Molecular Recognition on Demand

C. Sherrill, D. Marshall, C. Richmond, C. Scherrer and J.R. Prudent

*Eragen Biosciences, Inc.
Madison, Wisconsin, USA jprudent@eragen.com

ABSTRACT

DNA is the only chemistry that allows for "molecular recognition on demand". That is, unlike any other molecular recognition chemistry, DNA allows for the simple design and rapid synthesis of molecule sets that will recognize each other and self assemble into nanostructures. Expanding DNA chemistry to include additional base pairs, would allow for a more precise manipulation of nanostructures constructed with DNA. AEGIS (An expanded genetic information system) is that additional chemistry. Made up of four additional base pairs, AEGIS is used to self assemble more precise nanostructures in the commercially available branched DNA detection assay system[1]. Here the hydrogen bonding patterns that make AEGIS base pairs unique from natural DNA, allow for a simple means of transcending problems otherwise unsolvable.

Keywords: nucleic acids, molecular recognition, expanded genetic alphabet

1 OVERVIEW

In natural DNA, two complementary strands are joined by a sequence of Watson-Crick base pairs[2, 3]. These obey two rules of complementarity: size (large purines pair with smaller pyrimidines) and hydrogen bonding (hydrogen bond donors from one nucleobase pair with hydrogen bond acceptors from the other). The former is necessary to permit the structure that underlies enzyme recognition. The latter achieves the specificity that gives rise to the simple rules for base pairing ("A pairs with T, G pairs with C") that underlie genetics and molecular biology. No other class of natural products has reactivity that obeys such simple rules. Nor is it obvious how one designs a class of chemical substances that does so much so simply.

Some time ago, Benner and his group noticed that the DNA alphabet need not be limited to the four standard nucleotides known in nature [4, 5]. Rather, twelve nucleobases forming six base pairs joined by mutually exclusive hydrogen bonding patterns might be possible within the geometry of the Watson-Crick base pair (Figure 1). Chemical synthesis for many of these bases have been developed. Using these additional base pairs, we are designing and constructing new diagnostic tools. In particular, our molecular recognition coding systems (ERA-CODE™) allow for rapid and simple deconvolution of multiplexed reactions never before possible.

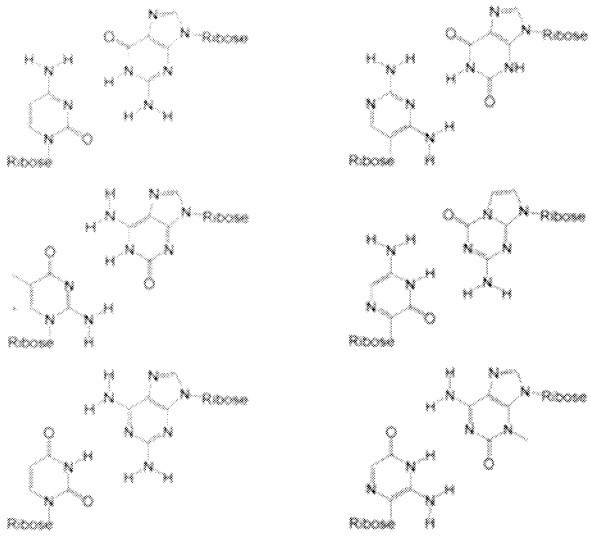

Figure 1. The twelve bases that make up AEGIS shown with their respective Watson-Crick base Pairs.

2 "ORTHOGONAL" BASE PAIRING

AEGIS provides a molecular recognition system that is "orthogonal" to that provided by natural DNA. The melting temperature data presented in Table 1 exemplify this point. Two conclusions are apparent from this data. First, as is well known in the literature, the Watson-Crick base pairing rules for natural DNA are not absolute. The matches and mismatches contribute or detract differently depending on their neighbors in the DNA duplex. [6]. Despite this, DNA performs its genetic functions remarkably well. This data illustrates how much quantitative deviation from the qualitative Watson-Crick rules a molecular recognition system can have and still be useful. An artificial system that meets the same quantitative standards as natures DNA is likely to be as useful. The second point is that AEGIS meets these quantitative standards. Table 1 is constructed so that the melting temperature values for duplexes obeying the extended Watson-Crick base pairing rules are found along the diagonal (values are underlined). In each case, the underlined value is higher than any of the melting temperature values associated with the mismatches. Especially important is the observation

that the melting temperature (Tm) values for mismatches with standard nucleobases are lower than the melting temperatures for mismatches with the correct AEGIS component. Thus, iC:iG and K:X form strong base pairs that are comparable in stability to the natural G:C and A:T base pairs respectively.

