Proc. Natl. Acad. Sci. USA
Vol. 93, pp. 10891–10896, October 1996
Genetics

NOMAD: A versatile strategy for *in vitro* DNA manipulation applied to promoter analysis and vector design

DMITRI REBATCHOUK, NIKOLAI DARASELIA, AND JONATHON O. NARITA†

Department of Biological Sciences, Laboratory for Molecular Biology, University of Illinois at Chicago, 900 South Ashland Avenue, Chicago, IL 60607

Communicated by Emanuel Margoliash, University of Illinois at Chicago, Chicago, IL, June 3, 1996 (received for review January 16, 1996)

ABSTRACT **Molecular analysis of complex modular structures, such as promoter regions or multi-domain proteins, often requires the creation of families of experimental DNA constructs having altered composition, order, or spacing of individual modules. Generally, creation of every individual construct of such a family uses a specific combination of restriction sites. However, convenient sites are not always available and the alternatives, such as chemical resynthesis of the experimental constructs or engineering of different restriction sites onto the ends of DNA fragments, are costly and time consuming. A general cloning strategy (nucleic acid ordered assembly with directionality, NOMAD; WWW resource locator http://Lmb1.bios.uic.edu/NOMAD/NOMAD.html) is proposed that overcomes these limitations. Use of NOMAD ensures that the production of experimental constructs is no longer the rate-limiting step in applications that require combinatorial rearrangement of DNA fragments. NOMAD manipulates DNA fragments in the form of "modules" having a standardized cohesive end structure. Specially designed "assembly vectors" allow for sequential and directional insertion of any number of modules in an arbitrary predetermined order, using the ability of type IIS restriction enzymes to cut DNA outside of their recognition sequences. Studies of regulatory regions in DNA, such as promoters, replication origins, and RNA processing signals, construction of chimeric proteins, and creation of new cloning vehicles, are among the applications that will benefit from using NOMAD.**

DNA of any living organism can be regarded as a modular structure consisting of individual elements such as regions coding for protein domains and protein-binding sites. All these elements, or modules, perform discrete functions and cannot be subdivided without loss of functionality. Modules are characterized by a certain degree of context independence and are often found in different combinations *in vivo*. For instance, regulatory regions such as promoters, replication origins, and RNA processing signals usually consist of many distinct protein-binding sites, and closely related or identical sites are commonly found in different regulatory regions.

To elucidate the function of a modular regulatory region in DNA, such as a promoter, it is not sufficient to merely identify individual regulatory modules. It is necessary to determine the spacing, orientation, and order of the constituent modules required for their normal interaction and, consequently, for the function of the region as a whole. This information can be obtained by systematically rearranging DNA modules *in vitro* and testing the resulting constructs. Similarly, RNA domains that regulate translational efficiency, subcellular localization, or processing can be dissected and tested by rearranging individual elements in experimental DNA constructs. The same approach can be applied to studies of multi-domain proteins, where the function of a complex protein is determined by the activities of the individual domains interacting with one another. Thus, modular analysis, or functional testing of experimental constructs having altered modular structure, is a powerful approach to a wide variety of biological problems. However, complications of a purely technical nature have made it difficult to consistently apply the concept of modular analysis. Traditional cloning techniques require a specific combination of restriction sites to be used for the creation of every new construct having an altered order of constituent elements. The required sites are not always available, and introduction of specific restriction sites onto the ends of DNA fragments becomes prohibitively time consuming as the number of required constructs increases. Significantly, rearranged modular constructs produced by the conventional technique are not always directly comparable to one another and may require additional control experiments, because the order of modules is not the only variable that changes. The sequence at module junctions is unavoidably altered as different adaptors are used. This can be critical, especially when short protein-binding regions are rearranged. Thus, limitations intrinsic to the traditional cloning strategy generally preclude comprehensive modular analysis of all but the simplest modular structures. The alternative strategy, direct chemical resynthesis of the required constructs, is resource intensive and thus acceptable only for relatively short regions and a limited number of test constructs.

The goal of this work was to develop a general strategy for combinatorial manipulation of DNA fragments *in vitro*. Nucleic acid ordered module assembly with directionality (NOMAD) permits the researcher to rapidly create families of DNA constructs of predetermined structure from a pool of fragments. NOMAD is well suited for modular analysis because DNA fragments containing individual regulatory or structural elements can be directionally combined in any desired order in the experimental construct. This strategy has numerous uses, including construction of chimeric proteins and analysis of DNA sequences containing multiple regulatory elements.

NOMAD is also applicable to the construction of cloning vectors. Common cloning vehicles consist of segments, which confer particular behaviors onto the vector, such as the ability to replicate in a particular host or render the host resistant to an antibiotic. DNA fragments containing origins of replication, resistance markers, reporter genes, and other similar elements are repeatedly used in different vectors. The ability to combine vector components with unlimited flexibility would streamline creation of "custom designed" vectors.

NOMAD uses two major elements: DNA fragments having standardized end structures, termed modules (Fig. 1*A*) and assembly vectors (Fig. 1 *B* and *C*) that allow for sequential and directional insertion of any number of modules, as well as their release individually or in blocks ("composite modules"). To

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: NOMAD, nucleic acid ordered module assembly with directionality; IIS REase, type IIS restriction endonuclease; HMG2, 3-hydroxy-3-methylglutaryl-CoA reductase 2; UTR, untranslated region; SIS, sequential insertion site; eSIS, excisable SIS.

