7414 *Biochemistry* **2010,** *49,* 7414–7416
DOI: 10.1021/bi101123g

BIOCHEMISTRY
including biophysical chemistry & molecular biology
Rapid Report

A Mutate-and-Map Strategy for Inferring Base Pairs in Structured Nucleic Acids: Proof of Concept on a DNA/RNA Helix

Wipapat Kladwang[‡] and Rhiju Das*[,‡,§]

[‡]*Department of Biochemistry and* [§]*Department of Physics, Stanford University, Stanford, California 94035*

Received July 14, 2010; Revised Manuscript Received July 30, 2010

ABSTRACT: We propose a rapid chemical strategy for identifying base pairs in structured nucleic acid systems. The approach goes beyond traditional chemical mapping approaches by monitoring perturbations of each residue's chemical accessibility in response to systematic mutagenesis of residues that are distant in sequence but nearby in three dimensions. As a proof of concept, we present high-throughput dimethyl sulfate accessibility data for a chimeric DNA/RNA system in which every possible sequence variation and deletion in a 20 bp region has been synthesized and tested. The data demonstrate that 88% of the system's base pairs can be robustly inferred, with A/A and T/C DNA/RNA mismatches giving the strongest signals. These results point to the feasibility of rapid base pair inference in larger and more complex nucleic acid systems with unknown structure.

Numerous fundamental processes in the cell involve nucleic acid−nucleic acid interactions, from telomere extension (*1*) to message splicing (*2*) to chromatin silencing (*3*, *4*). Despite remarkable advances in spectroscopy, crystallography, and phylogenetic covariation analysis, structurally characterizing the base pairs formed and broken in these assemblies remains a major challenge (*5*, *6*). Here, we propose that these structural data may be rapidly determined by an information-rich extension of a classic chemical approach.

Chemical accessibility (or "structure mapping" or "footprinting") profiles, read out through gel electrophoresis measurements, have been determined since the advent of nucleic acid sequencing (*7*). This class of approaches is routinely used to characterize folding of RNAs (*8*−*10*), ribonucleoprotein assemblies (*11*−*13*), and DNA−protein complexes (*14*), both in vitro and in vivo (*15*, *16*). The adoption of capillary electrophoresis sequencers and fluorescent primers rather than traditional slab gels and radiolabeled primers (*17*, *18*) continues to accelerate these techniques, even enabling the investigation of entire viruses (*18*, *19*).

In chemical accessibility measurements, RNA or DNA residues that form contacts are typically protected from modification while those that are unstructured or form certain noncanonical interactions are accessible to the modifiers. Such data do not directly give nucleic acid secondary or tertiary structure but can guide modeling algorithms (*20*, *21*). While these combined chemical−computational methods are being used widely, they are still not fully reliable (*20*, *21*), particularly for nucleic acid systems with non-Watson−Crick base interactions, with complex topologies such as pseudoknots, or with protections due to protein binding. Accurate and confident modeling of these systems (Figure 1a) would be facilitated by experimental "two-dimensional" information on which bases are partnered to each other, rather than the "one-dimensional" list of protected bases available from current chemical mapping approaches.

