Published in final edited form as:

Nat Methods. 2008 November ; 5(11): 959–964. doi:10.1038/nmeth.1260.

Decision Tree-Driven Tandem Mass Spectrometry for Shotgun Proteomics

Danielle L. Swaney^{1,*}, Graeme C. McAlister^{1,*}, and Joshua J. Coon^{1,2}

¹Department of Chemistry, 1101 University Avenue, University of Wisconsin, Madison, WI 53706

²Department of Biomolecular Chemistry, 1300 University Avenue, University of Wisconsin, Madison, WI 53706

Abstract

Mass spectrometry (MS) has become a key technology for modern large-scale protein sequencing. Tandem MS (MS/MS) — the process of peptide ion dissociation followed by mass-to-charge (m/z) analysis — is the critical component of MS approaches. Recent advances in mass spectrometry now permit two discrete, and complementary, types of peptide ion fragmentation — collision-activated dissociation (CAD) and electron transfer dissociation (ETD) on a single instrument. To exploit this complementarity and increase sequencing success rates, we designed and embedded a data dependent-decision tree algorithm (DT) to make unsupervised, real-time decisions of which fragmentation method to employ based on precursor charge (z) and m/z . Large-scale proteome analysis of *Saccharomyces cerevisiae* and human embryonic stem cells (hES) with the DT algorithm netted 53,055 peptide identifications — besting either CAD (38,293) or ETD (39,507) alone. That trend was maintained upon application of the DT method to phosphoproteomics, yielding 7,422 vs. either 2,801 (CAD) or 5,874 (ETD) phosphopeptides.

INTRODUCTION

Large-scale protein sequencing efforts employ enzymatic digestion of complex protein mixtures, e.g., cell lysates, to generate samples containing thousands of peptides.^{1–3} The resultant mixtures are rich with chemical diversity as the peptides vary in length, PTM status, and amino acid composition. Multi-dimensional chromatography can parse the peptides temporally over hours, or even days, so that downstream tandem MS instruments can autonomously interrogate as many peptides as possible.^{4–6} Still, sequence assignment requires a successful dissociation event, i.e., production of a sufficient number of informative fragment m/z peaks. Whatever the upstream separations, the chemical diversity exhibited by such peptide mixtures is problematic in this regard. Work by us and others demonstrate that CAD, the primary method of peptide cation dissociation, is most effective for small, lowly charged, unmodified peptide cations.^{7–9} Not all peptides fit this mold. ETD, a relatively new fragmentation method, is indifferent to either modification state or peptide mass, and shows preference for low m/z precursors.^{10–18}

*to whom correspondence should be addressed: jcoon@chem.wisc.edu

*these authors contributed equally to this work

AUTHOR CONTRIBUTIONS G.C.M., D.L.S., and J.J.C. designed research G.C.M. and D.L.S. performed research G.C.M. performed instrument modification G.C.M., D.L.S., and J.J.C. wrote the paper

COMPETING INTERESTS STATEMENT

J.J.C. is a co-inventor on two patent applications related, in part, to the material presented here.

Quadrupole ion trap mass spectrometers (QITs) outfitted with reagent anion sources are capable of performing both CAD and ETD.^{10, 11} Determination of the most appropriate dissociation method, however, requires a priori knowledge of key precursor attributes such as z and m/z ratio. QITs provide unmatched sensitivity and high scan cycle times (spectra/second) at unit m/z resolving power.¹⁹ Operation of these instruments at resolutions sufficient to determine precursor charge states is possible, but not practical as scan rates are substantially reduced. ETD-enabled quadrupole linear ion traps (QLT, a type of QIT), however, have recently been interfaced a variety of secondary high resolving power analyzers.²⁰⁻²³ We reasoned the m/z resolution afforded by such hybrid instruments could enable real-time generation of precursor z and m/z ratio information for intelligent selection between dissociation methods. Note previous works rely upon information present in tandem mass spectra to trigger an MS/MS/MS event, e.g., neutral loss-triggered MS³.^{11, 24} That approach applies the same fragmentation method twice to improve the probability of sequencing success. Our proposition is to utilize the information present in the full mass spectrum to make a priori decisions about which fragmentation method to apply to increase the probability of MS/MS scan success for all precursors. Such capabilities would represent a major advance for shotgun proteomics as online chromatographic separations may only present the instrument a single opportunity to interrogate a particular peptide precursor; hence, it is crucial that the foremost dissociation method be applied. Here we have developed an algorithm which exploits the high mass accuracy and resolution, achieved with orbitrap m/z analysis, to make real-time decisions of which dissociation method to employ in an unsupervised, data-dependent fashion. Large-scale proteome analysis of *Saccharomyces cerevisiae* and hES cells with the DT algorithm netted 53,055 peptide identifications — besting analysis by either CAD (38,293) or ETD (39,507) alone. That trend was maintained upon application of the DT method to phosphoproteomics. In total the DT method yielded 7,422 vs. either 2,801 (CAD) or 5,874 (ETD) phosphopeptides.

