APPLIED AND ENVIRONMENTAL MICROBIOLOGY, Oct. 1996, p. 3557–3559 Vol. 62, No. 10
0099-2240/96/$04.00+0

The Oligonucleotide Probe Database

ELIZABETH WHEELER ALM,[1] DANIEL B. OERTHER,[1] NIELS LARSEN,[2] DAVID A. STAHL,[3] AND LUTGARDE RASKIN[1]*

Environmental Engineering and Science, Department of Civil Engineering, University of Illinois, Urbana, Illinois 61801[1]; Department of Microbiology and Center for Microbial Ecology, Michigan State University, East Lansing, Michigan 48824[2]; and Department of Civil Engineering, Technological Institute, Northwestern University, Evanston, Illinois 60208[3]

The use of oligonucleotide hybridization probes and PCR primers has become widespread in microbial ecology and environmental microbiology (for reviews, see references 3, 5, 7, 17, and 21), and descriptions of probe applications are abundant in the literature. We have encountered, however, a number of difficulties when relying on the literature for information on probes and primers: (i) probe design, characterization, and application data are scattered throughout the literature and therefore are not easily available; (ii) probe nomenclature is not standardized, making it difficult to recognize a particular probe and evaluate results obtained with that probe; (iii) probes are often designed empirically and used without thorough experimental characterization, making it difficult to interpret experimental results; and (iv) information on the application of individual probes is often not published in detail in the original probe description since the value of some data becomes apparent only as a result of observations made subsequent to publication (e.g., hybridization buffer composition, formamide concentration, membrane supplier and lot number, target group specificity). We designed the Oligonucleotide Probe Database (OPD) to address these concerns.

The OPD centralizes information related to the design and use of oligonucleotide probes and PCR primers. The database was originally designed in Microsoft Access Version 2.0 and then converted to Hypertext Transfer Markup Language with PERL scripts. The current data set contains 96 hybridization probes and PCR primers used in microbial ecology and environmental microbiology, published by the authors as well as from direct on-line submissions to OPD. The majority of the probes in the current data set target rRNA, but the database is designed to accommodate probes targeting other gene families.

For each probe or primer, the information in the OPD includes design and characterization data important for probe and primer use, including a standardized name, probe sequence, nucleotide position within the target gene, optimal hybridization and wash conditions (or annealing conditions for PCR primers), intended target group, experimentally validated target group specificity, and original citations. Much of the experimental data available in the database were not included in the original publications describing the probes.

Standardization of oligonucleotide probe nomenclature. A source of much confusion and frustration during the use of probes or PCR primers designed and characterized in different laboratories has been the absence of a standardized probe nomenclature. Stahl and Amann (18) have previously attempted to address this problem for phylogenetic probes with a nomenclature consisting of three to five letters representing phylogenetic specificity, followed by a number indicating the 5′ position on the rRNA complementary to the 3′ end of an antisense probe or PCR primer or identical to the 5′ end of a sense primer. Limitations to this nomenclature system are apparent when several versions of a probe that have the same target specificity and nucleotide position exist. We modified the nomenclature originally utilized by Stahl and Amann to include multiple probe versions and also to provide additional identifying information. We suggest a method of standardizing the nomenclature for oligonucleotide probes and PCR primers that is both unambiguous and informative. The name for an oligonucleotide probe consists of seven components combined sequentially. These components are discussed below. An example demonstrating construction of probe nomenclature for a small-subunit rRNA-targeted probe is given in parentheses.

1. An indication of the target gene; use of a single uppercase letter is encouraged. For example, “S” for small-subunit rRNA (S).
1a. A hyphen (S-). Each character or group of characters in the probe name is set off with a hyphen to facilitate computer-aided manipulations.
2. An indication of the largest taxonomic group circumscribed by the oligonucleotide probe. Probes designed to target all organisms (universal probes) or unidentified environmental isolates and other probes that do not circumscribe a coherent taxonomic group will have a wild-card symbol “*” in this position rather than a letter. Designations for phylogenetically based probes include K, kingdom; D, domain; P, phylum; O, order; Sc, subclass; F, family; G, genus; S, species; Ss, subspecies; and St, strain. For example, “D” for a domain-level probe (S-D).
2a. A hyphen (S-D-).
3. An abbreviated description, limited to three to five letters, of the specific gene cluster, gene superfamily, or taxonomic or phylogenetic group circumscribed by the oligonucleotide probe. For example, “Bact” for the domain *Bacteria* (S-D-Bact).
3a. A hyphen (S-D-Bact-).
4. A number indicating the 5′ position on the nucleic acid sense strand complementary to the 3′ end of an antisense probe or PCR primer or identical to the 5′ end of a sense PCR primer (Fig. 1). The number will contain as many digits as the full-length gene. An example for a small-subunit rRNA-targeted probe (length of gene approximately 1,540 nucleotides or 4 digits) would be “0338” (S-D-Bact-0338). For rRNA-targeted probes, we use *Escherichia coli* numbering as a reference. If the 5′ position in the rRNA is in a sequence region that does not exist in *E. coli*, i.e., an insertion, then an

