# mFold, Delta G, and Melting Temperature

## Contents

## Abstract

For several years the value of Delta G ($\Delta G$) has been used to describe how 'bad' a duplexed structure derived from hairpin (mFold) and hetero and homo dimers analysis methods may be. Many commercially available oligo design software packages also use a $\Delta G$ profile as a metric of stickiness when examining a desired target for designing a probe or primer, where stickiness represents how thermodynamically stable a structure may be. Some feel that $\Delta G$ can give insight into the melting temperature ($T_m$) of the duplexed structure as if there is some kind of correlative relationship between the two.

Unfortunately, there is no real relationship between $\Delta G$ and $T_m$ where one can be used to identify the other. While it has been the standard to only use the $T_m$ value when comparing thermodynamic equivalency between oligos, whether it is for PCR, or microarray designs, or maximizing SNP discrimination, it is suggested that the $T_m$ value alone is also insufficient for thermodynamic comparison. The recommendation is that the two values, $T_m$ and $\Delta G$, together can provide much more qualitative thermodynamic understanding of a duplex or structure than either of the two values alone. Where as, the utilization of their principle components Enthalpy ($\Delta H$) and Entropy ($\Delta S$) can produce a quantitative thermodynamic understanding and is ideal for computational comparisons.

## Introduction

A Chemical Reaction is a process where one or more molecules are converted into one or more other molecules. In the case of oligos, it is a process where two single stranded oligo species come together to form a duplex. The transition from one state to another state results in a change in energy of the system, summarized by equation 1.

$$[Oligo_1] + [Oligo_2] \Leftrightarrow [Duplex] \hspace{3cm} \text{Equation 1}$$

where [Oligo$_1$] and [Oligo$_2$] represent the concentration of the two single stranded oligos in the system and [Duplex] represents the concentration of the two oligos hybridizing together to form the Duplex.

Delta G ($\Delta G$) is the change in Gibbs Free Energy (in units of kcal/mole) and is the net exchange of energy between the system and its environment and can be described by equation 2.

$$\Delta G = \Delta H - T \bullet \Delta S \qquad\qquad \text{Equation 2}$$

where $\Delta H$ (Enthalpy) represents the total energy exchange between the system and its surrounding environment (in units of kcal/mole), $\Delta S$ (Entropy) represents the energy spent by the system to organize itself (in units of cal/K•mole). Generally speaking a spontaneous system favors a more random system not an ordered system. Finally, T represents the absolute temperature of the system and is in units Kelvin (Celsius + 273.15).

The description of $\Delta G$ indicates that it is temperature dependent; hence a $\Delta G$ value at 25°C (298.15 Kelvin, room temperature) is not the same as what the $\Delta G$ value might be at 55°C (328.15 Kelvin, typical PCR reaction). At a given temperature a positive $\Delta G$ value indicates that the system will go in the direction of producing the reactants (non-spontaneous), for oligos this is single strandedness. A negative value indicates that the system will go in the direction of producing the product (spontaneous), for oligos this is double strandedness. Finally, a $\Delta G$ value of zero indicates that the system is in equilibrium and the rate at which the product is produced equals the rate of the reactants being produced. Thus it can be said that the temperature of a system containing two oligos (perfect complementary) that has a $\Delta G$ value of zero is its $T_m$. By rearranging the equation 2 and solve for T results in equation 3 and when $\Delta G$ is zero, equation 3 results in equation 4.

$$T = (\Delta H - \Delta G) / \Delta S \qquad\qquad \text{Equation 3}$$

$$\text{where } \Delta G = 0$$

$$T = \Delta H / \Delta S = T_m \qquad\qquad \text{Equation 4}$$

Melting temperature ($T_m$) represents a system in equilibrium where 50% of the molecular population is in one state and 50% of the population is in the other state. The above equation is the basic form of the $T_m$ equation, ignoring for now the effects of oligo and salt concentrations. In the case of oligos the two states are: single stranded (random coil) and double stranded (duplexed).

## Discussion

While Free Energy ($\Delta G$) and Melting Temperature ($T_m$) share primary components Enthalpy ($\Delta H$) and Entropy ($\Delta S$), $\Delta G$ and $T_m$ values are derived differently and have no correlative relationship. The only way to relate a given $\Delta G$ with a given $T_m$ value is to explicitly know the value of $\Delta H$ and $\Delta S$ from which they are derived.

