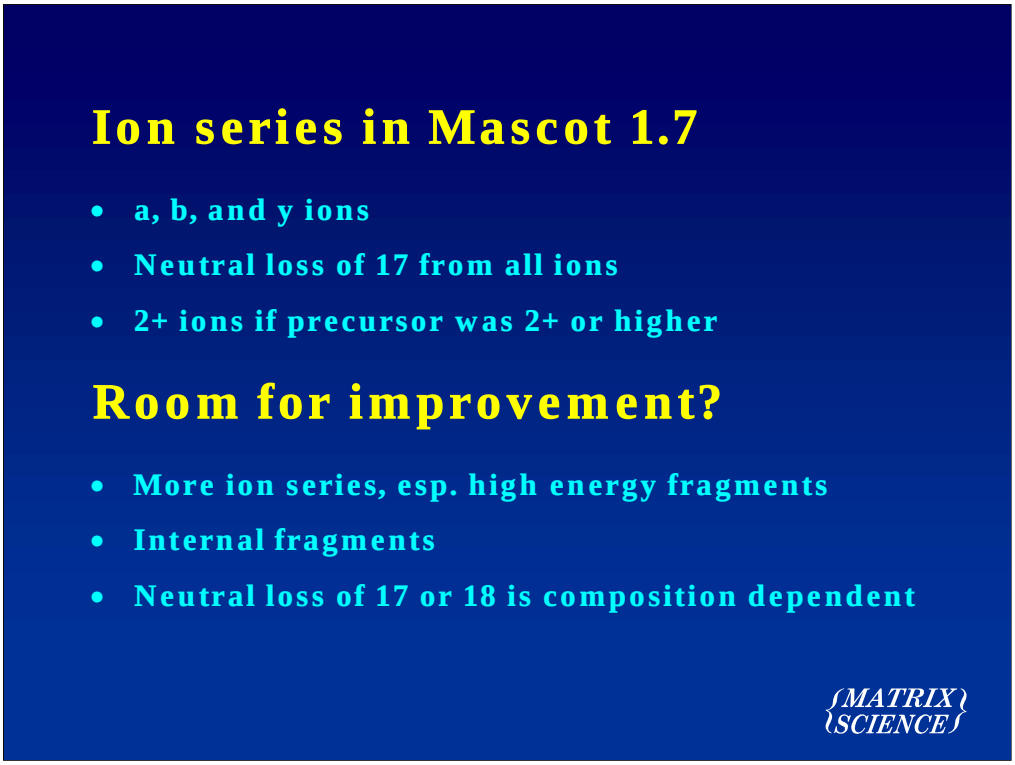
Instrument Specific MS/MS Ion Series Matching

MATRIX SCIENCE

This talk describes one of the new features in Mascot 1.8 that allows us to fine-tune the matching of MS/MS data

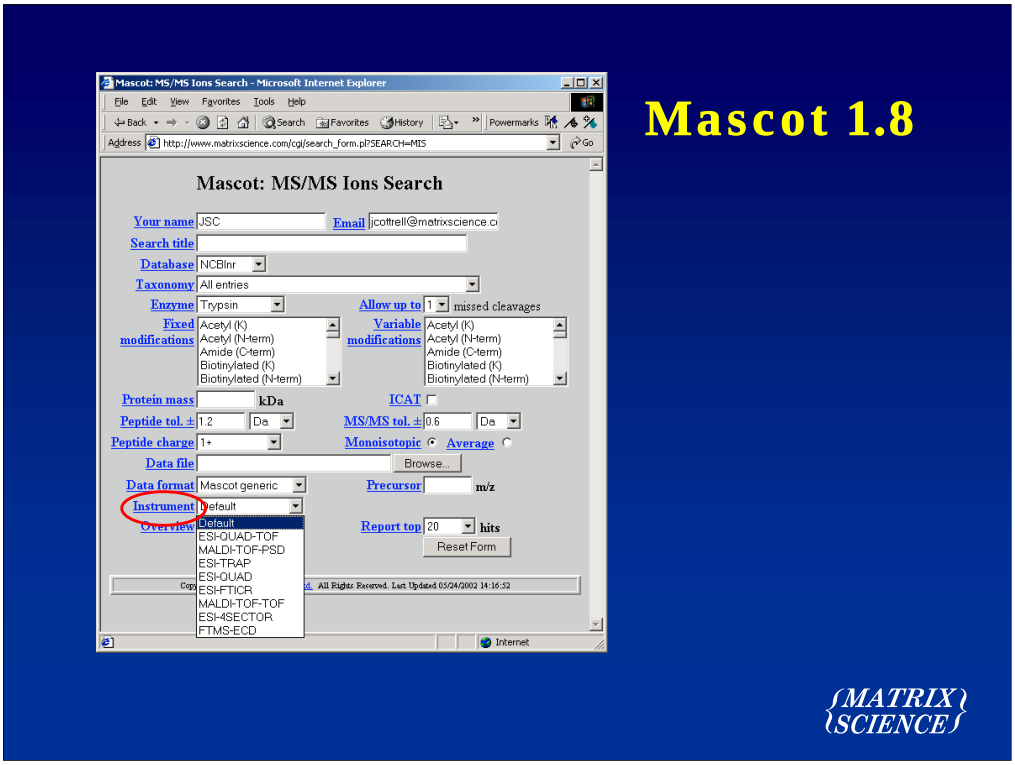


In Mascot 1.7, all MS/MS data was matched and scored using the a, b, and y ion series. In addition to the intact ion series, Mascot also looked for neutral loss of ammonia and, if the precursor charge was 2+ or higher, the 2+ ions series.

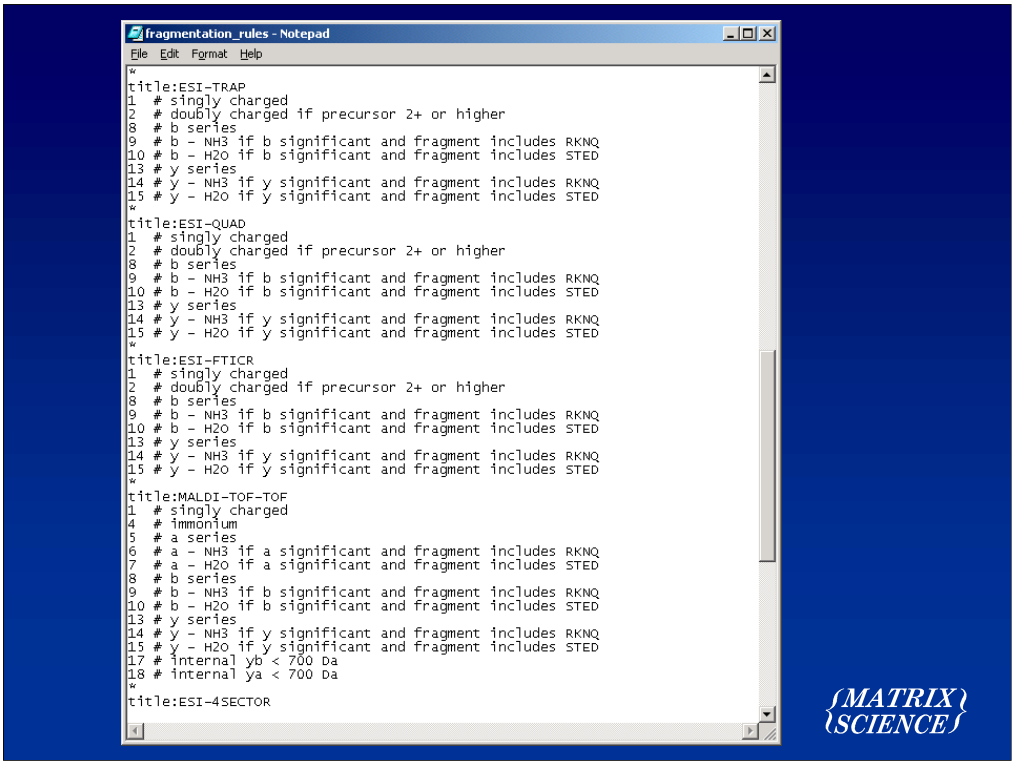
This works perfectly well for most data, such as electrospray on an ion trap or quadrupole TOF, or MALDI PSD. However, there are other cases where it would be desirable to have greater flexibility.

