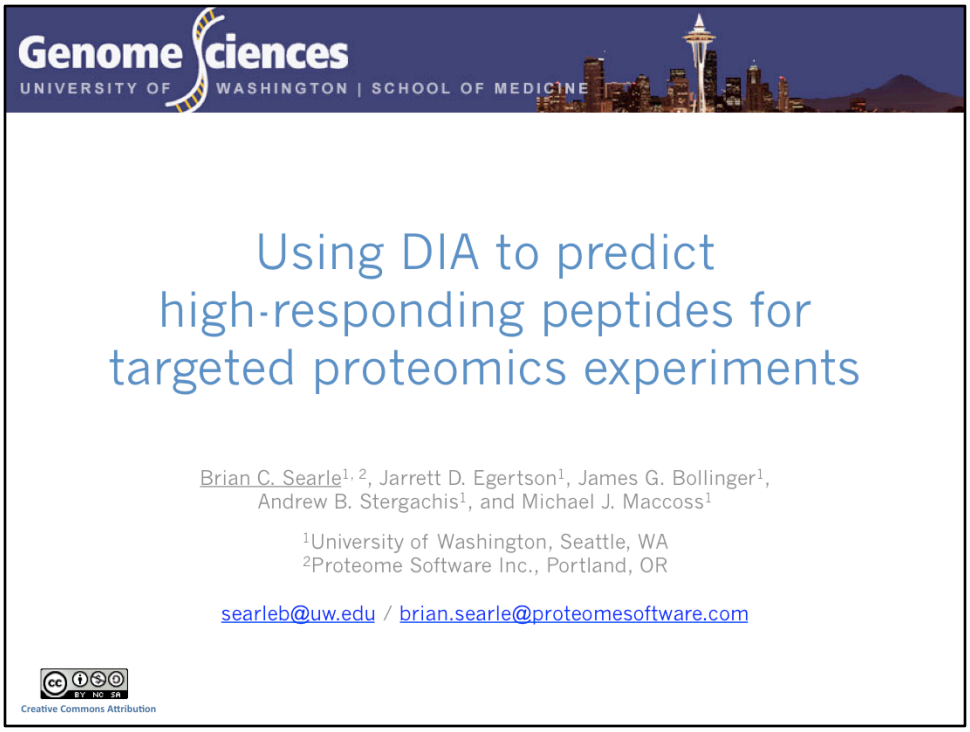
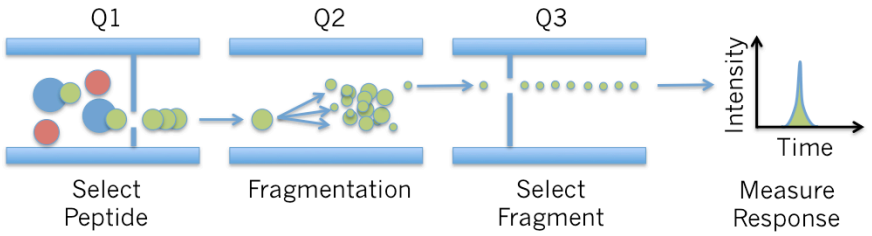


Hi, I'm Brian Searle. Many of you know me as a scientist from Proteome Software. Currently I wear two hats and I'm going to talk to you about a project from my second role at University of Washington on how to pick peptides for targeted experiments. Before we begin, these slides are creative commons licensed, so feel free to email me for a copy to reuse for whatever you'd like.

What is Selected Reaction Monitoring (SRM)?



- Like a mass spec-based western blot
- Select for peptides as a proxy for proteins
- Opportunity cost: only enough time to select for a few peptides/protein

Just a quick recap to get on the same page: selected reaction monitoring is typically performed using triple quads. Peptides are selected in the first quad, fragmented in the second, and then a few signature fragments are selected in the third quad. This works like a western blot, except you're scanning for peptides as a proxy for proteins. The key take home is that there's an implicit opportunity cost. There's only enough time to select for a few peptides per protein.

Which peptide do you pick?

>FLJ20321 (CASZ1 Castor Zinc Finger 1)

MVQPQGCSDE EDHAEEPSKD GGALEEKDSD GAASKEDSGP STRQASGEAS SLRDYAASTM
TEFLGMFGYD DQNTRDELAR KISFEKLHAG STPEAATSSM LPTSEDTLSK RARFSKYEEY
IRKLKAGEQL SWPAPSTKTE ERVGKEVVGT LPGLRLPSST AHLETKATIL PLPSHSSVQM
QNLVARASKY DFFIQKLKTG ENLRPQNGST YKKPSKYDLE NVKYLHLFKP GEGSPDMGGA
IAFKTGKVGR PSKYDVRGIQ KPGPAKVPPT PSLAPAPLAS VPSAPSAPGP GPEPPASLSF
NTPEYLKSTF SKTDSITTGT VSTVKNGLPT DKPAVTEDVN IYQKYIARFS GSQHCGHIHC
AYQYREHYHC LDPECNYQRF TSKQDVIRHY NMHKKRDNSL QHGFMRFSPL DDCSVYYHGC
HLNGKSTHYH CMQVGCNKVY TSTSDVMTHE NFHKKNTQLI NDGFQRFRAT EDCGTADCQF
YGQKTTHFHC RRPGCTFTFK NKCDIEKHKS YHIKDDAYAK DGFKKFYKYE ECKYEGCVYS
KATNHFHCIR AGCGFTFTST SQMTSHKRKH ERRHIRSSGA LGLPPSLLGA KDTEHEESSN
DDLVDFSALS SKNSSLSASP TSQQSSASLA AATAATEAGP SATKPPNSKI SGLLPQGLPG
SIPLALALSN SGLPTPTPYF PILAGRGSTS PPVGTPSLLG AVSSGSAASA TPDTPTLVAS
GAGDSAPVAA ASVPAPPASI MERISASKGL ISPMMARLAA AALKPSATFD PGSGQQVTPA
RFPPAQVKPE PGESTGAPGP HEASQDRSLD LTVKEPSNES NGHAVPANSS LLSSLMNKMS
QGNPGLGSLL NIKAEAEGSP AAEPSPFLGK AVKALVQEKL AEPWKVYLRR FGTKDFCDGQ
CDFLHKAHFH CVVEECGALF STLDGAIKHA NFHFRTEGGA AKGNTEAAFP ASAAETKPPM
APSSPPVPPV TTATVSSLEG PAPSPASVPS TPTLLAWKQL ASTIPQMPQI PASVPHLPAS
PLATTSLENA KPQVKPGFLQ FQEK

So the question is, if you're presented with a protein like CASZ1, which peptide do you pick to look for? In blue I've highlighted every tryptic peptide between 8 and 25 amino acids.

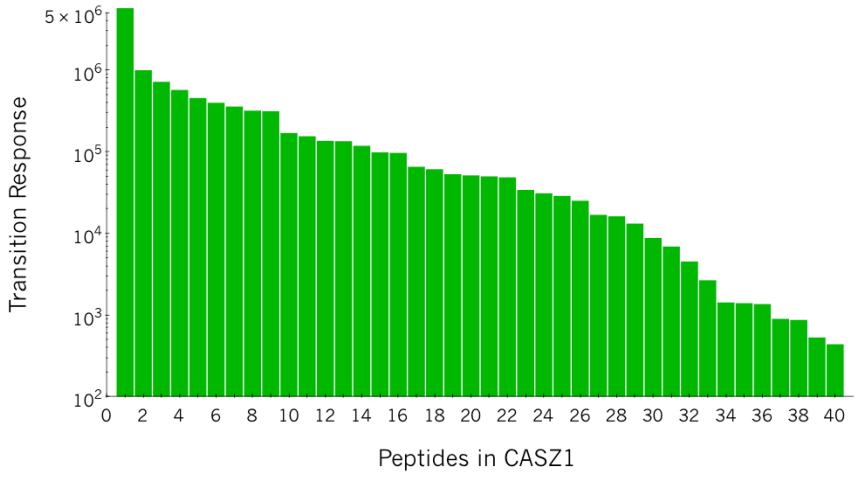
Which peptide do you pick?

>FLJ20321 (CASZ1 Castor Zinc Finger 1)

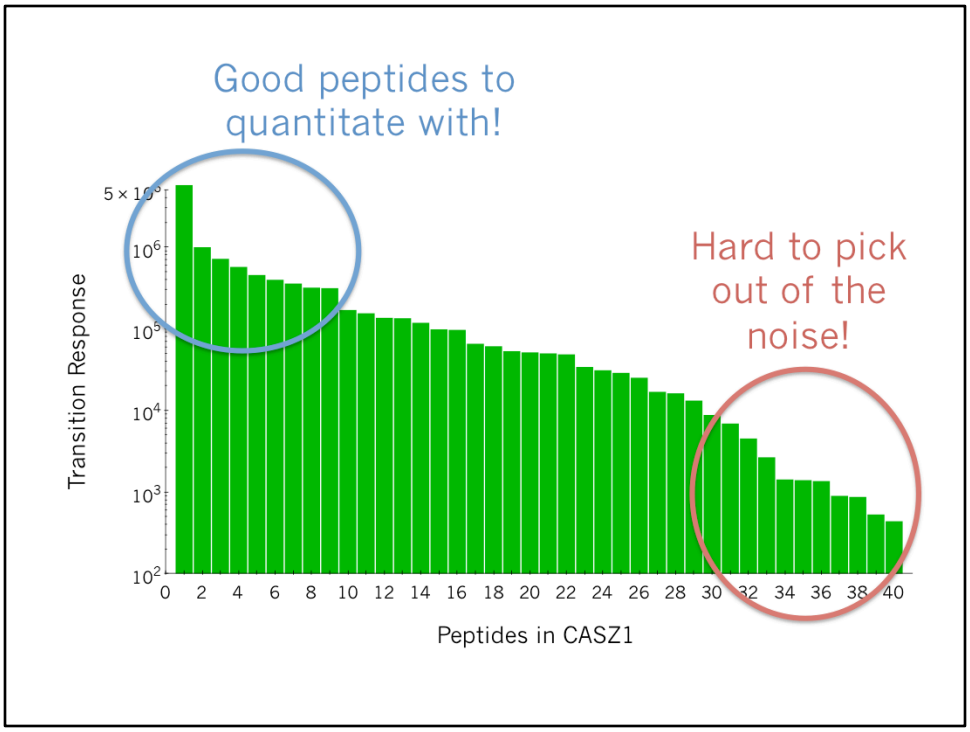
MVQPQGCSDE EDHAEEPSKD GGALEEKDSD GAASKEDSGP STRQASGEAS SLRDYAASTM
TEFLGMFGYD DQNTRDELAR KISFEKLHAG STPEAATSSM LPTSEDTLSK RARFSKYEEY
IRKLKAGEQL SWPAPSTKTE ERVGKEVVGT LPGLRLPSST AHLETKATIL PLPSHSSVQM
QNLVARASKY DFFIQKLKTG ENLRPQNGST YKKPSKYDLE NVKYLHLFKP GEGSPDMGGA
IAFKTGKVGR PSKYDVRGIQ KPGPAKVPPT PSLAPAPLAS VPSAPSAPGP GPEPPASLSF
NTPEYLKSTF SKTDSITTGT VSTVKNGLPT DKPAVTEDVN IYQKYIARFS GSQHCGHIHC
AYQYREHYHC LDPECNYQRF TSKQDVIRHY NMHKKRDNSL QHGFMRFSPL DDCSVYYHGC
HLNGKSTHYH CMQVGCNKVY TSTSDVMTHE NFHKKNTQLI NDGFQRFRAT EDCGTADCQF
YGQKTTHFHC RRPGCTFTFK NKCDIEKHKS YHIKDDAYAK DGFKKFYKYE ECKYEGCVYS
KATNHFHCIR AGCGFTFTST SQMTSHKRKH ERRHIRSSGA LGLPPSLLGA KDTEHEESSN
DDLVDFSALS SKNSSLSASP TSQQSSASLA AATAATEAGP SATKPPNSKI SGLLPQGLPG
SIPLALALSN SGLPTPTPYF PILAGRGSTS PPVGTPSLLG AVSSGSAASA TPDTPTLVAS
GAGDSAPVAA ASVPAPPASI MERISASKGL ISPMMARLAA AALKPSATFD PGSGQQVTPA
RFPPAQVKPE PGESTGAPGP HEASQDRSLD LTVKEPSNES NGHAVPANSS LLSSLMNKMS
QGNPGLGSLL NIKAEAEGSP AAEPSPFLGK AVKALVQEKL AEPWKVYLRR FGTKDFCDGQ
CDFLHKAHFH CVVEECGALF STLDGAIKHA NFHFRTEGGA AKGNTEAAFP ASAAETKPPM
APSSPPVPPV TTATVSSLEG PAPSPASVPS TPTLLAWKQL ASTIPQMPQI PASVPHLPAS
PLATTSLENA KPQVKPGFLQ FQEK

Even if we narrow that down to peptides that don't contain an oxidizable methionine, we're still forced to pick between a lot of peptides.