RESULTS

Characterization of peptide identification probabilities

To gather a training set of MS/MS spectra for probability calculations a whole cell yeast lysate was digested using the protease endo-LysC and then separated into 12 fractions by strong cation exchange chromatography (SCX, two biological replicates).^{24, 25} Each fraction was analyzed by online nanoflow reversed-phase liquid chromatography coupled to MS/MS (nLC-MS/MS), using a forty minute gradient with data-dependent precursor selection. Eluting peptide cation populations were analyzed using the orbitrap (i.e., MS¹ prescan, for high mass accuracy and resolution), while MS/MS product ion spectra were m/z analyzed in the QLT (for speed and sensitivity). Six separate analyses of each fraction were performed — three using CAD only and three with ETD only. Note the analyses were identical in all aspects save dissociation type (see methods for details). Imposing a false discovery rate of one percent via a target-decoy search,²⁴ the CAD and ETD-based analyses yielded 30,016 and 29,702 peptide identifications from 200,524 and 175,984 scans, respectively (Table 1, Supplementary Data Set 1). The 376,508 spectra were then binned by precursor z and m/z ratios and the probability of either a CAD or ETD scan generating a high confidence peptide sequence identification was calculated for each bin and plotted as a function of precursor m/z for precursor charges ranging from 2 to 7 (Fig. 1).

This data, which represent the largest comparison of ETD and CAD performed to date, confirmed the strong correlation between the probability of a successful sequencing event and the precursor attributes of z and m/z ratio previously reported by us and others.^{26, 27} To summarize, ETD was more effective at producing MS/MS spectra that correlated with high confidence to a candidate peptide sequence for precursor ions having low m/z ratios; CAD,

was most effective for medium to high m/z ratio precursors (Fig. 1). Precursor z is also important — as z increased the m/z ratio breakpoint at which CAD became more favorable than ETD increased. Peptide precursor cations having z equal to, or in excess of, six were rarely observed; hence, there were too few data points to produce robust probability calculations. For doubly charged species CAD was always more favorable than ETD. Note in a separate experiment we employed supplemental activation (ETcaD) to boost ETD efficiency for doubly charged precursors; however, the search algorithm failed to identify the additional c - and z -type products — H-atom rearrangement is elevated during ETcaD causing the expected product masses to deviate by 1 Da.²¹ Overall, the number of unique peptide identifications for the ETcaD method (4,585) was not an improvement over ETD alone (4,962 and 4,895, Supplementary Table 1b). However, it has recently been demonstrated that doubly charged precursors can be identified with good success via ETcaD when using a different searching algorithm, thus, we anticipate that in future implementations, it may be desirable to apply ETcaD to some portion of the double charged population as well.²⁷

To evaluate run-to-run irreproducibility, a common occurrence during shotgun proteomics experiments, we compared overlap between replicate analyses of the same dissociation type vs. that observed between ETD and CAD datasets (biological replicate 2). Approximately 61.9 % ($n = 3$, std. dev. = 4.2) of the sequences garnered from replicate runs using a single dissociation method overlapped (Fig. 2a-b). Comparison of the ETD and CAD-generated sequences revealed only 34.4 % overlap ($n = 4$, std. dev. = 2.9, Fig. 2c). From these data, and the probability distributions shown above, we conclude that the two dissociation methods are indeed complementary and that intelligently selecting between them will increase the number of identified peptides in a shotgun proteomics experiment.