In particular, new instruments such as the TOF/TOF, generate peaks in high energy ion series plus lots of immonium ions and internal fragments.

We also wanted to handle neutral losses more accurately, because loss of ammonia or water is actually composition dependent.

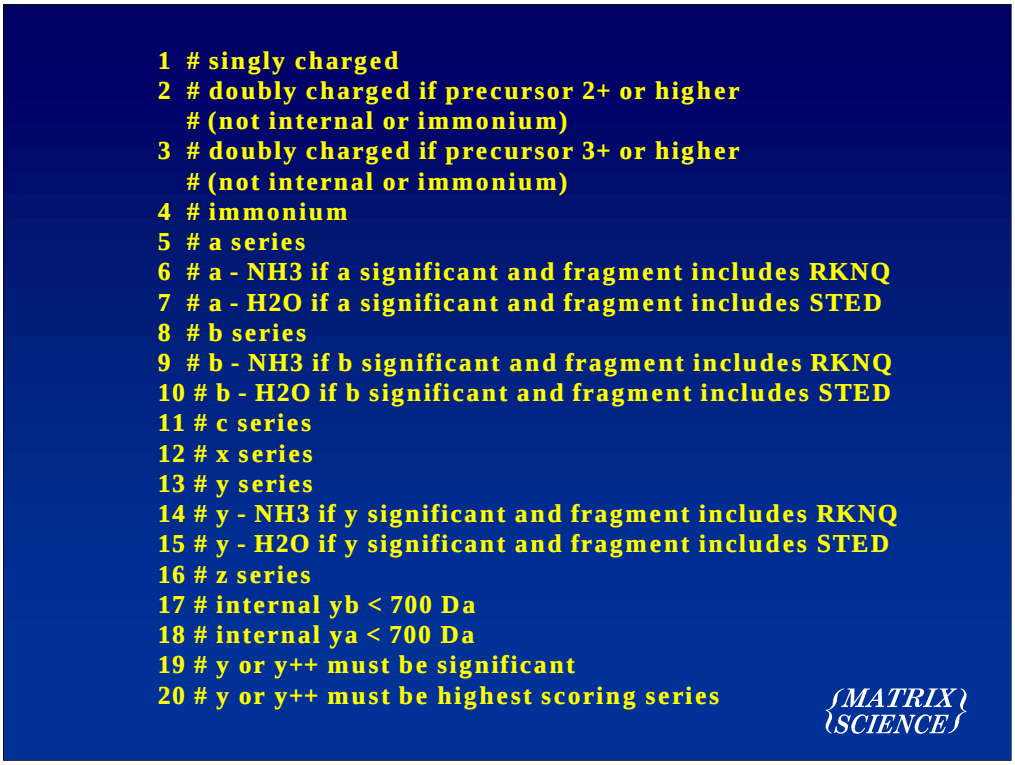


In Mascot 1.8, we now have an additional field on the search form: INSTRUMENT. This is a drop down list of all the common instrument configurations. For backward compatibility, the Default setting gives the same behaviour as Mascot 1.7



When you choose one of the instrument types, the ion series used in scoring are taken from a configuration file. Like other Mascot configuration files, this is a simple text file that you can edit to modify existing instrument settings or add new ones.

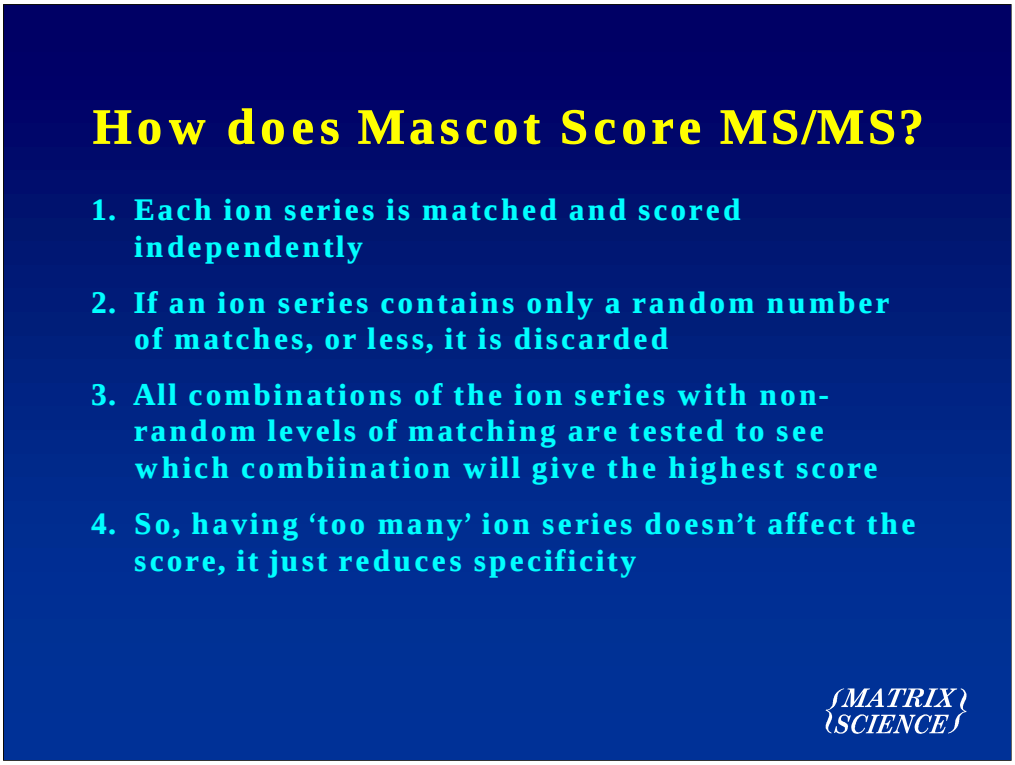
Here we can see some of the settings in the standard file. The ESI trap, for example, just looks at b and y ions



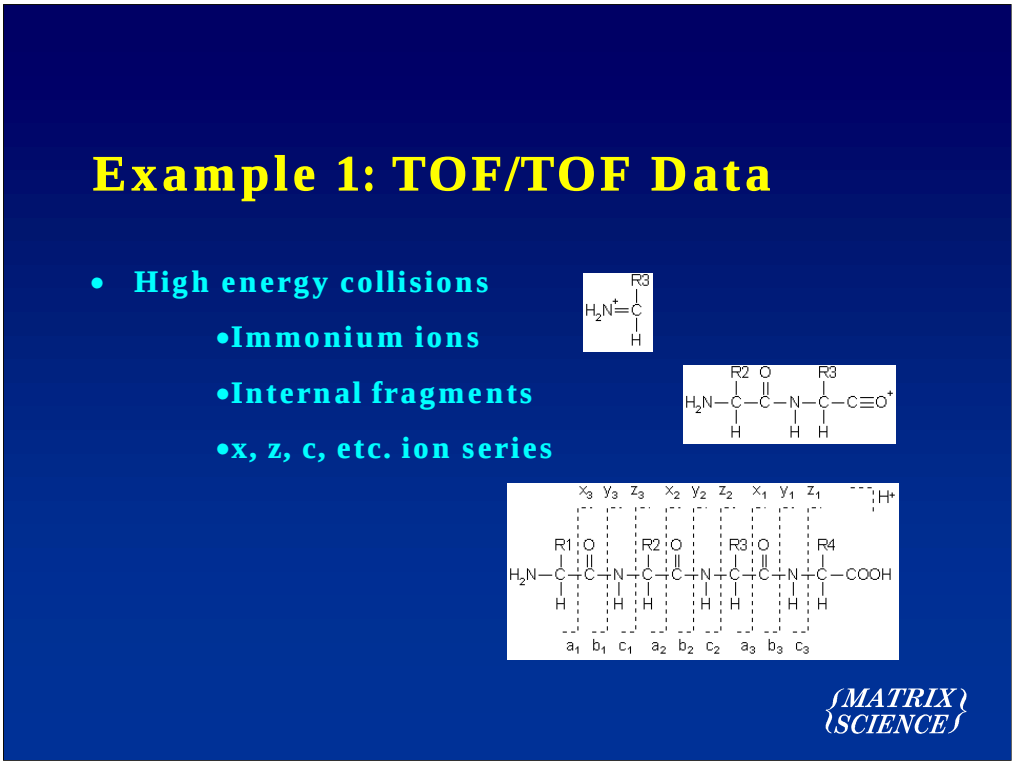
Here, you see the full list of choices for INSTRUMENT configuration in Mascot 1.8.