†To whom reprint requests should be addressed. e-mail: jon@uic.edu.

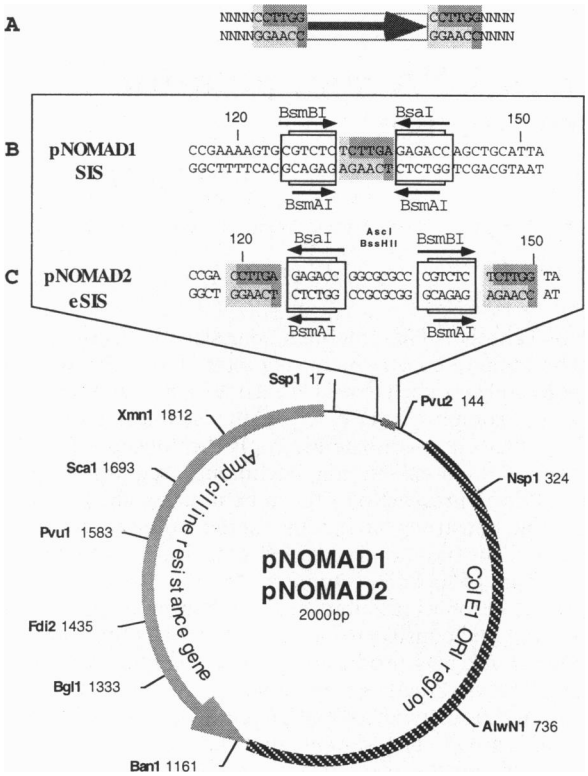

FIG. 1. (*A*) A CCTTGG-bracketed NOMAD module as a part of a larger DNA fragment, such as a PCR product. An arbitrarily selected "positive" orientation is shown by an arrow in the body of a module. pNOMAD compatible cohesive ends (on shaded background) are produced upon digestion with *Sty*I. (*B* and *C*) pNOMAD1 and pNOMAD2 are identical beyond the depicted regions: a sequential cloning site (SIS) (*B*) in pNOMAD1 and an excisable SIS (eSIS) (*C*) in pNOMAD2. *Bsa*I, *BsmA*I, and *Bsm*BI recognition sites are boxed; linkers that upon cleavage produce the *Sty*I (CCTTGG) compatible cohesive ends are on shaded background.

manipulate the modules, NOMAD relies on the properties of the type IIS restriction endonucleases (IIS REases). IIS REases efficiently cut DNA at a precise distance outside of their recognition sequence and can produce sticky ends of any sequence within the overall structure (e.g., a 4-nucleotide 5′ overhang). Digestion of a vector with a IIS REase, followed by insertion of a DNA fragment having compatible cohesive ends, leaves the type IIS site intact and unduplicated. The vector can be cut again by the same IIS REase, and another DNA fragment can be inserted next to the one inserted at the previous step. When the sticky ends used in joining are rotationally nonequivalent, the fragments are inserted directionally.

Several vectors that use IIS REases for manipulating DNA have been described (refs. 1–3; for review see ref. 4). However, all of these vectors were constructed for use in specialized applications, and no system as versatile as NOMAD has been previously proposed. In addition, vectors pNOMAD1 (Fig. 1*B*) and pNOMAD2 (Fig. 1*C*) incorporate several key features absent from the earlier constructs. Convergently oriented recognition sequences for two different type IIS REases (*Bsa*I and *Bsm*BI) flank the insertion site of pNOMAD1 (an SIS), allowing for sequential addition of new modules to either side of an initial insert (Fig. 2-1). The insert, which can consist of several modules, can be released as a whole by cutting with both IIS REases at the same time. This feature is notable because new constructs do not always have to be built by adding all of the modules one by one. Previously assembled multi-module blocks (composite modules) can be inserted in their entirety. In fact, creation of families of related constructs

