insert	IFN-β	neo^R	Clones
1	-	10	1	200
2	VIII	10	1	200
3	V_1	40	2	400
4	IV	40	2	400
5A	I	125	3	450
5B	I	100	N.D.	450
6	X (plant-MAR)	100	4	450
7	XI (control)	10	1	300

FIGURE 7: Transcriptional enhancements mediated by selected SAR fragments. SAR fragments have been cloned into the *Eco*RI site of vector pCATIFβ-a, i.e., into a position located between the huIFN-β and neo^r genes. Both orientations have been prepared and the resulting data from at least three experiments have been averaged. All data have been normalized by setting the insertless control to average values (experiment 1, underlined) and refer to pools from about the indicated number of clones. Average copy numbers in these pools were 1.2 in the case of transfection (all experiments except 5B) and 1.0 in the case of electropolation (5B). Analyses on individual clones proved that electroporation had produced single-copy integrates in all and transfection in the majority (80%) of cases. For the remainder, restriction analyses gave no indication for the head-to-tail integration mode, which is typical of protocols favoring multiple-copy integration events.

boxes", GTN(A/T)A(T/C)ATTNATNN(G/A). Although this consensus has been derived for the topoisomerase from *Drosophila melanogaster* and is rarely at all recognized by the corresponding enzymes from mammalian cells (Darby et al., 1986, Sperry et al., 1989), variants of this loosely defined box are widespread in SAR sequences of any source (see, for example, Figure 4). Our analysis (Table I) shows that the core of this sequence [A(T/C)ATTN] frequently coincides with a motif ATATTT, which has first been described by Cockerill and Garrard (1986). More recently, the relevance of oligo(A) tracts has been discussed, which tend to narrow the minor groove to yield a rigid stretch of DNA that might form a structure recognized by scaffold proteins (Adachi et al., 1989; Izaurralde et al., 1989; Käs et al., 1989). Other analyses (theoretical and experimental) showed that a number of SAR sequences share the property of DNA bending, which requires short oligo(dA) tracts phased relative to the DNA repeat (Anderson, 1986; Homberger, 1989). Clearly, the sequences described here are in accord with all these criteria as documented by the cursive numbers [tracts of oligo(dA)] and horizontal bars (tracts of $AAAN_{7-10}AAA$) in Figures 2, 4, and 6 (cf. also Figure 5 and Table I, which include the more frequent $AAN_{9-10}AA$ tracts).

If we compare SAR-type DNAs with a DNA of similar A + T content, the occurrence of ATATTT motifs, isolated or in the context of a Topo II box, appears to be more discriminating than the other features (Figure 4 and Table I). We have also noted that in an *Eco*RI–*Hin*dIII digest of phage λ DNA only fragments between 28.7 and 31.1 kb of the genetic map display a moderate affinity for the protein scaffold (not shown), whereas maxima of bending have been traced for two regions between 0 and 2 and between 20 and 22 kb (Anderson, 1986). While many SAR sequences may contain bent DNA, this is certainly not a sufficient criterion.

During our work, attempts have been undertaken to support the role of ATATTT stretches and Topo II boxes by reducing DNA length to the immediate environments of these motifs (Figure 2). All experiments designed along these lines agreed in that strong binding (which under standard conditions is comparable to fragments I or IV) is consistently observed as far as the resulting subfragments are longer than 0.3 kb and comprise several of these motifs. Subfragments between about 0.15 and 0.10 kb, are also bound, but only if major amounts of whole-length SARs are absent from the mixture (cf. subfragments VII and VIII). This property has been termed the "weak binding mode" and we have used fragment XII (Figure 5) as an internal reference to quantify this way of association. Below 0.1 kb, any segment of DNA becomes indistinguishable from plasmid DNA in that it looses its residual affinity for the scaffold (Figure 5 and several unpublished examples). Figure 2D shows that a 0.16-kb fragment (VIII) can be provided with a high binding affinity by simple dimerization, supporting the modular architecture of functional SARs. From Figure 2C (fragment IX) it is evident that it is not simply the length of DNA adjacent to the 0.16-kb SAR sequence that enables strong binding but the repetition of a given motif that effects the conversion between these modes of association.

A titration of SAR binding sites on nuclear scaffolds has been reported before by Cockerill and Garrard (1986) and by us (Bode & Maass, 1988). Our estimate of 3000 high-affinity sites per individual scaffold has been reproduced in Figure 3A. An extension of these studies shows that this value is typical for a SAR fragment around 2 kb since significantly higher figures are obtained for the shorter fragments. These results demonstrate that it is not a single structural feature that is recognized in fragments of an extended size but that long SAR sequences are able to interact by multiple-site attachment to a proteinaceous matrix. Fragment VI exemplifies a relatively short (0.31 kb) SAR fragment showing strong binding to 14 000 or more sites per cellular scaffold equivalent. As in the case of the 2.2-kb fragment I, the association is sigmoid in nature. This apparent cooperativity should be mechanistically different from the cooperative binding of soluble scaffold proteins to SAR DNA (Adachi et al., 1989), which is explained best by protein–protein contacts that reinforce binding to DNA after an initial association of some (originally separate) protein molecules.

