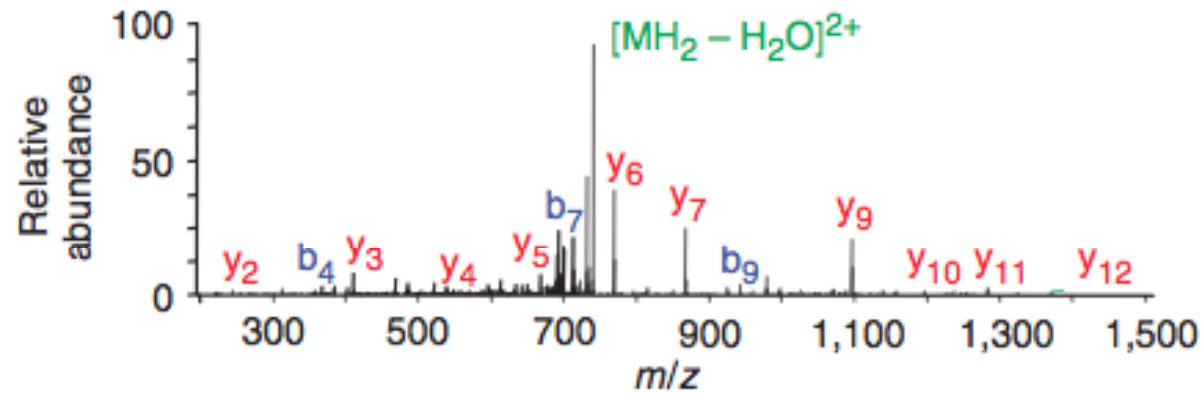


A PROBABILITY-BASED APPROACH FOR HIGH-THROUGHPUT PROTEIN PHOSPHORYLATION ANALYSIS AND SITE LOCALIZATION

SEAN A BEAUSOLEIL, JUDIT VILLÉN, SCOTT A GERBER, JOHN RUSH & STEVEN P GYGI

Phosphorylation

- Within the eukaryotic cell, approximately 30% of proteins are phosphorylated at any given point.
- Phosphorylation indispensably governs the cellular processes of differentiation, protein localization, and protein-protein interactions.
- Phosphorylation(+80Da) can occur on serine, threonine, or tyrosine residues.



$NH_2 - QSSVTQVTEQpSPK - COOH$

or

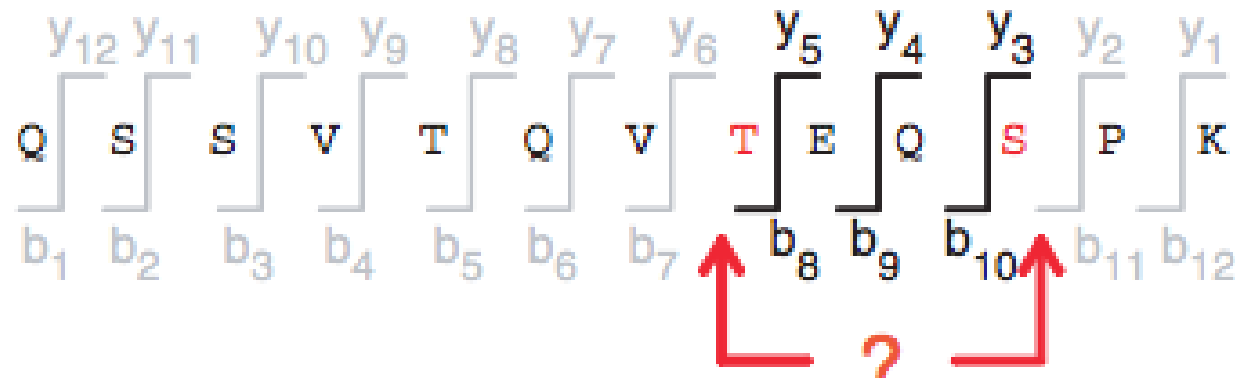
$NH_2 - QSSVTQVpTEQSPK - COOH$

?