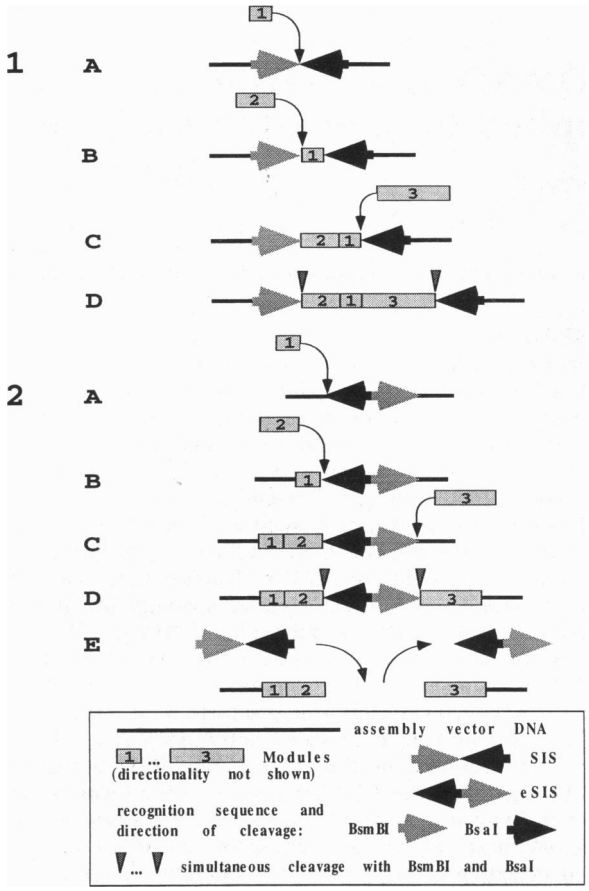

FIG. 2. (1) Sequential insertion of the *Sty*I (CCTTGG)-compatible modules into the assembly vector pNOMAD1. The vector is cut with *Bsm*BI, *Bsa*I, or *Bsm*AI to produce sticky ends compatible with *Sty*I (5′-CCTTGG-3′), and the first module is inserted (A). The second module can be inserted either 5′ or 3′ to the first. When the assembly vector is cut by *Bsm*BI, the second module is inserted 5′ to the first (B). Insertion of the third module 3′ to the previously inserted ones (C) is performed using *Bsa*I to cut the vector. Subsequent modules can be added 5′ or 3′ to the composite module. The composite module can be released by cutting with *Bsm*AI or by simultaneous cutting with *Bsm*BI and *Bsa*I (D). A modular construct can be disassembled into individual modules by digestion with *Sty*I and *Bsm*AI. Digestion with *Sty*I alone will release all modules but the "outermost." Insertion of modules is directional. (2) Use of pNOMAD2 as the basis for construction of custom assembly vectors. "Structural" modules are successively cloned into pNOMAD2 (A–C). eSIS is excised by simultaneous digestion with *Bsa*I and *Bsm*BI (or with *Bsm*AI alone) (D). An SIS-containing module replaces the eSIS, producing a custom assembly vector (E). All structural modules (cloned before the SIS-containing module was inserted) are physically separated from the "experimental" modules, which are inserted into the SIS and can be released as a composite module that does not include any structural components. Note that any module can be structural as well as experimental, depending on the details of a particular experiment.

should be planned in such a way as to maximize composite module reuse. For convenience, the insertion site of pNOMAD1 is designed so that a third type IIS REase (*Bsm*AI) recognizes both sequences flanking the insertion site and can release the insert by itself. Regardless of which SIS restriction enzymes are used, all released fragments have the same end structure and can be used as modules in subsequent assembly steps.

The second NOMAD series vector, pNOMAD2 (Fig. 1*C*), is designed to serve as a basis for construction of custom assembly vectors (see *Discussion*). pNOMAD2 uses the concept of excisable cassettes. Other implementations of the concept have been described (e.g., ref. 5). The important

difference, however, is that the predecessor vectors could not be used for sequential assembly because the excisable cassette contained two divergently oriented recognition sequences for a single enzyme. In a manner similar to pNOMAD1, pNOMAD2 permits sequential addition of modules (Fig. 2–2).

The recognition sequence for the type II restriction enzyme *Sty*I (C/CTTGG) is used to bracket the modules. Cutting the modules with *Sty*I produces sticky ends directionally compatible with pNOMAD vectors digested with any of the IIS REases used in NOMAD. Having *Sty*I sites at every module junction allows for convenient disassembly of composite modules.

The utility of NOMAD can be best realized if numerous research groups cooperate in creating commonly used modules. To facilitate the exchange of information on NOMAD and reduce the duplication of effort, a World Wide Web (WWW) site has been set up (see *Abstract*) that provides information on NOMAD, as well as the interface to a data base of available modules. We propose that the recognition site for *Sty*I (CCTTGG) is adopted as a standard module linking sequence. However, NOMAD assembly vectors are very flexible and can be easily modified to use four bp 5′ overhangs of any sequence (see *Discussion*), as required by particular applications.

In this paper the applications of NOMAD to promoter analysis and construction of fully modular vectors are presented as examples of the potentially broad applicability of NOMAD to manipulating DNA *in vitro*. The focus of our current research is transcriptional regulation of the genes involved in isoprenoid biosynthesis in tomato. These genes include the 3-hydroxy-3-methylglutaryl-CoA reductase 2 (HMG2). HMG2 is highly expressed at specific stages in tomato development. The promoter region of HMG2 has been sequenced and preliminary data from *in vitro* DNaseI footprinting experiments showed the existence of several protected areas (J.O.N. and T. Manzara, unpublished work). In agreement with the footprinting data, transient expression (bombardment) assays using promoter–reporter constructs at various stages of development showed that the 5′ untranslated leader region of HMG2 contains several elements important for high level expression (6). The effects of these regulatory elements on the expression of the tomato HMG genes can be conveniently studied using NOMAD. This work was initiated by making modules encompassing the entire promoter and 5′ untranslated leader of the HMG2, recreating a wild-type module arrangement and testing the resulting constructs in the transient expression assay. To conveniently test promoter modular constructs, as well as to demonstrate the general utility of NOMAD in vector construction, a series of plant reporter vectors has been created. These vectors were constructed by inserting general purpose modules—a constitutive promoter, the luciferase reporter gene, and a terminator fragment— into the minimal assembly vector, pNOMAD1.

MATERIALS AND METHODS

DNA restriction and modification enzymes used in this study were purchased from New England Biolabs and Life Technologies (Gaithersburg, MD) and were used according to the manufacturer's instructions. Oligonucleotides were from Universal DNA (Tigard, OR). KlenTaq and Pfu DNA polymerases were from AB Peptides (St. Louis) and Stratagene, respectively. A Sequenase v. 2.0 kit (United States Biochemical) was used for sequencing. Propagation and transformation of *Escherichia coli* were performed by conventional methods (7).