Thermal Melting Analysis – T_m in °C

	G	C	A	T	iG	iC	X
C	58	46	49	44	54	**44**	46
G	47	58	48	46	52	**44**	47
T	49	44	56	45	53	**41**	50
A	47	45	43	53	46	**43**	50
iC	45	44	42	41	62	**44**	43
iG	51	53	48	51	48	58	50
K	44	47	51	50	46	44	55

Table 1. The affinity between AEGIS components measured by solution phase melting. Melting temperatures of (5'-CACPACTTTCTCCT-3' and 5'-GGAGAAAGTQGTGT-3') oligonucleotide duplexes that contain AEGIS components. P = column, Q = row.

These observations form the basis for the assertion that AEGIS components interact orthogonally with AEGIS components and standard nucleobases. That is, a molecule that contains several AEGIS components will not bind to any natural DNA. This, in turn, means that AEGIS permits a non-standard molecular recognition system to function in the same test tube (or on the same chip, or within the same diagnostics kit) as one where natural DNA is present, without "cross-talk" or interference.

Additionally, nearest neighbor parameters can be generated for AEGIS. Attempts at this have led us to better understand how the additional bases will perform when they are incorporated into synthetic strands of DNA. Nearest neighbor parameters allow us to determine the thermodynamic parameters of each AEGIS base pair and mismatch under various conditions and at various concentrations.

The additional informational content that AEGIS builds into DNA allows for greater diversity when compared to their all natural counterparts. For example, if the four naturally occurring bases are used to make a library of six-mers, there are only 4^6 or 4096 possibilities. In contrast, a six or eight letter AEGIS system generates 46,656 and 262,144 different six-mers respectively.

The added bases also allow for construction of shorter sequences that assemble at lower temperatures. To build a library of 262,144 different sequences using just four bases, a library of 9-mers (always 50% greater in length) would need to be constructed versus the six-mer library discussed above. Yet within the libraries, sequences will assemble with members that may not be their direct complements. Therefore to build specificity, assembly typically takes place at temperatures near the Tm's. For example, the average melting temperature for a four-base library of 12-mers is around 36° degrees. This can be compared to an eight-base library of 8-mers with identical diversity, having an average melting temperature of 24° (~room temperature). In practice, specific assembly could occur at conditions much more amenable to nanostructure formation.

What may be to most important aspect to shorter sequences is the fact that shorter sequences tend to have less secondary and tertiary structure. Since single stranded nucleic acid sequences can fold into structures that can not be modeled, assembly to complementary sequences can be problematic [7]. In the simplest of cases, intra-molecular self-assembly forms due to hairpin or stem-loop formation. In these cases, assembly to complementary sequences is blocked. Occasionally, raising assembly temperature can help unblock the effect. Yet because they are intramolecular, the effective concentration is that of water (55 molar) and melting temperatures can be high. With shorter sequences, structure formation is minimized.

3 CODING

For capture and deconvolution of complex mixtures of nucleic acids, we developed a preliminary set of universal hybridization sequences (ERA-CODES™) incorporating the AEGIS™ nucleotides isoC and isoG. A computer algorithm called Era-Coder was implemented to design ERA-CODE™ sequences. The Era-Coder algorithm takes the following into consideration for ERA-CODE™ design:

1. Melting temperature (nearest neighbor methodology) near 22° C, room temperature.
2. Incorporation and distribution of AEGIS™ nucleotides in each ERA-CODE™.
3. Elimination of predicted cross hybridization between ERA-CODES™.
4. Elimination of predicted cross hybridization with natural DNA.
5. Elimination of predicted hairpins or bulges.

An initial set of ERA-CODE™ sequences, each 9-10 nucleotides in length and containing 2-3 non-standard bases was tested empirically for cross hybridization between ERA-CODES™ using the Luminex100 LabMAP system. With Luminex LabMAP technology, molecular reactions take place on the surface of color-coded polystyrene microspheres [8, 9]. A number of reports have shown that this platform can be used for multiplexed nucleic acid tests including genotyping [10-12], DNA quantitation [13], viral nucleic acid detection [14] and expression analysis [15]. Currently 100 unique bead types are commercially available from Luminex. Therefore, 100 bead types (assays) can be tested simultaneously and decoded using flow cytometry in a single reaction tube. Using this LabMap technology, we created a winnowing system in which fluorescently labeled complements of each ERA-CODE™ are tested against bead arrays containing sets of ERA-CODE™ sequences. This allows for rapid testing of large sets of coding sequences in relatively short periods of time. Fluorescent reporter signal associated with any non-complementary ERA-CODE™ are eliminated from the sets. The identity of ERA-CODE™ sequences that

specifically hybridize at room temperature without cross reactivity to other beads or natural DNA allows for the production of a liquid chip (Figure 2). So now, even in the presence of highly concentrated natural DNA sequences, the liquid chip performs specific molecular recognition at room temperature without washing in real-time. The hybridization can be reversed by raising the temperature by 10° to 15° C.