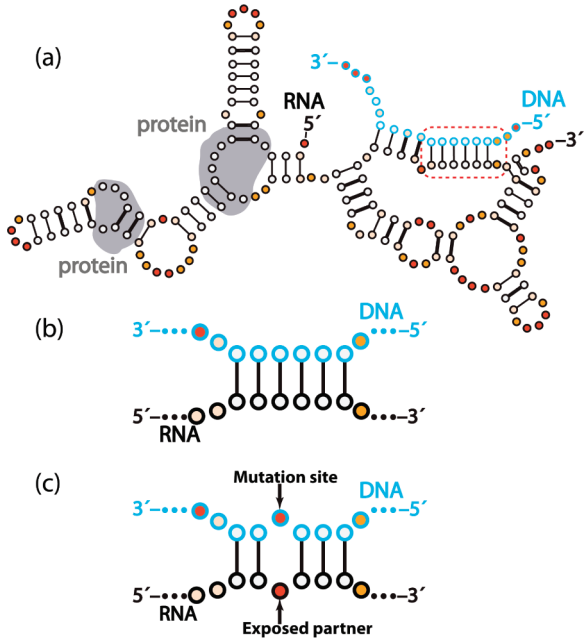

FIGURE 1: Mutate-and-map strategy. (a) Schematic of a complex of RNA (black circles), DNA (cyan circles), and proteins (gray shapes). The region boxed with a red dashed line is magnified in panel b. Chemical accessibility measurements at single-nucleotide resolution report on whether nucleic acid residues are exposed (orange or red circles) or protected (white circles) by contacts. (c) A mutate-and-map approach reveals interaction partners by monitoring whether the chemical accessibility at each residue is enhanced by systematic mutagenesis of other residues.

We reason that pairs of interacting nucleotides might be inferred by coupling mutagenesis to chemical accessibility measurements. Suppose we mutate the nucleic acid at each residue. Perturbations at certain sites may release interacting partners that are distant in sequence, leading to detectable changes in their modification by chemical reagents (Figure 1b,c). Not every sequence change will be informative: some mutations will be too conservative, permitting base pairing partners to remain structured, and some might induce drastic effects such as fully unfolding whole helices or larger domains. However, if even a subset of mutations yields precise information about interaction partners, this approach offers the prospect of determining the base pairing pattern rapidly and systematically.

To the best of our knowledge, this mutate-and-map approach has been applied only to the limited verification of individual

*To whom correspondence should be addressed. E-mail: rhiju@stanford.edu. Phone: (650) 723-5976. Fax: (650) 723-6783.

 Published on Web 08/02/2010

hypothesized interactions, e.g., in discoveries of the A302/−3u contact in the *Tetrahymena* group I intron catalytic core (*22*) and of a P7.1/9.1 helix in the bI3 group I intron (*23*). A full proof of concept, probing multiple base pairs in a nucleic acid system, has not been carried out. We are therefore conducting mutate-and-map measurements on RNA, DNA, and ribonucleoprotein molecules of known structure to systematically benchmark the approach. We report herein results of a first complete survey, using a well-defined 95-nucleotide nucleic acid complex.

We have designed a model system (Figure 2a) consisting of a 60-nucleotide RNA (called X-20) forming 20 bp with a 35-nucleotide DNA strand (called H-20). A DNA/RNA system was chosen for this proof of concept because of the low cost of chemical synthesis of multiple DNA sequences and the ease of residue-resolution chemical accessibility measurements on RNA.

The accessibility of A and C residues in the X-20 RNA system to alkylation by dimethyl sulfate (DMS) was measured at 0.06 μM RNA, in 50 mM Na-MOPS (pH 7.0) at 24 °C. In the presence of a 10-fold excess of H-20 DNA, the RNA showed a clear protection in the DNA/RNA base-paired region, as expected (in Figure 2c, compare WT to no H-20).

We then prepared each single mutant and single deletion of the H-20 DNA (80 variants total) and assayed changes in the chemical accessibility of each residue in the X-20 RNA (Figure 2b). None of the mutations or deletions that we tested resulted in a large conformational rearrangement of the H-20/X-20 DNA/RNA system at 24 °C [strand dissociation was observed at higher temperatures, 44 and 50 °C (Figure 1 of the Supporting Information)]. When displayed in sequential order, the raw data reveal diagonal perturbations corresponding to the exposure of each base-paired RNA residue by the mutation of its DNA base partner (Figure 2c).

Band quantification confirms that two mutations appear particularly informative about C/G and A/T base pairs. Mutating G9 to T in the DNA strand leaves a T/C mismatch (Figure 2b). The chemical accessibility of the C12 partner in the RNA strand rises to more than half the level seen in unpaired C residues (Figure 3a and Table 1 of the Supporting Information). A T10 to A mutation in the DNA gives a similarly increased exposure of its RNA partner (A11) (Figure 3b and Table 1 of the Supporting Information).

