

## Appendix C

# Equivalence between Multistrand’s thermodynamics model and the NUPACK thermodynamics model.

We now return to examine the thermodynamics model presented here (Section 3) and compare it to the NUPACK thermodynamics model [5]. As we noted previously, our energy model differs in exactly two ways: the lack of any symmetry terms in the energy of a complex, and the addition of the  $\Delta G_{volume}$  term. These, however, are merely natural consequences of a much more basic difference: the Multistrand energy model deals with systems of uniquely labelled strands in a stochastic regime, while the NUPACK energy model allows populations of strand species and operates in a mass action regime.

In this section we will examine these two models in greater detail. We introduce a particular class of macrostate which we call an “indistinguishability” macrostate, i.e. a way of grouping Multistrand’s system microstates (which have all strands uniquely identified) into macrostates that are equivalent to the states on which the NUPACK thermodynamics model is defined. We then show that using these macrostates our models are mathematically equivalent, that is to say, they predict the same partition function over these macrostates as well as the same probability of observing each macrostate.

In order to build up to this proof of equivalence, we will need to introduce some extra framework detailing the differences between the two models and discussing the main concepts necessary (such as the “indistinguishability” macrostates). To make this discussion easier, we note the following conventions: whenever we refer to a system microstate, we always mean a Multistrand system microstate, defined in section 2.3. Once we have defined indistinguishability macrostates, we will always refer to them as simply macrostates

(or perhaps ind. macrostates), with the understanding that they are a specific class of macrostates, rather than a general definition such as in section 6.2.

## C.1 Population Vectors

We previously introduced the set of (uniquely labelled) strands  $\Psi^*$ , but since we are now allowing duplicate strands we need a few more terms. Thus we have the set of strand species  $\Psi^0$ , which we can think of as the set of all the types of strands: one entry for each unique strand in the system, and we can then have multiple copies of that species in our system. We also have the set of strand complexes  $\Psi$ , where a strand complex is defined as some number of each strand species in a connected secondary structure. This is just the combinations of strand species we could have formed into a complex. Note that for a system with a finite number of strands, the set of possible strand complexes is also finite, though it can be quite large. Also,  $\Psi^0 \subset \Psi$ , and by convention two distinct species may not have the same sequence or label.

We can now think of an abstract representation of the current system's state: a population vector  $m \in \mathbb{N}^{|\Psi|}$ , where we note  $\mathbb{N} = \mathbb{Z}_{\geq 0}$ , which has an entry for each type of strand complex, indicating the number of complexes of that type in the current system's state. The initial population vector  $m^0 \in \mathbb{N}^{|\Psi^0|}$  indicates how many of each type of strand are present in the system. We relate the two by the strand matrix  $A \in \mathbb{N}^{|\Psi^0| \times |\Psi|}$ , whose entries  $A_{ij}$  correspond to the number of strands of species  $i$  in complex type  $j$ . By our previous definition of strand complex, the rows of  $A$  are distinct. Thus, if we have a system which starts with  $m^0$  of each strand species,  $\Lambda = \{m \mid Am = m^0\}$  is the set of all population vectors consistent with conservation of strand counts.

## C.2 Indistinguishability Macrostates

We now need to introduce a specific class of macrostate which we call **indistinguishability** macrostates, or **ind.** macrostates for short. The idea behind these is that in Multistrand, we assume every strand is uniquely labelled, but if we were given a set of strand species

and a mapping from our unique labels to a species, we then could partition our existing system microstates into groups representing those which would be indistinguishable when considering the strand species.

First, recall that every complex microstate  $c$  contained an ordering  $\pi^*(c)$  on the strand ids. While this wasn't used very much in our original discussion, we now need to define this further: the ordering  $\pi^*(c)$  is a *circular permutation* on the strand ids of  $c$  (which are  $ST(c)$ ). We note that *circular permutations* are the permutations of the objects which are distinct when they are laid out on a fixed circle without a defined starting point. E.g. for strands labelled 1, 2, 3 there are only two circular permutations on the strand ids: (1, 2, 3) and (1, 3, 2). As it will be used later, a *cyclic permutation* is a permutation which does NOT lead to a distinct ordering when laid out on a fixed circle – that is, they “rotate” the strands without rearranging their relative order. E.g. for a strand ordering (1, 2, 3), there are three cyclic permutations of that strand ordering: (1, 2, 3), (2, 3, 1) and (3, 1, 2). For a given set of  $n$  elements, there are  $(n - 1)!$  circular permutations and  $n$  cyclic permutations of the elements.