In our examples, SAR DNA is regarded as the ligand that binds to a high molecular protein matrix. This "scaffold" in turn may undergo conformational changes caused by a multiple-site interaction with sequences of more than 300 bp. These changes appear to be extremely slow, since the system requires 6 h or more to reach a thermodynamic equilibrium (Bode & Maass, 1988). The cooperative response is lost for the weak and competeable binding of sub-SAR fragments below a critical length (cf. sequence VIII in Figure 3B).

We consider the almost quantitative association of SAR DNA in the presence of a 10 000-fold excess of bacterial competitor DNA a strong argument for the specific mode of a SAR–scaffold interaction. This interaction must have

evolved early during the evolution of eukaryotes as it is already present in yeast (Hofmann et al., 1989). The addition of bulk eukaryotic DNA as a competitor clearly shifts the system to equilibrium with a retention of the cooperative features (Figure 3C,D).

Our hypothesis that ATATTT motifs participate in the high-affinity recognition process has been substantiated by a mutagenesis approach. Mutants were designed such that they did not interfere with possible alternative recognition modes based on bending and extended oligo(dA) tracts. Figure 6 displays the analogous results for two systems, i.e., the MAR sequence, that is located downstream from the IgG heavy-chain enhancer (containing three such boxes) and fragment V_1 from the huIFN-β gene (with five ATATTT tracts). In both cases it is possible to document the effect of mutating a single motif of this type (inserts A and B), although the remaining ones still allow for a significant binding potential. To our initial surprise, these mutation-dependent effects could not be amplified in the weakly bound subfragments, which harbor only the affected motif (Figure 6, insert C). At present we can only try to explain these observations in the light of a model that comprises several additional features (to be published):

Our choice of the particular aatATATTT sequences used for mutagenesis was biased by the observation that under superhelical tension this tract is a nucleation center for strand separation both in the context of the IgH-enhancer MAR (Kohwi-Shigematsu and Kohwi, submitted for publication) and the huIFN upstream SAR sequence (Kohwi-Shigematsu and Bode, manuscript in preparation).

Tsutsui et al. (1988) reported that the nuclear scaffold exhibits DNA-binding sites selective for supercoiled DNA. We could confirm their observations using plasmid pTZ18R and the same plasmid with fragment IV inserted into its *Eco*RI and *Hin*dIII sites. In both cases the association had to be classified as weak because it was efficiently competed for by linear SAR fragments.

Probst and Herzog (1985) demonstrated that DNA recovered from matrix preparations exposes single-stranded sites. We have found that by simple thermal denaturation any DNA fragment is converted to a species exhibiting a weak affinity for isolated scaffolds, although only in the absence of authentic (double-stranded) SAR elements.

Considering that all these species (supercoiled, single-stranded, and double-stranded, i.e., SAR-type, DNA) appear to compete for the same limited set of sites on the scaffold, it becomes likely that the recognition of single strands has to do with SAR function. Hence, the association of SAR elements may involve regions that are primed for strand separation (e.g., ATATTT tracts). If they are properly spaced, the intervening DNA may form a loop as a consequence of binding to scaffold proteins. This loop formation may induce a local unwinding of particular A + T rich sequences, especially in cases where the binding elements are displaced with respect to the helical repeat of DNA. Such a process would then reinforce the scaffold association. Short DNA segments would not permit such a loop formation, explaining their behaviour in the in vitro assay. In vivo, the association of superhelical and/or single-stranded DNA could serve as an auxiliary mechanism, which might support an intermediate and possibly transcription-dependent scaffold association during phases of gene expression (Tsutsui et al., 1988).