Single mutations that leave A/C, C/A, G/A, and other mismatches lead to weaker effects (Figure 2 of the Supporting Information; quantitative statistics are given in Table 1 of the Supporting Information), as does lowering the temperature to 0 °C (Figure 1 of the Supporting Information), suggesting that the perturbed bases remain stacked, perhaps forming noncanonical pairs, within the DNA/RNA double helix. Mutations that leave single deletions within the double helix also give weak effects, and these are straightforward to rationalize. Removing any one of a string of identical deoxynucleotides (e.g., T10, T11, T12, T13, and T14) results in the same variant DNA sequence. In response, the RNA can bulge out any of the string of A's across from the affected site, resulting in a 5-fold weaker exposure at any given A (Figure 2 of the Supporting Information).

We tested whether the entire data set could be used for robust inference of the system's base pairs by calculating Z scores, the number of standard deviations of each X-20 site's chemical accessibility relative to mean accessibility at that site across all H-20 variants. A Z score threshold of 2.5 gave 15 of the expected RNA/DNA base pairs and no incorrect base pairs at a resolution of one to two nucleotides (Figure 3c). The method thus recapitulates 88% of the 17 bp that have an A or C in the X-20 RNA strand. The two missed base pairs (C19 and C20) are at one end of the duplex; we hypothesize that their signals are weakened by fraying.