So, every complex microstate  $c$  contains an ordering  $\pi^*(c)$  that is a circular permutation on the strand ids. We define the set of *indistinguishable* orderings of  $c$  as the circular permutations of  $c$ 's strand labels (labels represent the strand species). For example, with a complex of 4 strands, where strand id's 1 and 2 have label A, and strand ids 3 and 4 have label B, the circular permutations on strand ids are  $\{(1, 2, 3, 4), (1, 2, 4, 3), (1, 3, 2, 4), (1, 3, 4, 2), (1, 4, 2, 3), (1, 4, 3, 2)\}$ , and the indistinguishable orderings are  $\{(A, A, B, B), (A, B, A, B)\}$ . Thus, we can think of a complex microstate as having the ordering  $\pi^*(c)$  from the circular permutations on the strand ids, but also having an ordering  $\pi(c)$  from the circular permutations on the strand labels which represents the indistinguishable ordering for the complex if strands with the same sequence are considered indistinguishable. This  $\pi(c)$  tells us which complex ind. macrostate this complex microstate is a member of.

We can now define the concept of two complex states being indistinguishable, and thus approach a definition for a macrostate of the system.

**Definition:**

The complex microstates  $c, c'$  are indistinguishable if the following properties hold:

1. The indistinguishable orderings  $\pi(c), \pi(c')$  correspond to the same indistinguishable ordering  $\pi_{eq}$ . Note that this is the case iff  $\pi(c)$  is a cyclic permutation of  $\pi(c')$ .
2. There exists a one-to-one mapping  $\xi$  between the strand ids of  $c$  and the strand ids of  $c'$  such that for every base pair  $(i_j \cdot k_l)$  in  $c$ , there is a base pair  $(i_{\xi(j)} \cdot k_{\xi(l)})$  in  $c'$ , and for every base pair  $(i_{\xi(j)} \cdot k_{\xi(l)})$  in  $c'$  there is a base pair  $(i_j \cdot k_l)$  in  $c$ . In other words, the mapping on strand ids induces a one-to-one mapping between the base pairs of each complex. Note that if  $\xi$  exists, it is not necessarily a unique mapping, with the cases where it is not corresponding to symmetric complexes, examined below.  $\square$

From this definition it is easy to see when two system microstates are indistinguishable: System microstates  $s, s'$  are indistinguishable if there exists a one-to-one mapping between the complex microstates of each system microstate, such that complex microstates map onto indistinguishable complex microstates.

We now put these all together to define an indistinguishability macrostate of the system (or, for want of a clunky name, a system ind. macrostate). A system ind. macrostate  $v$  is a set of system microstates from  $S$  (the set of all system microstates), such that for every pair of microstates  $s, s' \in v$  with  $s \neq s'$ ,  $s$  and  $s'$  are indistinguishable, and for any system microstate  $s'' \notin v$ ,  $s''$  is distinguishable from any system microstate  $s \in v$ . Note that in many cases when talking about ind. macrostates, we will use an alternate form, in which the ind. macrostate is given by a particular system microstate representative from the set. This is convenient for energy functions in particular, as every system microstate in an ind. macrostate has the same energy, and it is sometimes useful to know what a (microstate) energy function would give for a macrostate's energy. Note that this is **different** than the ind. macrostate's energy, as that is given by equation 6.2 from section 6.2.

We also have a notion of an ind. macrostate of a complex (a complex ind. macrostate, as used previously). This is defined similarly - a complex ind. macrostate  $v'$  is a set of

complex microstates from  $C$  (the set of all complex microstates for a particular complex), such that for every pair of complex states  $c, c' \in v'$  with  $c \neq c'$ ,  $c$  and  $c'$  are indistinguishable, and for any microstates  $c'' \notin v', c \in v'$ ,  $c''$  is distinguishable from  $c$ . Finally, a complex ind. macrostate has an additional property which will be used in energy functions, which is the symmetry factor. The symmetry factor of a complex ind. macrostate  $c$  (where we now use the representative complex microstate),  $R(c)$  is the number of cyclic permutations on the strand ids which map the base pairs of a representative complex state onto themselves. That is to say, it is the number of cyclic permutations  $\xi'$  such that for every base pair  $(i_j \cdot k_l)$  in  $s$ ,  $(i_{\xi'(j)} \cdot k_{\xi'(l)})$  is also a base pair in  $s$ .