Equimolar peptides have wide ranging responses

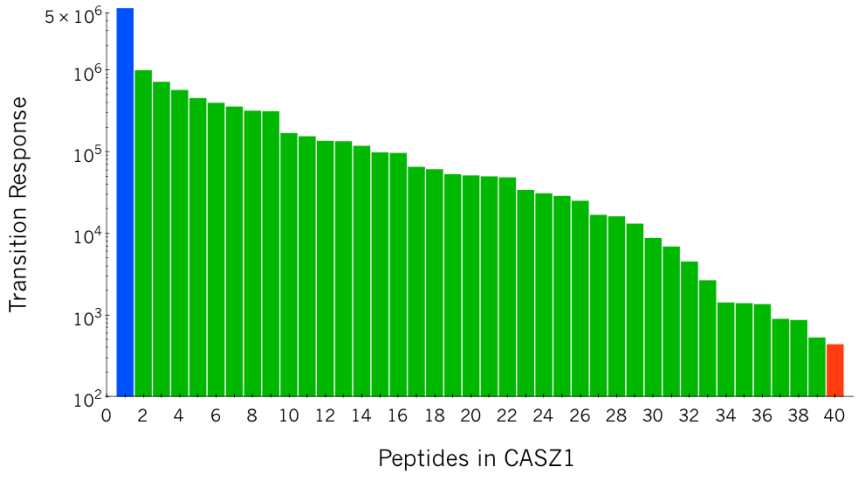


And if we measure them, these peptides all have widely different SRM responses. In this chart every bar is a peptide in CASZ1, sorted by highest y-ion transition intensity on a log scale.

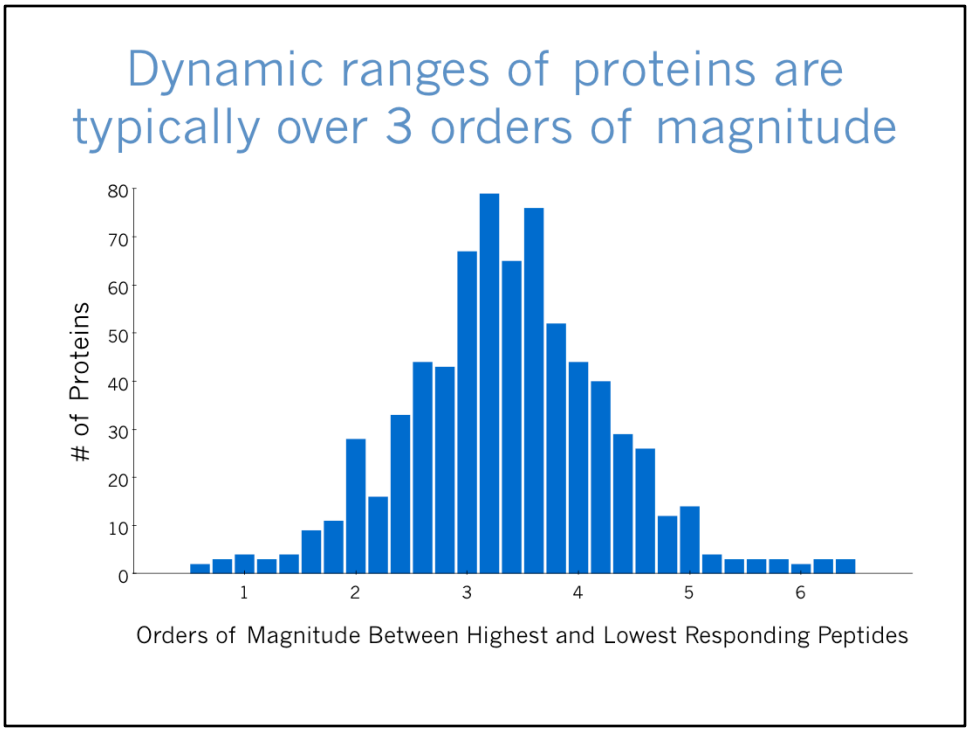


Clearly the peptides over on the left provide a much higher signal and the peptides on the right will be harder to pick out of the noise.

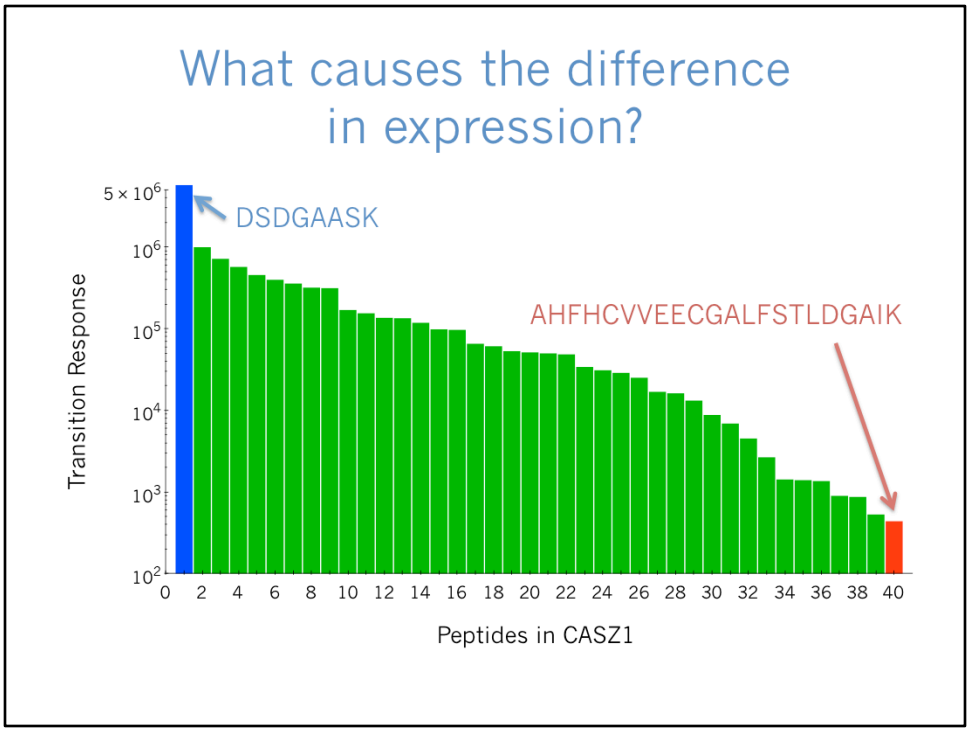
4 orders of magnitude between best and worst responding peptides



Between the best and worst peptides there's about four orders of magnitude in response.



And this range is not unusual. In an exhaustive experiment of over 700 proteins the dynamic range for each protein was typically between 3 and 4 orders of magnitude.



So what exactly makes the top peptide respond so much better than the bottom?

Many sources of variation cause equimolar peptides to produce different transition intensities

- Solubility problems
- Digestion efficiency
- Poor chromatography
- Variation in ionization
- Interference with matrix

  …

And the answer is that there are tons of confounding factors, involving prep-work, chromatography, ionization, and matrix competition.

How do we typically pick peptides?

- Prefer peptides with:
  - reasonable mass range (8-25 amino acids)
  - contain proline

Confronted with this complexity as a field we typically employ a handful of rules. First, peptides must fall in a reasonable mass range for detection. Peptides with prolines are preferred as they tend to direct fragmentation down certain predictable paths.

How do we typically pick peptides?

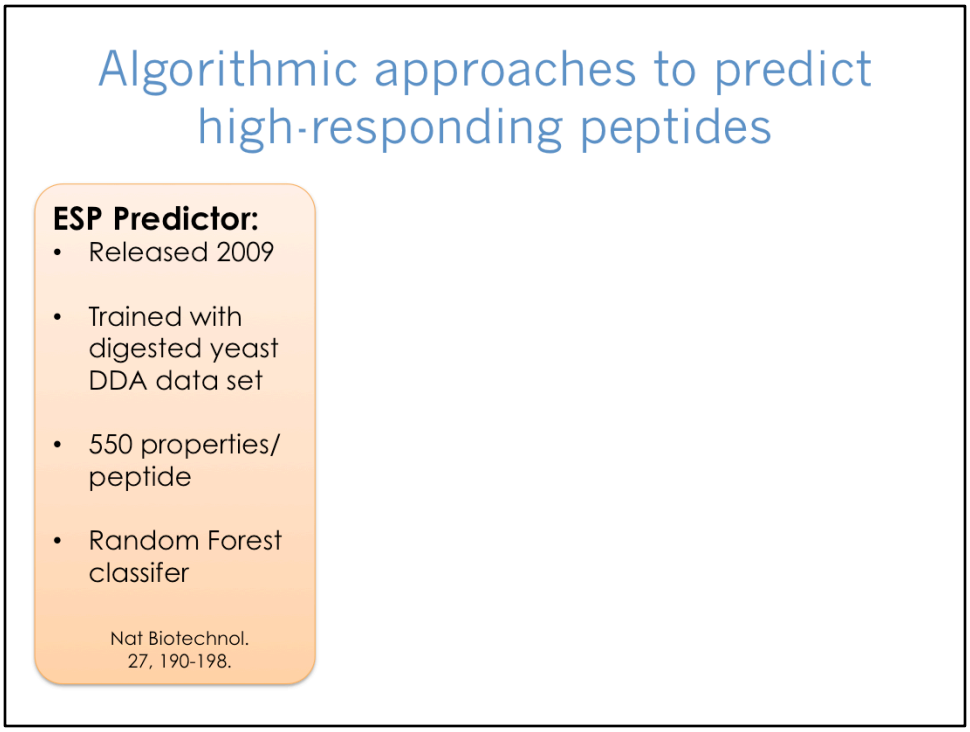
- Prefer peptides with:
  - reasonable mass range (8-25 amino acids)
  - contain proline
- Reject peptides with:
  - methionine (can be oxidized)
  - glutamine/asparagine (can be deamidated)
  - glutamine/glutamic acid/alkylated cysteine as the first residue (can cyclize)

Peptides with certain amino acids can undergo sample handling-based degredation, and so we tend to stay away from peptides with methionine, glutamine, and asparagine.

