Published in final edited form as:

Methods Enzymol. 2009 ; 468: 47–66. doi:10.1016/S0076-6879(09)68003-4.

Rapid quantification and analysis of kinetic •OH radical footprinting data using SAFA

Katrina Simmons, Joshua S. Martin, Inna Shcherbakova^{*}, and Alain Laederach

Developmental Genetics and Bioinformatics Wadsworth Center, Albany NY 12208

^{*}Department of Biochemistry, Albert Einstein College of Medicine 1300 Morris Park Avenue, Bronx, NY 10461

Abstract

The use of highly reactive chemical species to probe the structure and dynamics of nucleic acids is greatly simplified by software that enables rapid quantification of the gel images that result from these experiments. SAFA (Semi-Automated Footprinting Analysis) allows a user to quickly and reproducibly quantify a chemical footprinting gel image through a series of steps that rectify, assign, and integrate the relative band intensities. The output of this procedure is raw band intensities that report on the relative reactivity of each nucleotide with the chemical probe. We describe here how to obtain these raw band intensities using SAFA and the subsequent normalization and analysis procedures required to process this data. In particular, we focus on analyzing time-resolved hydroxyl radical (•OH) data, which we use to monitor the kinetics of folding of a large RNA (the L-21 *T. thermophila* group I intron). Exposing the RNA to bursts of •OH radicals at specific time-points during the folding process monitors the time-progress of the reaction. Specifically, we identify protected (nucleotides that become inaccessible to the •OH radical probe when folded) and invariant (nucleotides with constant accessibility to the •OH probe) residues that we use for monitoring and normalization of the data. With this analysis, we obtain time-progress curves from which we determine kinetic rates of folding. We also report on a data visualization tool implemented in SAFA that allows users to map data onto a secondary structure diagram.

1. Introduction

Chemical probes such as the •OH radical (Latham and Cech, 1989), dimethyl sulfate (DMS) (Tijerina, *et al.*, 2007), N-methylisatoic anhydride (NMIA) (Wilkinson, *et al.*, 2006), and various RNA modifying and cleaving enzymes (Donis-Keller, *et al.*, 1977) are commonly used to determine aspects of RNA structure and dynamics. The general principle behind all of these probes is their selective reactivity with nucleotides in different conformations or structural states (Felden, *et al.*, 1996). The information provided by these experiments indicates how the local structure of the nucleic acid affects the reactivity of the chemical probe. For example, •OH radicals will react more readily with nucleotides that are solvent exposed, reporting on the solvent accessibility of particular residues (Latham and Cech, 1989). DMS reacts selectively with adenines and cytosines that are not involved in base-pairing reactions (Lempereur, *et al.*, 1985), while NMIA creates 2'-O-Adducts with highly flexible nucleotides in the RNA (Badorrek and Weeks, 2005).

Chemical mapping experiments are relatively simple to carry out experimentally and readout is often accomplished using gel electrophoresis (Laederach, *et al.*, 2008). Nucleic acids are

readily separated out on electrophoretic gels with single nucleotide resolution. Large electrophoretograms such as that illustrated in Figure 1A will often contain several thousand individual bands, each representative of a nucleotide's relative chemical reactivity. Each lane in the gel can represent the RNA's reactivity under different conditions, or as is the case in Figure 1A, at different times during a folding reaction. The ability to probe RNA structure under different conditions and at different time points is a major motivation for obtaining quantitative band intensities from the gel image shown in Figure 1A.

The gel image shown in Figure 1A reports on the Mg^{2+} induced folding reaction of a mutant (L5b) of the L-21 *T. thermophila* group I intron (Laederach, *et al.*, 2007). We will use this data (along with a similar gel reporting on the folding reaction of the wild-type RNA) to illustrate the process of obtaining accurate and reproducible time-progress curves for analysis. We analyze the Mg^{2+} induced folding reaction as measured by $\bullet OH$ radical footprinting for both a wild type and mutant RNA, which have different folding rates (Shcherbakova and Brenowitz, 2005, Laederach, *et al.*, 2007). $\bullet OH$ radicals measure the solvent accessibility of nucleotides by selectively reacting with highly accessible residues, indicating the sites of protection and accessibility in the molecule (Latham and Cech, 1989). The radicals are generated either using synchrotron radiation, or the Fenton chemical reaction (Fenton, 1894, Shcherbakova, *et al.*, 2006, Shcherbakova and Brenowitz, 2008). The basic premise of the data we are analyzing is a relative change in the accessibility of nucleotides as the RNA folds upon the addition of Mg^{2+} . As shown in Figure 1B, certain bands become lighter (or darker) as a function of time, indicating folding of the RNA. By sequentially cleaving the RNA with $\bullet OH$ radicals at different time points during the reaction, we measure the fractional saturation of the reaction. We will use this data to generate time-progress curves, from which we can then obtain local rates.