In the upcoming sections we will examine the size of complex ind. macrostates, as well as energy functions which apply to complex microstates, system microstates, and ind. macrostates of both types. Finally, when talking about a single complex type  $j \in \Psi$ , we will occasionally refer to the set of all indistinguishable orderings of the strands,  $\Pi_j$ , as well as the complex ind. macrostates in a given indistinguishable permutation  $\pi \in \Pi_j$ , which is  $\Omega_j(\pi)$ .

### C.3 Macrostate Energy

Recall that we can calculate the probability of a specific system microstate  $s$  using the energy  $\Delta G_{box}^*(s)$  and the partition function  $Q_{kin} = \sum_i e^{-\Delta G_{box}^*(i)/RT}$  as  $Pr(s) = \frac{1}{Q_{kin}} e^{-\Delta G_{box}^*(s)/RT}$ . Macrostate probabilities can be calculated in the same manner, using the energy function for an ind. macrostate and the appropriate ind. macrostate partition function (or by summing the probabilities of the constituent microstates).

Recall that we previously gave the standard free energy of a complex microstate  $c$ , containing  $L$  strands as:

$$\overline{\Delta G}(c) = \left( \sum_{loop \in c} \Delta G(loop) \right) + (L - 1) \Delta G_{assoc}$$

And for our kinetic system, we use the following, which includes the volume-dependent

term in the complex's energy.

Energy of a complex microstate  $c$ , containing  $L$  strands:

$$\begin{aligned}\Delta G^*(c) &= \left( \sum_{loop \in c} \Delta G(loop) \right) + (L - 1) * (\Delta G_{assoc} + \Delta G_{volume}) \\ &= \overline{\Delta G}(c) + (L - 1)\Delta G_{volume}\end{aligned}$$

And combined, the energy of a system microstate  $s$  is:

$$\Delta G_{box}^*(s) = \sum_{\text{complex microstate } c \in s} \Delta G^*(c)$$

The volume term  $\Delta G_{volume}$  for the remainder of this discussion will be the quantity  $RT \log M_s$ , where  $M_s$  is the number of solvent molecules in the box. This will be a useful comparison later, as the NUPACK partition function uses the same terms.

Lastly, the energy of a complex ind. macrostate  $c$ , containing  $L$  strands with a symmetry factor  $R(c)$  as computed directly using the NUPACK energy model [5], is:

$$\Delta G(c) = \left( \sum_{loop \in c} \Delta G(loop) \right) + (L - 1)\Delta G_{assoc} + RT \log R(c) \quad (\text{C.1})$$

## C.4 Partition Function

We now wish to compare the partition function over our system microstates with the partition function  $Q_{box}$  examined in [5], which is over (our) system ind. macrostates.

First it is important to know what system we are working with. The general system is that of a box containing initial strands according to the labelled population vector  $m^0$ , thus we are considering a system with possibly indistinguishable strands. First, we define a system microstate as being consistent with a population vector  $m$  (this is a population vector over labelled complex types) if it contains the correct number of complexes of each type in the population vector, if we consider strands indistinguishably (according to strand species). Then, note that every individual system microstate must be consistent with some population vector  $m$  which satisfies  $Am = m^0$ . Let  $S$  be the set of all system microstates

which are possible with initial strands  $m^0$ , then we can use  $\Lambda = \{m \mid Am = m^0\}$  and let  $S(m)$  be the set of all microstates which have population vector  $m$ .

We can then write the system microstate partition function as:

$$Q_{kin} = \sum_{s \in S} e^{\Delta G_{box}^*(s)/RT}$$

And rewriting in terms of  $\Lambda$  as:

$$\begin{aligned} Q_{kin} &= \sum_{m \in \Lambda} \sum_{s \in S(m)} e^{\Delta G_{box}^*(s)/RT} \\ &= \sum_{m \in \Lambda} q_{kin}(m) \end{aligned}$$

Where we define  $q_{kin}(m)$  as  $\sum_{s \in S(m)} e^{\Delta G_{box}^*(s)/RT}$ .