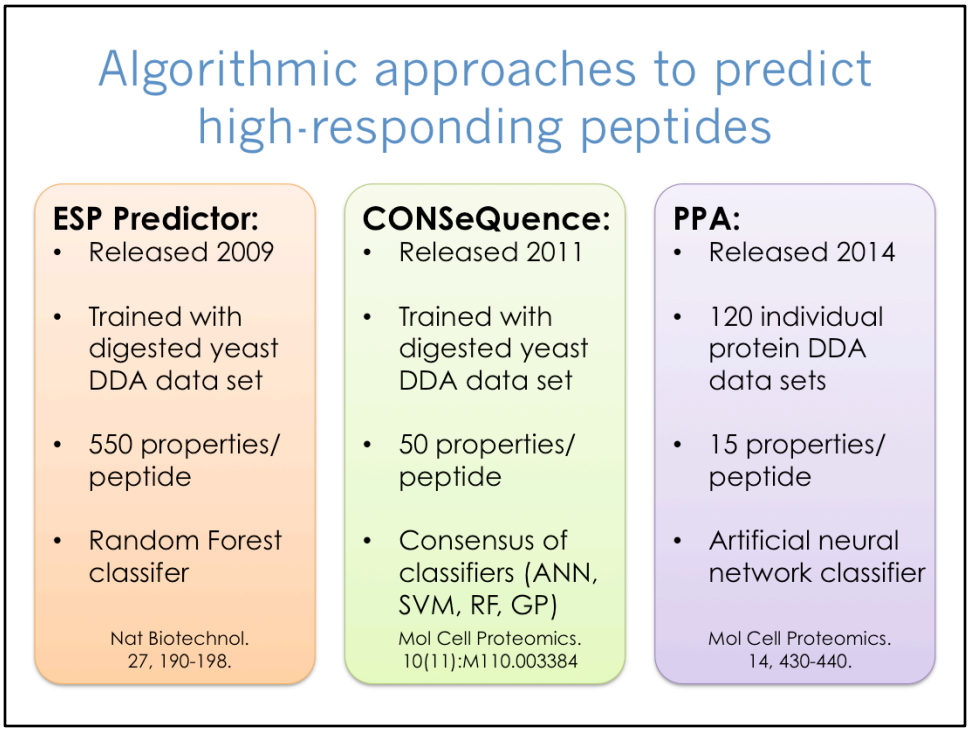
How do we typically pick peptides?

- Prefer peptides with:
  - reasonable mass range (8-25 amino acids)
  - contain proline
- Reject peptides with:
  - methionine (can be oxidized)
  - glutamine/asparagine (can be deamidated)
  - glutamine/glutamic acid/alkylated cysteine as the first residue (can cyclize)
- Randomly guess between what's left

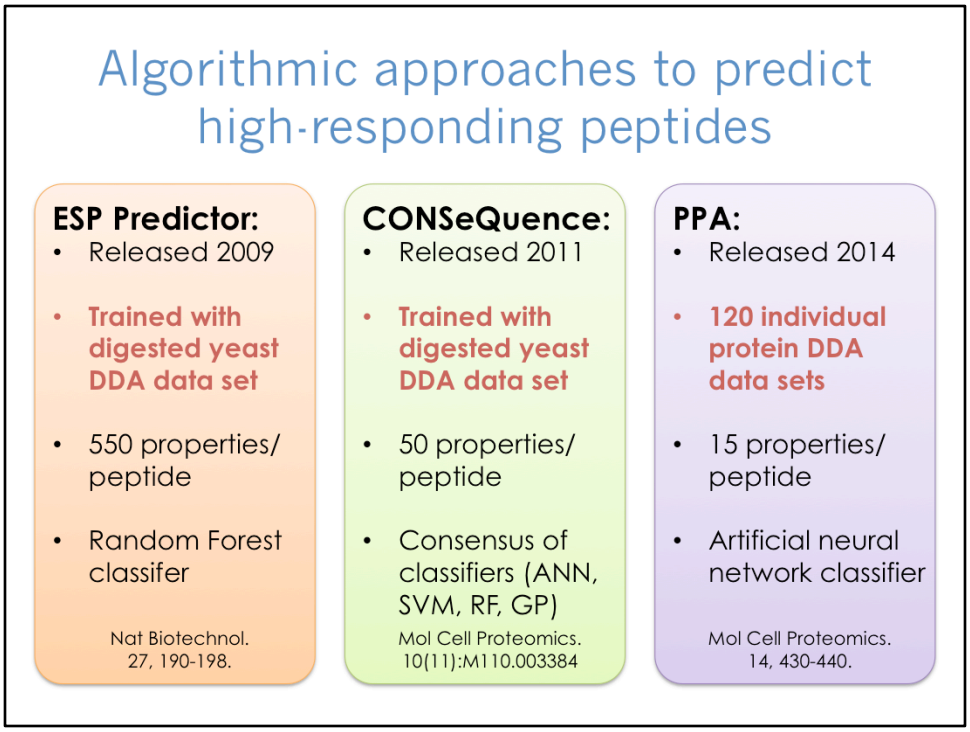
However, that often still leads several peptides to choose between and typically we just "guess and check".



Over the past few years several groups have attempted to predict so called “proteotypic”, or signature peptides, and a few have attempted to adopt this strategy for predicting good SRM peptides. ESP Predictor from the Broad was one early example in 2009, which was trained on a yeast DDA (or data dependent acquisition) data set, estimated 550 properties per peptide, and tried to label them as high and low responders using a random forest classifier.



More recently two newer methods have been published, CONSeQuence out of University of Manchester, and PPA out of Harvard, that purport to improve upon ESP using different methods and a reduced set of peptide properties.



One thing to note, though, is that all of these classifiers were trained using DDA data sets.

And I want to raise the question, are good SRM peptides the same as good DDA peptides?

Identifiable DDA peptides are not necessarily good SRM peptides

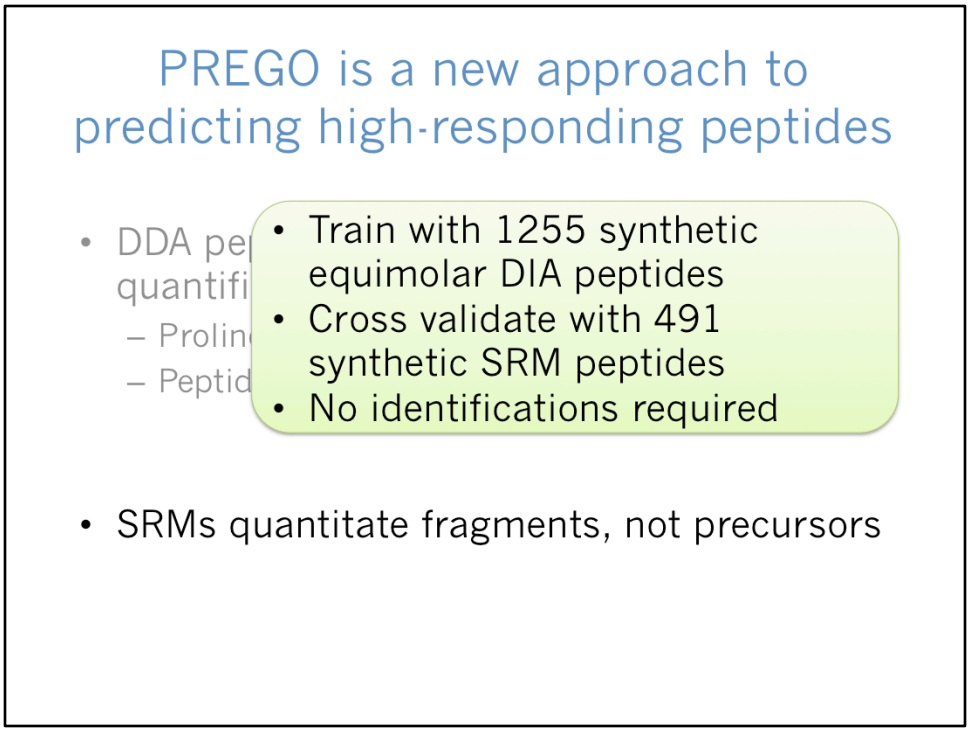
- DDA peptides must be identified before quantified
  - Proline makes great SRM transitions
  - Peptides with proline can be hard to ID

I would argue perhaps not. As we discussed earlier, peptides with proline generally make for great SRM transition targets, but on the other hand, those same strict fragmentation pathways make prolines hard to identify in DDA experiments.

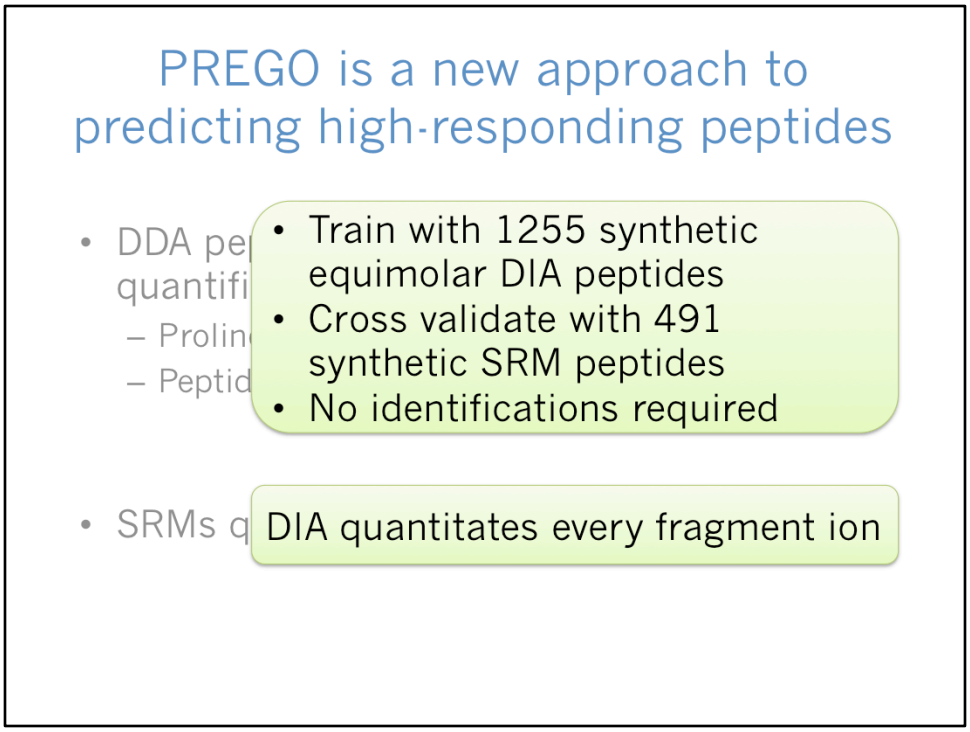
Identifiable DDA peptides are not necessarily good SRM peptides

- DDA peptides must be identified before quantified
  - Proline makes great SRM transitions
  - Peptides with proline can be hard to ID
- SRMs quantitate fragments, not precursors

Along those lines, in DDA experiments you quantitate on precursor ions, however in SRM experiments you quantitate on fragments. We've found that there can be an order of magnitude difference between these intensities.



We propose a new method to predict high-responding peptides, called PREGO. For this we developed a training data set of over 12 hundred synthetic, equimolar DIA peptides and a cross validation data set of almost 500 synthetic SRM peptides, both of which did not require identification.



In the DIA data sets we got quantitation of every viable y-ion fragment for free. In our 500 peptide SRMs we also every single viable y-ion.