SAFA is software designed to facilitate the analysis of RNA footprinting gels. We describe here the major steps in analyzing this data using the software, which can be downloaded along with the example files at <http://simtk.org/home/safa>. SAFA runs on Mac OS X, Windows XP, and is written using the Matlab (The Mathworks) programming language. Heavy users of the system should consider acquiring the Matlab software (which is often licensed campus-wide at most academic institutions) as performance of SAFA is significantly improved when used in conjunction with Matlab. When users choose to use SAFA in conjunction with Matlab, they must download the "Source" version of the software independently of whether they are running Matlab on a Windows or Macintosh computer. Furthermore, if users have the Windows Vista or Home edition, using Matlab will solve many compatibility issues.

2. Using SAFA

Using SAFA is relatively straightforward. Users will recognize familiar menus (such as the File Menu, indicated by a red arrow in Figure 2A). The basic workflow of SAFA is controlled through a series of buttons (marked by the green arrows) on the right side of the program in Figure 2A. Clicking on each button successively leads the user through the different steps of the procedure. Initially, the software requires as input both a gel image (either with a .gel or .tiff extension, 16 bit images are generally easier to analyze with the software, although 8-bit images are supported) and a sequence file (in FASTA, .fas format). The gel image will appear in the central SAFA window (Figure 2A) while reading in a sequence file will open the sequence selection dialog box (Figure 2B). In general, any keyboard commands or keyboard shortcuts are displayed above the main gel window in the software. Several example data sets are available for download from <http://simtk.org/home/safa> including all the data presented in this manuscript.

Our goal is to obtain quantitative intensities for all bands shown in Figure 1. As can be seen in Figures 1B and C, certain bands (in this case corresponding to nucleotides C177 and G160,

respectively) become significantly lighter as the folding reaction progresses. This indicates that as the molecule is folding these regions are becoming progressively more protected. By plotting the relative change in band intensity as a function of time, we can generate time-progress curves that report on the local rate of conformational change in the RNA molecule. Several steps are required to obtain this data from the gel image. The first step we describe here involves integrating the band intensity using a procedure called “single band peak fitting” (Takamoto, *et al.*, 2004).

Single band peak fitting fits a model to the raw image data, but requires accurate initial guesses of the relative positions of the bands on the gel. The inevitable imperfections in the gel matrix yield gel images with non-parallel lanes and “smiles” (or “frowns”) that make the problem of estimating band positions quite difficult. SAFA includes functionality to “rectify” the gel image, effectively correcting the imperfections so as to have parallel lanes, and no smiles (or frowns) in the gel image. Details of the procedure for gel rectification can be found in previous publications (Laederach, *et al.*, 2008) and a video demonstration of the process can be downloaded from <http://simtk.org/home/safa>. We recommend viewing the video tutorials, as the gel rectification process is fundamentally a visual process, requiring accurate user input. For the purposes of this Method, we will briefly outline the procedure for the example data set we use here.

2.1 Gel Cropping

The gel image must be initially cropped. In Figure 3A we show a screenshot of the main SAFA window in which we have loaded the gel image into SAFA. Using the cursor, we have drawn a cropping box around the right half of the gel image, which reports on the folding progress of the L5b mutant of the *T. thermophila* group I intron. We are careful to include in our cropping the T1 digest lane, which will help us assign the bands to the sequence in later steps. In general, a large enough portion of the gel image should be kept so that all lanes are encompassed.

2.2 Lane definition and gel rectification

The next step in the SAFA procedure involves defining lanes. This procedure is used to vertically align the gel by identifying lane boundaries. Such lane boundaries are generally trivial to identify by eye, but present a significant computational challenge when attempted in a fully automatic way. SAFA therefore uses a semi-automated approach to resolve this problem. The user initially defines at least one lane by drawing the boundaries of that lane using a click and draw procedure, analogous to a line tool in a computer graphics program. The user can then invoke an automated procedure for “guessing” the next lane boundaries, which uses the previous lane definitions as a starting point. This automated procedure is invoked by typing “G.”

The procedure of guessing lane boundaries is continued until all lanes are correctly selected, at which point the user invokes a vertical rectification procedure by typing “Z.” The vertical rectification procedure uses a linear interpolation routine that preserves the total pixel count in the gel image. The result of this procedure is illustrated in Figure 3B, where each lane is properly defined by red vertical lines. SAFA then identifies the center of the lanes as illustrated by dotted lines in Figure 3C. Once the lanes have been defined, the gel now requires horizontal alignment.