Design of Assembly Vectors pNOMAD1 and pNOMAD2. All changes to the parental plasmid pBluescript SK(±) (Stratagene) to produce pNOMAD vectors were made with a PCR-based mutagenesis approach using divergently oriented primers. To eliminate the recognition sites for the NOMAD enzymes present in the parental plasmid, C at pBluescript SK position 2117 was changed to A (GAGA*C* to GAGA*A*) and G at position 2878 was changed to C (*G*TCTC to *C*TCTC). Next, the unwanted portion of pBluescript was substituted with the sequential insertion site, resulting in pNOMAD1, or with the excisable sequential insertion site, resulting in pNOMAD2. pBluescript SK sequence from position 974 to position 2958, comprising the β-lactamase gene (which determines resistance to ampicillin) and the *Col*EI origin of replication, became incorporated into pNOMAD1. The sequence GTCTCTCTTGAGAGA, which comprises most of the SIS, was substituted for the rest of pBluescript SK(+). The eSIS was introduced at the same site, except that the pBluescript SK sequence from position 984 to position 2951 was incorporated into pNOMAD2. The sequence CCTTGAGAGACCGGCGCGCCCGTCTCTCTTGG, which comprises the eSIS, was substituted for the rest of pBluescript SK. Changes introduced by site-directed mutagenesis were confirmed by dideoxy sequencing. The above manipulations resulted in vectors with a deduced size of 2000 bp. PCR amplification may have resulted in more changes to the sequence of the parental vector. However, the ability of the vectors to confer resistance to ampicillin onto the bacterial host, as well as their ability to replicate in *E. coli*, remained apparently unchanged. The plasmid yield from strains harboring pNOMAD1 and pNOMAD2 is not significantly different from pBluescript SK.

Construction of pLucS. pUC119-based promoter-less luciferase reporter vector pLuc (8, 9) was modified to allow for directional insertion of modules. The *Xba*I site in the multiple cloning site of that vector was precisely replaced with an *Sty*I (CCTTGG) site. The derivative vector was designated pLucS (for pLuc–*Sty*I). All tomato HMG2 promoter constructs were assayed using pLucS.

Creation of Modules. Two methods were used to create modules: annealing of complementary synthetic phosphorylated oligonucleotides having *Sty*I (CCTTGG)-compatible 5′ extensions and PCR amplification with primer–adaptors. Tomato HMG2 TATA box and transcriptional start-site modules (designated, respectively, [TATA] and [trx]) were made of synthetic oligonucleotides. [TATA] was produced by annealing oligonucleotide 1 (CTTGGTTCCTCTATAAATACATTTC) and oligonucleotide 2 (CAAGGAAATGTATTTATAGAGGAAC), and [trx] by annealing oligonucleotide 3 (CTTGGTCTTCTCTTCTCCTCACATC) and oligonucleotide 4 (CAAGGATGTGAGGAGAAGAGAAGAC). Larger modules were created by PCR using plasmid DNA as a template. Primer–adaptors contained a four-nucleotide GC-rich "clamp" and the *Sty*I recognition site (CCTTGG or CCAAGG, depending on the orientation of the primer), followed by a region of exact homology to the template DNA. A PCR fragment produced with primer–adaptors was cut with *Sty*I and inserted into pNOMAD1. To minimize the possibility of PCR errors, the Long and Accurate PCR technique (10) was used. The module [5′Reg], which contains regulatory regions important for HMG2 high-level expression (6), and the first nine triplets of the protein coding sequence, was created using the cloned HMG2 gene (J.O.N., unpublished work) as a template. Oligonucleotides 5 (GAGGCCTTGGACTCTTCTTTTAACAATTATAC) and 6 (GAGGCCAAGGTCTTCAGATCTCCGGCGAAC) were used as primer–adaptors. The module [TATA-5′Reg], encompassing the tomato HMG2 TATA box, 5′ untranslated region (UTR), and the first nine triplets of the protein coding sequence, was created using oligonucleotides 6 and 7 (GGAGCCTTGGTTCCTCTATAAATACATTTCC) as primer–adaptors. The luciferase module [Luc+] was created using oligonucleotides 8 (GGAGCCTTGGCATGGAAGACGCCAAAAAC; initiator ATG underlined) and 9 (GGAGCCAAGGTTACACGGCGATCTTTCCG; complement of the terminator TAA codon underlined) as primer–adaptors. The [Luc+] module was

created from a modified version of luciferase enzyme known as luc+. The template vector pSP-luc+ was purchased from Promega. The 1.6-kb [Luc+] was sequenced at the termini and shown to encode functional luciferase in the transient expression assay. Note that the current version of [Luc+] contains three *Bsm*AI recognition sites. The *Agrobacterium tumefaciens* Ti plasmid nopaline synthase gene terminator was converted into a NOMAD module [nosT] by PCR amplification of a corresponding portion of the pLuc reporter plasmid with oligonucleotides 10 (GGAGCCTTGGATCGTTCAAACATTTGGC) and 11 (GGAGCCAAGGATCTAGTAACATAGATGAC) as primer–adaptors. To produce the [35S] module, 523 bp of the [35S] promoter of cauliflower mosaic virus were surrounded by *Sty*I (CCTTGG) sites using PCR with primer–adaptors oligonucleotide 12 (GGAGCCTTGGAGTCAAAAATTCAGATCGAGG) and 13 (GGAGCCAAGGATCCTCTCCAAATGAAATG), and the plasmid pDO432 (10) as a template. At present [35S] contains two *Bsm*AI recognition sites. [nosT] and [35S] were fully sequenced. Note that pDO432 is referred to descriptively as 35S-Luc throughout the text.