Figure 1 depicts all the oligos (2,603,379) that have been processed by mFold (25°C & 50mM [Na$^+$ Salt]) for production purposes to date. The oligos are plotted by mFold $T_m$ (0 to

100°C) and mFold $\Delta$G (2 to -41 kcal/mole), horizontal and vertical axis, respectively, with color representing number of oligos. Grey identifies the 2D space where oligos exist down to a single oligo, while yellow and magenta represent counts greater then 1 and 3,500 respectively. The insert is the same image rotated to indicate shape of the distribution in the third dimension of numerical count. What this demonstrates is the fact that any given $\Delta$G can give rise to a very large range of $T_M$ and similarly, any given $T_m$ can give rise to a range of $\Delta$G. Since the default temperature for mFold derivations is 25°C, the range of $\Delta$G at a given $T_m$ shrinks from right to left and $\Delta$G converges to zero when $T_m$ is 25°C. Hence it can be stated that a $\Delta$G value alone cannot represent how 'bad' something might be. Example: a $\Delta$G value of -5.0 kcal/mole relates to a $T_m$ value range of 33°C to 84°C. The isolated region between 70°C and 80°C (lower right) accounts for 4% of the population and appears to be 80% ddRNAi oligos.
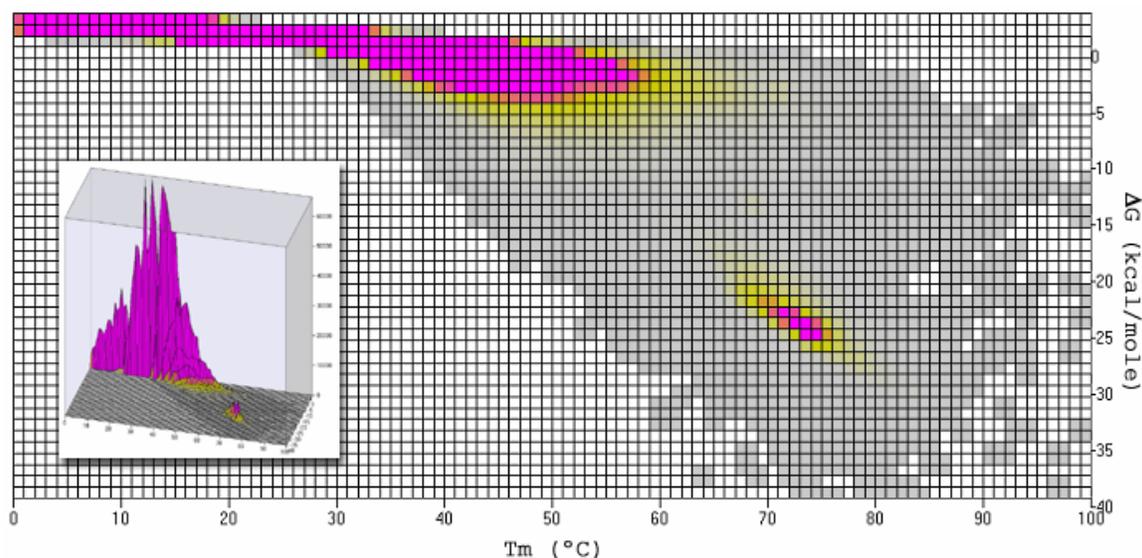