The horizontal alignment procedure in SAFA is analogous to the vertical one. In this case, horizontal lines are drawn through corresponding bands in the gel image as shown in Figure 3D. These lines should trace through the bands in the gel and will act as anchor lines for the horizontal rectification procedure. This procedure is manual, and the user should choose several well-resolved bands throughout the gel image as shown in Figure 3D. The user can zoom in

and out by right clicking to more accurately trace through the bands in the gel. The number of such anchor lines required and their position is dependent on the extent of horizontal deformation within the gel image. The ultimate result (Figure 3E) is a gel that is both vertically and horizontally aligned, making the process of band assignment straightforward.

2.3 Band assignment and use of a sequence file

SAFA requires two input files, a gel image and a FASTA sequence file. The FASTA file for the L-21 *T. thermophila* group I intron is shown below:

```
>Group I Intron, T. thermophila
GGAGGGAAAAGUUAUCAGGCAUGCACCUGGUAGCUAGUCUUUAAAACCAAUAG
AUUGCAUCGGUUUAAAAGGCAAGACCGUCAAAUUGCGGGAAAGGGGUCAACA
GCCGUUCAGUACCAAGUCUCAGGGGAAACUUUGAGAUGGCCUUGCAAAGGGU
AUGGUAAUAAGCUGACGGACAUGGUC
CUAACCCACGCAGCCAAGUCCUAAGUCAACAGAUCUUCUGUUGAUUAUGGAUGC
AGUUCACAGACUAAAUGUCGGUCGGGGAAGAUGUAUUCUUCUCAUAAGAUAU
AGUCGGACCUCUCCUUAUUGGGAGCUAGCGGAUGAAGUGAUGCAACACUGGA
GCCGCUGGGAACUAAUUUGUAUGCGA AAGUAUAUUGAUUAGUUUUGGAGU
```

In general, the first line of the FASTA file contains some descriptive text of the sequence preceded by a “greater than” character (>). It is also important to input the RNA sequence (replacing T with U) if the molecule being studied is in fact RNA. SAFA can also be used to study DNA, and therefore distinguishes between T and U in the sequence. Most importantly, the file should be saved as a “Western ASCII” text file, especially if it is being edited in a word processing application. SAFA assumes that the first nucleotide in the file corresponds to the first nucleotide in the sequence, unless an offset is set in the sequence selection tool (Figure 2B).

The sequence selection tool (Figure 2B) defines how the user will assign bands on the gel. It is invoked by clicking on “Load Sequence” (Figure 2A, green arrow marked 1) in the main SAFA window. The general concept behind the sequence selection tool is that one or several ladder lanes will be run in a separate lane on the gel being analyzed. In the case of RNA, T1 digests are often used to identify Guanosine nucleotides; this is the case in Figure 1. Alternatively, sequencing ladders can also be run to identify any (or all) of the other nucleotides (Tijerina, et al., 2007). Ultimately, with the gel rectified, the user will click on the bands they have chosen with the sequence selection tool. As is illustrated in Figure 2B, the user has selected “G” and set an offset of 21, since the first nucleotide in the sequence file corresponds to the 21st residue of the *T. thermophila* group I intron. The user will only need to identify bands corresponding to G residues, and SAFA will guess the positions of all the residues in between.

This semi-automated procedure allows the user to precisely define bands in the gel across all lanes rapidly. Furthermore, by adjusting the parameters in the sequence selection tool, the user has the option of selecting one or multiple nucleotides for manual definition during the assignment process. In general, selecting one nucleotide is sufficient, as SAFA is able to extrapolate the positions of the other nucleotides accurately during the assignment process. The sequence selection tool also allows the user to specify that the RNA is 3' labeled, indicating that the 5' band will be at the bottom of the gel. This option can also be used if the gel to be analyzed is the result of reverse transcription and the user wishes to use the RNA sequence. SAFA will use a single peak-fitting algorithm to fit the data, which can handle overlapping bands. As a result the user can choose to continue to assign bands in the overlapping region of the gel. How “high” to go in the gel will depend greatly on the quality of the gel image, desired quality of the data, and complexity of the experiment. Users will need to evaluate the reproducibility of the procedure on their own data to evaluate the degree to which they can

analyze highly overlapping regions in the gel. In the example we provide along with this manuscript, we analyze the gel conservatively, obtaining data for the first 100 bands.

2.4 Single peak fitting with SAFA

Peak fitting is a fully automated process in SAFA and is activated by clicking on the “Quantify” button in the main SAFA window (Figure 2A, green arrow marked 7). SAFA will only perform the peak fitting analysis if all the previous steps have been completed. Progress in the peak fitting procedure is monitored through a series of three graphs that appear when the peak fitting procedure is run (Figure 2C). SAFA proceeds through each lane of the gel successively, allowing the user to visualize the quality of the fit. Visualizing the resulting fit is key when troubleshooting the quantification procedure, as potential errors in the fitting procedure can easily be seen when looking at these results. SAFA will plot all of the fits again once the fitting procedure is complete, to encourage users to look carefully at the traces.