Implementation and application of the decision tree

To enable automated, real-time selection between dissociation methods we designed a probability-based decision tree (DT) algorithm. Because an ETD spectrum takes on average 14 % more time to collect than the corresponding CAD spectrum, the experimentally determined probabilities were corrected for scan duration. This places importance on the aggregate probability of the entire experiment rather than single scan success. From these data we constructed a DT algorithm, which is visually represented in Figure 3, and for which pseudo code is presented in the **Supplementary Methods**. This algorithm was written into the instrument control language and loaded onto the firmware of the modified orbitrap. When engaged, the z and m/z ratio of a targeted precursor (selected in the standard data-dependent fashion) is routed to the DT. Upon exiting the DT, the precursor is assigned the dissociation method most likely to result in an identification.

The efficacy of the DT logic was tested by triplicate analysis of the same twelve SCX fractions, under identical conditions, employed for collection of the training data set. From these data 41,719 peptides were identified (15,221 from CAD and 26,498 from ETD) at a false discovery rate of 1 % - a 39.0 % and 40.5 % increase over use CAD and ETD alone, respectively (Table 1, Supplementary Data Set 1). From the CAD- and ETD-only data sets the probability of any given MS/MS event resulting in a high confidence sequence assignment was 15.0 % and 16.9 %, respectively. With the DT data set this probability rose to 22.3 %, a 32 % and 49 % increase over the CAD- and ETD-only analyses. The utility of the DT logic stems from complementary nature of CAD and ETD; the pairing effectively accommodates the chemical diversity present in complex peptide mixtures as each is compatible with a distinct subset of the analyte population. To further explore these relationships we classified all identified peptides by precursor m/z ratio, m , z , and length (Fig. 4a-b and Supplementary Fig. 1). The DT logic combined the best of the CAD and ETD-only populations to produce an aggregate that was better than either alone. At the extremes of the precursor m/z plots, shown in Figure 4a-b, the

DT trace resembles either the CAD or ETD-only traces. This was due to the DT exclusively relying on either one dissociation method or the other at these outermost precursor m/z ratios, independent of charge state. Towards the middle of the plot, where both precursor z and m/z are critical factors, the DT algorithm resulted in sizeable gains.

The CAD- and ETD-only datasets contained 9,257 and 7,650 unique peptides. Overlap between these two was modest — 4,776 peptides (Fig. 4c). From the DT dataset (yellow portion of Fig. 4c) 11,237 unique peptide identifications were posted, a 21.4 % and 46.9 % increase over the CAD- and ETD-only datasets. We note the DT-based analysis netted over 90 % of the collective unique peptide identifications (11,237 of 12,134; 92.6 %) observed from the combined CAD- and ETD-only analyses in half the number of scans, acquisition time, and sample. Further, the identification sum of duplicate DT analyses was higher (8,939 unique) than any combination of CAD and ETD (e.g., CAD and CAD, ETD and ETD, or ETD and CAD, Fig. 2a-d, Table 1). The identified peptides from triplicate analyses of the three methods (CAD, ETD, DT) correlated to 2,187, 2,125, and 2,496 unique proteins, respectively (Supplementary Fig. 2). Collectively, the three methods identified 2,993 unique proteins — making this work among the largest yeast proteomics experiments reported.⁴⁻⁶

We note that one could perform sequential CAD and ETD tandem MS on each selected precursor and potentially obviate the need for the DT logic. We analyzed each yeast SCX fraction using this method (biological replicate 2) and identified 5,389 unique peptides with a very high scan success rate (23.2 %). This high success rate is the result of only interrogating the most intense precursor ions; the downside is a greatly reduced dynamic range as only half as many precursors are sampled. Sequential CAD and ETD identified fewer proteins than any of the methods (1,459) and was substantially worse than the DT (1,927 and 1,870 for the same sample, Supplementary Table 1b).

