DNA Fragment Analysis

ALLELE SIZE AND PLUS A IN CEQ™ FRAGMENT ANALYSIS SOFTWARE

Susan Stone
Beckman Coulter, Inc.

Most DNA polymerases frequently add a "non-templated" A to the ends of PCR products. When this happens, what results is a mixture of PCR products of different lengths from a single allele. Some of the products are the templated true length (also called "-A") and others are one nucleotide longer, or "+A." The ratio of the two types can be skewed in favor of the true-size allele or the +A size by the conditions of the PCR reaction and the design of the primers.

The presence of two different fragment sizes for a single allele complicates the interpretation of fragment analysis data. Instead of one peak for one allele, there are two that represent the same allele. For example, an allele with a true size of 107nt, when amplified, will give two fragments with sizes 107nt and 108nt. These will show up as two peaks in the fragment analysis data, spaced one nucleotide apart. Either the 107 or 108 peak may be larger, depending on the locus.

In the CEQ fragment analysis software, there are several places where information is entered regarding allele sizes and +A status. These affect where the software looks for peaks and how it assigns alleles to them.

On the **Allele List** dialog (accessed by selecting the **Allele List** button on the **Locus Tag Editor** dialog box), there is an **Allele List** at the bottom (Figure 1). This list is initially created by providing the following information and then clicking the **Generate List** button.

Maximum Fragment Size (on the **Locus Tag Editor** dialog)
Repeat Unit Length
No. of Repeats in Shortest Allele
Shortest Allele Size
Allele List IDs

The first four items mentioned above are used to generate a list of expected alleles (**True Sizes** and **Apparent Sizes** in the **Allele List** table), and the **# of Repeats.** The **IDs** are generic alphabetic or numeric names assigned in a logical sequential order. Numeric names generated increase by one for each repeat added to the smallest allele in the list. Single character alphabetic names are generated starting with "A" and incrementing through "Z." The user can adjust these names if desired.

The initial default numbers in the **Apparent Sizes** column are modified by the user to represent the sizes the alleles appear when analyzed on the CEQ. They will most likely *not* be exactly the same as the **True Sizes.** This is because the mobility of a fragment is somewhat affected by the sequence of the fragment, as well as its size. Sample fragments do not have exactly the same sequence as the fragments in the size standard ladder used as the basis for calculating size, so their calculated **Apparent Sizes** will differ slightly from their **True Sizes.** Note that this difference between apparent and true sizes is seen on all fragment analysis systems, not just the CEQ.

When the **Generate List** button is clicked, the **Apparent Sizes** of the generated alleles are initially set equal to the **True Sizes.** This is a temporary default. To determine the correct **Apparent Sizes** the user is expected to run test sample(s) of known alleles to determine the values to enter in this column. When the separations are run, the software

