Supporting Information for:

ShapeFinder: A software system for high-throughput quantitative analysis of nucleic acid reactivity information resolved by capillary electrophoresis

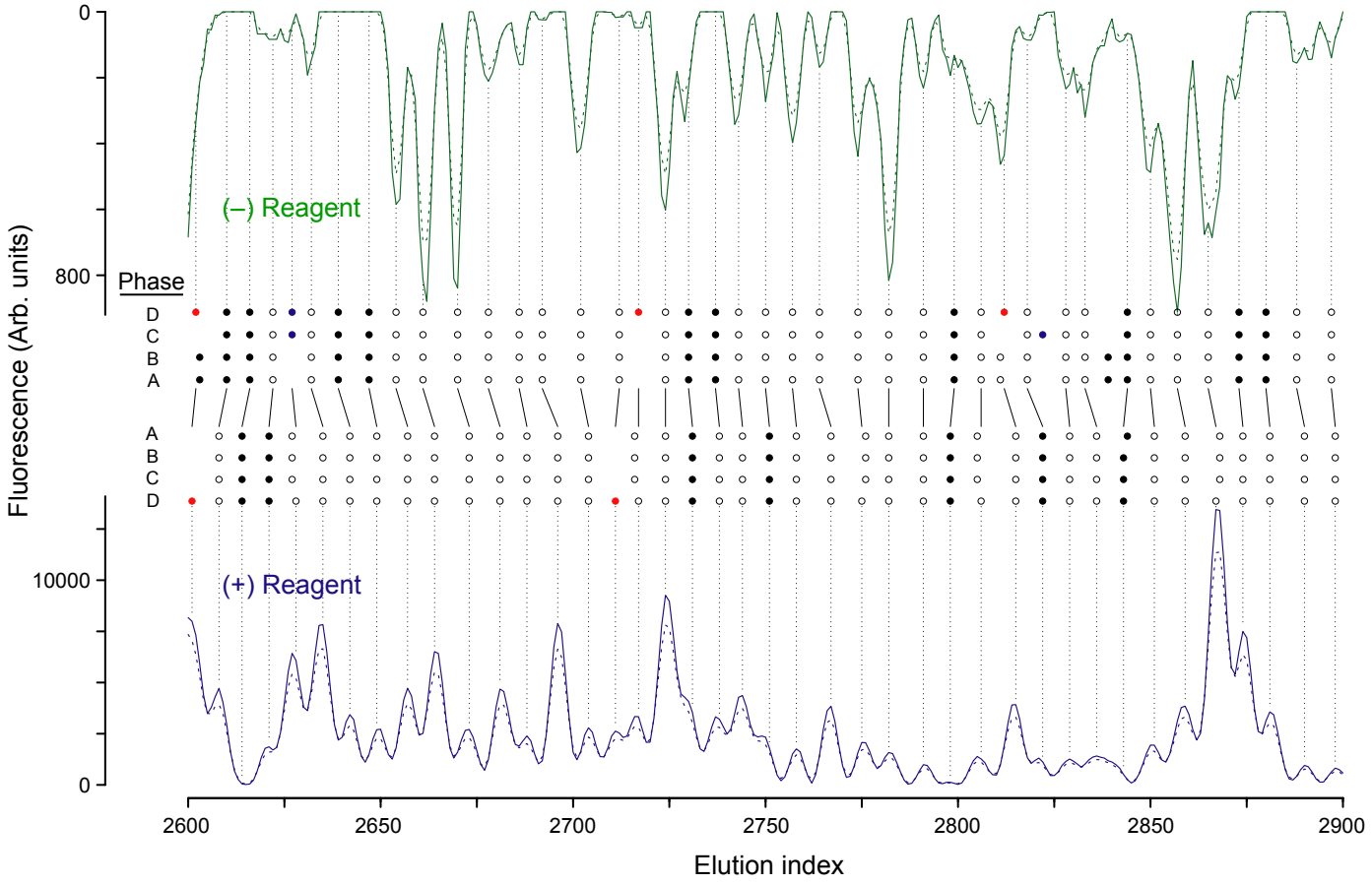
Suzy M. Vasa[1], Nicolas Guex[1], Kevin A. Wilkinson, Kevin M. Weeks*, Morgan C. Giddings*

[1] These authors contributed equally.