Ascore

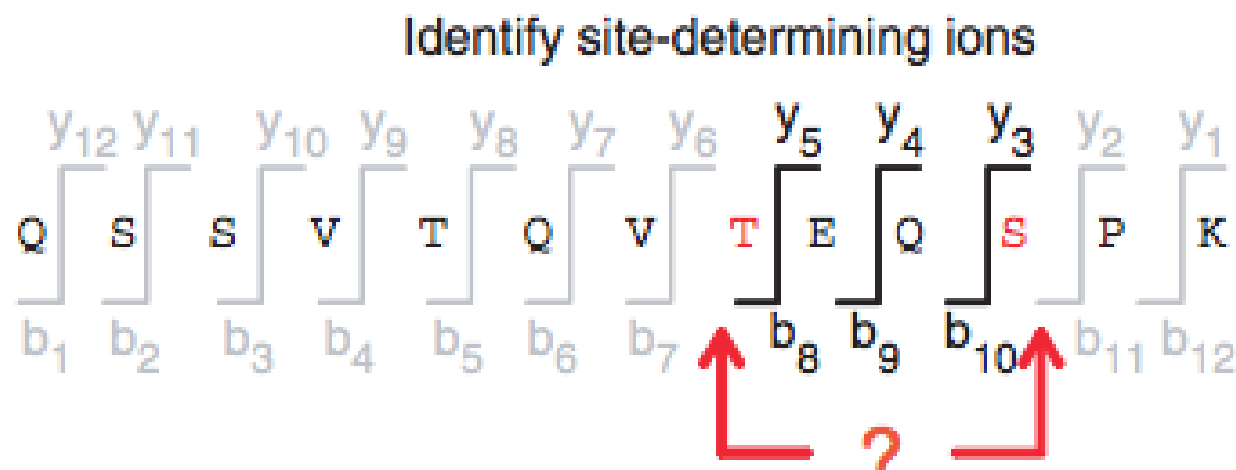
Identify site-determining ions

The precise site of phosphorylation can only be determined by the presence of site-determining ions.



Site-Determining Ions

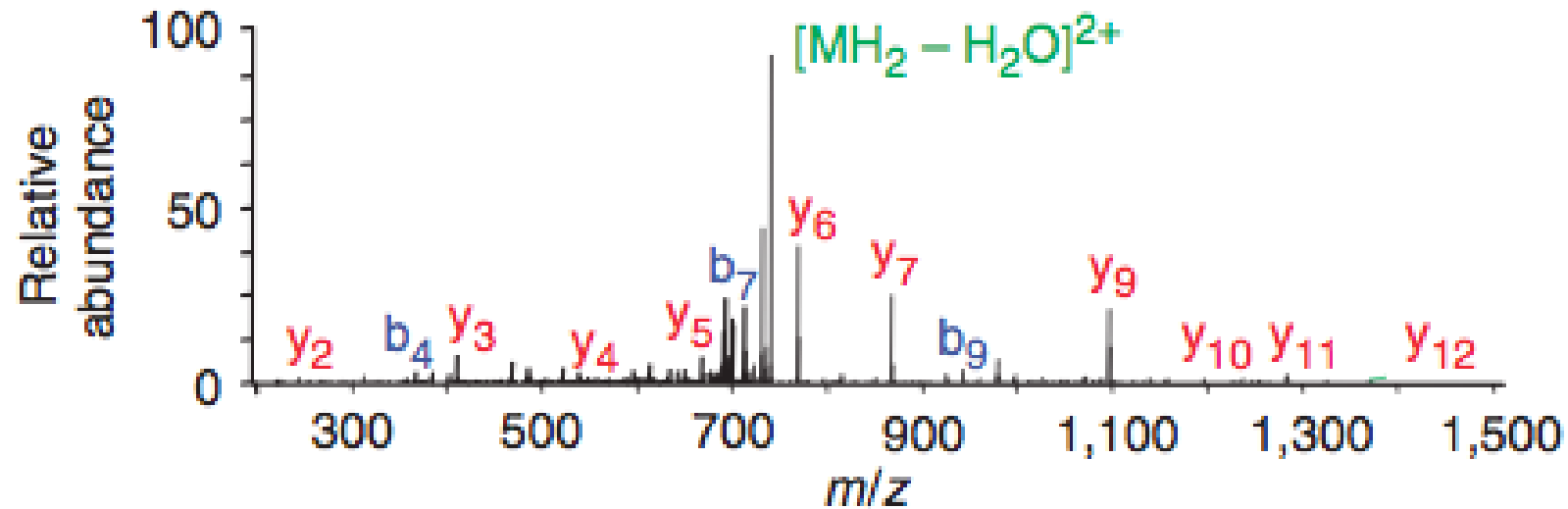
- Fragment ions that are exclusive to a specific site location and can be identified to uniquely assign a site to a specific residue.