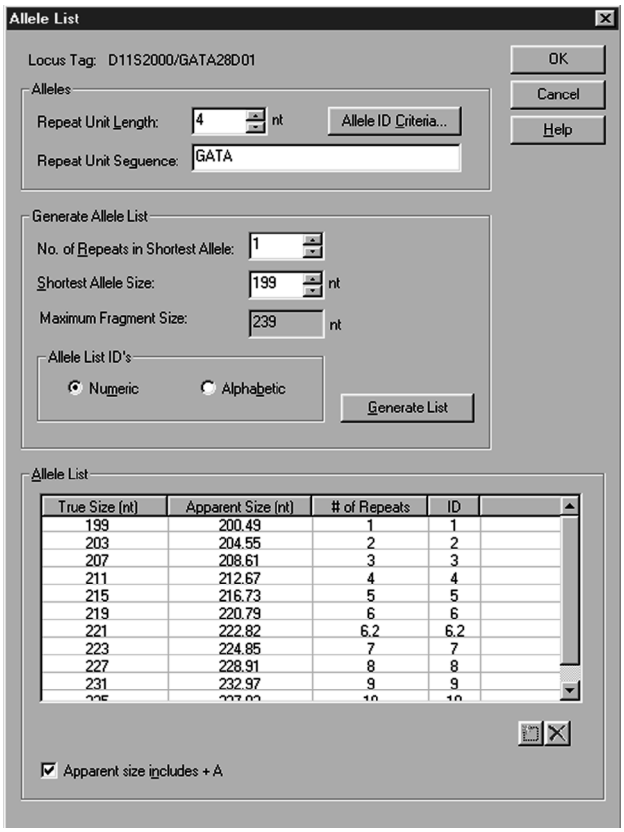

Figure 1. *Example of an Allele List for Locus D11S2000.*

assigns sizes to the allele peaks. The user records the values and returns to the **Allele List** screen to enter the observed **Apparent Sizes.** (The separation temperature may affect mobilities, so **Apparent Sizes** should be determined from known samples run using the same separation method that will be used for unknown samples.) The user also checks the **Apparent size includes +A** box at the bottom of the dialog box if the **Apparent Sizes** he chose to enter into the table are the +A versions of the alleles.

To expand on the example above, let's say the input is as follows:

Maximum Fragment Size:	119
Repeat Unit Length:	6
No. of Repeats in Shortest Allele:	4
Shortest Allele Size:	107
Allele List IDs:	Numeric

The default **Allele List** appears as follows:

True Size	*Apparent Size*	*# of Repeats*	*ID*
107	107	4	4
113	113	5	5
119	119	6	6

Apparent size includes +A not checked

Then the user runs multiple injections of two unknown alleles. The mean observed sizes for the -A and +A peaks for each allele are determined to be:

Allele ID	*True Size*	*-A peak*	*+A peak*
4	107	107.3	108.3
5	113	113.5	114.5

The third allele can be assumed to have the following characteristics, by assuming a linear relationship between the apparent size and the true size:

Allele ID	*True Size*	*-A peak*	*+A peak*
6	119	119.7	120.7

These assumed apparent sizes can be adjusted if necessary when the third allele is actually encountered.

Prior to updating the **Allele List,** it may be necessary to adjust the **Fragment Size Range** on the **Locus Tag Editor** dialog to accommodate apparent sizes outside the original range. The **Allele List** may be updated in one of two ways, depending on whether or not the **Use +A to call allele** checkbox was checked:

OPTION 1:

True Size	*Apparent Size*	*# of Repeats*	*ID*
107	107.3	4	4
113	113.5	5	5
119	119.7	6	6

Apparent size includes +A *not* checked

OPTION 2:

True Size	*Apparent Size*	*# of Repeats*	*ID*
107	108.3	4	4
113	114.5	5	5
119	120.7	6	6

Apparent size includes +A *is* checked

Now, let's move on to the **Allele ID Criteria** that are accessed by selecting the **Allele ID Criteria** button on the **Allele List** dialog box (see Figure 1). There are two criteria that have to do with +A detection. The first is a checkbox entitled **Detect +/-A.** If this box is checked, the software will identify in the Fragment List both the +A and –A peaks. It will identify one of them as the allele based on the **Use +A to call allele** checkbox described below and will designate the other as +A or –A accordingly.

If the **Detect +/-A** box is not checked, the software will identify one of the two peaks as the allele based on the **Use +A to call allele** checkbox. The

other peak will be included in the Fragment List as an Unknown Allele or it will be identified as a Spurious Peak if 1) **Detect Spurious Peaks** has been selected on the **Allele ID Criteria** dialog and 2) the peak height falls below the height designated in the **Maximum height for spurious peaks** selection.

The second +A criterion on the **Allele ID Criteria** dialog is a checkbox entitled **Use +A to call allele.** This box should be checked if the +A form is expected to be a higher peak, and not checked if the –A form is expected to be higher. If this box is checked, the software will identify the +A peak as the allele in the Fragment List. If it is not checked, it will identify the –A peak as the allele. In general, the reproducibility of size estimates is better for higher peaks than it is for very small peaks. In some cases, the smaller peak may be too small to be detected. So, it is good practice to direct the software to use the higher of the peaks associated with the allele to identify the allele.

Let's try some more examples. Assume the system has been set up with the OPTION 1 or the OPTION 2 allele list specified in an earlier example (it doesn't matter which one). Let's say this is the observed "raw data" for an unknown sample:

Peak	*Size*	*Peak Height*
Peak A	107.3	11,000
Peak B	108.3	3,000

Table 1 shows what will appear in the Fragment List based on the indicated parameter selections.