* Corresponding author. Mailing address: Department of Civil Engineering, University of Illinois at Urbana-Champaign, 3221 Newmark Civil Engineering Laboratory, 205 N. Mathews Ave., Urbana, IL 61801. Phone: (217) 333-6964. Fax: (217) 333-6968. Electronic mail address: lraskin@uiuc.edu.

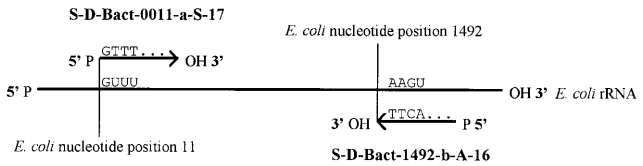

FIG. 1. Schematic representation illustrating the nucleotide position and orientation used in naming one sense and one antisense primer. Oligonucleotides targeting rRNA are used for illustrative purposes.

alternate reference organism should be indicated in parentheses following the number. Roller et al. (15, 16) determined that the gram-positive bacteria with a high DNA G+C content have an insertion of approximately 100 bases within the central part of the high-molecular-weight rRNA of the large subunit. They designed three species-specific probes to target a region within this insertion. For example, the probe for *Corynebacterium glutamicum* (16) would be renamed with our nomenclature as L-S-C.glut-1556 (*C. glutamicum*)-a-A-18 (the last three components of the name are discussed below).

4a. A hyphen (S-D-Bact-0338-).

5. A single lowercase letter indicating the version of the probe, for example, "a" for a first version, "b" for a second version that differs from version "a" only in its nucleotide composition (i.e., target group, position, orientation, and length are identical), and so forth (S-D-Bact-0338-a).

5a. A hyphen (S-D-Bact-0338-a-).

6. A single uppercase letter indicating whether the probe sequence is identical to the DNA sense (coding or nontemplate) strand (S) or to the antisense (template) strand (A) (19, 19a) (S-D-Bact-0338-a-A).

6a. A hyphen (S-D-Bact-0338-a-A-).

7. A number indicating the length of the probe in nucleotides (S-D-Bact-0338-a-A-18).

Examples of probe nomenclature for several rRNA-targeted probes in the database are provided in Table 1.

When organisms are reassigned to new taxonomic groups, the taxonomic designation for the target group of a probe may change and probe nomenclature will be updated to reflect these reassignments. To accommodate transition to new probe nomenclature, users may search the database for information on probes and primers by using either current or previously published nomenclature.

Characterization of oligonucleotide probes. A thorough characterization of oligonucleotide probes and PCR primers (3, 5, 14) and detailed recording of conditions used for characterization are important if probes and primers are to be used by multiple research groups. On the basis of our experience, we find the following experimental characterization requirements essential: (i) for membrane hybridizations, the type of membrane, compositions of hybridization and wash buffers, optimal wash temperature, incubation times, and identity of target and nontarget organisms used in characterization; (ii) for whole-cell hybridizations, method of fixation of reference organisms and samples, composition of hybridization buffer including optimal percent formamide, composition of wash buffer, identity of target and nontarget organisms used in characterization, and information on labeled competitor oligonucleotides (20); (iii) for PCR primer pairs, concentration of magnesium chloride and other components of the reaction mixture, optimal annealing temperature, and identity of target and nontarget organisms used in characterization. OPD provides this information on probes and PCR primers where possible. We encourage new submissions to adhere to these minimal requirements.

Access to OPD. OPD is accessible via the World Wide Web at http://www.cme.msu.edu/OPD. A forms- and tables-capable web browser should be used to allow full access to OPD functions. Contact with the staff of OPD can also be made at the Internet address OPD@uiuc.edu.