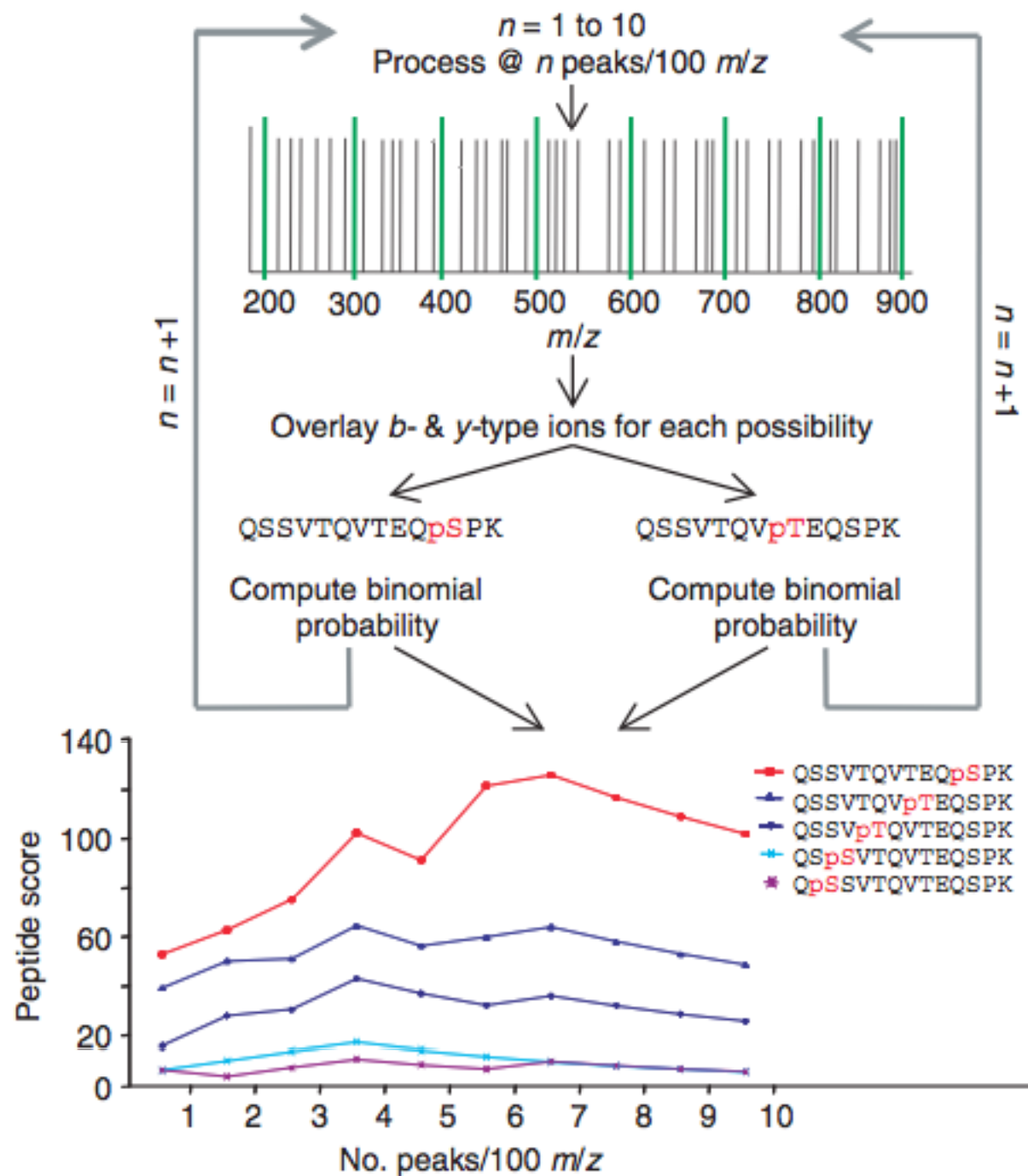


Step 1 of Ascore Method

- Peptide Score: Calculate a probability-based ion matching score for each potential phosphorylation site.

QpSSVTQVTEQSPK
QS_pSVTQVTEQSPK
QSSV_pTQVTEQSPK
QSSVTQV_pTEQSPK
QSSVTQVTEQ_pSPK





Cumulative binomial probability P

- P represents the probability of randomly matching at least the given number of fragment ions to the MS/MS spectrum
- The total number of trials (N) equaled the total number of fragment ions for the given peptide. The total number of successes (n) equaled the number of ions matched to the spectrum. (p) probability of matching a peak.

$$P(X) = \sum_{k=n}^N \binom{N}{k} p^k (1-p)^{N-k}$$

Step 2 of Ascore Method

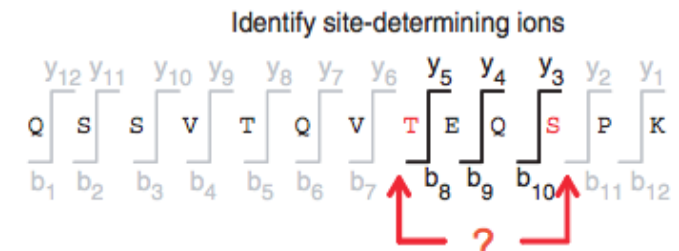
- Ascore: a probability-based metric that measures the likelihood that a difference in site-determining ions between two site positions was matched by random chance.

Determine peak depth from peptic
score for maximal difference
 $n = 6$ (panel b)