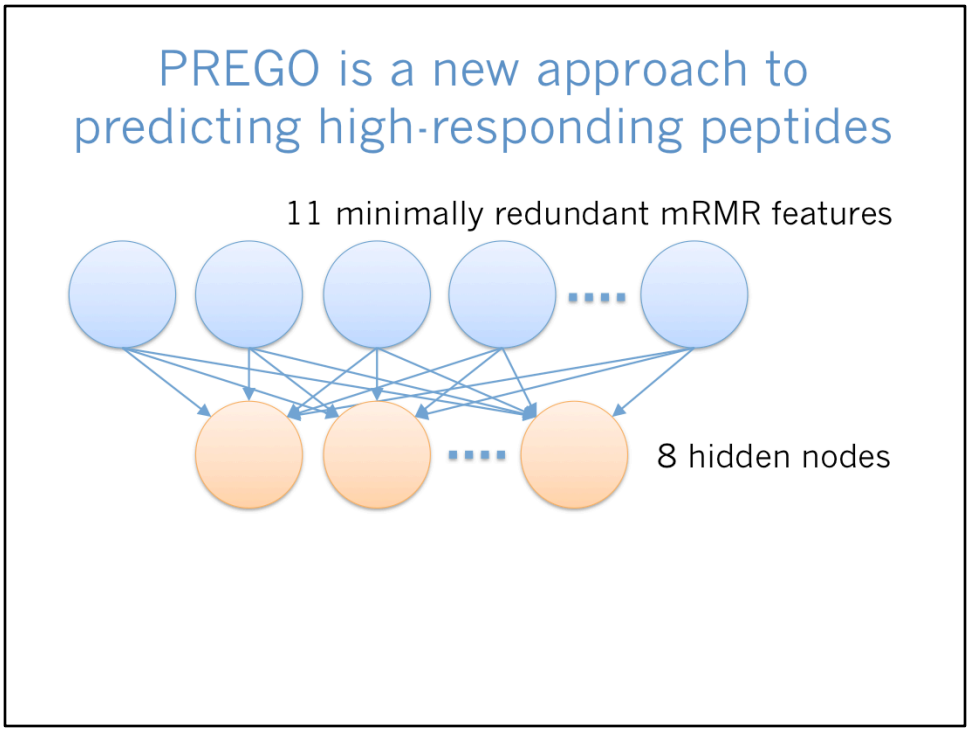
PREGO is a new approach to predicting high-responding peptides

11 minimally redundant mRMR features

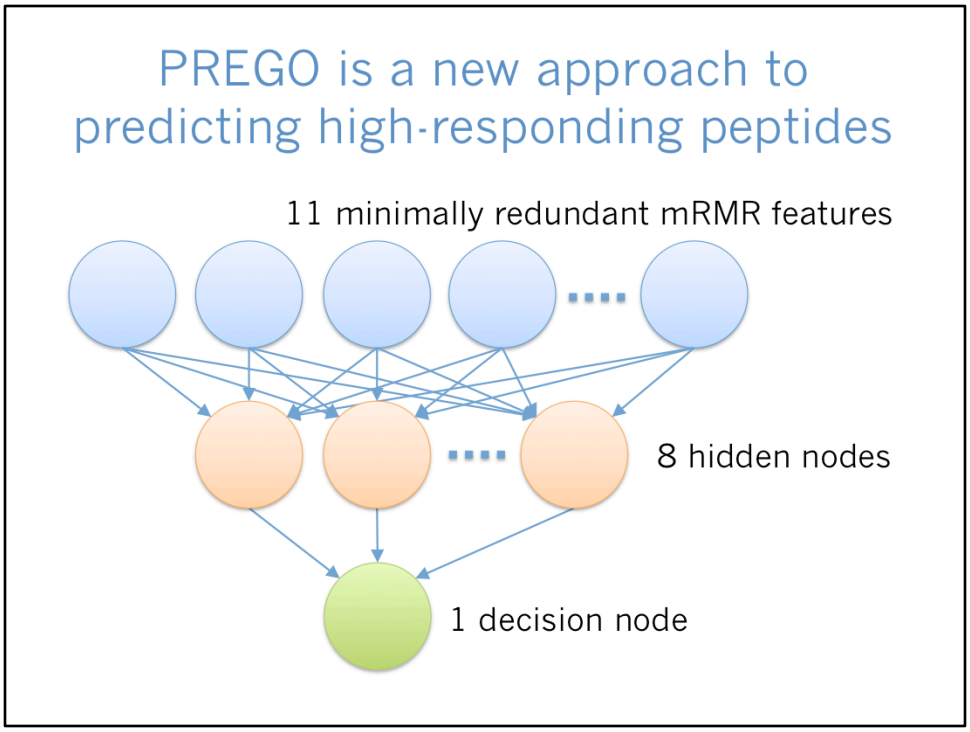
minimum Redundancy Maximum Relevance algorithm
for all 550 features:
a) calculate correlation to the transition intensities
b) keep feature that correlates the best
c) discard features that are similar
repeat until there are no features left

Ding C, Peng HJ. (2005) Bioinform Comput Biol. 3, 185-205.

Using the same 550 feature set all of the previous three groups employed, we narrowed down our list to 11 minimally redundant features using an mRMR algorithm originally developed to sift through gene expression data. This approach works by correlating every feature to the set of transition intensities, keeping the feature that correlates the best, and then discarding other features that correlate with that one. This process is repeated until there are no features left.



We fed these features into an artificial neural network



that predicted a response score for each peptide.

PREGO was validated using an enormous SRM data set

- 724 synthetic proteins
- generates 12973 peptides
- each protein analyzed independently
- exhaustive SRM sampling (every y_3 to y_{n-1} ion)

Stergachis AB, et al (2011) Nat Methods. 8, 1041-1043.

We validated PREGO using an exhaustive SRM experiment containing almost 13 thousand peptides from over 700 proteins. In this data set each protein every sequence specific y-ion was sampled, allowing us to know what the best responding transition was.

How do we measure success?

Our bar was relatively low:

We just wanted to do better than randomly guessing after using a few peptide selection rules

And ultimately, the bar for success is rather low. Really we just have to do better than randomly guessing after using the amino acid specific rules, such as preferring prolines.

How do we measure success?

PREGO

Which is actually how we named our approach, where we hoped that Prego...

How do we measure success?

>?

PREGO

was subtly better than Ragu...

How do we measure success?

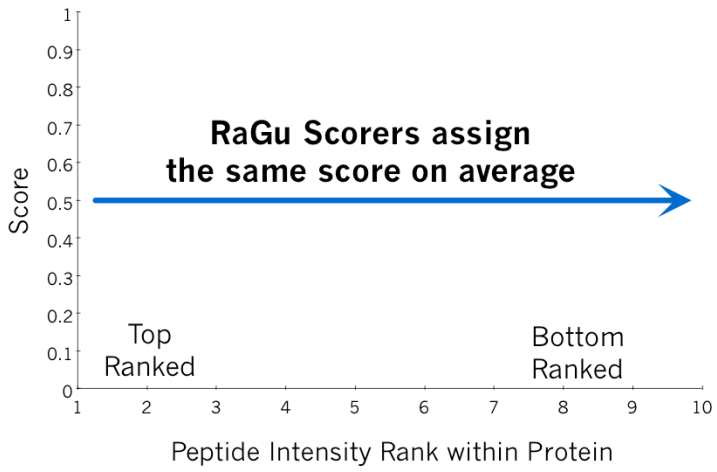
>?

PREGO

RaGu
(Randomly Guessing)

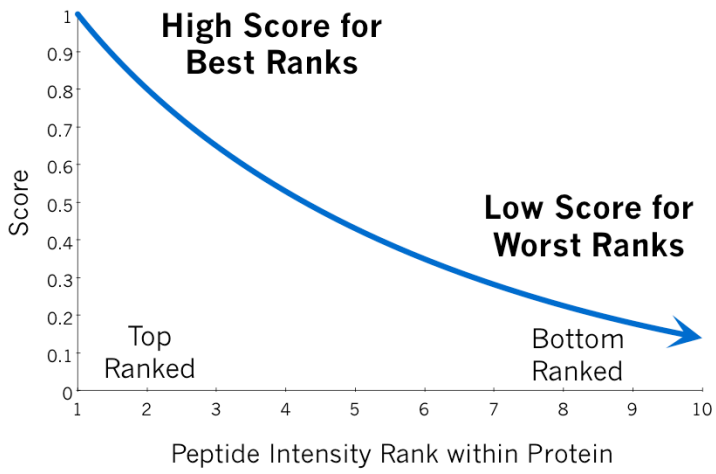
Our code for randomly guessing between the peptides.

Scoring peptides within a protein

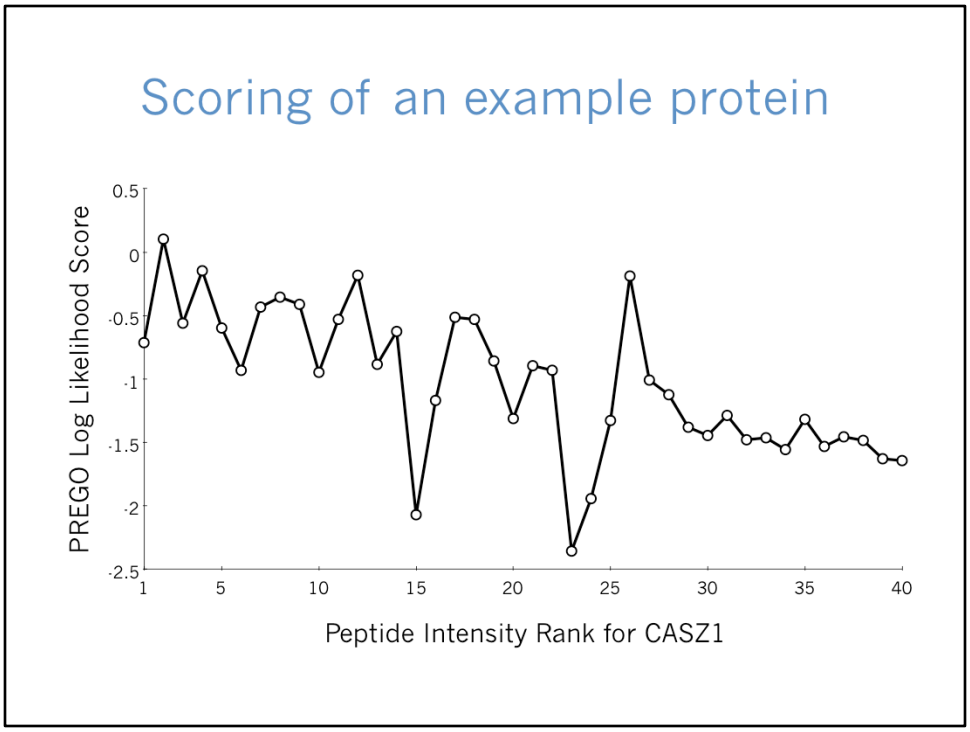


On average, RaGu scorers should assign every peptide the about the same score, no matter if they were top ranked, or bottom ranked peptides in our exhaustive SRM experiment.