Notice that some of the choices are more like rules. For example, #20 says that y or y++ must be the highest scoring series.

We may add more such rules in future releases, and would be very interested in hearing your suggestions

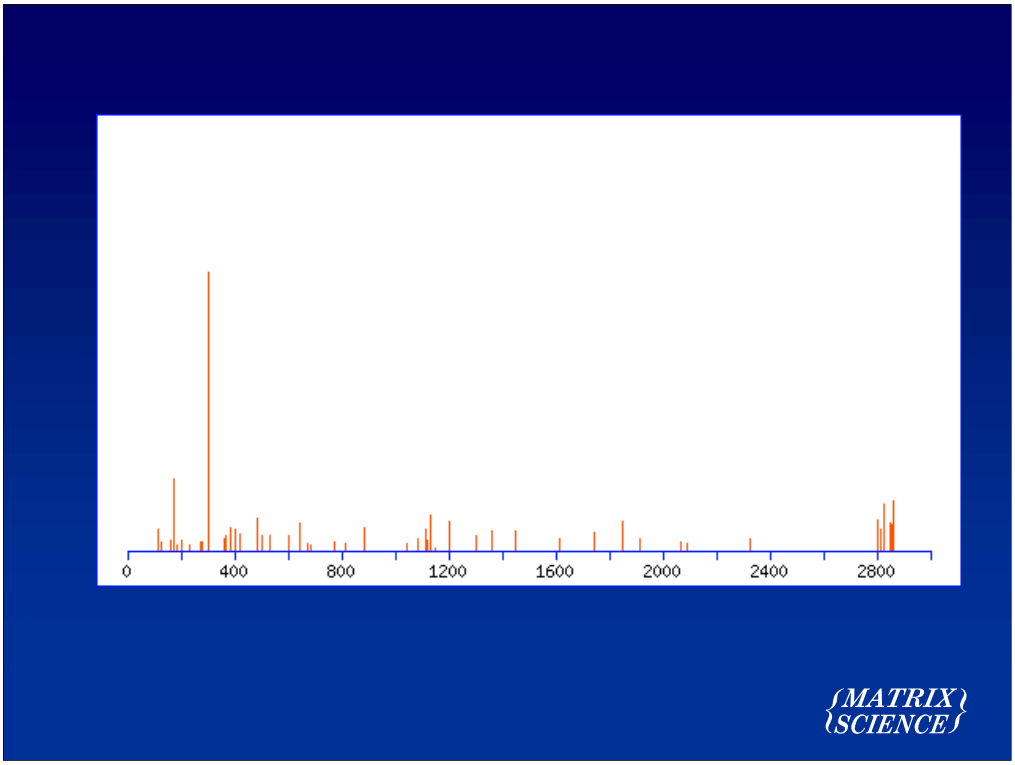


How exactly does Mascot score MS/MS data?

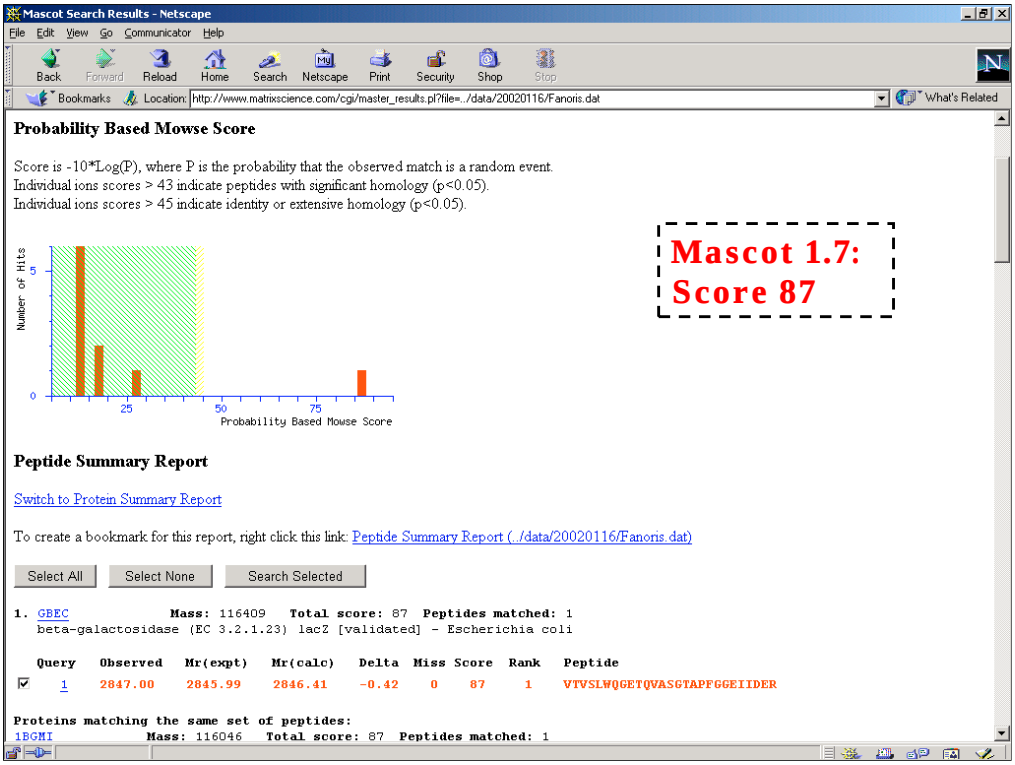


To illustrate how choosing a customised instrument setting can improve the Mascot score, I'd like to show a couple of examples.

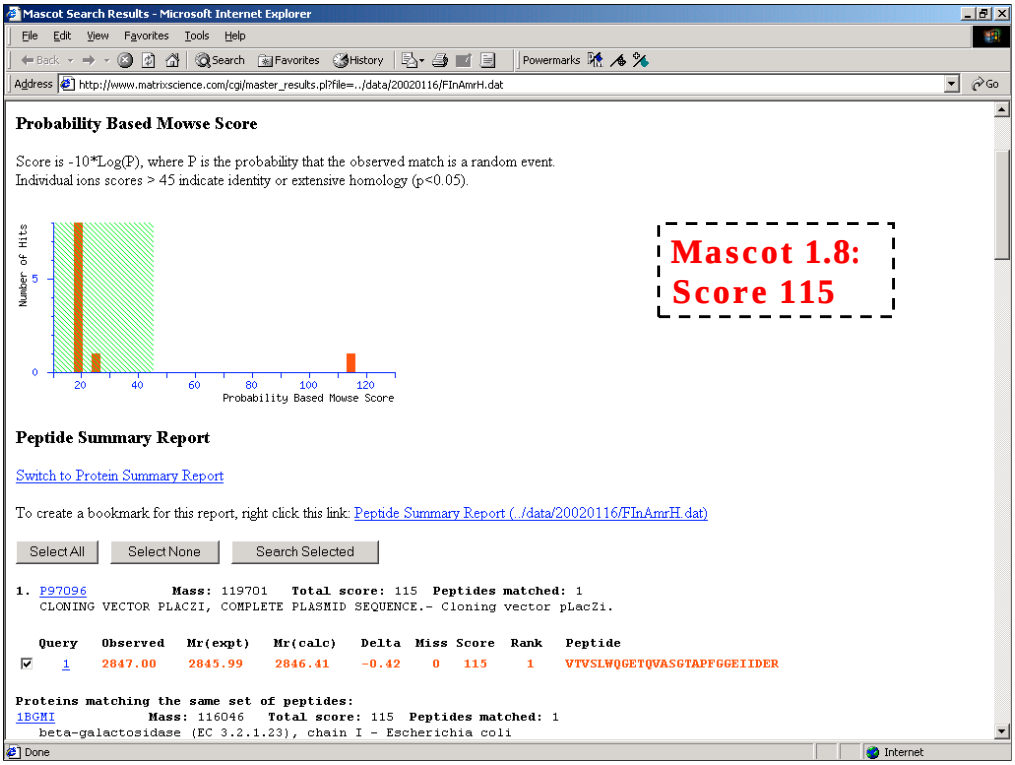
First, TOF/TOF data. Obviously, there is enormous interest in this new class of instrument. For Mascot, the most important consideration is that the fragmentation is high energy CID.