Cooperation with the Ribosomal Database Project. Integration of OPD and the Ribosomal Database Project (RDP [11]) provides additional benefits for users of rRNA-targeted probes. As new sequences are added to sequence databases, the specificity range for a probe may change. Therefore, links to RDP are provided from within OPD to allow the user to obtain up-to-date information on probe specificity. For exam-

TABLE 1. Examples of complete oligonucleotide probe names for several rRNA-targeted probes

Standardized probe name	Target group	Probe sequence from 5′ to 3′	Old probe name and/or reference
S-*-Univ-1392-a-A-15	Universal (all organisms)	ACGGGCGGTGTGTRC[a]	"universal" (13)
S-*-Univ-1392-b-A-15	Universal (all organisms)	ACGGGCGGTGTGTIC	22
S-*-Univ-1528-a-A-15	Universal (all organisms)	AAAGGAGGTGGTCCA	Eugen3R (10)
S-D-Arch-0915-a-A-20	Domain *Archaea*	GTGCTCCCCCGCCAATTCCT	Arch915 (2)
S-D-Bact-0011-a-S-17	Domain *Bacteria*	GTTTGATCCTGGCTCAG	8
S-Sc-aProt-0019-a-A-17	Subclass alpha of the *Proteobacteria* (and most spirochetes)	CGTTCGYTCTGAGCCAG[a]	ALF1b (12)
S-F-Dsv-0687-a-A-16	Family *Desulfovibrionaceae*	TACGGATTTCACTCCT	687 (6)
S-G-Fibr-0153-a-S-16	Genus *Fibrobacter*	CCGTGCCAACGCGCGG	Fibro (10)
S-S-B.dim-0023-a-A-18	Species *Brevundimonas diminuta* (formerly *Pseudomonas diminuta*)	TTCCACATACCTCTCTCA	Pdi23a (4)
S-Ss-F.s.suc-0628-a-A-22	Subspecies *Fibrobacter succinogenes* subsp. *succinogenes*	GATCCAGTTCGGACTGCAGAGC	Sub1 (9)
S-Ss-F.suc(2)-0628-a-A-22[b]	Subspecies *F. succinogenes* group 2	AACCCAGTTCGGACTGCAGGTC	Sub2 (9)
S-*-Dsb-0804-a-A-18[c]	*Desulfobacter* group of the delta subclass of the *Proteobacteria*	CAACGTTTACTGCGTGGA	804 (6)

[a] R, an A/G nucleotide degeneracy; Y, a C/T nucleotide degeneracy.
[b] A number is included in the abbreviated description of the taxonomic group to distinguish *F. succinogenes* group 2 from *F. succinogenes* group 3.
[c] Since this probe circumscribes several taxonomic groups (genera) within the delta subclass of the *Proteobacteria*, the wild-card symbol "*" substitutes for a letter indicating taxonomic specificity.

ple, target sequences for organisms that perfectly match a probe as well as nontarget sequences for organisms with a few mismatches can be obtained with RDP's program CHECK_PROBE, which is accessible for each probe from within OPD. The link to this program also will make it convenient to determine target organisms for individual probes in the multiple-probe approach recently proposed by Amann and coworkers (1). In this approach, probe combinations rather than individual probes define specificity, which increases flexibility in probe target group selection. Thus, the link to RDP's CHECK_PROBE will help to evaluate the specificity of a particular probe combination. In addition, those probes that do not individually circumscribe a taxonomic group, but that are designed to identify specific assemblages of organisms when used in combination with other probes, will be identified as such and will be linked to their partner probes in OPD.

Further integration of OPD and RDP will include: (i) specificity maps for all probes premade for each release of RDP and viewable through OPD; (ii) specificity maps of user-modified versions of probes in OPD generated "on the fly" by RDP; (iii) for each organism in RDP, a listing of all OPD probes that target this organism; (iv) alignment columns that include the subsequences to which a given probe specifically binds; and (v) probe target sites shown on the rRNA secondary structure. Similar to the connections between OPD and RDP, links to other relevant databases (such as PUMA [phylogeny of the unicellular organism's metabolic pathways and alignments, developed at Argonne National Laboratory] at http://www.mcs.anl.jov/home/compbio/PUMA/Production/puma_graphics.html) will be established in future releases of OPD.

Creation of the OPD was supported by grants from the National Science Foundation (BES 9410476) and the U.S. Department of Agriculture (95-37500-1911) to L.R. and by grants from the National Science Foundation, the Office of Naval Research, and the U.S. Department of Agriculture to D.A.S.

We acknowledge Lars Poulsen for original organization of temperature of dissociation data, Michael Wagner and Rudi Amann for helpful suggestions, and Carl Woese and Gary Olsen of RDP for providing computer space for development of the database.