We next wondered whether the DT branch points would be portable from sample to sample — that is, for the DT to be effective must one generate the probability calculations shown in Figure 1 for each sample? To answer this, proteins from another organism (human) were digested with LysC, fractionated by SCX, and sampled once each via the three methods (CAD, ETD, and DT). The DT method, which employed the unmodified parameters defined by the yeast training set, garnered 7,248 unique peptide sequences, topping both CAD (5,777) and ETD (6,142, Table 1). We conclude that for peptides generated by the same enzyme, DT parameters are effectively extended from one sample to the next.

Application of the decision tree to phosphoproteomics

Next we decided to test the performance of the DT algorithm on peptides having broadly different chemical characteristics. This was accomplished by enrichment of phosphorylated peptides that were generated following digestion with trypsin — the more conventional shotgun proteomics enzyme. Proteins, harvested from human embryonic stem cells (hES), were subjected to trypsin, SCX fractionation, and ion metal affinity chromatography (IMAC). Enriched phosphopeptides from each fraction were individually analyzed via the CAD and ETD methods, using the same parameters as described above. In 86,258 scans the ETD-based analyses sequenced 5,874 (3,062 unique) phosphopeptides — over twice that from CAD, 2,801 (1,791 unique) in 93,324 scans (both at 1 % false discovery rate, Table 1, Supplementary Data Set 1). These results are not surprising as phosphate modifications are labile when subjected to CAD.¹¹ Overlap between the phosphopeptide ETD and CAD datasets was only 17.9 %, suggesting a possible higher degree of complementarity of the two methods for phosphoproteomic analyses (Fig. 5a). We reasoned, however, that the optimal branches of the DT algorithm could be quite different (compared to Fig. 3) for these peptides as they were created by a different enzyme and contained at least one phosphorylation site. The DT branch

points for analysis of these tryptic, phosphopeptides are remarkably similar to those ideal for sequencing un-modified, LysC peptides (Supplementary Fig. 3).

Application of the DT algorithm to these phosphopeptide samples resulted in 7,422 (3,897 unique) phosphopeptide identifications (1 % false discovery rate) — a 118 % and 27 % increase over CAD or ETD alone (Table 1, Supplementary Data Set 1). When engaged, the DT logic resulted in a 64 % increase in scan success rate (7.5 vs. 4.8 overall) for these phosphopeptides. In a single analysis of all twelve SCX fractions, the DT algorithm identified 3,897 unique phosphopeptides, 95 % as many as sequenced by the combination of one CAD and one ETD analysis (4,115). This is accomplished with half the sample and analysis time — a major benefit as phosphopeptide enrichment strategies are highly labor intensive and demand large cell numbers. Still, application of duplicate DT analyses resulted in over 5,000 more phosphopeptide identifications (14,006) than summing one ETD and one CAD analysis (8,675) of the same sample (Fig. 5, Table 1). Such improvements increase identification confidence, the number of unique identifications, and total phosphoprotein hits (1,083, 1,665, and 2,001 from CAD, ETD, and DT, respectively, Supplementary Fig. 4). The increased number of identifications when using the DT, as compared to CAD or ETD alone, was independent of charge state, mass, length, or m/z (Supplementary Fig. 5). To our knowledge this work represents the largest hES cell phosphorylation study to date, identifying a total of 8,359 sites on 6,970 unique peptides from 2,958 proteins.

DISCUSSION

Large-scale protein sequencing experiments use chromatography and mass spectrometry to mine complex peptide mixtures that are lush with chemical diversity. Successful sequence analysis requires the production of informative m/z peaks upon MS/MS, a process that occurs several times per second. In this work we performed large-scale analyses of complex peptide mixtures that were created using multiple enzymes from multiple organisms. These peptides were sampled using two different and complementary methods of tandem MS — CAD and ETD. From these data we calculated the probability of an MS/MS event resulting in a high confidence sequence assignment for both methods as a function of the observed precursor peptide z and m/z ratio. From these data we reasoned the m/z resolution afforded by the ETD-enabled hybrid orbitrap system, which we recently constructed,^{21, 23} could enable real-time generation of precursor z and m/z ratio information for intelligent selection between dissociation methods — that is, to utilize the information present in the full mass spectrum to make a priori decisions about which fragmentation method to apply to increase the probability of MS/MS scan success.