We illustrate in Figure 4A a successful single-peak fitting procedure on Lane 4 of the gel shown in Figure 1A. In Figure 4A, SAFA plots the individual peaks in red, and the overall lane profile in blue. In this case, the individual Lorentzian peaks are evenly spaced, and have uniform widths. In general, SAFA optimizes the fit so as to have regularly spaced peaks with linearly varying widths. However, if the initial positions of the peaks as defined by the band assignment procedure (Figure 3F) are inaccurate, SAFA will find a solution with large differences in peak widths, as is illustrated in Figure 4B. In this case, the user has incorrectly assigned bands and the solution found by SAFA has irregularly spaced peaks and highly variable peak widths. SAFA will replot the fitted data as illustrated in Figure 4 for each lane after the fitting procedure, and it is recommended to visually inspect these plots to identify any fits that look like Figure 4B, as these will indicate potential inaccuracy in the results. Although SAFA will check for convergence of the fitting procedure and provide an error message if the convergence criteria are not met, it is often still possible to obtain convergence on an incorrect solution as is illustrated in Figure 4B. Visual inspection is thus an important final step of the fitting procedure.

2.5 Data output

At this stage, SAFA will output a text file with the raw band intensities determined through the single peak fitting procedure. The text file is tab delimited, with the first column containing the residue numbers, and the next columns the actual raw band areas, as determined by peak fitting. It should be noted that the data is output in scientific notation, and when the file is opened in a spreadsheet program like Excel, the nucleotide numbers are displayed in scientific notation as well. Interpretation of the raw data, however, requires further analysis, normalization and visualization. Although some users at this stage may prefer to perform these analyses in other software, SAFA offers several other tools for data visualization and normalization. The automated normalization functionality of SAFA is described in detail in a previous publication (Laederach, et al., 2008); we choose here instead analyze the time-resolved data manually to better illustrate the role of normalization in obtaining time-progress curves from these data (Shcherbakova, et al., 2006). We also turn our focus on a previously unpublished tool built into SAFA for the visualization/projection of chemical mapping data on RNA secondary structure diagrams. These two approaches are critical for the correct interpretation of chemical mapping data.

3. Data normalization

3.1 Data normalization using invariant residues

The first step of data normalization requires correcting for non-uniform loading of the RNA into each lane. This procedure involves identifying invariant residues, i.e. nucleotides where the relative density changes reflect the variations in the amount of RNA loaded per lane. The

choice of the invariant residues is based on visual inspection of the gel image to identify bands that demonstrate no visible systematic change in intensity. Furthermore any available structural information can be used to ensure that the selected residues belong to a region where structural changes are not expected. To verify whether the choice of invariant bands was correct, we plot the invariant band intensities as a function of reaction time to rule out a systematic increase or decrease in the band intensities. Usually a single set of invariant residues can be used for all the protections and sites of the enhanced reactivity on the same gel. In the case of our example, we chose nucleotides 113-115 (Figure 1D) which are in the P4 helix of L-21 *T. thermophila*, a highly stable region of the molecule that does not change conformation during folding. In general, helical segments work best as invariant residues.

To normalize the time-resolved data, we divide the intensities of the all the bands by the average intensity of the invariant residues. The result of this procedure is illustrated in Figure 5 where the raw data (Figure 5A) and invariant normalized data (Figure 5B) for the wild-type *Tetrahymena* ribozyme (black) and L5b mutant is plotted (red symbols). The effect of this simple procedure is dramatic, resulting in uniform time-progress curves which can be subjected to further analysis. We provide the Microsoft® Excel spreadsheet used to normalize this data at <http://simtk.org/home/safa>. It should be noted that small variations in sample loading are easily corrected in this way. However, if large variations in sample loading are observed, this will lead to much greater noise in the data regardless of which invariant residues are chosen.

The second step of data normalization aims to determine the extent of protection as a fractional value between 0 (initial state) and 1 (final state). This is accomplished by dividing the band intensity by the absolute difference in band intensity between the final and initial states. Following this transformation, the intensities are individually scaled to fractional saturation and this is illustrated in Figure 5C. This transformation creates uniform kinetic data, which allows us to compare changes in intensity between different molecules. In this case, it is clear that the wild-type and mutant RNA (black and red, Figure 5, respectively) have different rates of formation. For the three protections we illustrate in Figure 5, the L5b mutant appears to fold slower.