**Figure 1. mFold $T_m$ verses mFold $\Delta$G Energetics**

What the $\Delta$G value can provide is the temperature window size (or range) for which the transformation from a duplexed or structured state to a random coil state takes place. In other words, how many degrees Celsius does it take an oligo population to transition from its 100% structured state to its 100% random coil state? Fortunately, the laws of thermodynamics include more than just describing 50%:50% populations and at a given temperature the equilibrium of a system can be determined. Figure 2 contains three theoretical melting curves derived from the folding energetics of oligos with HT on the
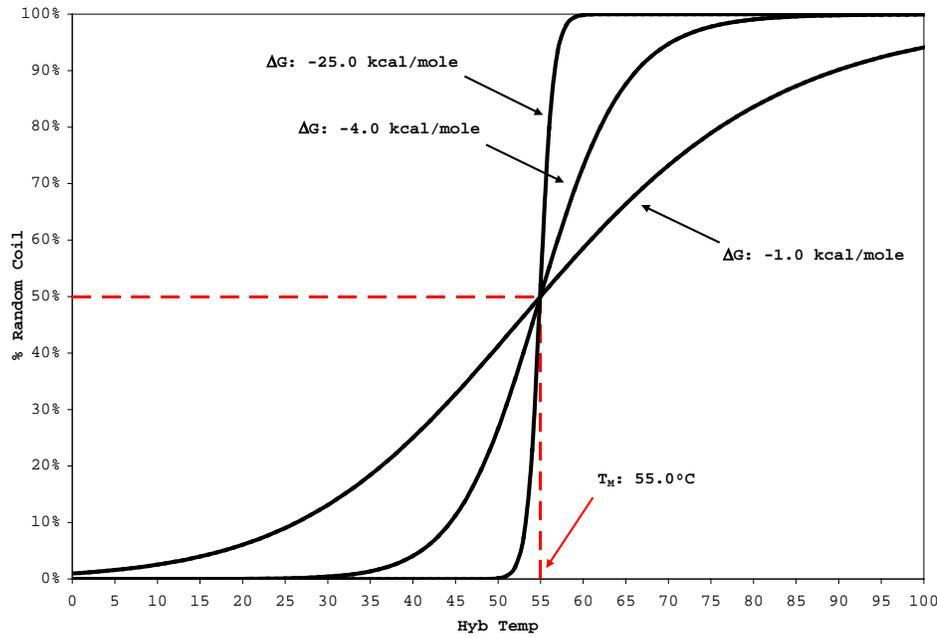
**Figure 2. Fixed mFold T$_m$**

x-axis and % Random Coil (% Single Stranded) on the y-axis. Each black curve depicts, at a given temperature, what the percent of the population is in the single stranded state. By drawing a line from the 50% tick mark, the point where the curve intersects this line determines the mFold T$_m$ of the oligo's secondary structure. In the example, these three oligos have the same mFold T$_m$ at 55.0°C but have very different mFold ΔG value: -1.0, -4.0, and -25.0 kcal/mole. Based on Figure 2, if an experiment using these three oligos not at 55°C but at 60°C, the predictive % Random Coil for each population would not be 50% but 58.6%, 73.1%, and 99.8% for mFold ΔG values -1.0, -4.0, and -25.0 kcal/mole, respectively. Due to the oligo with the mFold ΔG value of -25.0 kcal/mole having such a steep transition curve that it only takes an increase of 5°C to go from 50% duplexed to effectively 0% duplexed.

By fixing the mFold ΔG value and comparing a range of mFold T$_m$ values (Figure 3) the rate of transition is relatively similar at the different mFold T$_m$s, with a slight decrease in the transition rate as the mFold T$_m$ increases. The reason for this drift in transition rate is due to the transition rate being truly related to ΔH (Enthalpy). By taking a melting curve as depicted in the two sets of melt curves (Figures 2 & 3) and performing a van't Hoff plot (ln K$_{eq}$ verses 1/T) on the data, the plot can yield a linear least-squares fit resulting in:

- A straight line of slope    = −ΔH/R    where R is the Gas Constant
- An intercept    = ΔS/R

Even with this drift in the transition rate the correlation of mFold ΔG with respect to the transition rate (van't Hoff plot Slope) is 0.918, where a value of 1.000 is perfect correlation. Hence, mFold ΔG can be used to effectively identify the severity of an oligo's secondary structure transition rate with the ability to always fall back on ΔH to exactly identify the transition rate for quantitative computational needs.
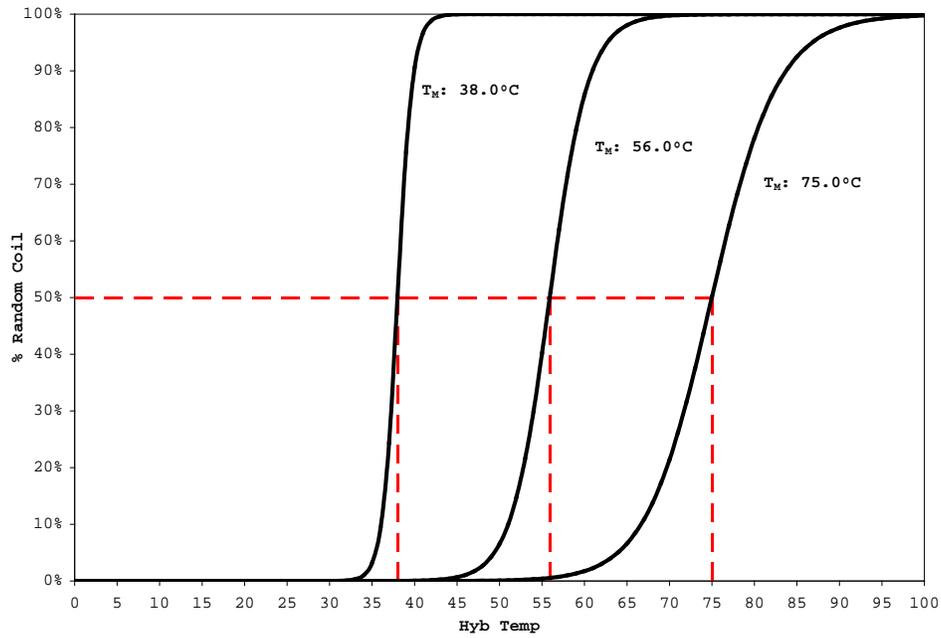
**Figure 3. Fixed mFold ΔG**