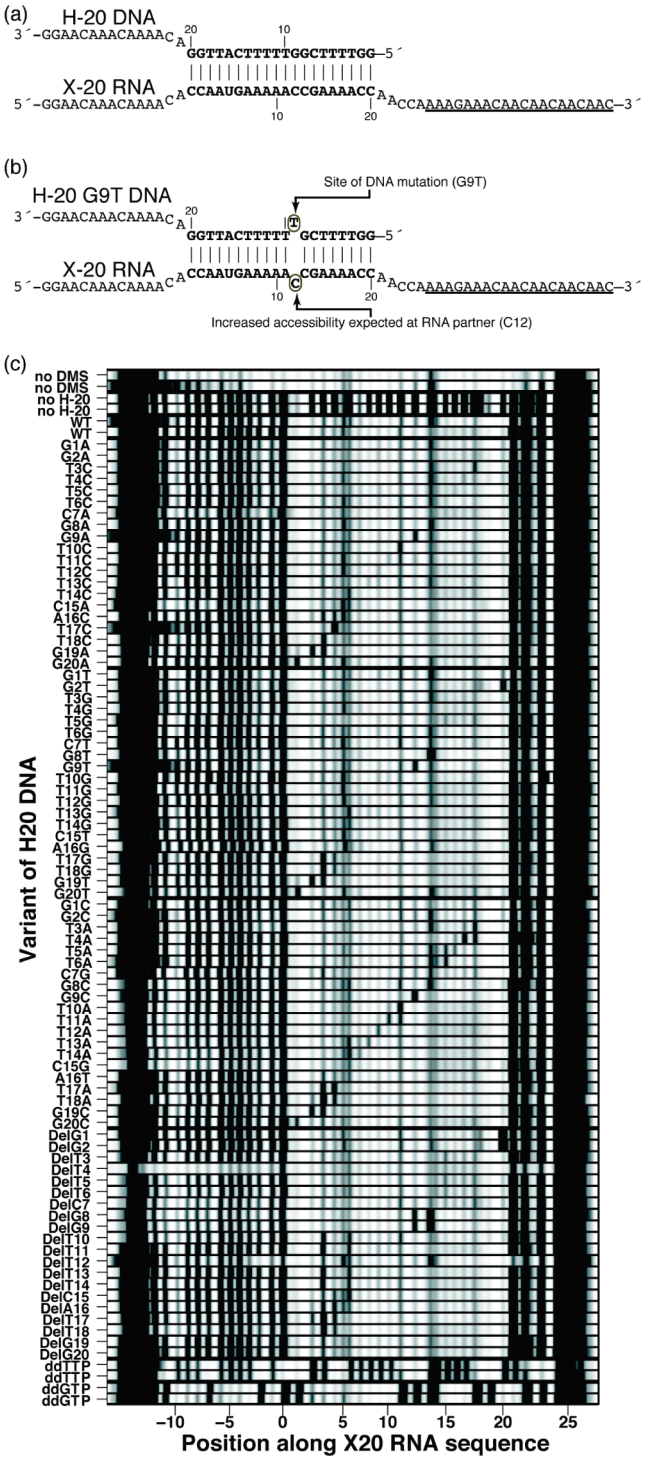

FIGURE 2: Model system for the mutate-and-map experiment. (a) Chimeric duplex of H-20 DNA and X-20 RNA. (b) Predicted effect of the G9T mutation in the DNA. (c) Chemical accessibilities of X-20 read out by high-throughput reverse transcription with 5′-fluorescently labeled radiolabeled primers and capillary electrophoresis. Raw fluorescence data (arbitrary units) are shown after automated alignment of traces and normalization to mean intensity. Shorter products (higher electrophoretic mobility) appear on the right.

This study establishes a proof of concept for identifying nucleic acid base pairs via a two-dimensional extension of a classic

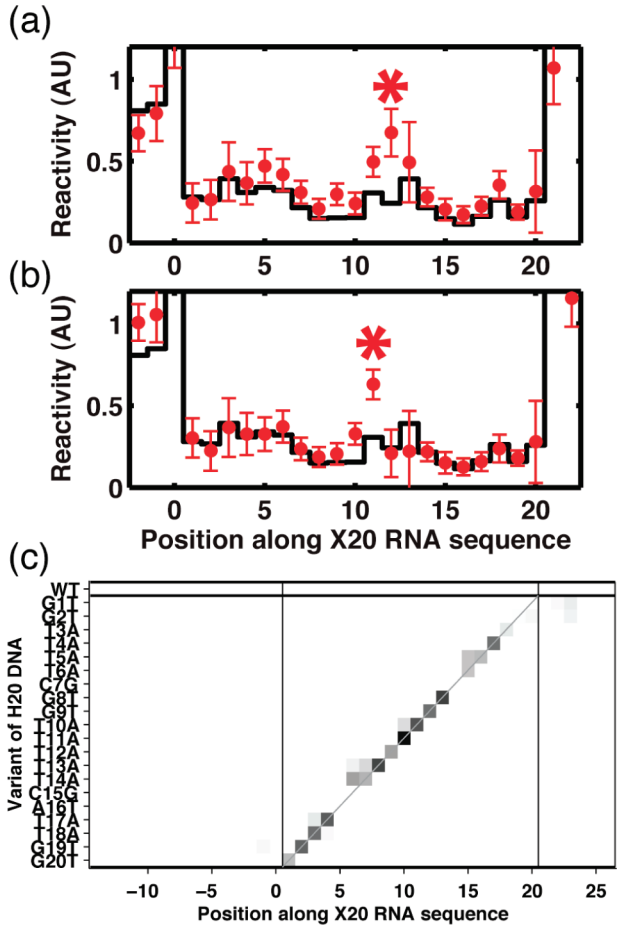

FIGURE 3: Inferring long-range base pairs from differences in chemical accessibility between mutated and nonmutated duplexes. Quantified band intensities for the X-20 RNA (red) upon mutations of the H-20 DNA G9T (a) and T10A (b). Error bars give standard deviations per site across all measured variants. Reference data are shown for RNA protected by the fully complementary H-20 DNA (black). (c) *Z* scores of band intensities delineate base pairs between the RNA and DNA. Scores between 1.5 (white) and 4.5 (black) are shown, averaged between two measurements (see also Figure 2 of the Supporting Information). RNA bases paired to mutated DNA sites are denoted with asterisks (a and b) and a diagonal line (c).

nucleotide-resolution chemical approach. The technique provides direct experimental support for each base pair and does not rely on phylogenetic analysis or models of base pairing energetics, which may be unavailable or inaccurate. The method is rapid and affordable; obtaining the data takes approximately 1 week, and the material costs are similar to current costs of capillary sequencing. The method is expected to be useful for characterizing DNA/RNA structures that can be toggled by changing solution conditions or partner molecules, e.g., telomerase (*1*) and deoxyribozymes (*24*), in advance of more arduous high-resolution crystallographic or spectroscopic approaches. In future extensions, modification chemistries besides DMS alkylation, including the facile SHAPE method (*25*), may provide signals for contacts between chemical moieties besides the Watson−Crick edges of A and C. Mutagenesis and mapping of riboswitches, viral RNAs, and long noncoding RNAs are currently under investigation. The high-throughput RNA preparation and purification methods required for these extensions appear feasible and inexpensive.