3.2 Generating time-progress curves and obtaining rate constants

The normalized data is best visualized in Figure 5C by plotting the fractional saturation of the individual sites that demonstrate changes in the reactivity as a function of time. The reaction time is calculated as a sum of the quench-flow delay time and half of the cleavage (or modification) reaction time. In order to visualize the details of the fastest events the x-axis is often set to a logarithmic scale.

The collection of the individual time points can be fit to an exponential function to discern rate constants for the processes that are reflected in the time-resolved changes in the local reactivity to the footprinting reagent (Figure 5C, lines). These changes are best described in terms of an exponential function:

$$\bar{Y} = 1 - \sum_{i=1} \alpha_i \exp(-k_i \cdot t),$$

where α_i and k_i are the amplitude and rate constant, respectively, of the i^{th} kinetic phase. When fitting the data we start with a single exponential parameter ($i=1$). If a systematic error arises as a result of the fitting two rate constants are fit. We usually report lower and upper limits calculated at the 65% confidence interval that corresponds to one standard deviation of the data.

3.3 Data interpretation by inspection of the time progress curves and comparison of rate constants

In many cases a working hypothesis can be discerned from a simple inspection of the time progress curves and the comparison of the rate constants. For example, for the limited number of time-progress curves illustrated in Figure 5 we conclude that not all the time-progress curves demonstrate the same kinetic behavior. The significant offset in the reactivity (or burst) between initial state and the first data point for the wild type data (Figure 5, black curves) indicates that the onset of the folding process is faster than the dead time of the experiment. In this case better time resolution would be beneficial to characterize the fastest events of the folding under this condition. Improved time resolution is achieved by Fast Fenton footprinting (Shcherbakova, et al., 2006) and will reveal the multiple kinetic phases of the local changes in the reactivity at shorter time-points. The data we present here, however, was collected using synchrotron radiation to generate •OH radicals, which did not have the time-resolution to resolve the fastest events in the folding of the wild-type molecule (Ralston, *et al.*, 2000). We can nonetheless conclude from this analysis that the mutation of the L5b tetraloop causes significant changes in the folding mechanism, which yield smaller rates (Figure 5C). The presence of a distinct hierarchical folding mechanism for the wild type ribozyme is gleaned from the comparison of the rate constants and the amplitudes of the fastest folding events: formation of the hinge region of the P4-P6 domain, reported by increase in the reactivity for nucleotide 122, precedes to the formation of junction region between P4-P6 domain and the catalytic core (nucleotides 109-112), whereas the catalytic core (nucleotides 95-97) is the slowest to fold. A complete analysis of these data is described in (Laederach, et al., 2007) where we model the folding pathways of the RNA based on these experiments.

4. Data visualization

SAFA outputs a large table of numbers, which alone can be difficult to interpret. If the data measures time-progress of the folding reaction, time-progress curves can be generated and visualized as described above (Figure 5). Often, however, the ability to project the data onto the two-dimensional structure of the RNA is valuable in identifying regions of the molecule that are highly reactive to a particular probe. We have therefore developed a way to quickly and clearly map the quantified data generated through SAFA analysis onto the molecule's secondary structure. Our idea is based on the fact that each lab has their favorite secondary structure diagram, and that a tool able to map data directly onto this diagram will be most valuable in interpreting the data within the structural context of the experiment. Other tools such as VARNA (Ferre, *et al.*, 2007) and S2S (Jossinet and Westhof, 2005) generate automated secondary structure representations on which users may map data. These representations however do not always correspond to the context in which the data is being visualized, and as such we decided a tool that can facilitate plotting of data onto a user-defined diagram would have value. This tool can be accessed from the Secondary Structure Plot menu item under Visualize Data in SAFA.

The visualization tool requires two inputs, SAFA data and a high-resolution (greater than 150 dpi) image of the secondary structure. SAFA reads either JPEG, GIF, or TIFF formatted pictures with the preferred format TIFF (without any form of compression). The image should have some space reserved on one side for a color legend. This procedure allows the user to manually define the relative positions on the image once and plot any data onto the image thereafter. It is therefore worthwhile obtaining a high quality image before carrying out the manual parts of the procedure.

There are four steps required to generate a figure.