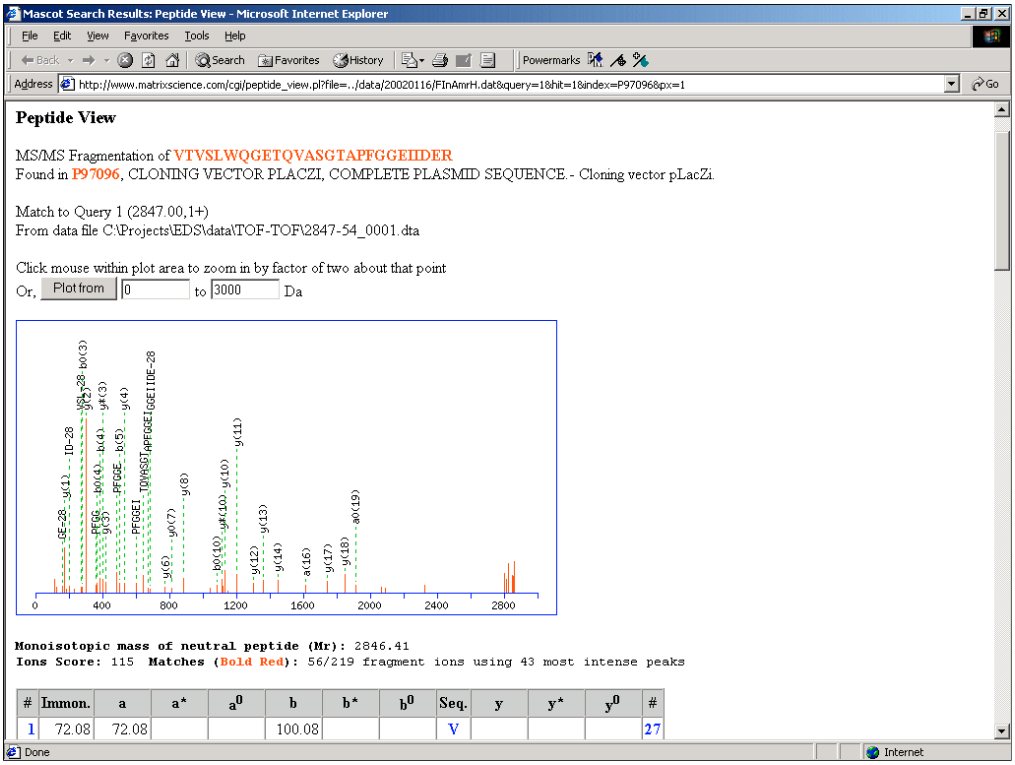
At first sight, some TOF/TOF spectra don't look very promising. However, the information content is actually very high.



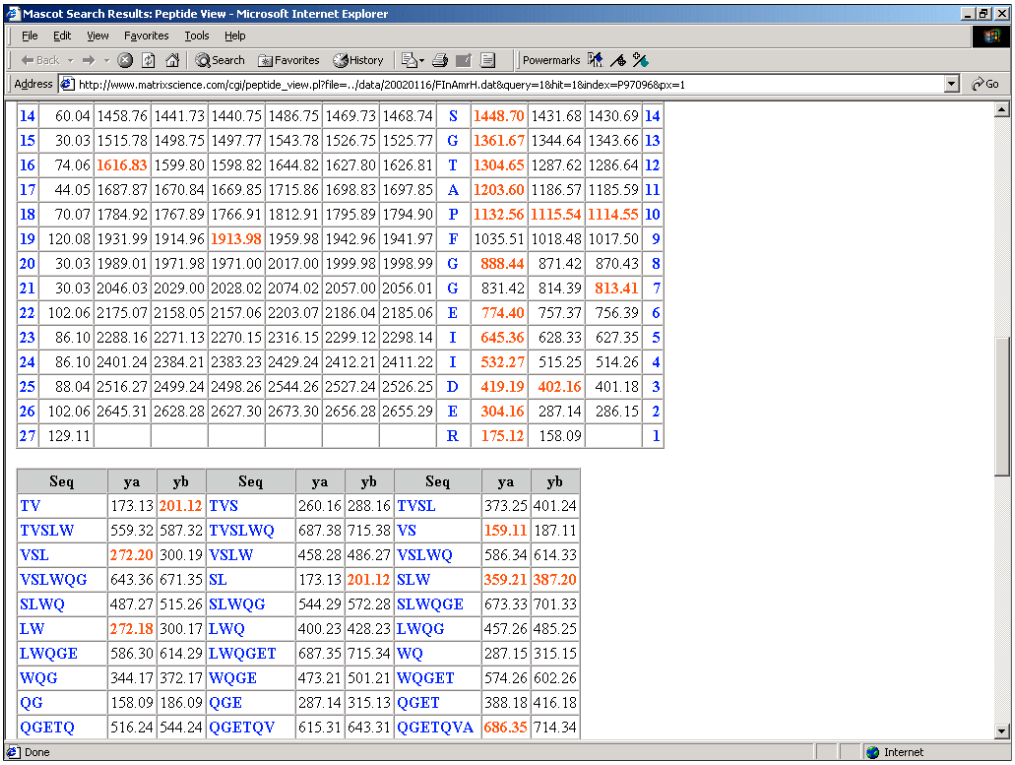
When this spectrum was searched using Mascot 1.7, the ions score was excellent: 87



However, searching with Mascot 1.8, where we can account for the internal fragments and the immonium ions increases this score to 115