These studies have been initiated to answer the question of whether or not an in vitro scaffold-reassociation assay is able to predict the potential of SAR sequences for a transcriptional enhancement. The plasmids we have designed to this end contain a marker gene (huIFN-β) and a selector gene (neo^r), which are physically linked to enable various modes of gene transfer. In principle, such an approach tends to minimize the effects of an integral SAR sequence: the expression of the selector gene in the basic construct (experiment 1 in Figure 7) depends on favorable integration sites in the host's genome to a larger extent than in a SAR construct. This argument has been forwarded by Blasquez et al. (1989) and is supported by their data. Our observations (Figure 7), that the number of clones obtained after transfection of a SAR-free plasmid is low (as only few integration sites enable levels of neo^r expression sufficient for G-418 resistance) and relatively high in case of SAR-containing constructs (because they enhance the capacity of mediocre sites), tend into the same direction. As a consequence, we will compare a relatively small cellular population with a priori favorable integration sites (experiments 1 and 2 in Figure 7) with a larger population based on SAR-improved sites. Despite this experimental design, it is quite evident that potent SAR sequences of a sufficient length are able to stimulate the expression of the marker gene 10–12-fold and that of the selector gene up to 4-fold; these enhancements are independent of SAR orientation. The localization of the SAR sequence between the two coding regions might cause their functional separation in the sense that both belong to separate domains. Since we select only for the expression of the neo^r gene, this may be the reason for a much stronger enhancement of a physically linked but functionally separate marker gene.

Appropriate controls have been included into this series of experiments. Figure 7 shows that a 1.3-kb intronic MAR element from plants, which has been selected solely by its in vitro binding properties, behaves as efficiently as an authentic 2.2-kb SAR element in our assay. Its origin makes it highly unlikely that these effects are due to the presence of other features such as cryptic promoter or enhancer sequences. It also supports our notion that SAR functions depend on the presence of specific sequences like ATATTT tracts and/or "Topo II-boxes", which are present in this class of A + T rich elements and not on other features that are more or less a property of any A + T rich DNA (Figure 7, cf. experiments 6 and 7). A comparison of the data derived from fragments I and X suggests that maximum efficiencies may be achieved with SAR fragments around 1 kb. Efforts will therefore continue to obtain such fragments by screening from the repertoire of several species.

In conclusion, we have selected efficient SAR elements on the basis of their in vitro properties. In order to be efficient, they must bind to the scaffold in the high-affinity mode, which requires a certain length of SAR DNA and a repetition of elements of the ATATTT type. In vitro, these elements associate with the scaffold in a cooperative manner. A useful criterion for screening functional SAR elements is their high-affinity binding in the presence of known fragments, which include at least one standard SAR sequence. This will direct any weakly binding fragment to the supernate fraction and allow, by simple comparison, for convenient information about the relative binding strength of the remaining species. It is predicted that any fragment that is recovered by this approach will enhance gene expression after stable integration and that two SAR elements flanking a transcription unit will be most efficient (Klehr et al., submitted for publication).

ACKNOWLEDGMENTS

We thank A. Pucher for sequencing the huIFN-β upstream SAR element. Thanks are also due to R. Duggleby for his

assistance in adapting his program, DNRP53. We further appreciate numerous intriguing discussions with D. Klehr and the expert technical assistance of K. Maass.

