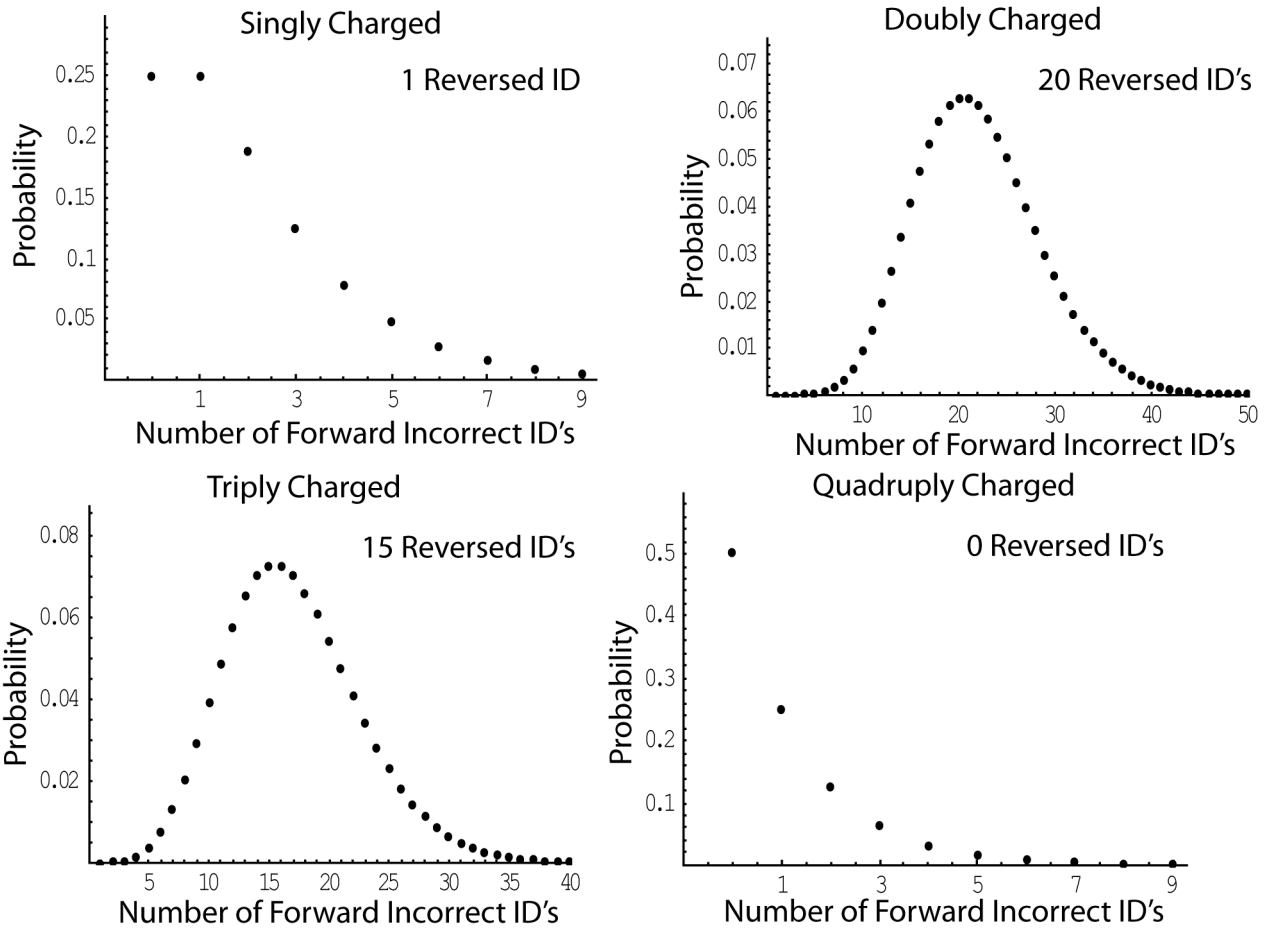
CALCULATIONS FOR COMBINED CHARGE STATES
(See Table 2)

Commonly when a reversed database strategy is employed to determine false positive rates in a large proteomics dataset, peptides are often separated by charge state. After setting appropriate false positive thresholds, the populations are then recombined and grouped by parent protein. To define the error in false positive rate associated with this approach, we need to determine the numbers of incorrect peptide identifications that are likely to be present in the combined dataset. Listed below is an explanation of how these calculations were done.

1. Use the model presented in this paper to calculate the probability distributions representing the numbers of forward incorrect peptide identifications given the number of reversed peptide identifications found for each charge state.



2. Combine distributions from singly and doubly charged distributions to obtain a composite distribution representing the probabilities associated with all numbers of incorrect forward peptide identifications. This is done through multiplication of the two distributions.

   Single charge distribution: {P(0), P(1), P(2), P(3), …}
   Double charge distribution: {P`(0), P`(1), P`(2), P`(3), …}
   Composite distribution: {P``(0), P``(1), P``(2), P``(3), …}

   a. Each distribution is essentially a list of probabilities. We can derive composite probabilities for specific events via multiplication. For example:

      Find the probability of there being zero incorrect forward peptide identifications in both singly and doubly charged peptides:

      P``(0) = P(0) * P` (0) (they're independent events)

      Find the probability of there being one incorrect forward peptide identification in either the singly or doubly charged peptides, and none in the other group:

      P``(1)= P(1)*P`(0) + P(0)*P`(1)

      Find the probability of there being a total of two incorrect peptide identifications among the singly and doubly charged peptides:

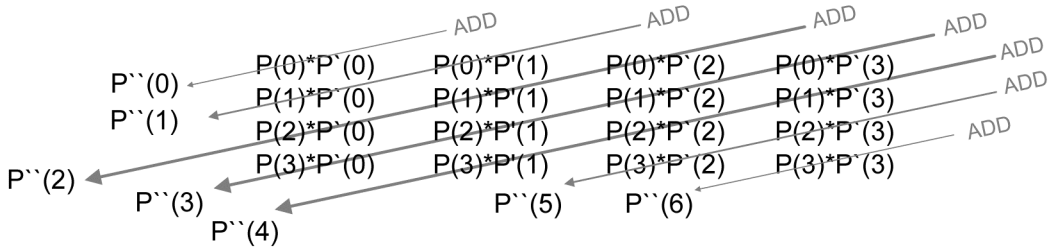
      P``(2)=P(2)*P`(0) + P(1)*P`(1) + P(0)*P`(2)

      …and so on for all likely values

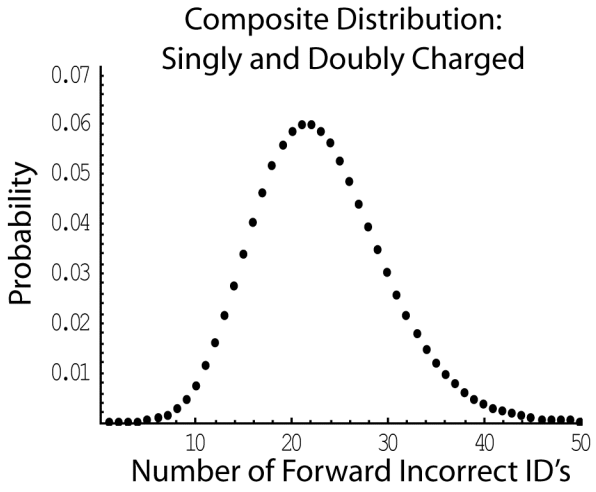
   b. In practice, these can be easily calculated together as a matrix of probabilities for all pairs of values from each distribution:

		Doubly Charged			
		P`(0)	**P`(1)**	**P`(2)**	**P`(3)**
Singly Charged	**P(0)**	P(0)*P`(0)	P(0)*P'(1)	P(0)*P`(2)	P(0)*P`(3)
	P(1)	P(1)*P`(0)	P(1)*P'(1)	P(1)*P`(2)	P(1)*P`(3)
	P(2)	P(2)*P`(0)	P(2)*P'(1)	P(2)*P`(2)	P(2)*P`(3)
	P(3)	P(3)*P`(0)	P(3)*P'(1)	P(3)*P`(2)	P(3)*P`(3)

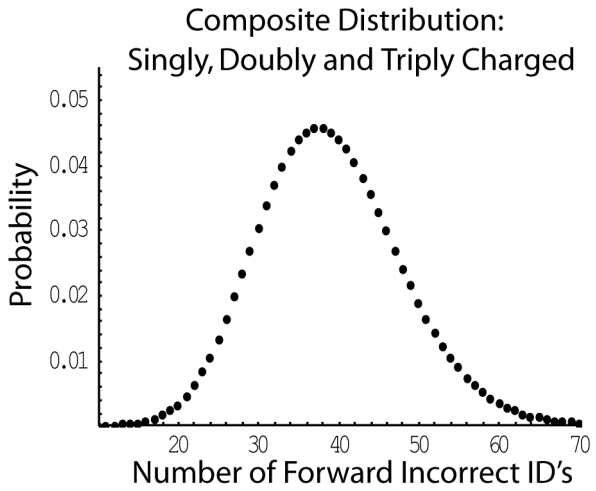
c. After calculating a matrix of all individual probabilities, we add those entries together that represent the same total number of incorrect identifications among singly and doubly charged peptides. Conveniently they lie along diagonals pointing downward to the left. By adding the sums along each of these diagonals, we have calculated all the terms for the composite distribution.



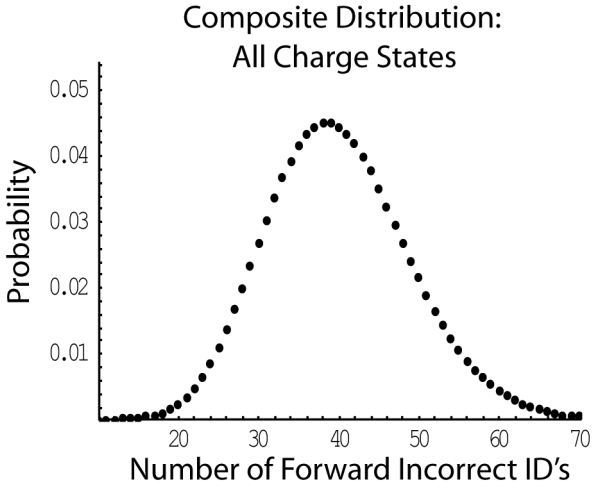
d. Plotted below is the composite distribution observed after combining the distributions from singly and doubly charged peptides, respectively.



3. Now we repeat the process, this time combining our composite distribution from step two with the triply charged peptide distribution



4. Finally, we multiply our composite distribution from step 3 with the quadruply charged peptide distribution.



5. Use this final composite distribution to determine the 95% confidence limits for the number of forward incorrect peptide identifications in the combined dataset

6. The numbers of peptides given for combined charge states are simply the sums of forward, forward-incorrect, and reversed peptide identifications seen for each charge state.

7. Divide the numbers of incorrect forward peptide identifications expected by the total numbers of forward peptide identifications to obtain confidence limits expressed as false positive rates.