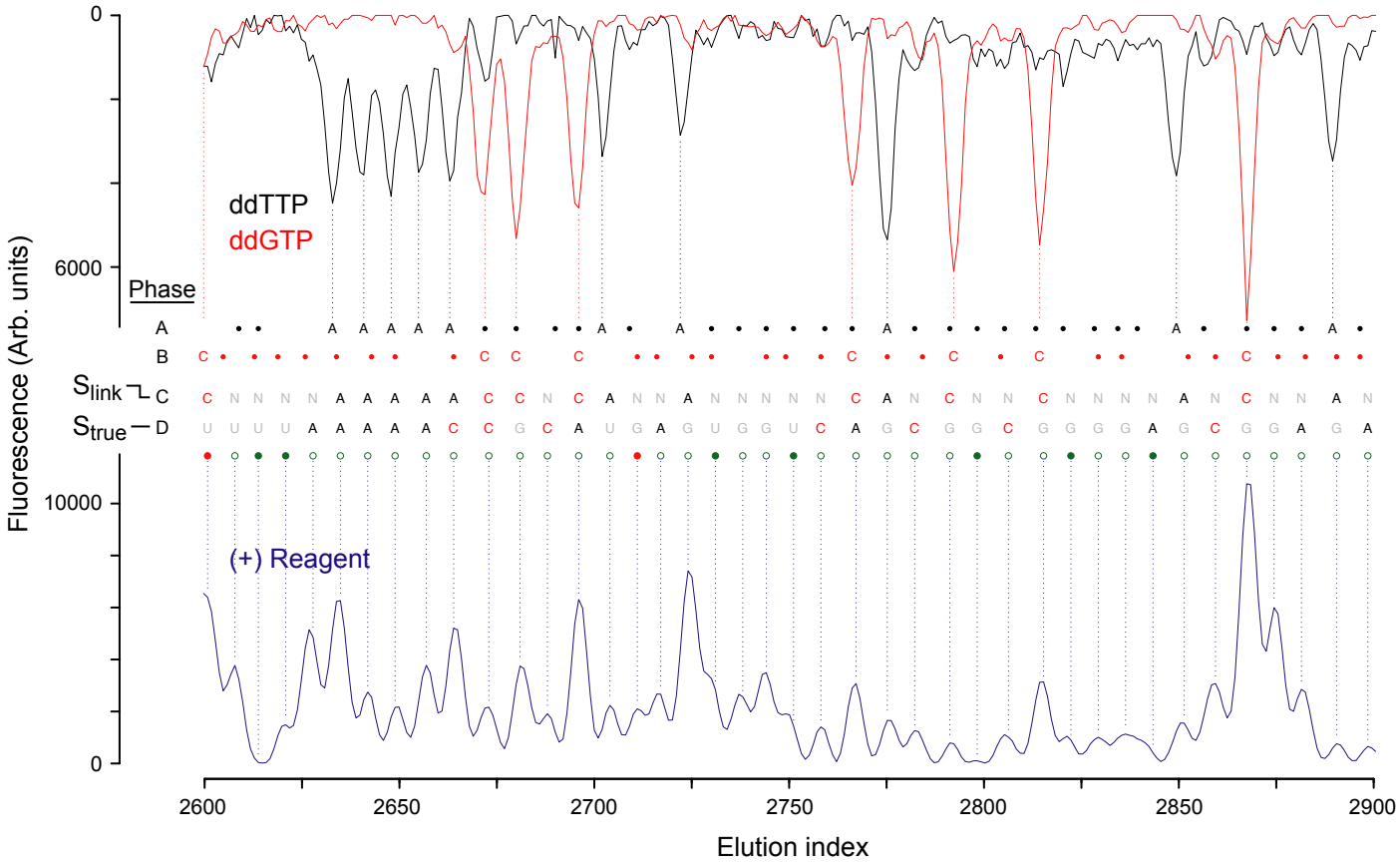
* Correspondence: weeks@unc.edu and giddings@unc.edu

Figure S1. Peak finding. The electropherogram shows the (+) and (–) reagent channels (blue and green, respectively). The (–) reagent intensities have been inverted and are plotted on an expanded scale to facilitate visualization of peak synchronization. Preprocessed channels are shown as solid lines, channels smoothed over a 3-point window are dashed. (A) Identification of peak positions by analysis of (*i*) signal amplitude versus (*ii*) interpolation are illustrated by open and closed circles, respectively. (B) Refinement of interpolated peak positions. (C) Automatic addition of missing peaks (blue circles) after comparison of the (+) and (–) reagent channels. (D) Incorporation of peaks added (red circles) or deleted by the user and subsequent refinement of peak positions. Positions of synchronized (+) and (–) reagent peaks that will be used during the integration phase are emphasized with solid lines.

Figure S2. Sequence alignment. Electropherogram showing the (+) reagent (bottom) and the ddNTP (top) channels. Sequencing channels have been inverted to facilitate visualization of peak synchronization with the reagent channel. Reagent peaks assigned in the alignment step (Figure S1) are highlighted with blue dotted lines. Results of the four phases of sequence assignment are shown explicitly. (A,B) Detection of peaks corresponding to the first and second sequencing channels, respectively. Peaks not accepted as valid sequencing positions are shown with black and red filled circles. (C) Assignment of input sequence with the identified sequencing peaks. (D) Complete alignment of the input RNA sequence. This alignment is offset by one nucleotide to reflect that dideoxy sequencing fragments are 1 nucleotide longer than the cDNA fragments that identify 2'-*O*-adduct sites.



Vasa et al. / Figure S1



Vasa et al. / Figure S2