Next we wish to examine the system ind. macrostate partition function  $Q_{box}$ , which considers the system ind. macrostates with initial population vector  $m^0$ . We will explain the components of  $Q_{box}$  in detail, as presented in [5]:

$$Q_{box} \approx Q_{ref} \sum_{m \in \Lambda} q(m)$$

This is the basic equation<sup>1</sup>.  $Q_{ref}$  is the reference state of the box, which includes terms related to the volume and the reference state  $s$  for which  $\Delta G_{box}(s) = 0$ , as will be discussed later.

$$q(m) \equiv \prod_{j \in \Psi} \frac{M_s^{m_j} Q_j^{m_j}}{m_j!} \quad (\text{C.2})$$

This is the partition function over the macrostates of a particular population vector  $m$ , and it is the product of the partition functions of the individual complex types ( $Q_j$ 's), corrected for solvent (the  $M_s^{m_j}$  term) and symmetry (the  $m_j!$  term, as well as the particular  $Q_j$  used).

---

<sup>1</sup>We note that this approximation is due to an assumption that the number of solvent molecules in the box is much greater than the number of strands.

$$Q_j = \sum_{\pi \in \Pi_j} Q_j(\pi)$$

This is the partition function over all the macrostates (of a complex) for a particular type of complex  $j$ . Since each type of complex could have many different orderings of the strands, this is a sum over all the indistinguishable orderings for that complex type ( $\Pi_j$ ), of the partition function for that type of complex for a particular indistinguishable ordering  $Q_j(\pi)$ .

$$Q_j(\pi) = \sum_{c \in \Omega_j(\pi)} \exp(-\Delta G(c)/RT)$$

Finally, this is the partition function over all the macrostates of a complex of type  $j$ , in indistinguishable ordering  $\pi$ , and it is just the sum over the symmetry corrected energy for each macrostate.

## C.5 Proof of equivalence between Multistrand’s partition function and the NUPACK partition function

### C.5.1 Proof Outline

Our goal now is to show the equivalence between the partition function over system microstates,  $Q_{kin}$  and that for system ind. macrostates  $Q_{box}$ . While we would like to also show that the probability of observing a given system ind. macrostate is equivalent between the two models, it is beyond the scope of this already quite long proof. However, the same general structure used for the partition function proof can be used to show the equivalence of the probability between the two models. The two key insights for the probability proof are to count the number of system microstates composing a particular ind. macrostate, and to derive from [5] the probability of a specific ind. macrostate as in that work the smallest component considered in the “box” partition function is the probability distribution for a particular population vector  $m$ .



It is helpful to understand what direction we take to prove the equivalence between these two formulations of the partition function. We first examine entire systems where there is only a single complex of a particular type and the complex is always connected throughout time. This allows us to work out the microstate partition function for a very small case where it is easy for us to count the number of microstates in each macrostate. This also leads to examining the difference in energy functions used, and explains why we match the symmetry corrections even though the microstate energy function doesn't include any symmetry terms.

Once we have established the very small case, we can examine a larger system, where we have any number of complexes of a single type, and that number doesn't change. This is a building block step towards being able to write out the microstate partition function for any population vector  $m$ , at which point we can use the  $Q_{kin}$  equation above to handle any system (recall that individual states in a macrostate must always be within the same population vector).

The key here is to remember that in each of these three steps, the **system** we are dealing with is a different one, whose microstates are those possible given a particular choice of population vector  $m$ .

### C.5.2 System with a single complex

Consider a kinetic system whose states are entirely in the population vector  $m : m_j = 1, m_i = 0 \forall i \neq j$ . This corresponds to a system where there is a single complex, which is never allowed to disassociate. Let  $S$  be the set of microstates of the system (which satisfy population vector  $m$ ). This is equivalently the  $S(m)$  discussed before, but we simplify the notation here as we are always working with a single  $m$ . We then have:

$$q_{kin}(m) = \sum_{s \in S} e^{-\Delta G_{box}^*(s)/RT}$$

Note that for this simple system, the microstates  $s$  are each a set containing a single

corresponding complex state  $c$ . It is interesting to note that for this system, since we have restricted the complex to always be connected, the system is not necessarily connected by kinetic moves in the state space  $S$ . Thus  $\Delta G_{box}^*(s) = \Delta G^*(c)$  for these microstates. Note that they can still be from different indistinguishable permutations of the strands, so we group them by the indistinguishable permutations  $\Pi$ , letting  $S(\pi)$  be the set of microstates which are consistent with the indistinguishable ordering  $\pi$ :

$$q_{kin}(m) = \sum_{\pi \in \Pi} \sum_{\substack{s \in S(\pi) \\ s = \{c\}}} e^{-\Delta G^*(c)/RT}$$