To test this hypothesis we devised and embedded a DT algorithm into the orbitrap firmware. The logic exploits the high mass accuracy and resolution afforded by the orbitrap full MS prescan to make real-time assignments of which dissociation method to employ in an unsupervised, data-dependent fashion. In all instances studied, the automated tailoring of dissociation method to precursor afforded by the DT method substantially increased the probability of sequence identification. We note the decision tree logic described here can easily be expanded upon to consider other inputs (e.g., precursor signal intensity, etc.) that would trigger different outputs (e.g., selection of analyzer, reaction time, amount of precursor necessary, etc.). Further, other dissociation methods besides ion trap CAD and ETD can likewise be readily implemented. As each method brings a subset of the proteome into view, we anticipate the gains shown here will be further multiplied upon expansion of the DT approach.

METHODS

Sample preparation, nanoHPLC, mass. spectrometry, and database searching. See Supplementary Methods.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGEMENTS

We are grateful to J. Griep-Raming, S. Horning, O. Lange, A. Makarov, J. Schwartz, M. Senko, G. Stafford, and J. Syka, all of Thermo Scientific, for providing advice and technical assistance during the development of the ETD-Orbitrap instrumentation. We also thank B. Craig, S. Herbst, H. Coon, and K. Thompson for culture and harvest of the yeast and Jessica Antosiewicz-Bourget and James Thomson for culture of the hES cells (University of Wisconsin, Madison, WI). We gratefully acknowledge J. Marto and S. Ficarro for advice with the Waters nUPLC. Finally, we thank G. Barrett-Wilt, D. Good, J. Keith, D. Phanstiel, and A. Huhmer for helpful discussions. The University of Wisconsin, the Beckman Foundation, Eli Lilly, and the National Institutes of Health (1R01GM080148 to J.J.C.) provided financial support for this work. G.C.M. and D.L.S. acknowledge support from the National Institutes of Health pre-doctoral fellowships — (Biotechnology Training Program, NIH 5T32GM08349, to G.C.M.; the Genomic Sciences Training Program, NIH 5T32HG002706, to D.L.S.).

REFERENCES

1. Aebersold R, Mann M. Mass spectrometry-based proteomics. *Nature* 2003;422:198–207. [PubMed: 12634793]
2. Hunt DF, Yates JR, Shabanowitz J, Winston S, Hauer CR. Protein Sequencing by Tandem Mass-Spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 1986;83:6233–6237. [PubMed: 3462691]
3. Coon JJ, Syka JEP, Shabanowitz J, Hunt DF. Tandem mass spectrometry for peptide and protein sequence analysis. *Biotechniques* 2005;38:519–+. [PubMed: 15884666]
4. Washburn MP, Wolters D, Yates JR. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnology* 2001;19:242–247.
5. Peng JM, Elias JE, Thoreen CC, Licklider LJ, Gygi SP. Evaluation of multidimensional chromatography coupled with tandem mass spectrometry (LC/LC-MS/MS) for large-scale protein analysis: The yeast proteome. *Journal of Proteome Research* 2003;2:43–50. [PubMed: 12643542]
6. de Godoy LMF, et al. Status of complete proteome analysis by mass spectrometry: SILAC labeled yeast as a model system. *Genome Biology* 2006;7
7. Good DM, Wirtala M, McAlister GC, Coon JJ. Performance Characteristics of Electron Transfer Dissociation Mass Spectrometry. *Mol Cell Proteomics* 2007;6:1942–1951. [PubMed: 17673454]
8. Dongre AR, Jones JL, Somogyi A, Wysocki VH. Influence of peptide composition, gas-phase basicity, and chemical modification on fragmentation efficiency: Evidence for the mobile proton model. *Journal of the American Chemical Society* 1996;118:8365–8374.
9. Huang YY, et al. Statistical characterization of the charge state and residue dependence of low-energy CID peptide dissociation patterns. *Analytical Chemistry* 2005;77:5800–5813. [PubMed: 16159109]
10. Coon JJ, Syka JEP, Schwartz JC, Shabanowitz J, Hunt DF. Anion dependence in the partitioning between proton and electron transfer in ion/ion reactions. *International Journal of Mass Spectrometry* 2004;236:33–42.
11. Syka JEP, Coon JJ, Schroeder MJ, Shabanowitz J, Hunt DF. Peptide and protein sequence analysis by electron transfer dissociation mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 2004;101:9528–9533. [PubMed: 15210983]
12. Coon JJ, et al. Protein identification using sequential ion/ion reactions and tandem mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 2005;102:9463–9468. [PubMed: 15983376]