For biological purposes scientists have only been interested in an oligo's $T_m$ value to rate the effectiveness to perform at a given hybridization temperature (abbreviated henceforth as HT). In the case of PCR if the $T_m$ of an oligo is much higher than the specified HT then one would say that it would very likely participate in off-target hybridization events (bad), similarly if the specified HT is much greater than the $T_m$ then one would say that the oligo would more than likely not participate at all (bad). Hence one wishes to design an oligo with a $T_m$ very close to the HT.

As more genomes are sequenced more targets are available to interrogate the potential for oligos to participate in off-target interactions. As more work is done on SNPs sites the importance of understanding and maximizing the discrimination between perfect match and its single base mismatch increases. The use of only the $T_m$ value is becoming insufficient as a test, because of the effects of the aforementioned $\Delta G$ value.

Figure 4 depicts the exact-complement duplex thermodynamics (instead of the mFold secondary structure thermodynamics) of all the oligos (14,483,088) that are in the IDT oligo system from November 1999 to date. The conditions were 250 nM [Oligo], 50 mM [$Na^+$ Salt], and $\Delta G$ Temp was 25°C. The oligos are charted by $T_m$ (0 to 100°C) and $\Delta G$ (11 to -170 kcal/mole), vertical and horizontal axis, respectively, with color representing number of oligos. Grey identifies the 2D space where oligos exist down to a single oligo, while yellow and magenta represent counts greater then 1 and 20,000 respectively. The insert is the same image rotated to indicate shape of the distribution in the third dimension.

The value of DG used was not the value that is stored in the system, because it doesn't account for [Oligo] or [$Na^+$ Salt]. The method used to derive the $\Delta G$ value required the derivation of the $T_m$ value, then recalculate $\Delta S$ with $\Delta H$ being constant, and finally using the

new ΔS value along with the ΔH value to derive the new ΔG.  The reason for this recalculation is due to IDT's current salt correction being applied solely to the T$_m$ value and not its components ΔH and ΔS.
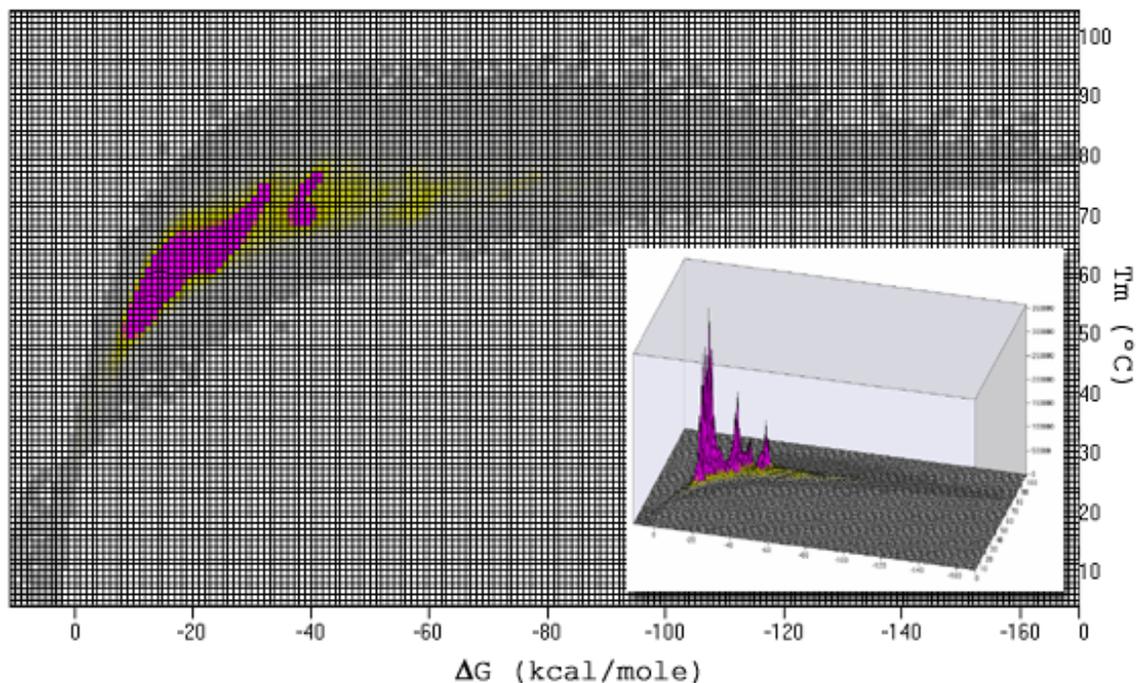