We now consider the macrostates of the complex  $j$ ,  $s' \in \Omega_j(\pi)$ . We wish to know how many microstates  $s \in S(\pi)$  are contained in the macrostate  $s'$  (or put differently, indistinguishable from the macrostate  $s'$ , if we treat the macrostate as a representative of the set). Let  $R(s')$  be the symmetry factor of the macrostate. First, we note that for  $R(s') = 1$ , any permutation on the indistinguishable strands will lead to a unique microstate that is indistinguishable from  $s'$ . Thus there are  $\prod_{i \in \Psi^0} (Am)_i!$  microstates corresponding to each non-symmetric macrostate  $s'$ . (Note that they must all have the same energy.) In symmetric states, some of these permutations (the circular ones with the appropriate symmetry) lead to equivalent microstates, and thus the total number of microstates corresponding to a symmetric macrostate is lower, by the symmetry factor  $R$ . Thus for either case we have the general formula of  $\frac{1}{R(s')} \prod_{i \in \Psi^0} (Am)_i!$  microstates for each macrostate  $s'$ .

We then rewrite  $q_{kin}$  in terms of the macrostates and  $L$ , the total number of single strands used in the system ( $L(m) = \sum_{i \in \Psi^0} (Am)_i$ ):

$$\begin{aligned}
q_{kin}(m) &= \sum_{\pi \in \Pi} \sum_{\substack{s' \in \Omega(\pi) \\ s' = \{c'\}}} \left( \frac{1}{R(s')} \prod_{i \in \Psi^0} (Am)_i! \right) e^{-\Delta G^*(c')/RT} \\
&= \sum_{\pi \in \Pi} \sum_{\substack{s' \in \Omega(\pi) \\ s' = \{c'\}}} \left( \prod_{i \in \Psi^0} (Am)_i! \right) e^{-\overline{\Delta G}(c')/RT - (L(m)-1) \log(M_s) - \log(R(c'))} \\
&= \sum_{\pi \in \Pi} \sum_{\substack{s' \in \Omega(\pi) \\ s' = \{c'\}}} \left( \prod_{i \in \Psi^0} (Am)_i! \right) \frac{M_s}{M_s^{L(m)}} e^{-\Delta G(c')/RT} \\
&= \left( \prod_{i \in \Psi^0} (Am)_i! \right) \frac{M_s}{M_s^{L(m)}} Q_j
\end{aligned}$$

And so we have shown that the case of a single complex comes out to exactly the ind. macrostate partition function  $q(m)$  (Equation C.2) of the complex, with an extra factor of  $(\prod_{i \in \Psi^0} (Am)_i!) \frac{1}{M_s^{L(m)}}$  due to reference state and volume.

### C.5.3 System with multiple copies of a single complex type

We now build on this result by using it as the base case for induction to cover a system composed entirely of complexes of a single type. In particular, we are examining a system where  $m : m_j = n, m_i = 0 \forall i \neq j$ , and our inductive hypothesis is:

$$q_{kin}(m) = \left( \prod_{i \in \Psi^0} (Am)_i! \right) \frac{M_s^{m_j} Q_j^{m_j}}{M_s^L m_j!}$$

We have shown the base case ( $n = 1$ ), so now we need to show the inductive step. Again, we define  $S$  as the set of microstates consistent with  $m$ , and our normal definition of  $q_{kin}$  is:

$$q_{kin}(m) = \sum_{s \in S} e^{-\Delta G_{box}^*(s)/RT}$$

To cover this case we need to look at an expanded set of states, where the difference is that we add a “marking” to some part of the microstate (complexes).

We expand the system as follows:

$$\begin{aligned} S' &= \bigcup_{s \in S} \text{Mark}(s) \\ \text{Mark}(s) &= \{(s, c) | c \in s\} \end{aligned}$$

Consider now  $q_{exp}(m) = \sum_{(s,c) \in S'} e^{-\Delta G_{box}^*(s)/RT}$ .  $\text{Mark}(s)$  expands every original state into  $m_j$  new states, all of which have the same energy for the system microstate component. Thus we know that  $q_{kin}(m) = \frac{1}{m_j} q_{exp}(m)$ .