ACKNOWLEDGMENT

We thank members of the Das laboratory and D. Herschlag for comments on the manuscript.

SUPPORTING INFORMATION AVAILABLE

Additional methods, statistical analysis of mismatch signals, temperature variation studies, and comparison of replicate data sets. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Collins, K. (2006) *Nat. Rev. Mol. Cell Biol. 7*, 484–494.
2. Staley, J. P., and Guthrie, C. (1998) *Cell 92*, 315–326.
3. Panning, B., Dausman, J., and Jaenisch, R. (1997) *Cell 90*, 907–916.
4. Moazed, D. (2009) *Nature 457*, 413–420.
5. Cruz, J. A., and Westhof, E. (2009) *Cell 136*, 604–609.
6. Noller, H. F. (2005) *Science 309*, 1508–1514.
7. Peattie, D. A., and Gilbert, W. (1980) *Proc. Natl. Acad. Sci. U.S.A. 77*, 4679–4682.
8. Latham, J. A., and Cech, T. R. (1989) *Science 245*, 276–282.
9. Pollard, K. S., Salama, S. R., Lambert, N., Lambot, M. A., Coppens, S., Pedersen, J. S., Katzman, S., King, B., Onodera, C., Siepel, A., Kern, A. D., Dehay, C., Igel, H., Ares, M., Jr., Vanderhaeghen, P., and Haussler, D. (2006) *Nature 443*, 167–172.
10. Das, R., Karanicolas, J., and Baker, D. (2010) *Nat. Methods 7*, 291–294.
11. Krol, A., and Carbon, P. (1989) *Methods Enzymol. 180*, 212–227.
12. Tijerina, P., Mohr, S., and Russell, R. (2007) *Nat. Protoc. 2*, 2608–2623.
13. Garcia, I., and Weeks, K. M. (2004) *Biochemistry 43*, 15179–15186.
14. Tullius, T. D. (1991) *Free Radical Res. Commun. 12−13* (Part 2), 521–529.
15. Adilakshmi, T., Lease, R. A., and Woodson, S. A. (2006) *Nucleic Acids Res. 34*, e64.
16. Schroeder, R., Grossberger, R., Pichler, A., and Waldsich, C. (2002) *Curr. Opin. Struct. Biol. 12*, 296–300.
17. Mitra, S., Shcherbakova, I. V., Altman, R. B., Brenowitz, M., and Laederach, A. (2008) *Nucleic Acids Res. 36*, e63.
18. Wilkinson, K. A., Gorelick, R. J., Vasa, S. M., Guex, N., Rein, A., Mathews, D. H., Giddings, M. C., and Weeks, K. M. (2008) *PLoS Biol. 6*, e96.
19. Watts, J. M., Dang, K. K., Gorelick, R. J., Leonard, C. W., Bess, J. W., Jr., Swanstrom, R., Burch, C. L., and Weeks, K. M. (2009) *Nature 460*, 711–716.
20. Mathews, D. H., Disney, M. D., Childs, J. L., Schroeder, S. J., Zuker, M., and Turner, D. H. (2004) *Proc. Natl. Acad. Sci. U.S.A. 101*, 7287–7292.
21. Deigan, K. E., Li, T. W., Mathews, D. H., and Weeks, K. M. (2009) *Proc. Natl. Acad. Sci. U.S.A. 106*, 97–102.
22. Pyle, A. M., Murphy, F. L., and Cech, T. R. (1992) *Nature 358*, 123–128.
23. Duncan, C. D., and Weeks, K. M. (2008) *Biochemistry 47*, 8504–8513.
24. Baum, D. A., and Silverman, S. K. (2008) *Cell. Mol. Life Sci. 65*, 2156–2174.
25. Wilkinson, K. A., Merino, E. J., and Weeks, K. M. (2006) *Nat. Protoc. 1*, 1610–1616.