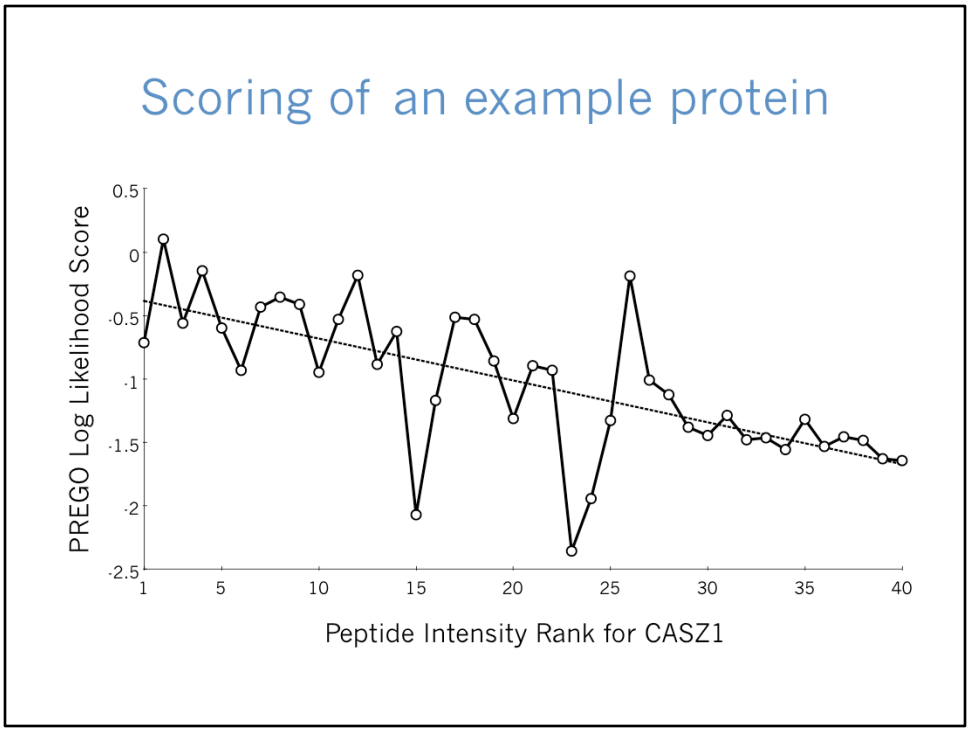
Scoring peptides within a protein



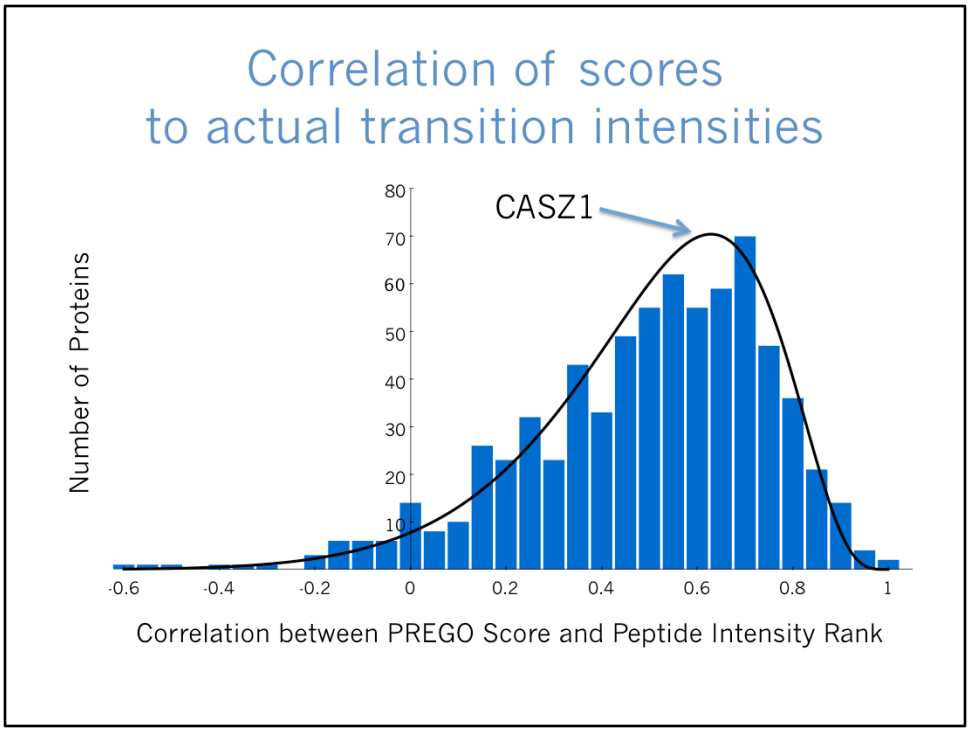
On the other hand, a successful method should assign high scores to top ranked peptides, and low scores to bottom ranked peptides.



Here are the 40 appropriate peptides in CASZ1 scored by PREGO.

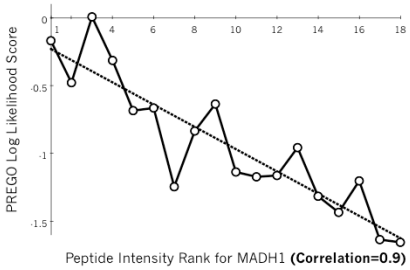


Now, there's a lot of variability in this scoring system, which underlines the large complexity of this problem, but the trend is that top ranked peptides generally get higher scores than bottom ranked peptides.



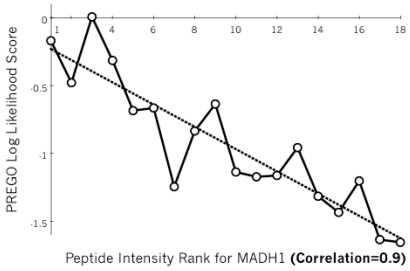
Looking at a correlation analysis between PREGO scores and true intensity ranks, PREGO's success with CASZ1 is pretty typical.

Correlation of scores to actual transition intensities

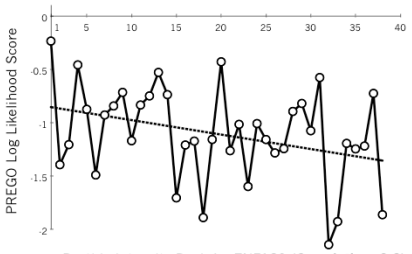


To give you an idea that I'm not cherry picking well behaved proteins, here's a protein at the top end of the correlation spectrum,

Correlation of scores to actual transition intensities

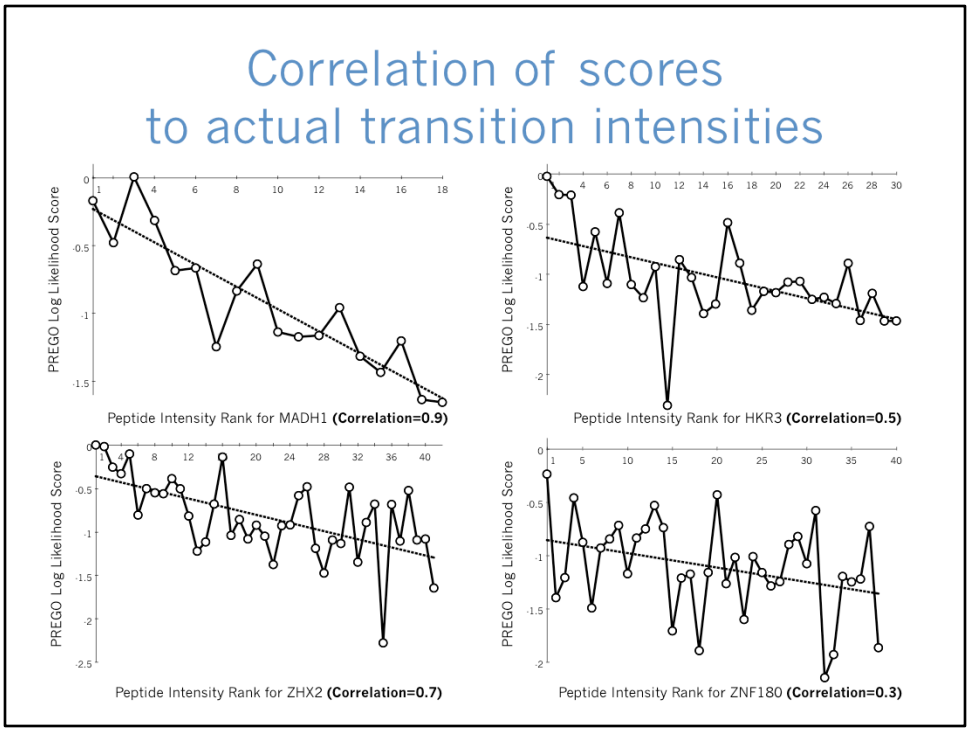


Peptide Intensity Rank for MADH1 **(Correlation=0.9)**

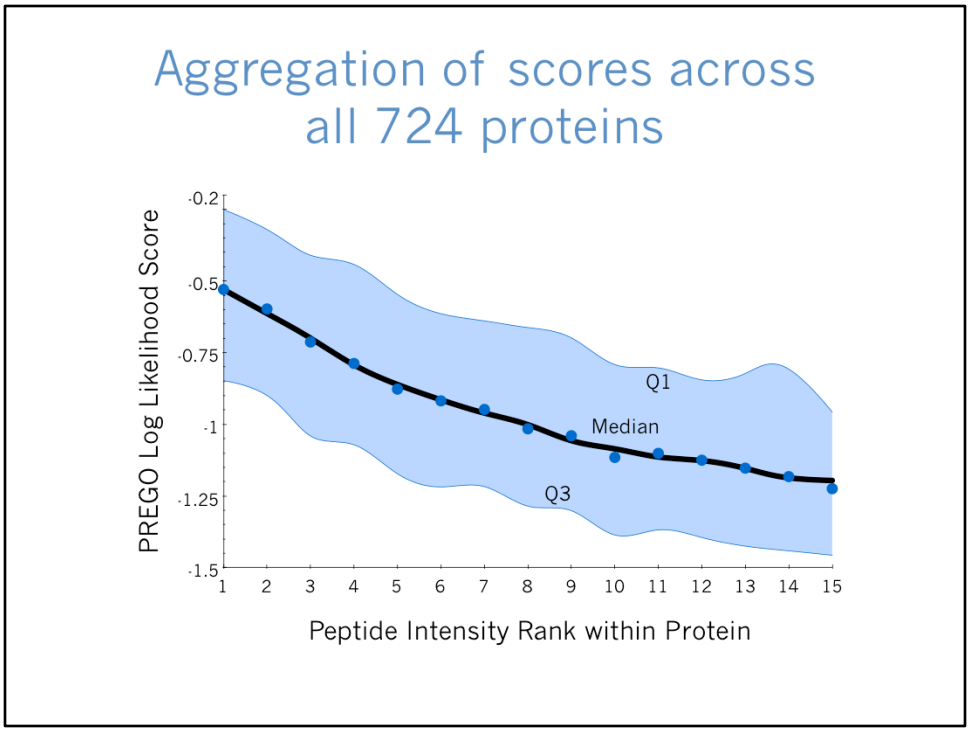


Peptide Intensity Rank for ZNF180 **(Correlation=0.3)**

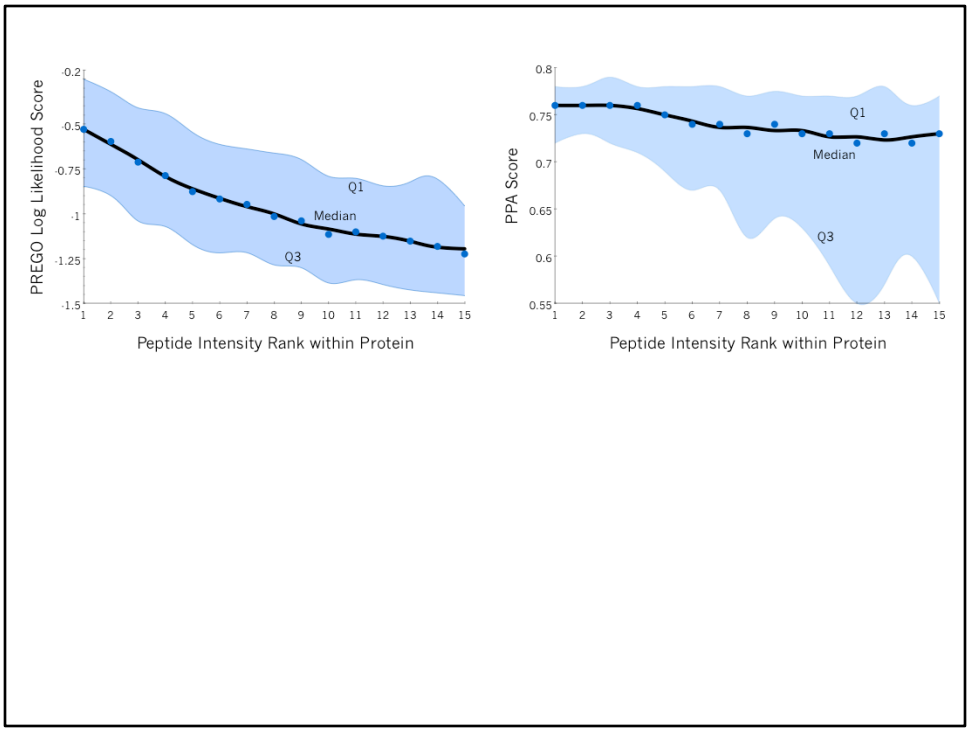
and here's a protein at the bottom end.