Cloning of Modules into pNOMAD1. pNOMAD1 was cut with either *Bsm*AI, *Bsm*BI, or *Bsa*I prior to insertion of the first module. To clone additional modules into a modular construct, *Bsm*BI was used to insert the new module 5′, and *Bsa*I 3′ to the previously inserted module (Fig. 2–1). To prevent self-ligation, the linearized assembly vector was treated with calf intestinal phosphatase. Cloned modules were excised from the corresponding vectors by digestion with *Bsm*AI or by simultaneous digestion with *Bsa*I and *Bsm*BI, and agarose gel purified by conventional methods (7). To 20 ng (≈15 fmol) of dephosphorylated vector, a gel-purified module (in the case of the modules being recloned and modules produced by PCR) or a mixture of annealed phosphorylated oligonucleotides was added at a 5-fold molar excess, and ligation was performed in a total volume of 20 μl using T4 DNA ligase, for 1–3 hr at room temperature. Competent DH5α cells were transformed using standard techniques (7), and bacteria were incubated overnight at 37°C. Recombinant colonies were screened by PCR. The [TATA][trx][5′Reg] composite module was assembled in pNOMAD1 and recloned into pLucS for use in the transient assay. The [TATA-5′Reg] module was directly cloned into pLucS. Component modules of the plant expression vector pNPR1 were successively cloned into pNOMAD1. The final tHMG2 promoter/5′Reg constructs were fully sequenced. Larger modular constructs created from previously characterized modules were sequenced across module junctions.

Transient Expression Assay. Transient expression assays using the BioListic Particle Bombardment System (Bio-Rad) are described elsewhere (6). Bombardments were performed in triplicate, and all experiments were repeated at least twice. Young tomato fruits (≈7 mm in diameter) and tomato leaves were used in bombardment experiments with the HMG2 and 35S-driven transient expression, respectively.

RESULTS

Construction of Modular Assembly Vectors pNOMAD1 and pNOMAD2. Assembly vectors pNOMAD1 and pNOMAD2 were derived from the phagemid pBluescript SK (Stratagene) so as to include only the β-lactamase gene (which confers resistance to ampicillin) and the *Col*EI origin of replication. The rest of pBluescript was substituted by the SIS in the case of pNOMAD1 and by the eSIS in the case of pNOMAD2.

pNOMAD1 SIS (Fig. 1*B*) consists of the *Bsm*BI recognition sequence (CGTCTC), a spacer (TCTTGA), and the *Bsa*I recognition sequence in the opposite orientation (GAGACC). Note that the *Bsm*AI recognition sequence (GTCTC) is part of both the *Bsa*I and *Bsm*BI sites. Digestion of pNOMAD1 with any of the three restriction endonucleases having recognition sequences within the SIS (*Bsa*I, *Bsm*AI, and *Bsm*BI) cuts the TCTTGA spacer and forms cohesive ends of the same structure (3′-GAAC-5′ and 5′-CTTG-3′). The sticky ends are rotationally nonequivalent and hence permit directional insertion of a module with complementary sticky ends into the vector (Fig. 2–1). Another module can be inserted 5′ or 3′ to the first, by cutting the plasmid with *Bsm*BI and *Bsa*I, respectively. The procedure is repeated until a recombinant molecule of the desired structure is assembled. When necessary, all inserted modules can be released from the assembly vector as a group by cutting with *Bsm*AI, or by simultaneous cutting with *Bsm*BI and *Bsa*I.

The eSIS of pNOMAD2 (Fig. 1*C*) consists of the following parts, in 5′ to 3′ order: a spacer CCTTGA that produces *Sty*I-compatible sticky ends, a *Bsa*I recognition site oriented so as to cut at the spacer, an *Asc*I/*Bss*HII recognition site (not used in NOMAD), a *Bsm*BI site, and an adjoining spacer TCTTGG that forms *Sty*I-compatible sticky ends when cut with *Bsm*BI. Cleavage with *Bsa*I or *Bsm*BI allows for successive insertion of modules into pNOMAD2 (Fig. 2–2). Simultaneous cleavage of pNOMAD2-based vectors with *Bsa*I and *Bsm*BI releases the eSIS, which can be replaced by any module, including a module containing the sequential insertion site. The primary role for pNOMAD2 is to serve as the foundation for custom assembly vectors.

Modular Assembly of the Tomato HMG2 Promoter/5′ Regulatory Region. The utility of NOMAD for promoter analysis was demonstrated using the tomato HMG2 promoter/5′ regulatory region (5′Reg). The promoter was reconstructed from three modules (Fig. 3). The modules were designed so as to replace parts of the endogenous sequence with *Sty*I recognition sites, rather then to insert them. Thus, the spacing between the TATA box, transcription start site, and the ATG initiation codon remained unchanged. Two of the modules ([TATA], which encompasses the HMG2 TATA box and [trx], the gene's transcription start site region) are each 25 bp long. These structural elements are both highly conserved in all of the tomato HMG genes (6). The third module ([5′Reg]) is 138 bp long. [5′Reg] encompasses the entire 5′ UTR, as well as the first nine triplets of the HMG2 coding region. Functional characterization of this region is currently in progress. Available data indicate that it contains elements that are necessary and sufficient for high-level expression of tHMG2 in transient expression assays.