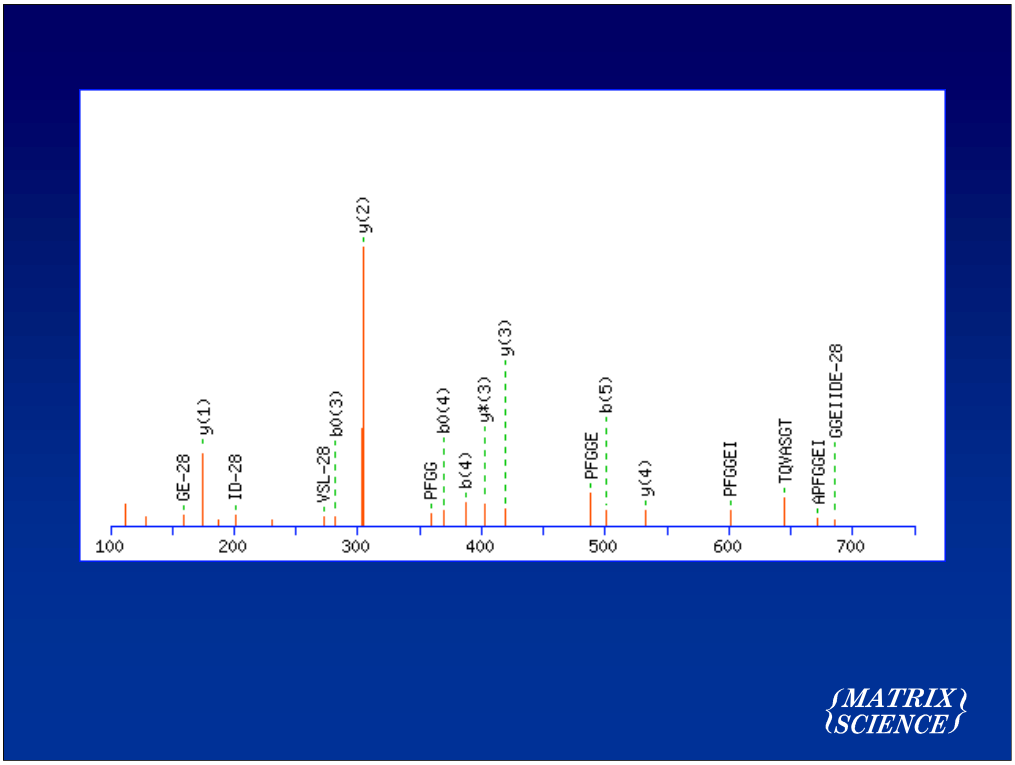
The reason the score is so high is that pretty much every single peak is a real peak, that can be assigned to a fragment ion.



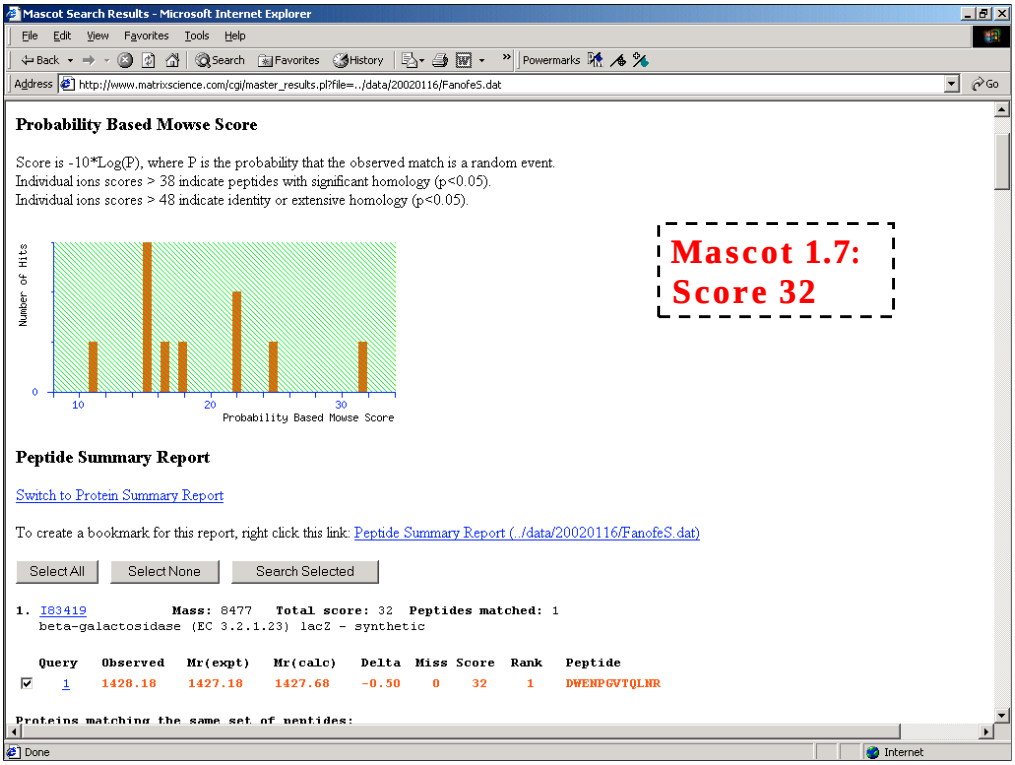
14	60.04	1458.76	1441.73	1440.75	1486.75	1469.73	1468.74	S	1448.70	1431.68	1430.69	14
15	30.03	1515.78	1498.75	1497.77	1543.78	1526.75	1525.77	G	1361.67	1344.64	1343.66	13
16	74.06	1616.83	1599.80	1598.82	1644.82	1627.80	1626.81	T	1304.65	1287.62	1286.64	12
17	44.05	1687.87	1670.84	1669.85	1715.86	1698.83	1697.85	A	1203.60	1186.57	1185.59	11
18	70.07	1784.92	1767.89	1766.91	1812.91	1795.89	1794.90	P	1132.56	1115.54	1114.55	10
19	120.08	1931.99	1914.96	1913.98	1959.98	1942.96	1941.97	F	1035.51	1018.48	1017.50	9
20	30.03	1989.01	1971.98	1971.00	2017.00	1999.98	1998.99	G	888.44	871.42	870.43	8
21	30.03	2046.03	2029.00	2028.02	2074.02	2057.00	2056.01	G	831.42	814.39	813.41	7
22	102.06	2175.07	2158.05	2157.06	2203.07	2186.04	2185.06	E	774.40	757.37	756.39	6
23	86.10	2288.16	2271.13	2270.15	2316.15	2299.12	2298.14	I	645.36	628.33	627.35	5
24	86.10	2401.24	2384.21	2383.23	2429.24	2412.21	2411.22	I	532.27	515.25	514.26	4
25	88.04	2516.27	2499.24	2498.26	2544.26	2527.24	2526.25	D	419.19	402.16	401.18	3
26	102.06	2645.31	2628.28	2627.30	2673.30	2656.28	2655.29	E	304.16	287.14	286.15	2
27	129.11							R	175.12	158.09		1

Seq	ya	yb	Seq	ya	yb	Seq	ya	yb
TV	173.13	201.12	TVS	260.16	288.16	TVSL	373.25	401.24
TVSLW	559.32	587.32	TVSLWQ	687.38	715.38	VS	159.11	187.11
VSL	272.20	300.19	VSLW	458.28	486.27	VSLWQ	586.34	614.33
VSLWQG	643.36	671.35	SL	173.13	201.12	SLW	359.21	387.20
SLWQ	487.27	515.26	SLWQG	544.29	572.28	SLWQGE	673.33	701.33
LW	272.18	300.17	LWQ	400.23	428.23	LWQG	457.26	485.25
LWQGE	586.30	614.29	LWQGET	687.35	715.34	WQ	287.15	315.15
WQG	344.17	372.17	WQGE	473.21	501.21	WQGET	574.26	602.26
QG	158.09	186.09	QGE	287.14	315.13	QGET	388.18	416.18
QGETQ	516.24	544.24	QGETQV	615.31	643.31	QGETQVA	686.35	714.34

You can see here that we have very good coverage of the y series, and many of the low mass peaks are internal fragments