13. Hogan JM, Pitteri SJ, Chrisman PA, McLuckey SA. Complementary structural information from a tryptic N-linked glycopeptide via electron transfer ion/ion reactions and collision-induced dissociation. *Journal of Proteome Research* 2005;4:628–632. [PubMed: 15822944]
14. Pitteri SJ, Chrisman PA, Hogan JM, McLuckey SA. Electron transfer ion/ion reactions in a three-dimensional quadrupole ion trap: Reactions of doubly and triply protonated peptides with SO₂ center dot. *Analytical Chemistry* 2005;77:1831–1839. [PubMed: 15762593]
15. Good DM, Coon JJ. Advancing proteomics with ion/ion chemistry. *Biotechniques* 2006;40:783–789. [PubMed: 16774122]
16. Chi A, et al. Analysis of phosphorylation sites on proteins from *Saccharomyces cerevisiae* by electron transfer dissociation (ETD) mass spectrometry. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104:2193–2198. [PubMed: 17287358]
17. Khidekel N, et al. Probing the Dynamics of O-GlcNAc Glycosylation in the Brain using Quantitative Proteomics. *Nature Chem. Biol* 2007;3:339–348. [PubMed: 17496889]
18. Lecchi S, et al. Tandem phosphorylation of Ser-911 and Thr-912 at the C terminus of yeast plasma membrane H⁺-ATPase leads to glucose-dependent activation. *Journal of Biological Chemistry* 2007;282:35471–35481. [PubMed: 17932035]
19. Schwartz JC, Senko MW, Syka JEP. A two-dimensional quadrupole ion trap mass spectrometer. *Journal of the American Society for Mass Spectrometry* 2002;13:659–669. [PubMed: 12056566]
20. Xia Y, et al. Implementation of ion/ion reactions in a quadrupole/time-of-flight tandem mass spectrometer. *Analytical Chemistry* 2006;78:4146–4154. [PubMed: 16771545]
21. Swaney DL, et al. Supplemental Activation Method for High-Efficiency Electron-Transfer Dissociation of Doubly Protonated Peptide Precursors. *Anal. Chem* 2007;79:477–485. [PubMed: 17222010]
22. Palczewski K, et al. Characterization of a Truncated Form of Arrestin Isolated from Bovine Rod Outer Segments. *Protein Science* 1994;3:314–324. [PubMed: 8003967]
23. McAlister GC, et al. A Proteomics Grade Electron Transfer Dissociation-Enabled Hybrid Linear Ion Trap-Orbitrap Mass Spectrometer. *J. Proteome Res* 2008;7:3127–3136. [PubMed: 18613715]
24. Beausoleil SA, et al. Large-scale characterization of HeLa cell nuclear phosphoproteins. *PNAS* 2004;101:12130–12135. [PubMed: 15302935]
25. Villen J, Beausoleil SA, Gerber SA, Gygi SP. Large-scale phosphorylation analysis of mouse liver. *Proceedings of the National Academy of Sciences of the United States of America* 2007;104:1488–1493. [PubMed: 17242355]
26. Good DM, Wirtala M, McAlister GC, Coon JJ. Performance characteristics of electron transfer dissociation mass spectrometry. *Mol Cell Proteomics*. 2007
27. Molina H, Matthiesen R, Kandasamy K, Pandey A. Comprehensive comparison of collision induced dissociation and electron transfer dissociation. *Analytical Chemistry* 2008;80:4825–4835. [PubMed: 18540640]

Figure 1.

Comprehensive probability of a high confidence peptide identification. The probability of either a ETD or CAD tandem MS event resulting in a high confidence sequence identification for peptide cations having various z as a function of precursor m/z ratio ($n = 376,508$). To evaluate the importance of any one m/z ratio bin, the percentage of all precursors observed having the specific z and m/z ratio are given below. Error bars indicated ± 1 s.d. ($n = 3$). Note that despite using LysC instead of trypsin, the bulk of all sampled peptide precursor cations (39.2 %) were doubly charged.