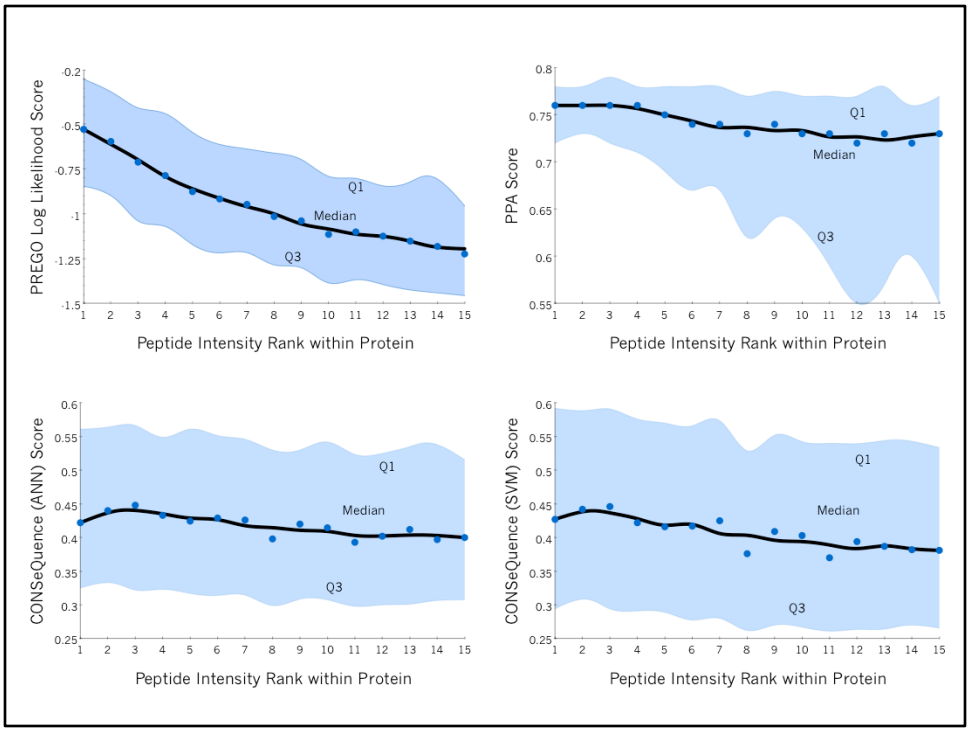
These proteins are somewhere in between. You can see that they all still maintain a downward trend relative to real SRM data.



If we aggregate the top 15 peptides for all 700 proteins, we get this trace, where I've drawn the downward sloping median and interquartile range. There's clearly a lot of variability in the PREGO scores, but the trend is clear.



Now if I compare that against PPA we see a different trend. PPA shows some downward trend, but it still tends to assign high scores to peptides of low rank.

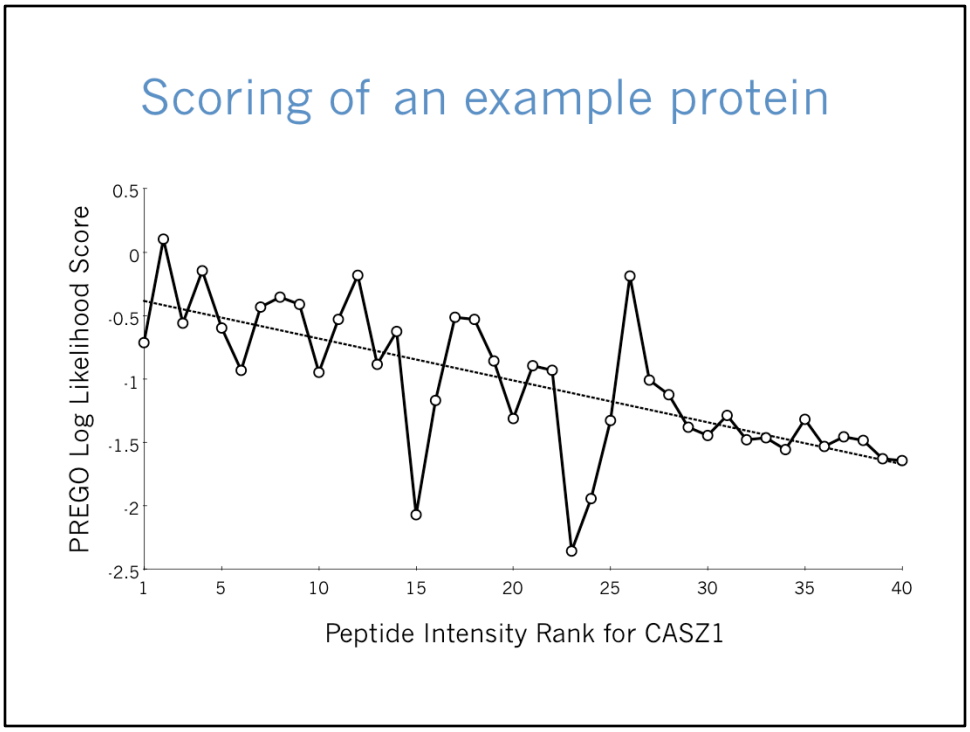


Similarly, CONSeQuence tends to have a slight downward trend, but a significantly high variability.

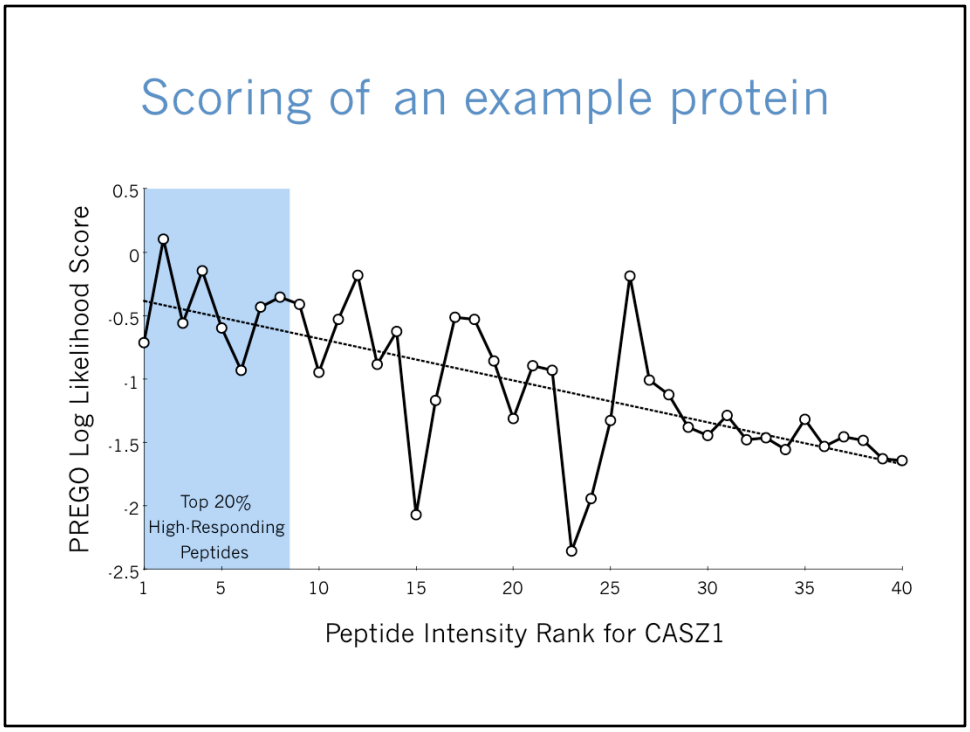
How do we pick peptides for SRM experiments?

- Pick 2-5 peptides per protein
- Run some test experiments
- Hope that one works well

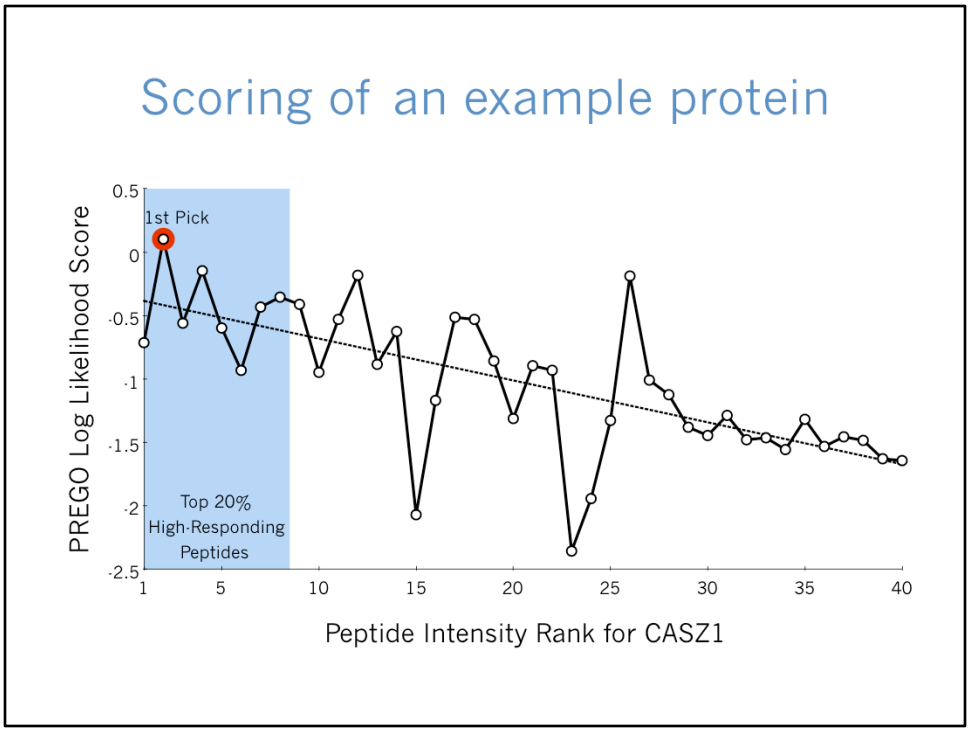
So this is great, PREGO generates scores that are somewhat meaningful. But how does this relate to real SRM experiments? Typically in these experiments we pick a handful of peptides per protein, run some tests, and hope that at least one of them produces a reasonable signal.