To permit testing of modular promoter constructs prior to creation of a fully modular plant reporter vector (pNPR1, see below), an *Sty*I (CCTTGG)-compatible reporter vector pLucS was derived from pLuc. All tomato HMG2 constructs are translational fusions with luciferase in the pLucS reporter vector. The control construct, HMG2WT-Luc, is a translational fusion of HMG2 to luciferase in pLuc, the parental plasmid of pLucS. HMG2WT-Luc (6) includes the same region of the tomato HMG2 gene as do the modular constructs

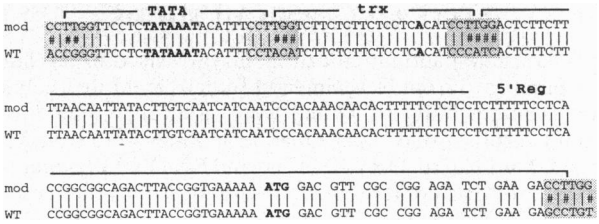

FIG. 3. Alignment of the wild-type and modular reconstitution of the tomato HMG2 promoter/5′Reg. mod, sequence of the composite module [TATA][trx][5′Reg]; WT, wild-type HMG2 sequence. [TATA-5′Reg] is the same as WT except that it is bracketed with the *Sty*I sites. Mismatches are highlighted by # signs. *Sty*I (CCTTGG) sites are shown on shaded background, the TATA box, the transcriptional start site (adenine at +1), and initiator ATG are in boldface type.

[TATA-5'Reg]-LucS and [TATA][trx][5'Reg]-LucS. Introduction of *Sty*I (CCTTGG) sites at the borders of the tomato HMG2 promoter/5'Reg region did not change its ability to drive expression of a reporter vector compared with an analogous construct not bracketed by *Sty*I sites (compare [TATA-5'Reg]-LucS and HMG2WT-Luc in Fig. 4*A*). Furthermore, it was shown that reconstruction of the tomato HMG2 promoter/5'Reg region from modules containing the TATA box, the transcriptional start site region, and the 5'Reg did not result in changes in the expression of the luciferase reporter gene in transient expression assays (compare [TATA][trx][5'Reg]-LucS to HMG2WT-Luc on Fig. 4*A*). Removal of the TATA box reduced expression of the luciferase reporter gene to background level (data not shown).

Construction of a Fully Modular Reporter Vector. The utility of NOMAD in constructing useful vectors was demonstrated by creating a modular reporter vector for transient expression in plants. pNPR1 (pNOMAD1 plant reporter 1) was created by inserting a modified luciferase (Luc+) reporter gene and the nopaline synthase gene terminator modules into the pNOMAD1 assembly vector. The luciferase module can accommodate both N-terminal translational fusions and transcriptional fusions. In transcriptional fusions, the endogenous ATG initiator codon of the luciferase gene is utilized. The sequence around the initiator codon satisfies most requirements of the Kozak consensus sequence (11): adenine is present in the position +4 and a purine (G) is in position −3 (where A in the ATG codon is +1). At the same time, N-terminal translational fusions with [luc+] can be conveniently created, since the *Sty*I (CCTTGG) site in front of the ATG does not introduce a termination codon.

To test the functionality of pNPR1, a module of a constitutive promoter was constructed. The 35S promoter of cauliflower mosaic virus is known to be highly active in many plant tissues (12). Consequently, it is often used as a constitutive control promoter in transient expression and stable transformation experiments. PCR-based mutagenesis was used to introduce *Sty*I (CCTTGG) recognition sequences flanking a 523-bp 3' region of the 35 S promoter. The resulting module was designated [35S]. The [35S] module is at least as active in driving luciferase expression as the nonmodular full-length 35S promoter (data not shown). [35S] was inserted in front of the luciferase gene in pNPR1, which resulted in a modular transcriptional fusion construct designated [35S]-pNPR1. In transient expression ("bombardment") experiments [35S]-pNPR1 produced high-level expression (Fig. 4*B*; compare [35S]-pNPR1 and 35S-Luc). 35S-Luc (pDO432, Ref. 9) is the 1.6-kb 35 S promoter region inserted in front of the luciferase gene in the pUC119-Luc. The observed difference in expression may be due to better performance of the "original" luciferase gene compared with the modified Luc+ gene in transient expression assay in tomato. It should be noted that pNPR1 does not itself code for any extended 5' untranslated leader. This property is intentional and allows for insertion of any desired 5' UTR in front of the luciferase gene.