1. The image (in JPEG, GIF, or TIFF format) of the molecule's secondary structure is loaded using the *SecStruct Image* button (Figure 6A, red arrow) When the image loads correctly, "Loaded" should appear to the right of the *SecStruct Image* button, and the image will appear in the main window as shown in Figure 6B.
2. The *Pick Points* (Figure 6B, blue arrow) button will become active once an image is loaded. Clicking the *Pick Points* button will begin the procedure for indicating the location of nucleotides on the secondary structure diagram. The user can place squares over each residue in the corresponding region of the molecule. SAFA assumes that the first square to be placed corresponds to nucleotide number one, unless the user specifies an offset (Figure 6B, green arrow). Each box added by a left click corresponds to a particular numbered nucleotide starting from the offset and increasing by increments of one. To assist in determining which box corresponds to what nucleotide, the next box number is indicated by the orange arrow on Figure 6C. To place successive squares, the left button is clicked. The right button on the mouse zooms in and out and inaccurately placed boxes are removed by either clicking the middle button on the mouse or moving the crosshairs over the box and typing "E" on the keyboard. The size of the squares can be set with the *Square Width* button (magenta arrow, Figure 6B). The value must be a whole number. An example of the pick points procedure on the 5' region of the *T. thermophila* group I intron is shown in Figure 6D. Once all the boxes have been matched with their corresponding nucleotides (as illustrated in Figure 6E), clicking either Z or Q will end the pick points procedure. The figure will return to its original window. Here the red highlighted boxes will become blue and this signifies the picking points procedure is done. Figure 6F illustrates the result of the completed procedure.
3. Once all the points have been picked, clicking the *Protection Data* button (indicated by a black arrow on Figure 6F) will invoke a dialog to input the text file. A new window will appear with a color-plot of the data (Figure 6G). The user can then choose the lane corresponding to the data that he/she wishes to visualize on the secondary structure diagram. The colors used to represent the data are defined in the following manner: The white nucleotides correspond to "zero" intensity. The value of the zero level is initially set to be the average of the minimum and maximum intensities in the data. Each residue that has higher reactivity will be colored red with intensity on a linear scale from this zero-level to maximum value. The maximum value is also a default value and is plotted in solid red. Similarly, residues with reactivity values lower than the zero-value will be colored blue with intensity on a linear scale from the zero-level to a minimum value. The default minimum value is plotted as solid blue. The maximum, minimum, and zero-values are adjusted by clicking the appropriate *Set Max/Zero/Min Value* buttons (Figure 6F, purple arrows).
4. When the structure image, residue locations, and accessibility data have all been loaded or assigned, the *Generate Figure* button will become activate. This generates a full-sized figure showing the reactivity levels plotted on the secondary structure of the molecule (Figure 6H). The plot protection legend checkbox controls whether the color legend is displayed in the preview plot as well as in the full-sized figure. If a legend is being displayed, its location is indicated before saving the figure or exporting to the file format of choice. Any nucleotides where no data is available will be covered in a gray box (as illustrated in Figure 6H).

The secondary structure image and the points picked for the residue locations on that image can be loaded independently from previously saved data files using the *SecStruct Image* and *Load Pickpoints* buttons, and all fields can be saved and later loaded together from those data files using the *Load/Save SecStruct Data* buttons (Figure 6F, light blue arrows).

5. Conclusion

In conclusion, SAFA is a software package specifically designed for obtaining quantitative and reproducible data from a gel image of a chemical mapping reaction. It was originally intended for use mostly with •OH radical footprinting gels (Gross, *et al.*, 1998, Brenowitz, *et al.*, 2002, Shcherbakova, *et al.*, 2006, Woodson, 2008), but is now being used for many other applications in nucleic acid chemical mapping (Wilkinson, *et al.*, 2005, 2006, Vicens, *et al.*, 2007). Since most chemical mapping experiments use similar principles to •OH radical footprinting for the analysis of the reactions, it is easy to adapt SAFA to a wide range of applications. The sequence browser (Figure 2B) is instrumental in defining how the band assignment procedure will behave, and by using the correct sequence and selection of nucleotides; the user can define any arbitrary combination of nucleotides. This flexibility, along with the gel rectification procedure, provide a series of tools for the rapid and accurate quantification of chemical mapping gels (Das, *et al.*, 2005).

Acknowledgments

This work is supported the US National Institutes of Health (NIGMS) through grant R00-79953 to A.L.