Below is another example, with the +A form as higher and "raw data" as follows:

Peak	*Size*	*Peak Height*
Peak A	107.3	10,000
Peak B	108.3	15,000

Table 2 shows what will appear in the Fragment List based on the indicated parameter selections. The peak identifications are the same as above, except in one circumstance where the second peak is identified as an Unknown Allele instead of a Spurious Peak because it is above the Spurious Peak threshold.

The lesson from these illustrations is that it is generally best to check the **Detect +/-A** option so the software will correctly categorize peaks. If the option is incorrectly configured, the alleles are still identified, but the +A or –A peaks are frequently categorized as Unknown Alleles and the software will report "Too many alleles" in the Comment column of the fragment data table.

Three of the parameters discussed above are key to determining where the software looks when identifying alleles. They are **Apparent Size, Apparent size includes +A,** and **Use +A to call allele.** Table 3 elucidates this relationship.

In summary, here are the three +A parameters in the CEQ fragment analysis software and their functions.

- **Apparent size includes +A** (on **Allele List** dialog) is used by the operator to indicate whether the sizes he entered into the **Apparent**

Table 1

	Parameter Selections		
Detect +/-A	***Use +A to Call Allele***	***Detect Spurious Peaks***	***Fragment List***
Checked	Checked	Doesn't matter	Peak A ➜ Minus A Peak B ➜ Allele 4
Checked	Not Checked	Doesn't matter	Peak A ➜ Allele 4 Peak B ➜ Plus A
Not Checked	Checked	Checked/50%	Peak A ➜ Unknown Allele Peak B ➜ Allele 4
Not Checked	Checked	Not Checked	Peak A ➜ Unknown Allele Peak B ➜ Allele 4
Not Checked	Not Checked	Checked/50%	Peak A ➜ Allele 4 Peak B ➜ **Spurious Peak***
Not Checked	Not Checked	Not Checked	Peak A ➜ Allele 4 Peak B ➜ Unknown Allele

*Because peak is below Spurious Peak threshold.

Table 2

	Parameter Selections		
Detect +/-A	***Use +A to Call Allele***	***Detect Spurious Peaks***	***Fragment List***
Checked	Checked	Doesn't matter	Peak A ➜ Minus A Peak B ➜ Allele 4
Checked	Not Checked	Doesn't matter	Peak A ➜ Allele 4 Peak B ➜ Plus A
Not Checked	Checked	Checked/50%	Peak A ➜ Unknown Allele Peak B ➜ Allele 4
Not Checked	Checked	Not Checked	Peak A ➜ Unknown Allele Peak B ➜ Allele 4
Not Checked	Not Checked	Checked/50%	Peak A ➜ Allele 4 Peak B ➜ **Unknown Allele***
Not Checked	Not Checked	Not Checked	Peak A ➜ Allele 4 Peak B ➜ Unknown Allele

*Unknown Allele because peak is above the Spurious Peak threshold.

Table 3

	Parameter Selections	
Apparent Size Includes +A	***Use +A to Call Allele***	***Where Software Looks***
Checked	Checked	At Apparent Size
Checked	Not Checked	At Apparent Size –1nt
Not Checked	Checked	At Apparent Size +1nt
Not Checked	Not Checked	At Apparent Size

Size column were for +A or –A fragments. This parameter affects where the software looks for peaks when identifying alleles.

- **Detect +/-A** (on **Allele ID Criteria** dialog) is used to select whether both the +A and –A peaks are identified in the Fragment List, or just one peak.
- **Use +A to call allele** (on **Allele ID Criteria** dialog) is used to indicate whether the +A or –A form of the allele will be more prevalent. It also indicates whether the +A or –A form of the allele will be identified in the Fragment List as the allele. This parameter affects where the software looks for peaks when identifying alleles.

www.sciex.com/ce

View SCIEX products at www.sciex.com
Find your local office at www.sciex.com/offices

AB SCIEX Headquarters
500 Old Connecticut Path | Framingham, MA 01701 USA
Phone 508-383-7700
www.absciex.com