**Figure 4.  Duplex T$_m$ verses Duplex ΔG Energetics**

Again this kind of plot demonstrates that any given ΔG can give rise to a very large range of range of T$_m$ and vice versa.  Since the default temperature for ΔG derivations is 25°C, the range of ΔG at a given T$_m$ shrinks from right to left and ΔG converges to zero when T$_m$ is 25°C.

Figure 5 contains three theoretical melting curves derived from the energetics of three oligos and their exact-compliment with HT on the x-axis and % Random Coil (% Single Stranded) on the y-axis. Each black curve depicts, at a given temperature, what the percent of the population is in the single stranded state.  By drawing a line from the 50% tick mark to the curve and dropping it down to the HT shows the T$_m$ of the oligo bound to its target.  In the example, these three oligos have the same T$_m$ at 55.0°C but have very different mFold
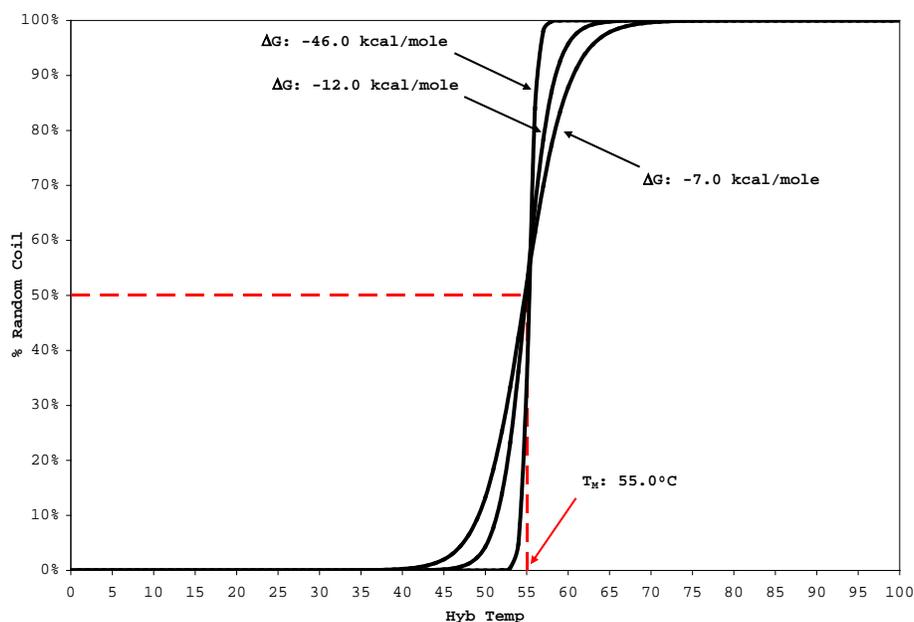
**Figure 5. Fixed Duplex T$_m$**

$\Delta$G values; -7.0, -12.0, and -46.0 kcal/mole. Hence, an experiment that used these three oligos at a hybridization temperature not at 55°C but at 53°C, the predictive % Random Coil for each population would not be 50% but 33.3%, 23.2%, and 0.5% for $\Delta$G values -7.0, -12.0, and -46.0 kcal/mole, respectively. Due to the oligo with the $\Delta$G value of -46.0 kcal/mole having such a steep transition curve that it only takes a decrease of 2°C to go from 50% duplexed to effectively 100% bound to its compliment.

By fixing the $\Delta$G value and comparing a range of T$_m$ values (Figure 6) the rate of transition is again relatively similar at the different melting temperautres, with a slight decrease in the transition rate as the T$_m$ increases. The correlation of $\Delta$G with respect to the transition rate (van't Hoff plot Slope) is again very high, 0.983. Hence, $\Delta$G for the formation and melting of a duplex can also be used to effectively identify the severity of an oligo's transition rate from random coil to duplexed.