7. Bibliography

- Badorrek CS, Weeks KM. RNA flexibility in the dimerization domain of a gamma retrovirus. *Nat Chem Biol* 2005;1:104–111. [PubMed: 16408007]
- Brenowitz M, Chance MR, Dhavan G, Takamoto K. Probing the structural dynamics of nucleic acids by quantitative time-resolved and equilibrium hydroxyl radical “footprinting”. *Curr Opin Struct Biol* 2002;12:648–653. [PubMed: 12464318]
- Das R, Laederach A, Pearlman SM, Herschlag D, Altman RB. SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. *RNA* 2005;11:344–354. [PubMed: 15701734]
- Donis-Keller H, Maxam AM, Gilbert W. Mapping adenines, guanines, and pyrimidines in RNA. *Nucleic Acids Res* 1977;4:2527–2538. [PubMed: 409999]
- Felden B, Florentz C, Westhof E, Giege R. Usefulness of functional and structural solution data for the modeling of tRNA-like structures. *Pharm Acta Helv* 1996;71:3–9. [PubMed: 8786997]
- Fenton HJH. What species is responsible for strands scission in the reaction of [FEIIEDTA]₂ with H₂O₂ with DNA? *J Am Chem Soc* 1894;6:899.
- Ferre F, Ponty Y, Lorenz WA, Clote P. DIAL: a web server for the pairwise alignment of two RNA three-dimensional structures using nucleotide, dihedral angle and base-pairing similarities. *Nucleic Acids Res* 2007;35:W659–668. [PubMed: 17567620]
- Gross P, Arrowsmith CH, Macgregor RB Jr. Hydroxyl radical footprinting of DNA complexes of the ets domain of PU.1 and its comparison to the crystal structure. *Biochemistry* 1998;37:5129–5135. [PubMed: 9548743]
- Jossinet F, Westhof E. Sequence to Structure (S2S): display, manipulate and interconnect RNA data from sequence to structure. *Bioinformatics* 2005;21:3320–3321. [PubMed: 15905274]
- Laederach A, Das R, Vicens Q, Pearlman SM, Brenowitz M, Herschlag D, Altman RB. Semiautomated and rapid quantification of nucleic acid footprinting and structure mapping experiments. *Nature protocols* 2008;3:1395–1401.
- Laederach A, Shcherbakova I, Jonikas MA, Altman RB, Brenowitz M. Distinct contribution of electrostatics, initial conformational ensemble, and macromolecular stability in RNA folding. *Proc Natl Acad Sci U S A* 2007;104:7045–7050. [PubMed: 17438287]
- Latham JA, Cech TR. Defining the inside and outside of a catalytic RNA molecule. *Science* 1989;245:276–282. [PubMed: 2501870]
- Lempereur L, Nicoloso M, Riehl N, Ehresmann C, Ehresmann B, Bachellerie JP. Conformation of yeast 18S rRNA. Direct chemical probing of the 5' domain in ribosomal subunits and in deproteinized RNA

- by reverse transcriptase mapping of dimethyl sulfate-accessible. *Nucleic Acids Res* 1985;13:8339–8357. [PubMed: 2417197]
- Ralston CY, Sclavi B, Sullivan M, Deras ML, Woodson SA, Chance MR, Brenowitz M. Time-resolved synchrotron X-ray footprinting and its application to RNA folding. *Methods Enzymol* 2000;317:353–368. [PubMed: 10829290]
- Shcherbakova I, Brenowitz M. Perturbation of the hierarchical folding of a large RNA by the destabilization of its Scaffold's tertiary structure. *J Mol Biol* 2005;354:483–496. [PubMed: 16242711]
- Shcherbakova I, Brenowitz M. Monitoring structural changes in nucleic acids with single residue spatial and millisecond time resolution by quantitative hydroxyl radical footprinting. *Nature protocols* 2008;3:288–302.
- Shcherbakova I, Mitra S, Beer RH, Brenowitz M. Fast Fenton footprinting: a laboratory-based method for the time-resolved analysis of DNA, RNA and proteins. *Nucleic Acids Res* 2006;34:e48. [PubMed: 16582097]
- Takamoto K, Chance MR, Brenowitz M. Semi-automated, single-band peak-fitting analysis of hydroxyl radical nucleic acid footprint autoradiograms for the quantitative analysis of transitions. *Nucleic Acids Res* 2004;32 in press.
- Tijerina P, Mohr S, Russell R. DMS footprinting of structured RNAs and RNA-protein complexes. *Nature protocols* 2007;2:2608–2623.
- Vicens Q, Gooding AR, Laederach A, Cech TR. Local RNA structural changes induced by crystallization are revealed by SHAPE. *Rna* 2007;13:536–548. [PubMed: 17299128]
- Wilkinson KA, Merino EJ, Weeks KM. RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. *J Am Chem Soc* 2005;127:4659–4667. [PubMed: 15796531]
- Wilkinson KA, Merino EJ, Weeks KM. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nature protocols* 2006;1:1610–1616.
- Woodson SA. RNA folding and ribosome assembly. *Current opinion in chemical biology* 2008;12:667–673. [PubMed: 18935976]

Figure 1.

A) Gel image of a typical chemical mapping experiment on the L5b mutant of the *T. thermophila* group I intron analyzed by $\bullet\text{OH}$ radical cleavage. Each band in the gel corresponds to a specific nucleotide and each lane, in this case, a different time point along during the folding reaction. Nucleotides where a visual change in band intensity is seen as a function of time are magnified in B) and C). D) Illustrates a zoomed in version of the gel image on the invariant residues (113-115) used for normalization. This manuscript focuses on the SAFA software, which is designed to rapidly quantify the relative band intensities from such a gel image (Das, et al., 2005).