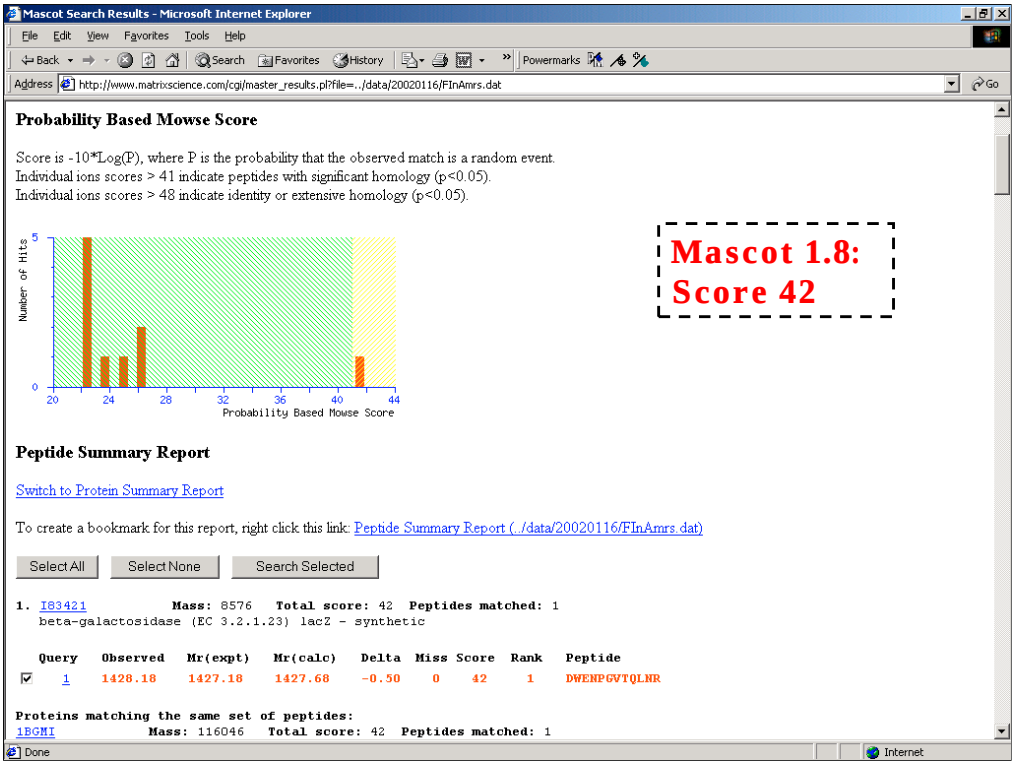


Here is an expanded view of the low mass end of this spectrum. You can see what I mean about almost every peak being accounted for.

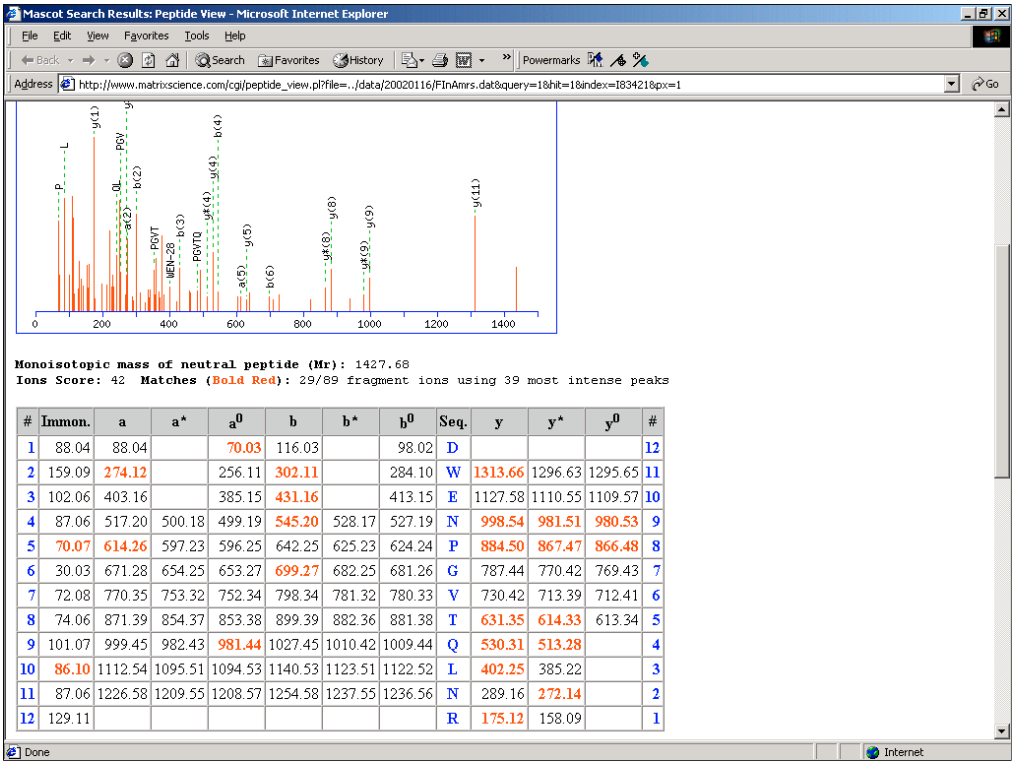


Now, although its great to see a score of 87 increased to 115, what matters more is whether we can turn a failed match into a marginal one or a marginal match into a good one.

Here, for example, we have an inconclusive search result from Mascot 1.7

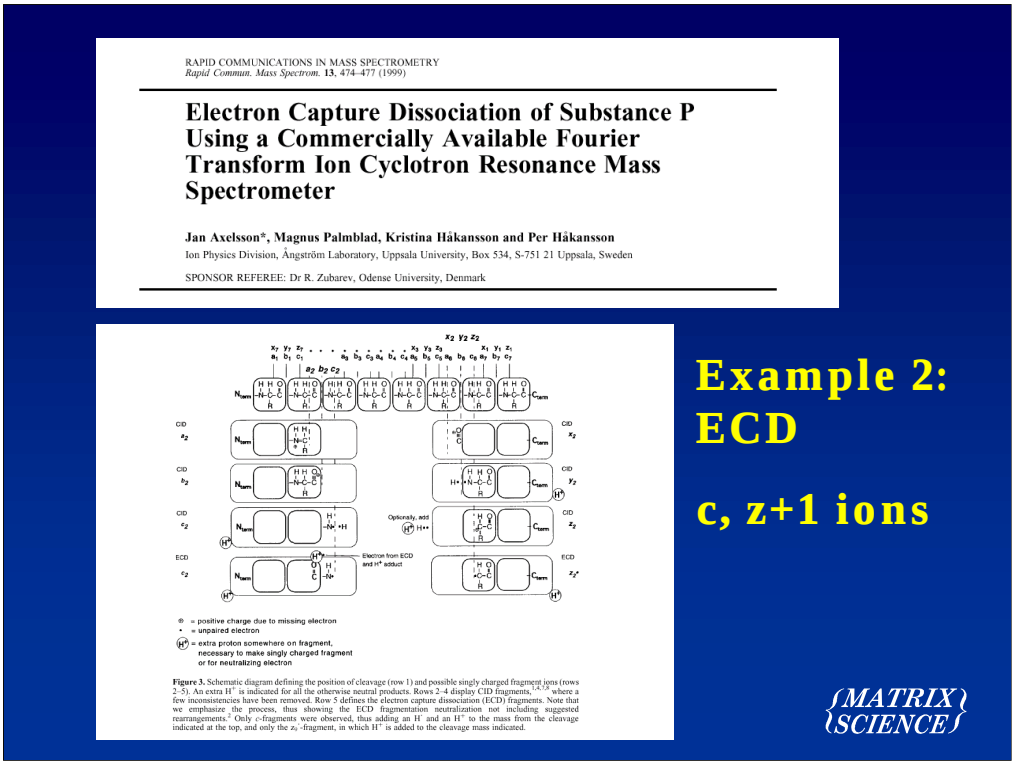


With Mascot 1.8, the more accurate ion series modelling pulls out a reasonably good match. This is very pleasing to see, particularly when you see the spectrum ...