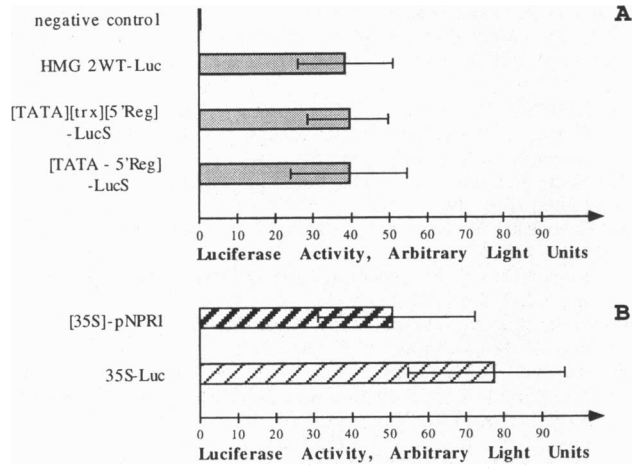

FIG. 4. Activity of modular constructs in transient expression (bombardment) experiments. Modular constructs [TATA][trx][5'Reg]-LucS and [TATA-5'Reg]-LucS show luciferase activity levels practically identical to the nonmodular construct HMG2WT-Luc (*A*). A fully modular reporter construct [35S]-pNPR1, which contains the modules [35S], [Luc+], and [nosT] in the assembly vector pNOMAD1, is highly expressed in transient expression assays (*B*). Positive control: 35S-Luc (pDO432, Ref. 10). Negative control: pUC119, which does not have a luciferase gene.

DISCUSSION

NOMAD allows for rapid creation of DNA constructs from compatible modules, providing at the molecular level the degree of flexibility in handling standardized "building blocks" that is reminiscent of Lego. The modules can be sequentially added to a construct in any desired order (Fig. 2), and the groups of modules (composite modules) can be released as a whole and recloned into another modular construct. The collection of modules will grow with the continued use of NOMAD: any module need only to be created once, and than it can be used in any construct. In effect, NOMAD has the potential to eliminate *in vitro* DNA manipulation as the rate-limiting step in many applications that require families of constructs composed of rearranged DNA fragments. Modular analysis of regulatory regions, engineering of chimeric proteins using protein-coding DNA segments, and vector construction will be facilitated by NOMAD.

Application of NOMAD to Promoter Analysis. To illustrate the applicability of NOMAD to promoter analysis we isolated as individual modules the TATA box, the transcription start region, and the 5' UTR of the tomato HMG2 gene, and then reconstructed the HMG2 promoter/5' UTR from these modules. The activity of the reconstructed promoter in the transient expression assay did not differ from the nonmodular, wild-type promoter. This result shows that individual promoter elements apparently can reside in different modules (i.e., be separated by *Sty*I recognition sites), and still interact normally.

Ideally, selection of modules for the analysis of a regulatory region should be based upon the data obtained by methods, such as DNA footprinting, that allow for at least tentative identification of the important elements. Alternatively, interesting sequence elements such as palindromes or homologies to known protein-binding elements may be isolated as modules and tested in reporter constructs. A regulatory region under investigation can be recreated from synthetic modules or from modules produced by PCR. Both of these methods were exploited in this paper. An alternative, as yet untested, approach would be to introduce the *Sty*I sites by site-directed mutagenesis. The construct can then be disassembled (broken down into component modules) by digestion with *Sty*I. Individual modules would then be cloned for subsequent use in assembly of experimental constructs. Whatever the means of their isolation, individual elements isolated as modules can be joined in any required combination and order to create altered versions of the regulatory region under study. The overall structure of a region, as well as the contribution of each individual component, can be rapidly analyzed by assessing the activity of NOMAD reporter constructs.

Application of NOMAD to Vector Construction. NOMAD was also used to construct a fully modular transient expression vector [35S]-pNPR1. A constitutive promoter, luciferase reporter gene, and transcription termination signal were converted into modules and cloned into a minimal assembly vector pNOMAD1. The resulting construct gave high-level expression in transient expression assays. Creation of [35S]-pNPR1

shows that NOMAD can be used for flexible construction of cloning vectors out of modules.

pNOMAD1 and pNOMAD2 were created as alternative foundations for creating custom modular vectors. Specifically, pNOMAD1 and pNOMAD2 allow for different treatment of structural and experimental modules. Structural modules such as a replication origin or selection marker are only required to confer the desired properties onto the vector. In contrast, experimental DNA segments such as promoter components are directly subjected to investigation. When pNOMAD1 is used as the basis for an experimental construct, both kinds of modules are cloned into the same SIS. The advantage of this arrangement is that new modules can be added both 5′ and 3′ to the structural modules. In the case of pNPR1, experimental modules can be cloned in front of the luciferase reporter gene, as well as downstream from the nopaline synthase terminator module [nosT]. However, it is not possible to release only experimental modules from pNPR1, because any composite module will also include [Luc+] and [nosT]. pNOMAD2 (Fig. 1*C*) was designed to permit separable handling of structural and experimental modules. The eSIS of pNOMAD2 allows for sequential addition of modules in much the same way as pNOMAD1 does (Fig. 2–2). The crucial feature of pNOMAD2 is that after all the structural components of the vector have been added, the eSIS is removed by simultaneous digestion with *Bsa*I and *Bsm*BI, and a module containing the SIS is inserted. From this moment on, all general purpose modules are located outside the SIS. The new vector now permits sequential addition of experimental modules, as well as release of the composite experimental module for insertion into a different vector, without any structural vector components attached. Importantly, the difference between structural and experimental modules is purely operational, and the same module can play either of these roles in different applications.

Use of NOMAD in the Assembly of Protein-Coding Regions. Implementation of NOMAD in manipulating protein-coding DNA sequences will likely depend on the sensitivity of particular experimental systems to amino acid sequence alterations. For instance, in joining domains originating in different proteins, wherein substitution of a small number of amino acids at the junctions may not be critical, a uniform DNA (and consequently amino acid) junction sequence can be used. The use of a standard domain junction allows for maximal flexibility and decreased experimental variability.