Figure 2.

Screenshots of the SAFA software. A) The main SAFA window used to control all aspects of the software. The analysis procedure is carried out by successively clicking on each of the buttons to the right side the main window in which the gel image is displayed. B) The Sequence Selection tool allows the user to define how the semi-automated procedure of band assignment will be carried out. In this case, the sequence selection tool is setup to use a T1 ladder, which indicates the positions of Guanosines on the gel. C) The quantification progress window appears when the software is automatically fitting lanes and is used to monitor the progress and accuracy of the single peak fitting procedure (Takamoto, et al., 2004).

Figure 3.

Illustration of the gel rectification procedure implemented in SAFA. A) Screenshot of the main SAFA window with a gel image loaded; the red box delineates the user defined cropping boundary for subsequent analysis. B) Screenshot of the lane definition procedure. A red line is manually drawn that marks the boundaries of each lane. Once a few lanes have been manually defined, SAFA is able to “guess” the outline of the following lanes, which is done by typing the letter “G”. C) Screenshot of the completed and recorded defined lanes that occur after the user types “Z” or “Q”. The color of the lines changes from red to green and a dotted line marks the middle of the lane when the procedure is completed. D) Screenshot of the horizontal alignment procedure. The gel is aligned horizontally to ensure the bands are parallel. Similar to the lane definition process, the user identifies bands across the gel image with a horizontal line. Only one line should be drawn per band. The left button of the mouse begins the line and marks sequential points throughout the row and the right button ends the line. The number of rows to mark is at the users discretion, however, the more rows used will result in a more accurate an adjustment of the image. E) Screenshot of the finished horizontally aligned gel. F) Screenshot of the band assignment procedure.

Figure 4.

SAFA performs a peak fitting procedure to determine the relative band intensity in the gel image. When the fitting procedure has converged, SAFA will plot the results as shown here. The red lines represent individual Lorentzians that when summed yield the lane profile shown in blue. A) A successful and accurate single-peak fitted graph for lane 15 of the gel shown in Figure 1. B) Visual inspection of the fitted profile can be used to determine if the user was not accurate in assigning bands (Figure 3F). In this example there are large differences in the peak widths and irregularly spaced peaks. These are signs of inaccurate results.

Figure 5.

Time-progress curves following SAFA quantification. The wild-type data are shown in black, and the L5b mutant in red. A) When the data that is not normalized is plotted as a function of time, it is difficult to identify any trends in the changes in reactivity. B) When invariant residue normalization is applied (in this case based on invariant residues 113-115), much clearer changes in reactivity are revealed. C) When further normalization procedures are applied based on setting the initial and final fractional saturation to zero for $t=0$ and one for the final state, we are able to compare the data between experiments and determine rate constants for the transitions.

Figure 6.

The data visualization component allows users to map their data onto secondary structure diagrams. A) Screenshot of the starting window that opens when *Visualize Data -> Secondary Structure Plot* is selected. Selecting the *SecStruct Image* button (highlighted by the red arrow) allows the user to load a JPEG, GIF, or TIFF image of their secondary structure. B) Once the image is loaded, it will appear in the window and the next step is to pick points on the image. The *Pick Points* button indicated by the blue arrow begins the procedure where the user defines the location of each individual nucleotide on the diagram. The magenta arrow indicates the *Square Width* button, which determines the size of the box that is placed over the selected residues, and the green arrow indicates the *Offset* button that allows the user to account for any offset in the start nucleotide of the nucleic acid. C) A new window appears when the *Pick Points* button is pushed. Indications for keyboard shortcuts are printed above the upper x-axis along with the next box number indicated by an orange arrow. D) An example of pick-points procedure on the *T. thermophila* group I intron secondary structure diagram. E) An example of a completed procedure on the same figure. Each nucleotide in the diagram is covered by a square. F) Once the procedure of picking points is completed, the image and corresponding boxes are saved and can also be reloaded. The boxes change color from red to blue signifying the points have been saved. The light blue arrows mark the *Load SecStruct Data* and *Save SecStruct Data* buttons that allow the user to save their progress and load their data. The black arrow highlights the *Protection Data* button that initiates the next step in the process where the user chooses the data to be displayed on the figure. The purple arrows indicate the *Set Max/Zero/Min Value* buttons that allow the user to adjust these values. G) Representation of the SAFA quantified data for the gel image shown in Figure 1 in a colorplot allowing the user to choose which lane to plot on the secondary structure. H) Mapping of Lane 15 onto the secondary structure of the *T. thermophila* group I intron indicating nucleotides that are protected in blue and cleaved by $\bullet\text{OH}$ radicals in red.