We now use the markings to group microstates from  $S'$ . Consider the following grouping of microstates, based on a complex microstate  $c$  (note that  $c$  by definition includes a specific set of strand ids used in the complex):

$$F(c) = \{s | (s, c) \in S'\}$$

Thus  $F(c)$  is all system microstates in  $S'$  that contain a complex microstate  $c$  (with specific set of strand ids). Now consider the sum over all the microstates within one grouping:

$$\sum_{s \in F(c)} e^{-\Delta G_{box}^*(s)/RT} = e^{-\Delta G^*(c)/RT} * \sum_{s \in F(c)} e^{-\Delta G^*(s \setminus c)/RT}$$

where we let  $s \setminus c$  represent the microstate that would result if the marked complex microstate  $c$  were removed from the system microstate  $s$ . Looking at the resulting summation, note that it is over all the microstates for a system with one less complex of type  $j$ ! To see why it is exhaustive, note that if it were not, the original system would be missing valid microstates consistent with  $m$ , similarly, it cannot overcount any states, as that would imply duplicated states in the original  $S$ . Thus we can use our inductive hypothesis to express the result:

$$\sum_{s \in F(c)} e^{-\Delta G_{box}^*(s)/RT} = e^{-\Delta G^*(c)/RT} * q_{kin}(m')$$

Where  $m' : m'_j = m_j - 1, m'_i = m_i = 0 \forall i \neq j$  is our smaller system. Note that the choice of strand ids only matters for the complex in state  $c$ , not for the rest, as in general we can use any set of strand ids corresponding to the correct numbers of strand types. Let us expand on this idea to break down our base  $q_{exp}$  in terms of  $F(c)$  and combinations of strand ids. We introduce a new set  $E'$ , which will be the set of all valid combinations of strand id's for a complex of type  $j$  ( $j$  is a labeled complex). It is helpful to note the size of this set, which will be the number of ways for each strand species, to select the appropriate number of strand ids of that species for a single complex of type  $j$ . This is, in two different forms:  $|E'| = \prod_{i \in \Psi^0} \binom{m_i^0}{A_{ij}} = \prod_{i \in \Psi^0} \binom{(Am)_i}{(Am')_i}$

For example, if our complex of type  $j$  is two A strands, and we have 4 A strands with ids  $\{1, 2, 3, 4\}$ , then  $E' = \{\{1, 2\}, \{1, 3\}, \{1, 4\}, \{2, 3\}, \{2, 4\}, \{3, 4\}\}$  are the valid sets of strand ids. Finally, let  $U(E)$  be the set of all microstates for a single complex of type  $j$  which use only the set of strand id's  $E$ . We can then rewrite  $q_{exp}$  in terms of these, noting again that we are not overcounting any states in  $S'$  nor leaving any out (for the same reasons as before, if either were the case it would indicate a defect in the original  $S$ ):

$$\begin{aligned} q_{exp}(m) &= \sum_{E \in E'} \sum_{c \in U(E)} \sum_{s \in F(c)} e^{-\Delta G_{box}^*(s)/RT} \\ &= \sum_{E \in E'} \sum_{c \in U(E)} e^{-\Delta G^*(c)/RT} q_{kin}(m') \end{aligned}$$

Note that  $m'$  is defined as before:  $m' : m'_j = m_j - 1, m'_i = m_i \forall i \neq j$ . Finally, note that the sum over  $U(E)$  is just the sum over all microstates of a single complex, and is thus equivalent to our base case!

$$q_{exp}(m) = \sum_{E \in E'} q_{kin}(m'') * q_{kin}(m')$$

where  $m'$  is as before, and  $m'' : m''_j = 1, m''_i = 0 \forall i \neq j$ . Now we see that the  $q_{kin}$  in the sum don't depend on  $E'$ , and thus we have:

$$q_{exp}(m) = |E'| * q_{kin}(m'') * q_{kin}(m')$$

where  $|E'|$  is the number of ways to choose strand ids for one complex out of the strand ids for the whole system:  $|E'| = \prod_{i \in \Psi^0} \binom{(Am)_i}{(Am'')_i} = \prod_{i \in \Psi^0} \frac{(Am)_i!}{(Am'')_i! * (Am')_i!}$ . Putting it all together:

$$\begin{aligned}
q_{exp}(m) &= \left( \prod_{i \in \Psi^0} \frac{(Am)_i!}{(Am'')_i! * (Am')_i!} \right) * \left( \prod_{i \in \Psi^0} (Am'')_i! \right) \frac{M_s}{M_s^{L(m'')}} Q_j * q_{kin}(m') \\
&= \left( \prod_{i \in \Psi^0} \frac{(Am)_i!}{(Am')_i!} \right) * \frac{M_s}{M_s^{L(m'')}} Q_j * q_{kin}(m') \\
&= \left( \prod_{i \in \Psi^0} \frac{(Am)_i!}{(Am')_i!} \right) * \frac{M_s}{M_s^{L(m'')}} Q_j * \left( \prod_{i \in \Psi^0} (Am')_i! \right) \frac{M_s^{m_j-1}}{M_s^{L(m')}} \frac{Q_j^{m_j-1}}{(m_j-1)!} \\
&= \left( \prod_{i \in \Psi^0} (Am)_i! \right) * \frac{M_s}{M_s^{L(m'')}} Q_j * \frac{M_s^{m_j-1}}{M_s^{L(m')}} \frac{Q_j^{m_j-1}}{(m_j-1)!} \\
&= \left( \prod_{i \in \Psi^0} (Am)_i! \right) * \frac{M_s^{m_j}}{M_s^{L(m)}} \frac{Q_j^{m_j}}{(m_j-1)!} \\
q_{kin}(m) &= \frac{1}{m_j} * q_{exp}(m) \\
q_{kin}(m) &= \left( \prod_{i \in \Psi^0} (Am)_i! \right) * \frac{M_s^{m_j}}{M_s^{L(m)}} \frac{Q_j^{m_j}}{m_j!}
\end{aligned}$$

And the inductive step holds. Note again that  $L(m) = \sum_{i \in \Psi^0} (Am)_i$ , which is the total number of strands in the system with population vector  $m$ .

#### C.5.4 Full System

We now have the components necessary to give a formula for  $q_{kin}(m)$  for a general  $m$ , which we can prove via induction. We wish to show for a general  $m$ :

$$q_{kin}(m) = \left( \prod_{i \in \Psi^0} (Am)_i! \right) * \frac{1}{M_s^{L(m)}} * \prod_{j \in \Psi} \frac{M_s^{m_j} Q_j^{m_j}}{m_j!}$$

We proceed via induction on the number of non-zero entries in  $m$ , and have already shown the base case for a single non-zero entry in  $m$ , above. In order to break down the

general case into smaller components so that we can apply the inductive hypothesis, we wish to follow a process similar to the previous marking, in which we separate out a particular component of the system microstate. This component  $t$  is the system microstate corresponding to a system with population vector  $m'$ , which has  $m'_j = m_j > 0, m'_i = 0 \forall i \neq j$ , using a particular set of strand ids. Thus we can define a 'marked' set of states  $S'$  as before, using  $T(m')$  as the set of all system microstates corresponding to a system with population vector  $m'$  and an arbitrary choice of strand ids:

$$S' = \{(s, t) | s \in S, t \in T(m'), s.t. \forall c \in t, \text{if } t \subseteq s \text{ then } c \in s\}$$

Note that we must pick a  $j$  such that  $m_j > 0$ , and that this is an arbitrary choice which remains for the rest of this step ( $j$  never changes, we just use it to pick a complex type which appears in  $m$ .) We can now group the system microstates as follows:

$$F(t) = \{s \in S | (s, t) \in S'\}$$

Again, if we sum over the states in each grouping, we get:

$$\begin{aligned} \sum_{s \in F(t)} e^{-\Delta G_{box}^*(s)/RT} &= \sum_{s \in F(t)} e^{-\Delta G_{box}^*(t)/RT} * e^{-\Delta G_{box}^*(s \setminus t)/RT} \\ &= e^{-\Delta G_{box}^*(t)/RT} \sum_{s \in F(t)} e^{-\Delta G_{box}^*(s \setminus t)/RT} \\ &= e^{-\Delta G_{box}^*(t)/RT} q_{kin}(m'') \\ &= e^{-\Delta G_{box}^*(t)/RT} q_{kin}(m - m') \end{aligned}$$

where  $m'' : m''_j = 0, m''_i = m_i \forall i \neq j$ . We proceed exactly as before, letting  $E'$  be the set of all combinations of strand ids which could be used in the complexes of type  $j$ , and  $U(E)$  the set of all microstates  $T(m')$  which use the set of strand ids  $E$ . Note that the following decomposition doesn't involve a  $q_{exp}$ , because in this marking we didn't expand our system at all (each  $s \in S$  corresponds to exactly one  $(s, t) \in S'$ ). It must also be exhaustive and non-redundant for the same reasons as before, if it were otherwise it would imply either a