Figure 2.

Overlap between various combinations of duplicate analyses of yeast peptides. Venn diagrams displaying the overlap of: CAD and CAD (**a**), ETD and ETD (**b**), DT and DT (**c**), CAD and ETD (**d**) trials. The number of unique identifications corresponding to each diagram is shown above it, while the percentage of overlap is displayed within the diagram.

Figure 3.

Schematic representation of the probabilistic decision tree. The plots in Figure 1 were the basis for this decision tree. Note, m/z is the mass-to-charge ratio of the precursor selected for MS/MS interrogation.

Figure 4. Distribution of identified peptides using either the CAD (red), ETD (blue), or DT (yellow) methods. (a) and (b) display the number of identified peptides for each method as a function of precursor m/z ratio for the yeast dataset — (d) and (e) show the same for the hES cell phosphopeptide dataset. (c) and (f) display the redundancy and complementary between the unique peptides identified during the CAD- and ETD-only analyses for the yeast and hES datasets, respectively. The yellow area present in each indicates the coverage provided by the DT logic, which is ~ 90 % of that observed upon combination of the CAD- and ETD-only data sets. For further details see Table 1.

Figure 5. Overlap between various combinations of duplicate analyses of hES cell phosphopeptides. Venn diagrams displaying the overlap of: DT and DT (**a**) and CAD and ETD (**b**). The number of unique identifications corresponding to each diagram is shown above it, while the percentage of overlap is displayed within the diagram.

Table 1

Summary peptide identifications. The number of identifications derived from the four analysis methods (CAD, ETD, DT, and sequential CAD and ETD) for the yeast and hES cell experiments.

Yeast (All 3 Trials)	CAD	ETD	Decision Tree
Identifications (1% FP)	30,016	29,702	41,719
Total Scans	200,524	175,984	186,964
Scan Success Rate	14.97%	16.88%	22.31%
Unique Peptides (1% FP)	9,257	7,650	11,237
Unique Proteins (1% FP)	2,187	2,125	2,496
Yeast	Sum 1 ETD and 1 CAD Trial (Biorep2, Trial 1)	Sum 1 ETD and 1 CAD Trial (Biorep2, Trial 2)	Sum 2 Decision Tree Trials (Biorep2, Trial 1 and 2)
Identifications (1% FP)	23,694	23,629	32,374
Total Scans	127,632	130,144	128,651
Scan Success Rate	18.56%	18.16%	25.16%
Unique Peptides (1% FP)	8,052	7,842	8,939
Unique Proteins (1% FP)	2,035	2,012	2,210
hES cells (1 trial each)	CAD	ETD	Decision Tree
Identifications (1% FP)	8,277	9,805	11,336
Total Scans	71,082	58,816	59,548
Scan Success Rate	11.64%	16.67%	19.04%
Unique Peptides (1% FP)	5,777	6,142	7,248
Unique Proteins (1% FP)	2,777	3,076	3,329
hES cell phosphopeptides	CAD (Trial 1)	ETD (Trial 1)	Decision Tree (Trial 1)
Identifications (1% FP)	2,801	5,874	7,422
Total Scans	93,324	86,258	94,794
Scan Success Rate	3.00%	6.81%	7.83%
Unique Peptides (1% FP)	1,791	3,062	3,897
Unique Proteins (1% FP)	1,083	1,665	2,001
Phosphorylation sites on Unique Peptides (1% FPR)	2,206	3,646	4,641
hES cell phosphopeptides	Decision Tree (Trial 2)	Sum 1 ETD and 1 CAD Trial (Trial 1)	Sum 2 Decision Tree Trials (Trial 1 and 2)
Identifications (1% FP)	6,584	8,675	14,006
Total Scans	92,917	179,582	187,711
Scan Success Rate	7.09%	4.83%	7.46%
Unique Peptides (1% FP)	3,707	4,115	5,338
Unique Proteins (1% FP)	1,968	2,166	2,825
Phosphorylation sites on Unique Peptides (1% FPR)	4,461	4,996	6,474