REFERENCES

1. **Amann, R. I.** Personal communication.
2. **Amann, R. I., L. Krumholz, and D. A. Stahl.** 1990. Fluorescent-oligonucleotide probing of whole cells for determinative, phylogenetic, and environmental studies in microbiology. J. Bacteriol. **172:**762–770.
3. **Amann, R. I., W. Ludwig, and K.-H. Schleifer.** 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. Microbiol. Rev. **59:**143–169.
4. **Amann, R. I., B. Zarda, D. A. Stahl, and K.-H. Schleifer.** 1992. Identification of individual prokaryotic cells using enzyme-labeled, rRNA-targeted oligonucleotide probes. Appl. Environ. Microbiol. **58:**3007–3011.
5. **Atlas, R. M., G. Sayler, R. S. Burlage, and A. K. Bej.** 1992. Molecular approaches for environmental monitoring of microorganisms. BioTechniques **12:**706–718.
6. **Devereux, R., M. D. Kane, J. Winfrey, and D. A. Stahl.** 1993. Genus- and group-specific hybridization probes for determinative and environmental studies of sulfate-reducing bacteria. Syst. Appl. Microbiol. **15:**601–609.
7. **Holben, W. E., and J. M. Tiedje.** 1988. Applications of nucleic acid hybridization in microbial ecology. Ecology **69:**561–568.
8. **Kane, M. D., L. K. Poulsen, and D. A. Stahl.** 1993. Monitoring the enrichment and isolation of sulfate-reducing bacteria by using oligonucleotide hybridization probes designed from environmentally derived 16S rRNA sequences. Appl. Environ. Microbiol. **59:**682–686.
9. **Lin, C., B. Flesher, W. C. Capman, R. I. Amann, and D. A. Stahl.** 1994. Taxon-specific hybridization probes for fiber-digesting bacteria suggest novel gut associated *Fibrobacter*. Syst. Appl. Microbiol. **17:**418–424.
10. **Lin, C., and D. A. Stahl.** 1995. Taxon-specific probes for the cellulolytic genus *Fibrobacter* reveal abundant and novel equine-associated populations. Appl. Environ. Microbiol. **61:**1348–1351.
11. **Maidek, B. L., N. Larsen, M. J. McCaughy, R. Overbeek, G. J. Olsen, K. Fogel, J. Blandy, and C. R. Woese.** 1994. The ribosomal database project. Nucleic Acids Res. **17:**3485–3487.
12. **Manz, W., R. Amann, W. Ludwig, M. Wagner, and K.-H. Schleifer.** 1992. Phylogenetic oligodeoxynucleotide probes for the major subclasses of *Proteobacteria*: problems and solutions. Syst. Appl. Microbiol. **15:**593–600.
13. **Pace, N. R., D. A. Stahl, D. J. Lane, and G. J. Olsen.** 1986. The analysis of natural microbial populations by ribosomal RNA sequences. Adv. Microb. Ecol. **9:**1–55.
14. **Raskin, L., J. M. Stromley, B. E. Rittmann, and D. A. Stahl.** 1994. Group-specific 16S rRNA hybridization probes to describe natural communities of methanogens. Appl. Environ. Microbiol. **60:**1232–1240.
15. **Roller, C., W. Ludwig, and K.-H. Schleifer.** 1992. Gram-positive bacteria with a high DNA G+C content are characterized by a common insertion within their 23S rRNA genes. J. Gen. Microbiol. **138:**1167–1175.
16. **Roller, C., M. Wagner, R. Amann, W. Ludwig, and K.-H. Schleifer.** 1994. *In situ* probing of Gram-positive bacteria with high DNA G+C content using 23S rRNA-targeted oligonucleotides. Microbiology **140:**2849–2858.
17. **Stahl, D. A.** 1988. Phylogenetically based studies of microbial ecosystem perturbation, p. 373–390. *In* P. A. Hedin, J. J. Menn, and R. M. Hollingworth (ed.), Biotechnology for crop protection. American Chemical Society, Washington, D.C.
18. **Stahl, D. A., and R. Amann.** 1991. Development and application of nucleic acid probes, p. 205–248. *In* E. Stackebrandt and M. Goodfellow (ed.), Nucleic acid techniques in bacterial systematics. John Wiley & Sons Ltd., Chichester, United Kingdom.
19. **Stryer, L.** 1995. Biochemistry, 4th ed., p. 843. W. H. Freeman and Company, New York.

19a. **Stryer, L.** 1996. Personal communication.

20. **Wagner, M., G. Rath, R. Amann, H.-P. Koops, and K.-H. Schleifer.** 1995. In situ identification of ammonia-oxidizing bacteria. Syst. Appl. Microbiol. **18:**251–264.
21. **Ward, D. M., M. M. Bateson, R. Weller, and A. L. Ruff-Roberts.** 1992. Ribosomal RNA analysis of microorganisms as they occur in nature. Adv. Microb. Ecol. **12:**219–286.
22. **Zheng, D., E. W. Alm, D. A. Stahl, and L. Raskin.** Unpublished data.