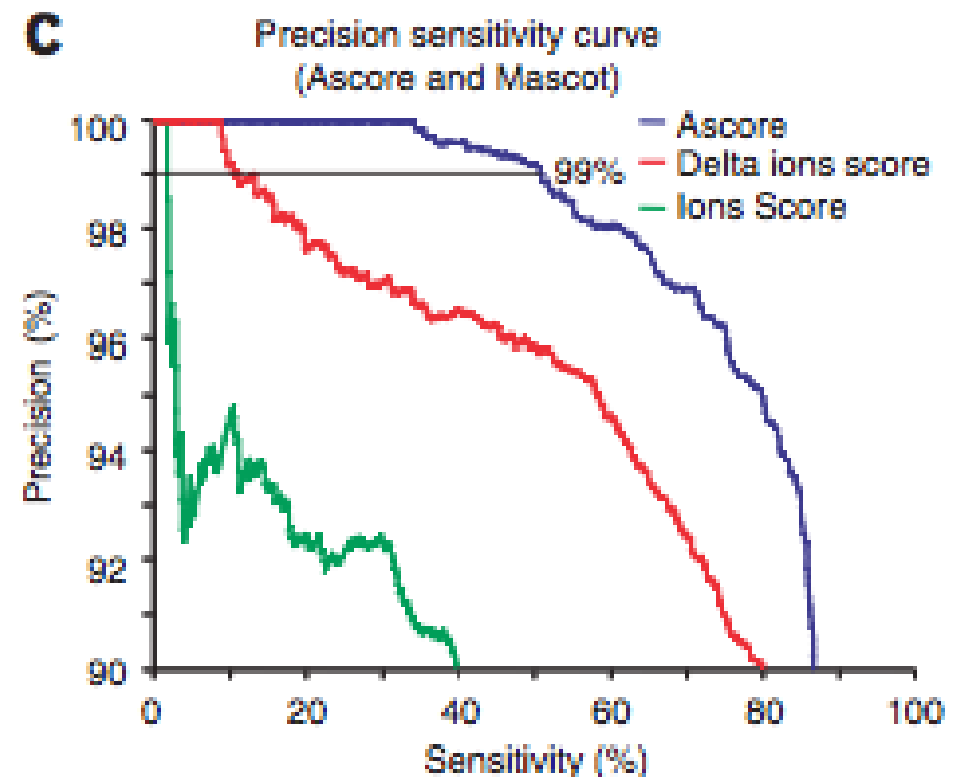
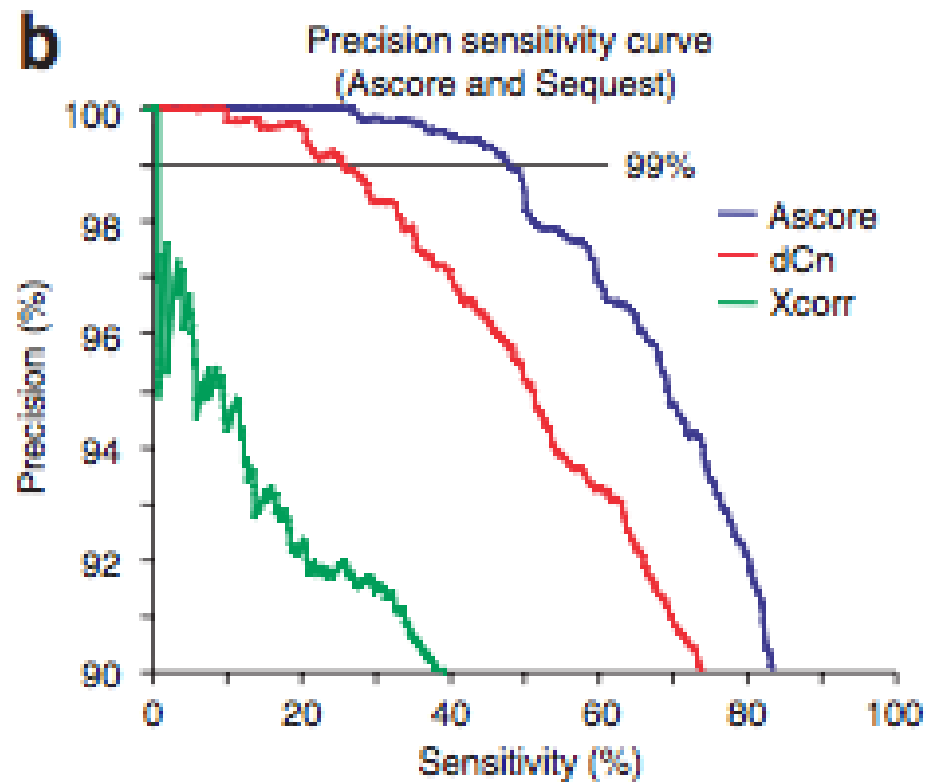
Calculate cumulative binomial probability using only
site-determining ions at highest-scoring peak depth

$$P(x) = \sum_{k=n}^N \binom{N}{k} p^k (1-p)^{N-k}$$



Phosphopeptide	QSSVTQVTE Q p S PK	QSSVTQ V p T EQSP K
Trials (N)	6 ($y_3, y_4, y_5, b_8, b_9, b_{10}$)	6 ($y_3, y_4, y_5, b_8, b_9, b_{10}$)
Successes (n)	5 ($y_3, y_4, y_5, b_9, b_{10}$)	0
p (6 peaks / 100 m/z)	0.06	0.06
P	0.0000044	1.0
Score [$-10 \times \log(P)$]	53.57	0
Ascore = ambiguity score (difference of the top two candidates)	$53.57 - 0 = 53.57$	

Comparision with Sequest and Mascot



nocadazole-arrested HeLa cell lysate

Summary

- Phosphopeptides with an Ascore ≥ 19 always produced $\geq 99\%$ certainty for correct phosphorylation site localization regardless of the data set.
- A substantial improvement over both Sequest and Mascot at the same confidence level.