REFERENCES

Adachi, Y., Käs, E., & Laemmli, U. K. (1989) *EMBO J, 8*, 3997–4006.
Anderson, J. N. (1986) *Nucleic Acids Res. 21*, 8513–8533.
Blasquez, V. C., Xu, M., Moses, S. C., & Garrard, W. T. (1989) *J. Biol. Chem. 264*, 21183–21189.
Bode, J., & Maass, K. (1988) *Biochemistry 27*, 4706–4711.
Bode, J., Pucher, H. J., & Maass, K. (1986) *Eur. J. Biochem. 158*, 393–401.
Chomczynski, P., & Sacchi, N. (1987) *Anal. Biochem. 162*, 156–159.
Cockerill, P. N., & Garrard, W. T. (1986) *Cell 44*, 273–282.
Cockerill, P. N., Yuen, M. H., & Garrard, W. T. (1987) *J. Biol. Chem. 262*, 5394–5397.
Darby, M. K., Herrera, R. E., Vosberg, H. P., & Nordheim, A. (1986) *EMBO J, 5*, 2557–65.
Dinter, H., & Hauser, H. (1987) *Eur. J. Biochem. 166*, 103–109.
Dirks, W., Mittnacht, S., Rentrop, M., & Hauser, H. (1989) *J. Interferon Res. 9*, 125–133.
Duggleby, R. G. (1984) *Comput. Biol. Med. 14*, 447–455.
Duggleby, R. G. (1988) *J. Theoret. Biol. 130*, 123–124.
Forrester, W. C., Novak, U., Gelinas, R., & Groudine, M. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86* (14), 5439–5443.
Gasser, S. M., & Laemmli, U. K. (1986) *EMBO J. 5*, 511–518.
Gasser, S., & Laemmli, U. K. (1987) *Trends Genet. 3*, 16–22.
Gross, D. S., & Garrard, W. T. (1987) *Trends Biochem. Sci. 12*, 293–297.
Hofmann, J. F.-X., Laroche, T., Branas, A. H., & Gasser, S. M. (1989) *Cell 57*, 725–737.
Homberger, H. P. (1989) *Chromosoma 98*, 99–104.
Izaurralde, E., Mirkovitch, J., & Laemmli, U. K. (1988) *J. Mol. Biol. 200*, 111–126.
Izaurralde, E., Käs, E., & Laemmli, U. K. (1989) *J. Mol. Biol. 210*, 573–585.
Jarman, A. P., & Higgs, D. R. (1988) *EMBO J. 7*, 3337–3344.
Käs, E., & Chasin, L. A. (1987) *J. Mol. Biol. 198*, 677–692.
Käs, E., Izaurralde, E., & Laemmli, U. K. (1989) *J. Mol. Biol. 210*, 587–599.
Klehr, D., & Bode, J. (1988) *Mol. Gen. (Life Sci. Adv.) 7*, 47–52.
Lutfalla, G., Blanc, H., & Bertolotti, R. (1985) *Somatic Cell Mol. Gen. 113*, 223–238.
Mirkovitch, J., Mirault, M. E., & Laemmli, U. K. (1984) *Cell 39*, 223–232.
Phi-Van, L., & Strätling, W. H. (1988) *EMBO J. 7*, 655–664.
Probst, H., & Herzog, R. (1985) *Eur. J. Biochem. 146*, 167–171.
Scheuermann, R. H., & Chen, U. (1989) *Genes Dev. 3*, 1255–1266.
Sperry, A. O., Blasquez, V. C., & Garrard, W. T. (1989) *Proc. Natl. Acad. Sci. U.S.A. 86*, 5497–5501.
Stief, A., Winter, D. M., Strätling, W. H., & Sippel, A. E. (1989) *Nature (London) 341*, 343–345.
Stockhaus, J., Eckes, P., Blau, A., Schell, J., & Willmitzer, L. (1987) *Nucleic Acids Res. 158*, 3479–3491.
Tsutsui, K., Tsutsui, K., & Muller, M. T. (1988) *J. Biol. Chem. 263*, 7235–7241.
van Assendelft, G. B., Hanscombe, O., Grosveld, F., & Greaves, D. R. (1989) *Cell 56*, 969–977.
Wigler, M., Silverstein, S., Lee, L.-S., Pellicer, A., Cheng, Y.-C., & Axel, R. (1977) *Cell 11*, 223–232.
Wirth, M., Bode, J., Zettlmeissl, G., & Hauser, H. (1988) *Gene 73*, 419–426.
Xu, M., Hammer, R. E., Blasquez, V. C., Jones, S. L., & Garrard, W. T. (1989) *J. Biol. Chem. 264*, 21190–21195.

DNA Solution Conformation via Infrared Circular Dichroism: Experimental and Theoretical Results for B-Family Polymers†

Wenxin Zhong, Miriam Gulotta, Dixie J. Goss, and Max Diem*
Department of Chemistry, City University of New York, Hunter College, 695 Park Avenue, New York, New York 10021
Received November 13, 1989; Revised Manuscript Received March 28, 1990

ABSTRACT: Infrared (vibrational) circular dichroism (VCD) has been observed for the DNA models $d(CG)_5$, poly(dG-dC)·poly(dG-dC), poly(dG)·poly(dC), poly(dA-dT)·poly(dA-dT), and poly(dA)·poly(dT) in the B-conformation in buffered, aqueous solution. The observed results are quantitatively interpreted in terms of the exciton model for coupled carbonyl stretching vibrational states.

Circular dichroism in vibrational transitions, observed in the infrared spectral region, has recently been applied to the elucidation of nucleic acid solution conformation (Annamalai & Keiderling, 1987; Gulotta et al., 1989). For nucleic acids, both electronic CD and VCD signals arise from the dipolar coupling of (electronic or vibrational) transitions localized within the bases. These transitions interact, or couple, and produce absorption and circular dichroism features that are governed by the exciton theory (Tinoco, 1963).

However, VCD offers the advantage over conventional electronic CD, observed in the visible/UV spectral region, of more and narrower spectral features. In addition, for certain transitions, such as carbonyl stretching vibrations, the vibrational modes giving rise to the spectral features are well localized, and the dipole change occurs in a direction that is nearly parallel to the bond connecting the carbon and oxygen

†This research was supported by the National Science Foundation (CHE 86-07070 to D.J.G.), an American Heart Association–NYC Established Investigatorship (D.J.G.), the National Institutes of Health (GM 28619 to M.D.), and several City University of New York Faculty Research Awards. The VCD instrument utilized was constructed with funds derived from a previous NSF grant to M.D. (CHE 86-07934). DNA structural computations were carried out on the departmental molecular modeling facility, purchased under RCMI Grant RR 03037.

0006-2960/90/0429-7485$02.50/0