A more conservative approach to the introduction of the *Sty*I-based NOMAD module junctions in protein-coding regions would be to use endogenous two or three amino acid sequences at the domain junctions that approach most closely the sequences that can be accommodated by CCTTGG. For instance, when the *Sty*I (CCTTGG) module junction sequence is used in reading frame 3, as in pLucS and pNPR1 (NNC CTT GGN, where N is any nucleotide), 15 combinations of 3 amino acids at the junction site can be accommodated. These are (* Leu Gly), where the asterisk signifies any amino acid except Gln, Glu, Lys, Met, and Trp. Therefore, most Leu-Gly sequences are potential NOMAD junction sites; alternatively, a junction can be built around an endogenous Leu or Gly.

In situations when absolutely no alterations to the amino acid sequence are acceptable, NOMAD assembly vectors can be modified as described below to incorporate a module junction sequence required by the particular application.

NOMAD Vectors Can Be Easily Modified to Use Cohesive Ends Other than CTTG 5′ Overhang. In fact, any 4-nucleotide 5′ overhang can be used. This can be achieved by cutting an assembly vector with *Bsa*I, *Bsm*AI, or *Bsm*BI and making the cohesive ends blunt. A blunt-ended double-stranded linker can be ligated to the vector, and the sequence of this linker will determine the structure of the cohesive ends that form upon digestion of the new vector by the enzymes flanking the SIS.

Use of NOMAD Assembly Vectors for Seamless Gene Assembly. Combinatorial flexibility in manipulating DNA fragments is the ultimate goal of NOMAD. However, pNOMAD1 can also be used for seamless gene assembly. To do this, blunt-ended DNA fragments that are to form a large construct are created with a 4-bp overlap that is unique for each junction to be formed. On the next step, the fragments are individually cloned into blunted pNOMAD1. Digestion of the resulting constructs with NOMAD enzymes will result in creation of DNA fragments having unique 4-nucleotide 5′ overhangs. They can be joined to form a "seamless" DNA construct. Using this technique, a long DNA segment having a known sequence can be assembled from shorter fragments.

Adapting Existing DNA Constructs for Use with NOMAD. The optimal strategy for making a particular assay system compatible with NOMAD depends upon whether the recognition sequences for the NOMAD enzymes (primarily *Bsa*I and *Bsm*BI) are present in the corresponding vectors. If elimination of the existing restriction sites is not practical or desirable, a bipartite system can be used. In this implementation of NOMAD, there is "division of labor" between an assembly vector such as pNOMAD1, where the modules are put together in the required order, and an assay vector such as pLucS, which allows for insertion of (composite) modules and their functional testing. The advantage of such a bipartite system is that minimal modification to the assay vector is necessary.

Data base of the Available Modules and Modular Constructs. The advantages of NOMAD can be exploited most effectively if there is coordination among laboratories working in a particular field. Most of the commonly desired modules (e.g., reporter genes, enhancer elements, or antibiotic resistance genes) could be rapidly created if several groups split the effort and then made their modules available to others. To facilitate this process, a World Wide Web site describing NOMAD in detail and providing access to the authors' data base of the currently available modules ("NOMADbase") was established (http://Lmb1.bios.uic.edu/NOMAD/NOMAD.html). Researchers who use NOMAD in their projects are encouraged to share new NOMAD modules, as well as general information on the technique.

We wish to thank Drs. David Stone, Don Morrison, Lon Kaufman, and Thianda Manzara for helpful discussion and critical reading of the manuscript and Dr. Don Morrison for computer hardware.

1. Kim, S. C. & Szybalski, W. (1988) *Gene* **71,** 1–8.
2. Gayle, R. B., Auger, E. A., Gough, G. R., Gilham, P. T. & Bennett, G. N. (1987) *Gene* **54,** 221–228.
3. Stahl, S., Sjolander, A., Hansson, M., Nygren, P. A. & Uhlen, M. (1990) *Gene* **89,** 187–193.
4. Szybalski, W., Kim, S. C., Hasan, N. & Podhajska, A. J. (1991) *Gene* **100,** 13–26.
5. Kegler-Ebo, D. M., Docktor, C. M. & DiMaio, D. (1994) *Nucleic Acids Res.* **22,** 1593–1599.
6. Daraselia, N., Tarchevskaya, S. & Narita, J. O. (1996) *Plant Physiol.* **112,** in press.
7. Ausubel, F. M., Brent, R., Kingston, R., Moore, D., Seidman, J. G., Smith, J. A. & Struhl, K., eds. (1995) *Current Protocols in Molecular Biology* (Wiley, New York).
8. Ow, D. W., Wood, K. V., De Luca, M., deWet, J. R., Helinski, D. R. & Howell, S. H. (1986) *Science* **234,** 856–859.
9. Montgomery, J., Goldman, S., Deikman, J., Margossian, L. & Fischer, R. L. (1993) *Proc. Natl. Acad. Sci. USA* **90,** 5939–5943.
10. Barnes, W. M. (1994) *Proc. Natl. Acad. Sci. USA* **91,** 2216–2220.
11. Kozak, M. (1991) *J. Biol. Chem.* **266,** 19867–19870.
12. Odell, J. T., Nagy, F. & Chua, N. H. (1985) *Nature (London)* **313,** 810–812.