missing system microstate in  $S$ , or the duplication of some microstate in  $S$ . We then get:

$$\begin{aligned}
q_{kin}(m) &= \sum_{E \in E'} \sum_{t \in U(E)} \sum_{s \in F(t)} e^{-\Delta G_{box}^*(s)/RT} \\
&= q_{kin}(m'') * \sum_{E \in E'} \sum_{t \in U(E)} e^{-\Delta G_{box}^*(t)/RT} \\
&= q_{kin}(m'') * q_{kin}(m') * \sum_{E \in E'} 1 \\
&= q_{kin}(m'') * q_{kin}(m') * |E'|
\end{aligned}$$

where  $m'$  is as stated above:  $m' : m'_j = m_j, m'_j = 0 \forall i \neq j$ . Note that the second step is possible because  $q_{kin}(m'')$  is dependent on the choice of  $j$ , but not  $E$  or  $t$ , and similarly for the third step. These terms are again the base case for the induction ( $q_{kin}(m')$ ), the inductive hypothesis ( $q_{kin}(m'')$ ) and the number of ways to pick the strand ids for each component. We continue, substituting for  $|E'| = \prod_{i \in \Psi^0} \frac{(Am)_i!}{(Am'')_i! * (Am')_i!}$ . Putting it all together:

$$\begin{aligned}
q_{kin}(m) &= q_{kin}(m'') * q_{kin}(m') * \prod_{i \in \Psi^0} \frac{(Am)_i!}{(Am'')_i! * (Am')_i!} \\
&= \left( \prod_{i \in \Psi^0} \frac{(Am)_i!}{(Am'')_i! * (Am')_i!} \right) * \left( \prod_{i \in \Psi^0} (Am')_i! \right) * \frac{M_s^{m_j}}{M_s^{L(m')}} \frac{Q_j^{m_j}}{m_j!} * q_{kin}(m'') \\
&= \left( \prod_{i \in \Psi^0} \frac{(Am)_i!}{(Am'')_i!} \right) * \frac{M_s^{m_j}}{M_s^{L(m')}} \frac{Q_j^{m_j}}{m_j!} * q_{kin}(m'') \\
&= \left( \prod_{i \in \Psi^0} \frac{(Am)_i!}{(Am'')_i!} \right) * \frac{M_s^{m_j}}{M_s^{L(m')}} \frac{Q_j^{m_j}}{m_j!} * \left( \prod_{i \in \Psi^0} (Am'')_i! \right) * \frac{1}{M_s^{L(m'')}} * \prod_{k \in \Psi} \frac{M_s^{m'_k} Q_k^{m'_k}}{m'_k!} \\
&= \left( \prod_{i \in \Psi^0} (Am)_i! \right) * \frac{M_s^{m_j}}{M_s^{L(m')}} \frac{Q_j^{m_j}}{m_j!} * \frac{1}{M_s^{L(m'')}} * \prod_{k \in \Psi} \frac{M_s^{m'_k} Q_k^{m'_k}}{m'_k!} \\
&= \left( \prod_{i \in \Psi^0} (Am)_i! \right) * \frac{1}{M_s^{L(m')}} * \prod_{k \in \Psi} \frac{M_s^{m_k} Q_k^{m_k}}{m_k!}
\end{aligned}$$



And the inductive step holds. We conclude that with  $Q_{ref} = \prod_{i \in \Psi^0} (Am)_i! * \frac{1}{M_s^L}$  (which is independent of  $m$  for  $m \in \Lambda$ ), we have:

$$Q_{kin} = \sum_{m \in \Lambda} q_{kin}(m) = Q_{ref} * \sum_{m \in \Lambda} \prod_{j \in \Psi} \frac{M_s^{m_j} Q_j^{m_j}}{m_j!} = Q_{ref} \sum_{m \in \Lambda} q(m) = Q_{box}$$

Unsurprisingly, [5] uses the same reference state as the our model, where every strand is a separate complex with no internal base pairings. ( $\Delta G_{box}^*(nobp) = 0$ ), so it is unsurprising that the  $Q_{ref}$  is exactly the same for both, and thus we have shown agreement between Multistrand's microstate partition function formulation and the NUPACK partition function formulation.