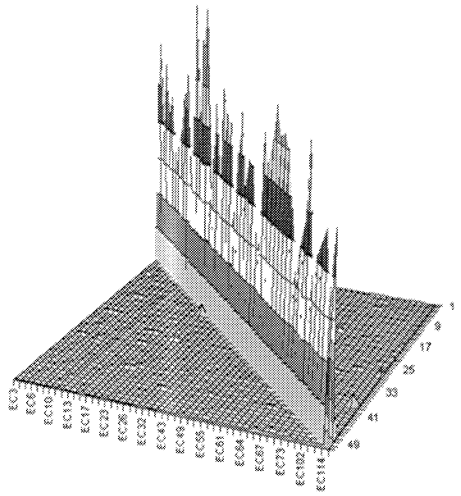

Figure 2. ERA-CODES™. Specific hybridization is shown on the diagonal and any non-specific hybridization is seen off the diagonal. X axis, ERA-CODE bead array; Y axis, ERA-CODE complement; Z axis, fluorescent signal.

4 CONCLUSION

DNA has found its way into nanotechnology because it can form molecular switches, assemble into 3-dimensional memory elements, create high density molecular recognition bio-chips and assemble into defined structures. Because of the simple rules used to create intricate DNA architectures and the tools used to screen and manipulate more complex DNA nanostructures, DNA is an obvious addition to nanotechnology. The expansion of the four letter code broadens the scope that DNA will have in the field. We have greatly simplified the use of the expanded genetic alphabet and the chemistry used to incorporate isoC and isoG into DNA. Compared to earlier reports, our chemistry greatly increases the stability of these bases. We have found no problems with deamination [16]. Our efforts to make this chemistry available and affordable should facilitate the expansion of DNA's use in nanotechnology.

REFERENCES

1. Collins, M.L., et al., *A branched DNA signal amplification assay for quantification of nucleic acid targets below 100 molecules/ml.* Nucleic Acids Res, 1997. **25**(15): p. 2979-2984.
2. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids. A structure for deoxyribose nucleic acid. 1953.* Ann N Y Acad Sci, 1995. **758**: p. 13-4.
3. Watson, J.D. and F.H. Crick, *Genetical implications of the structure of deoxyribonucleic acid. 1953.* Jama, 1993. **269**(15): p. 1967-9.
4. Piccirilli, J.A., et al., *Enzymatic incorporation of a new base pair into DNA and RNA extends the genetic alphabet.* Nature, 1990. **343**(6253): p. 33-37.
5. Switzer, C.Y., S.E. Moroney, and S.A. Benner, *Enzymatic incorporation of a new base pair into DNA and RNA.* J AM Chem Soc, 1989. **111**: p. 8322-8323.
6. SantaLucia, J., Jr., H.T. Allawi, and P.A. Seneviratne, *Improved nearest-neighbor parameters for predicting DNA duplex stability.* Biochemistry, 1996. **35**(11): p. 3555-62.
7. Hermann, T. and D.J. Patel, *Adaptive recognition by nucleic acid aptamers.* Science, 2000. **287**(5454): p. 820-5.
8. Keij, J.F. and J.A. Steinkamp, *Flow cytometric characterization and classification of multiple dual-color fluorescent microspheres using fluorescence lifetime.* Cytometry, 1998. **33**(3): p. 318-23.
9. Kettman, J.R., et al., *Classification and properties of 64 multiplexed microsphere sets.* Cytometry, 1998. **33**(2): p. 234-43.
10. Cai, H., et al., *Flow cytometry-based minisequencing: a new platform for high-throughput single-nucleotide polymorphism scoring.* Genomics, 2000. **66**(2): p. 135-43.
11. Chen, J., et al., *A microsphere-based assay for multiplexed single nucleotide polymorphism analysis using single base chain extension.* Genome Res, 2000. **10**(4): p. 549-57.
12. Iannone, M.A., et al., *Multiplexed single nucleotide polymorphism genotyping by oligonucleotide ligation and flow cytometry.* Cytometry, 2000. **39**(2): p. 131-40.
13. Spiro, A., M. Lowe, and D. Brown, *A bead-based method for multiplexed identification and quantitation of DNA sequences using flow cytometry.* Appl Environ Microbiol, 2000. **66**(10): p. 4258-65.
14. Smith, P.L., et al., *A rapid, sensitive, multiplexed assay for detection of viral nucleic acids using the FlowMetrix system.* Clin Chem, 1998. **44**(9): p. 2054-6.
15. Yang, L., D.K. Tran, and X. Wang, *BADGE, BeadsArray for the Detection of Gene Expression, a High- Throughput Diagnostic Bioassay.* Genome Res, 2001. **11**(11): p. 1888-98.
16. Switzer, C.Y., S.E. Moroney, and S.A. Benner, *Enzymatic recognition of the base pair between isocytidine and isoguanosine.* Biochemistry, 1993. **32**(39): p. 10489-10496.