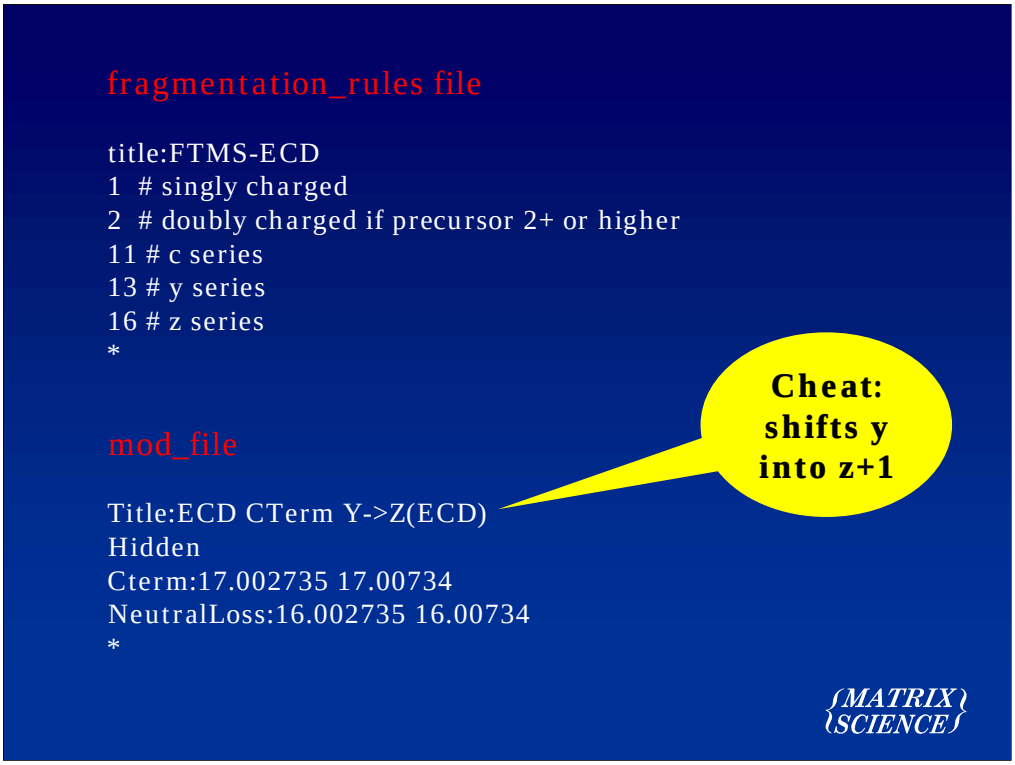
#	Immon.	a	a*	a^0	b	b*	b^0	Seq.	y	y*	y^0	#
1	88.04	88.04		70.03	116.03		98.02	D				12
2	159.09	274.12		256.11	302.11		284.10	W	1313.66	1296.63	1295.65	11
3	102.06	403.16		385.15	431.16		413.15	E	1127.58	1110.55	1109.57	10
4	87.06	517.20	500.18	499.19	545.20	528.17	527.19	N	998.54	981.51	980.53	9
5	70.07	614.26	597.23	596.25	642.25	625.23	624.24	P	884.50	867.47	866.48	8
6	30.03	671.28	654.25	653.27	699.27	682.25	681.26	G	787.44	770.42	769.43	7
7	72.08	770.35	753.32	752.34	798.34	781.32	780.33	V	730.42	713.39	712.41	6
8	74.06	871.39	854.37	853.38	899.39	882.36	881.38	T	631.35	614.33	613.34	5
9	101.07	999.45	982.43	981.44	1027.45	1010.42	1009.44	Q	530.31	513.28		4
10	86.10	1112.54	1095.51	1094.53	1140.53	1123.51	1122.52	L	402.25	385.22		3
11	87.06	1226.58	1209.55	1208.57	1254.58	1237.55	1236.56	N	289.16	272.14		2
12	129.11							R	175.12	158.09		1

The score is marginal because the signal to noise is not as good as the first example. You wouldn't want to do de novo on a spectrum like this. Even finding a sequence tag wouldn't be easy. But probability based matching with Mascot has been successful.



My second example concerns electron capture dissociation, which is often performed on FTMS instruments.

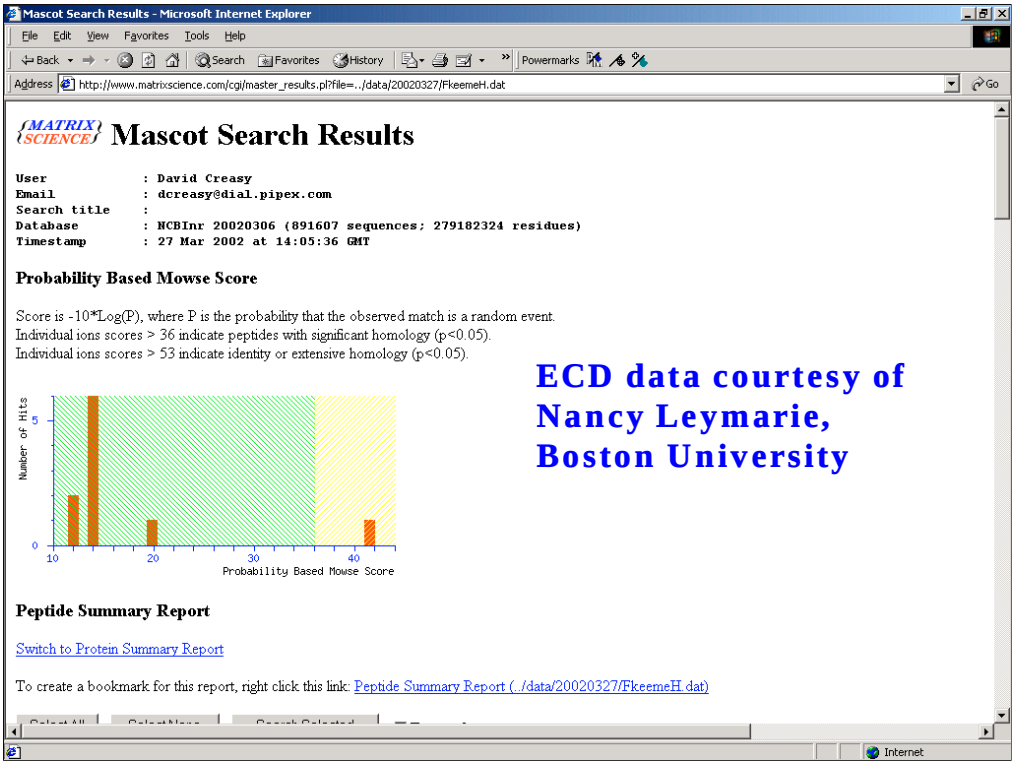
ECD is attracting increased levels of interest because it can provide good sequence coverage for very low level samples. The most prominent ion series in ECD are c and z+1 ions, that is z plus a proton



To support ECD, we should just need to add a new section to the fragmentation_rules file.

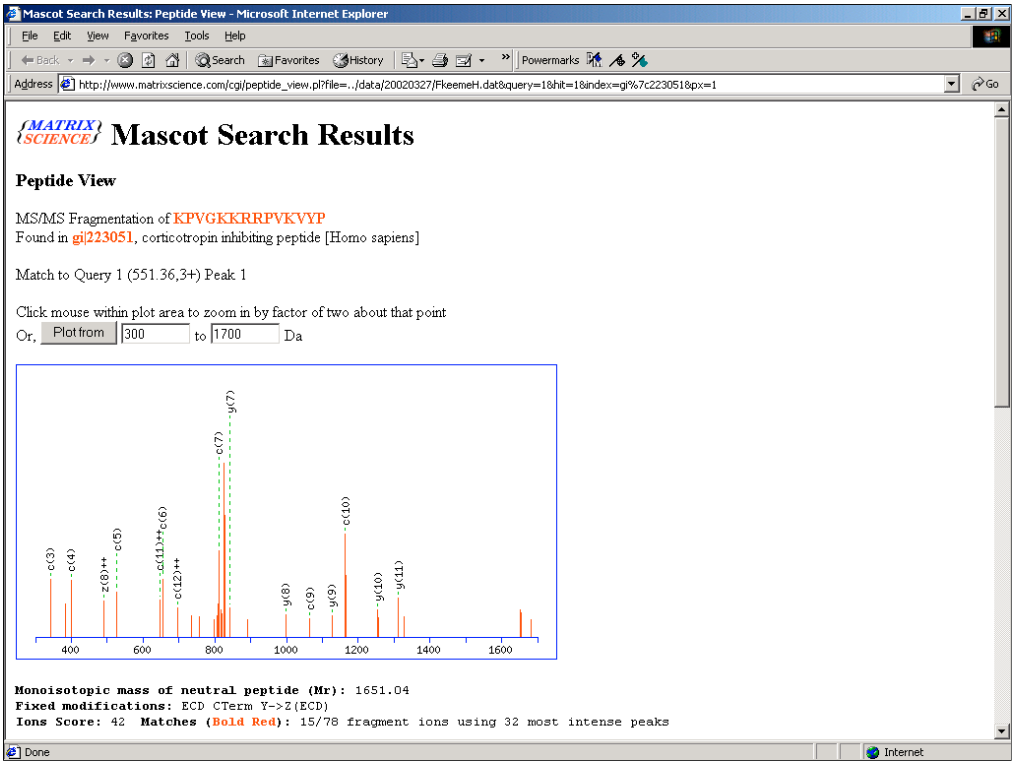
Unfortunately, when we released Mascot 1.8, we didn't realise that the z+1 series was important for ECD, and we only included support for the z series. However, Mascot is very flexible, and we can work around this by using a modfication.