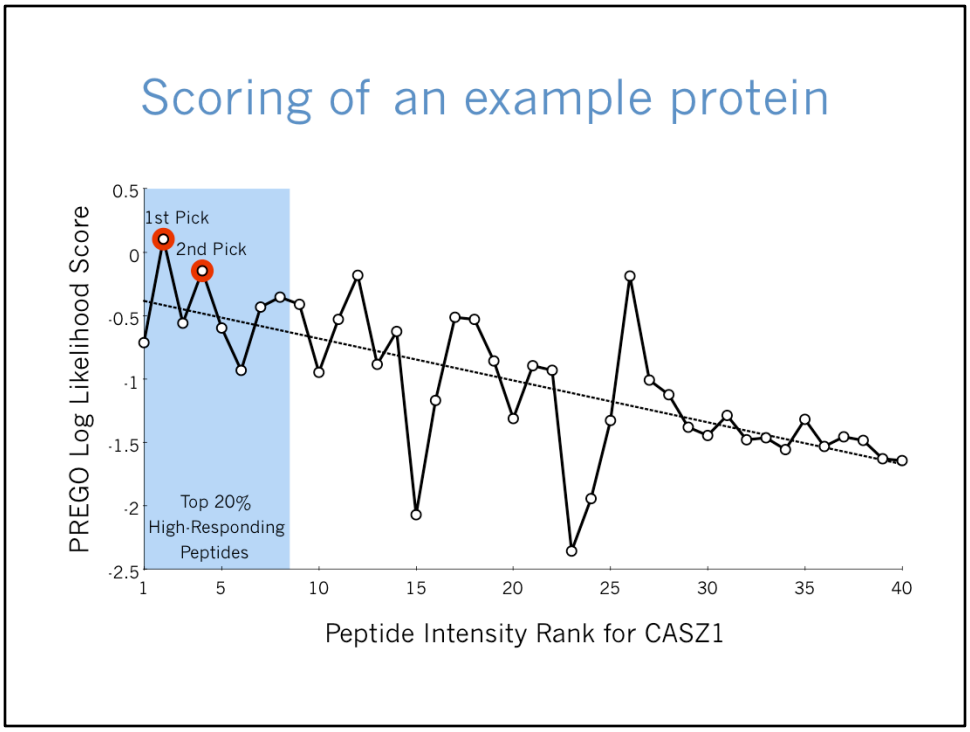
if we go back to CASZ1,



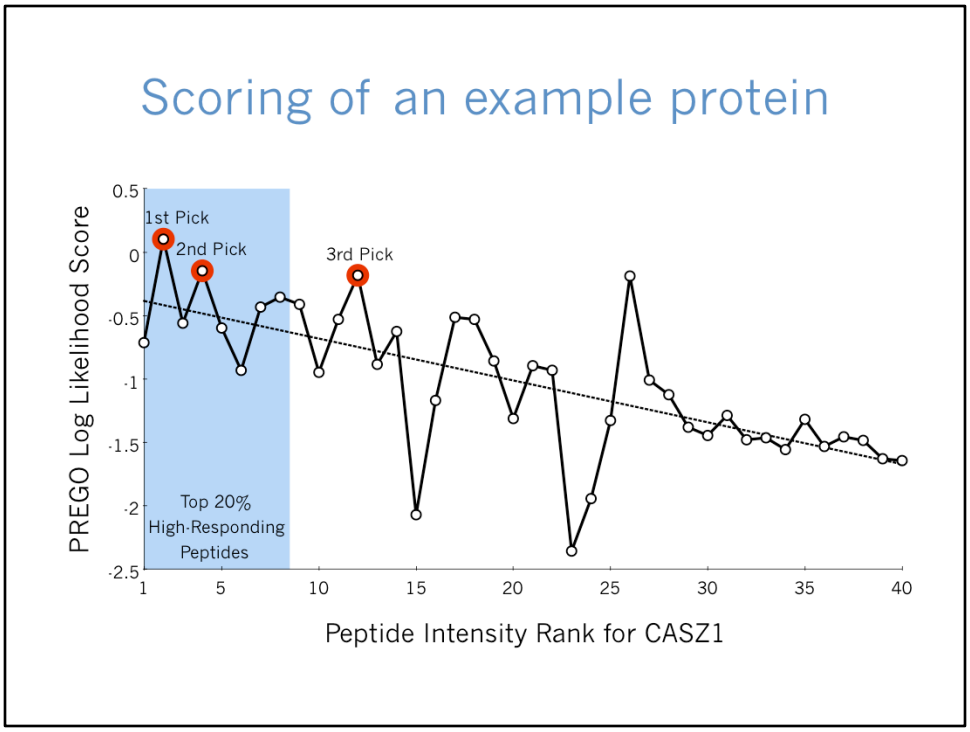
Let's say that peptides with good signal are in the top 20%.



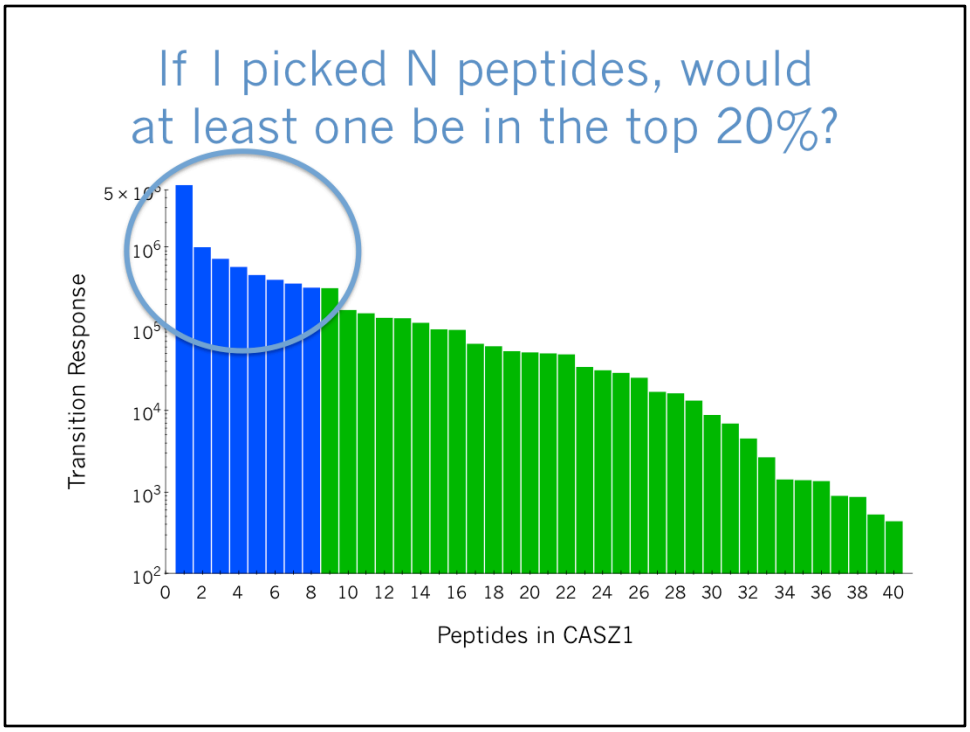
If we go in order of PREGO scores, the top pick is in this range,



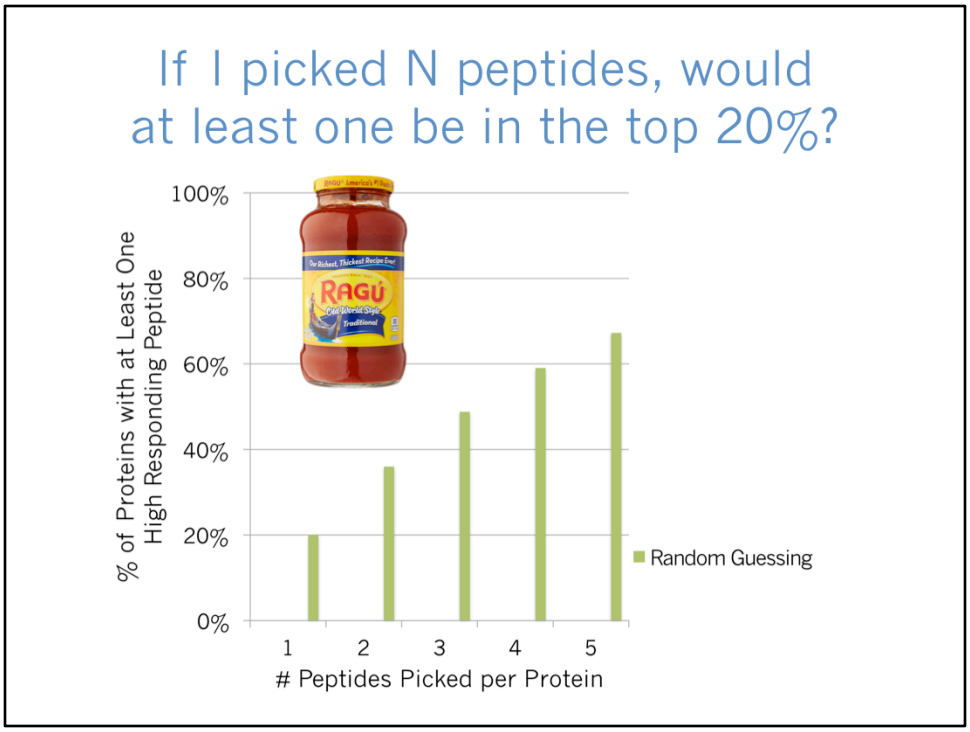
and so is the second,



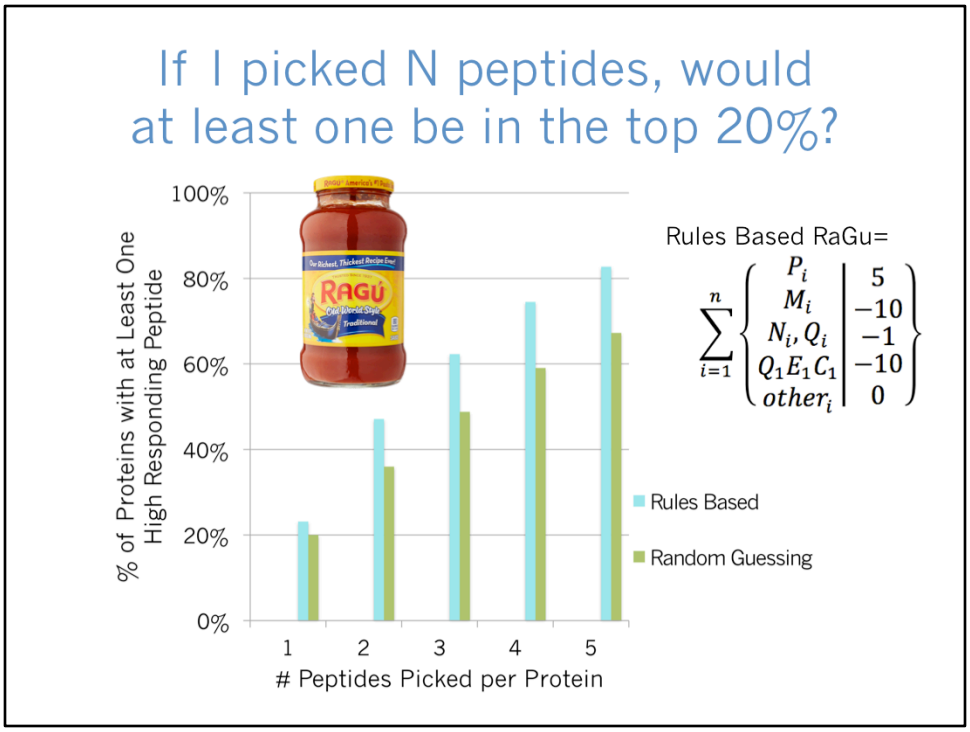
but the third we would classify as not a “top responder”. In this particular protein PREGO would get 2 out of 3 and do pretty well.



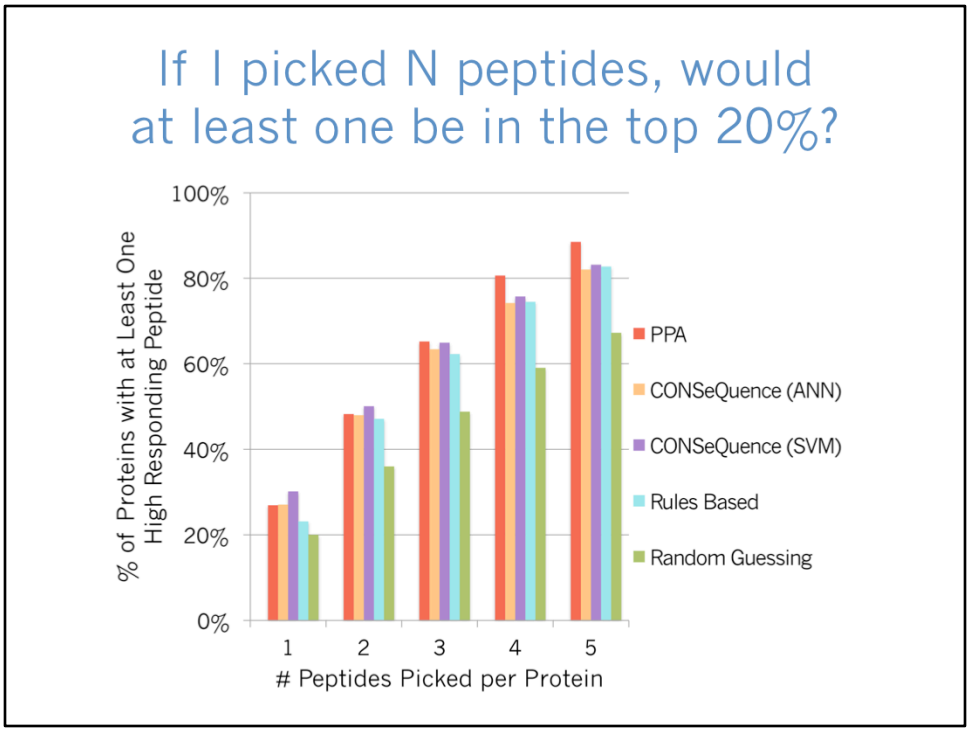
So let's ask the question across all proteins, if we picked N peptides, would at least one be in the top 20%?