## Conclusion

Examined was the relationship between $\Delta$G and T$_m$ for both uni-molecular interactions (mFold) and bi-molecular interactions (Oligo - Target). I showed that $\Delta$G alone cannot be used to properly indicate a level of stickiness for secondary structure and for
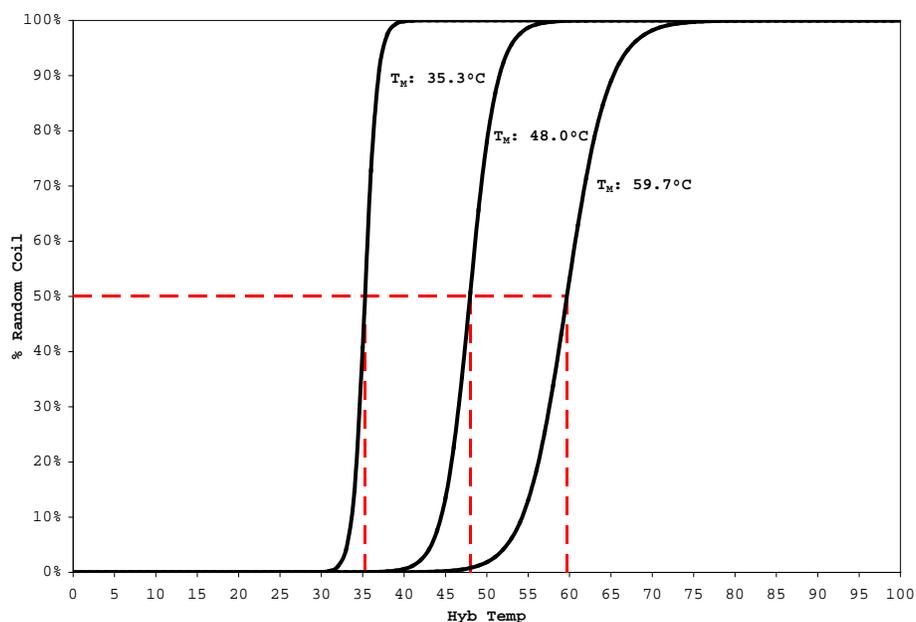
**Fig. 6.  Fixed duplex ΔG**

duplex formation and melting.  It was found that utilizing both ΔG and $T_m$ is important when interrogating the thermodynamic potential of a duplex or structure and that $T_m$ alone isn't truly sufficient.  Depending on the task, PAGE purification, analytical CE, PCR, or other hybridization experiments like microarray will determine a particular $T_m$ and ΔG characteristic to look for.  Generally speaking, for PAGE or CE the steeper the transition (more negative ΔG) the better, locking the oligo in one state or the other, while for PCR and other hybridization experiments the steep transition is undesirable due to the small window of functional effectiveness.  A couple degrees too high and nothing will bind where as a couple degrees too low and off-target interactions can result.

Therefore it is recommended that both $T_m$ and ΔG values together be used to qualitatively assess the tendency to be potentially 'bad' for the formation of secondary structure or undesired duplex formations under given experimental conditions.  Both values should also be used when identifying PCR primer pairs that will behave similarly under a given condition, and for high throughput hybridization assays like Microarrays one needs to balance for thermodynamic equivalence and not just $T_m$ equivalents.

There is also the ability to more accurately quantify the percent of an oligo population that may be in a random coil state or in a duplexed or structured state when assessing how an oligo may behave under a given assay or experimental temperature.  By solving for the % Random Coil state one can identify the population that is potentially single stranded (% Random Coil) and double stranded or structured (1 – % Random Coil).  The following equations are used to derive the % Random Coil value.

$$\% \text{ Random Coil} = 1 / (K_a + 1) \qquad\qquad \text{Equation 5}$$

where $K_a$ is the association constant of a formation and is defined by equation 6,

$$K_a = e^{(-\Delta G / RT)} \qquad\qquad\qquad \text{Equation 6}$$

where $\Delta G$ is the Gibbs Free Energy defined by equation 2, R is the Gas Constant (1.986 kcal/mol), and T is the absolute temperature of the environment (Kelvin). In the case of duplex formation the $\Delta S$ value will already have the effects of Oligo concentration and Salt concentrations accounted for by solving for the adjusted $\Delta S$ from the derived $T_m$ in the desired Oligo and Salt concentrations.