The difference between a y series ion and a z series ion is just 17 Da. So, if we have a fixed modification at the C-terminus which does nothing except have a neutral loss of 16 Da, we can shift the y series into the z+1 series!

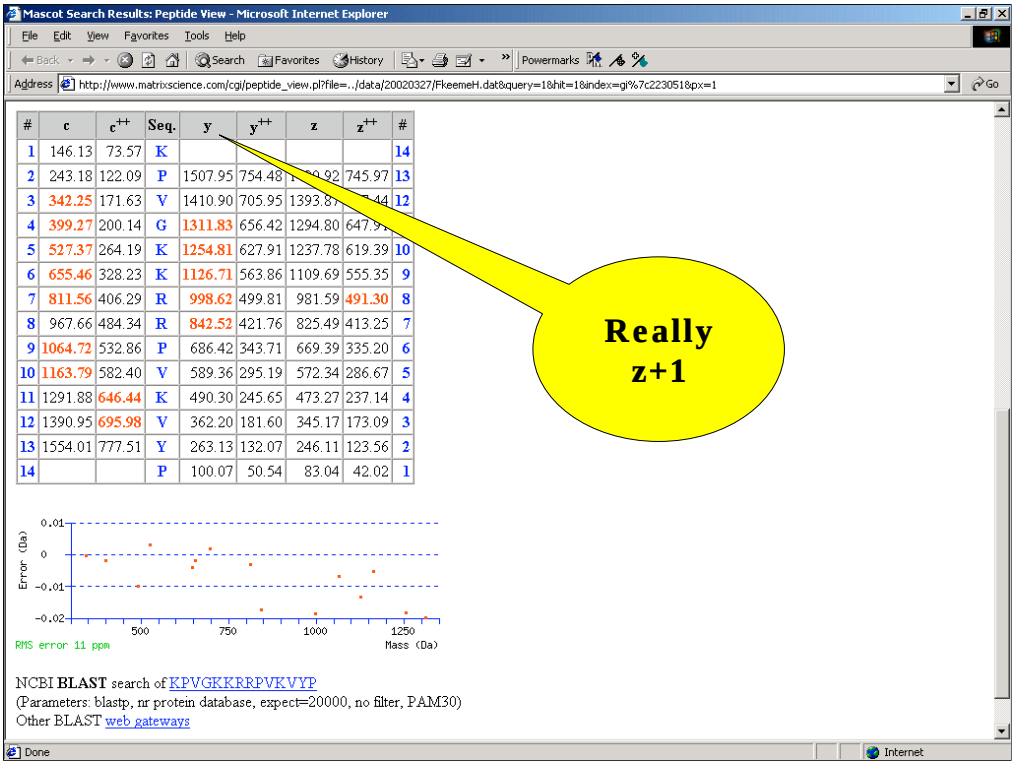


We are grateful to Nancy Leymarie of Boston University for pointing this out to us, and for giving us some examples of ECD data from an FTMS instrument.

Here is an example of a match to ECD data

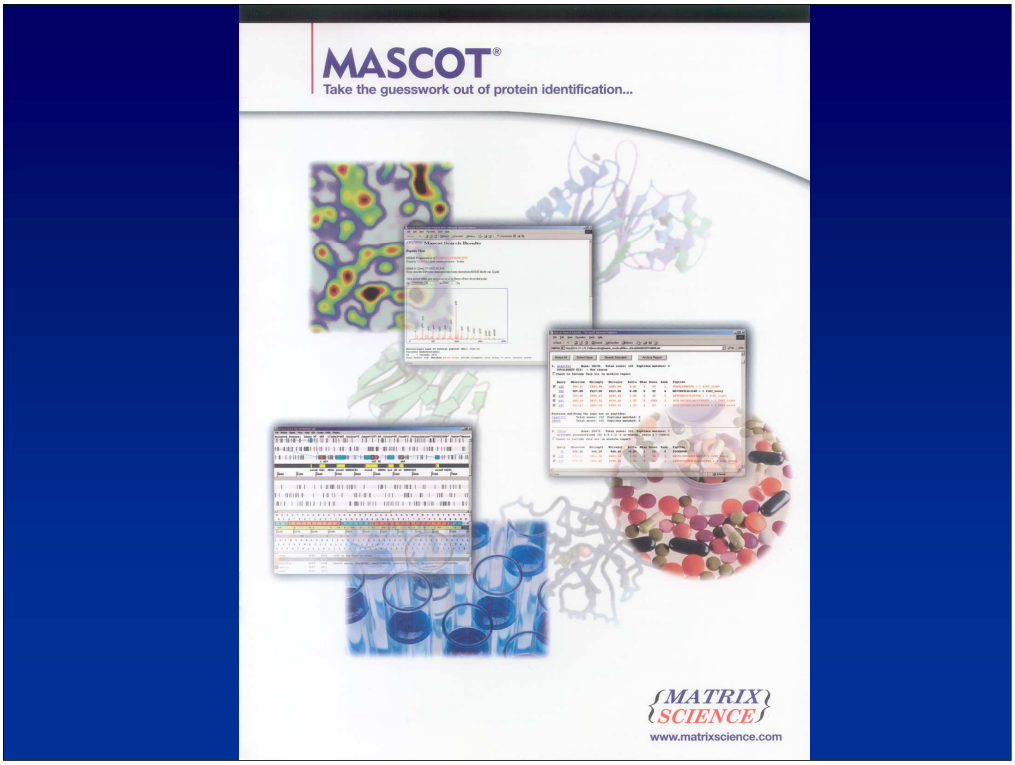


The spectrum is pretty sparse



#	c	c^{++}	Seq.	y	y^{++}	z	z^{++}	#
1	146.13	73.57	K					14
2	243.18	122.09	P	1507.95	754.48	[illegible]	745.97	13
3	342.25	171.63	V	1410.90	705.95	1393.87	[illegible]	12
4	399.27	200.14	G	1311.83	656.42	1294.80	647.9[illegible]	
5	527.37	264.19	K	1254.81	627.91	1237.78	619.39	10
6	655.46	328.23	K	1126.71	563.86	1109.69	555.35	9
7	811.56	406.29	R	998.62	499.81	981.59	491.30	8
8	967.66	484.34	R	842.52	421.76	825.49	413.25	7
9	1064.72	532.86	P	686.42	343.71	669.39	335.20	6
10	1163.79	582.40	V	589.36	295.19	572.34	286.67	5
11	1291.88	646.44	K	490.30	245.65	473.27	237.14	4
12	1390.95	695.98	V	362.20	181.60	345.17	173.09	3
13	1554.01	777.51	Y	263.13	132.07	246.11	123.56	2
14			P	100.07	50.54	83.04	42.02	1

And the matches that matter are those to the c and z+1 series. Without the facility to define which ion series are to be used, it would be difficult to get a positive match to this spectrum.



If you have any suggestions for new or improved instrument settings, or new rules, please let us know. We believe this new functionality is a significant enhancement to Mascot, especially for new types of instrument and unusual ionisation techniques.