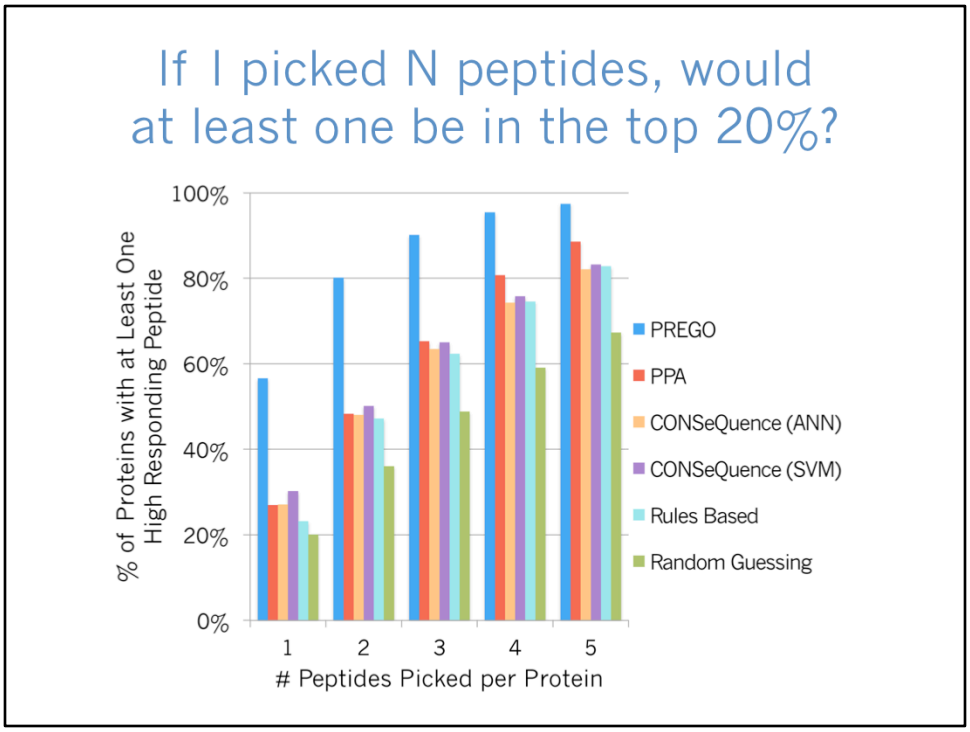
Here is the mathematical probability of picking a peptide in the top 20% guessing completely at random if we were to pick 1 through 5 peptides.



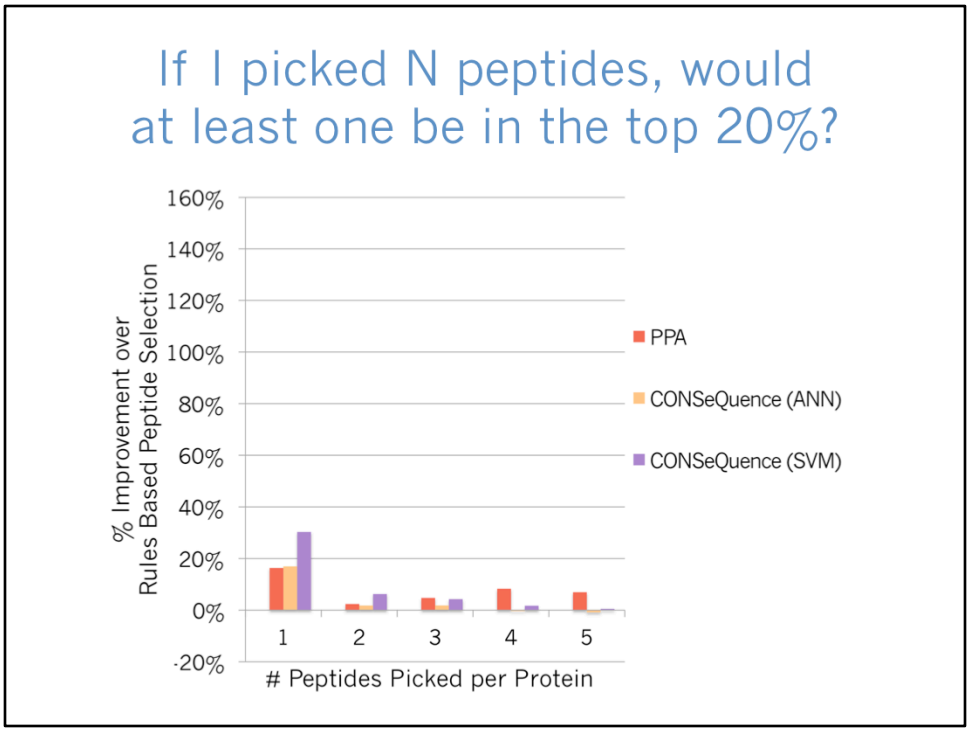
Honestly, though, this is a straw-man comparison, as we typically use rules to pick peptides as we discussed at the beginning. I've generated a basic rules-based random guesser that scores peptides based on some basic rules, and picks between the top scoring peptides. This models a typical SRM assay designing workflow, and as one would hope, this does better than guessing truly at random.



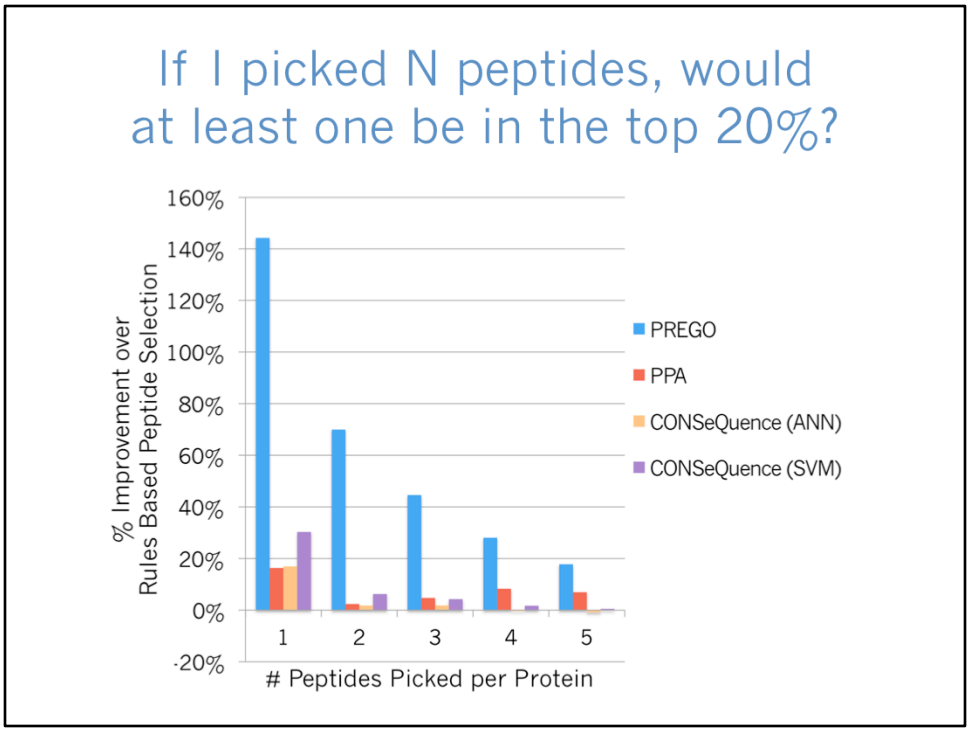
Now if I compare that to PPA and CONSeQuence, it's not clear that either of these algorithms perform any better than picking based on simple rules. This probably explains why adoption of algorithmic peptide picking strategies have relatively low penetration in real SRM assay development.



PREGO, on the other hand, does considerably better than rules-based random guessing or the competing algorithms.



Let me show this as the improvement over a rules based model. PPA and CONSeQuence perform slightly better on the first guess than picking at random, but it's a wash after that.



PREGO maintains improvement over picking with rules throughout the top 5 peptides. I want to stress that I don't think PPA and CONSeQuence are worthless tools. The extremely similar behavior from these two very differently architected tools strongly suggests that while they don't perform well at the task they were INTENDED to solve, i.e. picking high responding peptides for SRM quantitation, they probably perform very well at the task they were TRAINED to do, which was to pick the best peptides for DDA identification. If this result is correct, then it's probably also not good to use spectral counts or spectral libraries as a proxy for good SRM peptides either.

Why does PREGO do so well?

- ANN design and mRMR features help but are probably minor contributors

So why does PREGO perform so well? Realistically tuning the neural network and feature selection probably helped a little, but not that much.

Why does PREGO do so well?

- ANN design and mRMR features help but are probably minor contributors
- DIA fragment intensities are a better model for SRM fragment intensities than DDA

We feel the biggest improvement towards picking peptides for SRM experiments is to use training data that actually models SRM experiments.

Why does PREGO do so well?

- ANN design and mRMR features help but are probably minor contributors
- DIA fragment intensities are a better model for SRM fragment intensities than DDA
- Equimolar training peptides avoids ranking within proteins

An additional factor is probably that using equimolar training peptides helped avoid the problem of protein intensities confounding peptide responses.

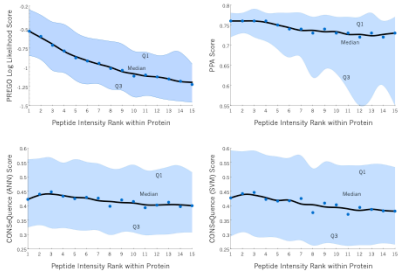
Why does PREGO do so well?

- ANN design and mRMR features help but are probably minor contributors
- DIA fragment intensities are a better model for SRM fragment intensities than DDA
- Equimolar training peptides avoids ranking within proteins
- SRM data was used for cross validation

Finally, using actual SRM data in the training process again was key to generating a successful classifier.

Take homes:

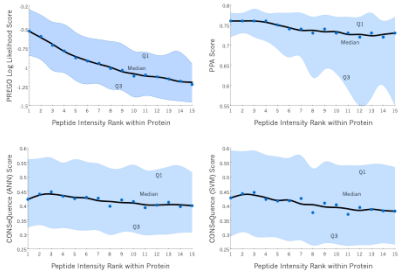
- Machine learning architecture is less important than training data set quality



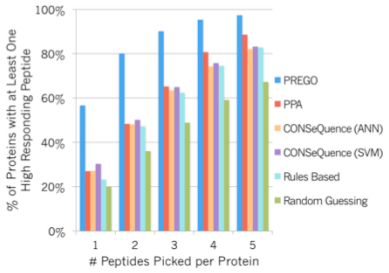
If I may be so bold as to present some basic machine learning take home messages, I would posit that the architecture used for machine learning is significantly less important than the

Take homes:

- Machine learning architecture is less important than training data set quality



- Training data must generalize to the desired problem



training data sets ability to generalize to the desired problem.

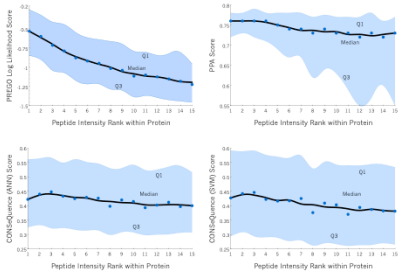
Acknowledgements

In particular:
Mike MacCoss
Jarrett Egertson
Jim Bollinger
Andrew Stergachis
Rich Johnson
Vagisha Sharma

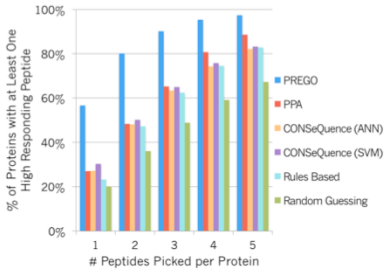
I'd like to thank the MacCoss lab for hosting me, in particular my co-authors, Mike, Jarrett, Jim, and Andrew, as well as Rich and Vagisha for helping with the lab and computer work, respectively.

Take homes:

- Machine learning architecture is less important than training data set quality



- Training data must generalize to the desired problem



With that I'd like to take any questions.