ABRF-PRG04: Differentiation of Protein Isoforms

D. P. Arnott[1], M. A. Gawinowicz[2], J. A. Kowalak[3], W. S. Lane[4], T. A. Neubert[5], K. D. Speicher[6], C. W. Turck[7], and K. A. West[8]

[1]Genentech, Inc., South San Francisco, CA, [2]Columbia University, New York, NY, [3]NIH, Bethesda, MD, [4]Harvard University, Cambridge, MA, [5]New York University School of Medicine, New York, NY, [6]The Wistar Institute, Philadelphia, PA, [7]Max Planck Institute of Psychiatry, Munich, Germany, [8]Cleveland Clinic Foundation, Cleveland, OH

MISSION

The mission of the Proteomics Research Group (PRG) is to assist ABRF members in evaluating their capabilities to identify "unknown" proteins in order to establish realistic expectations for this technology.

ABSTRACT

Accurate protein identification sometimes requires careful discrimination between closely related proteins. Protein isoforms and/or protein homologues may differ by as little as a single amino acid substitution or posttranslational modification, yet these differences can have profound effects on the structure and function of the proteins. The ABRFPRG04 sample was designed to study a simple mixture of 3 closely related proteins, each at 3 picomoles. The Proteomics Research Group (PRG) sent the sample to interested laboratories in the form of intact proteins, and participating laboratories were asked to identify the proteins and report their results. The primary goal of the PRG04 Study is to give participating laboratories a chance to evaluate their capabilities and practices with regards to sample fractionation (1D- or 2D PAGE, HPLC, or none), protein digestion methods (in-solution, in-gel, enzyme choice), and approaches to protein identification (instrumentation, use of software and/or manual techniques to facilitate interpretation), as well as determination of amino acid or posttranslational modifications. Compilation of submitted data will allow a comparison of the strategies used and aid in optimization of these techniques.

INTRODUCTION

A common quest in proteomics core facilities, aside from the usual identification of proteins, is to locate differences within protein(s) of interest. These may simply be cross-species differences or modifications associated with a vital role, for example in protein function or activity. The investigator may have only minimal information regarding the type or location of these differences. These considerations make the ability to locate such variations very important. To make this determination, a core facility must investigate the protein in depth, going beyond the identification of the protein. This study will help in evaluating a core facility's ability to identify closely related proteins and determine where the differences exist between them. Therefore, the primary goals of this study are to give each laboratory a chance to evaluate its capabilities and practices with regards to:

- •Protein Digestion methods (solution-based, in-gel, choice of enzymes)
- •Protein identification methods
- •Methods for the determination of amino acid differences between protein isoforms
- •Amino acid sequence coverage of the identified proteins

as well as to obtain data that would allow a comparison of the strategies used and aid in optimization of these techniques.

With this in mind, we provided laboratories that requested samples (106 labs) with a mixture of three intact proteins: two bovine carbonic anhydrase isoforms differing by a single AA substitution plus a human carbonic anhydrase. The identity of the respondents was not known to members of the PRG. Since the sequences of these proteins are in the public databases, we required that the participants supply proof of the differences ascertained by MS/MS or some other means they may have used. This study relates to those previous PRG studies in that proteins in a mixture will be identified and AA modifications/substitutions will be analyzed. However, in this study, intact proteins, rather than a predigested mixture, were examined. Furthermore, the analysis involved searching for differences between 3 very similar proteins, rather than modifications within the same protein. Also, because the participants did their own digests, this study addressed, to some extent, the issue of front end sample preparation. We asked that participants return proof of the differences they have determined, along with the completed questionnaire.

METHODS

Three carbonic anhydrase proteins (C6165, C2522, and C3640) were purchased from Sigma-Aldrich CO (St Louis, MO). Small amounts of each protein were weighed out on a Cahn microbalance and an appropriate amount of 1% acetic acid was added to give a final concentration of 1 μg/μL (stock solutions). Three microliters of each protein solution was placed into an amino acid analysis tube, each sample was prepared in triplicate, with 3 μL of 1% acetic acid as a control. The samples were lyophilized in a vacuum centrifuge, sealed by wrapping in parafilm and analyzed for amino acid content, which determined the protein content of the solid proteins to be approximately 60%.

Fresh stock solutions (1 μg/μL) of each of the three proteins were prepared in 1% acetic acid as were fresh working solutions at a concentration of 3 picomoles/μL. Samples were prepared by aliquoting 1 μL of each protein solution into 0.5 μL Eppendorf tubes containing 10 μL of 1% acetic acid. All samples were lyophilized in a vacuum centrifuge, and sealed by wrapping in parafilm. Lyophilized samples were then sent to 85 requesting laboratories in the USA and 21 labs in Europe and Israel for protein identification.

Table 1. Sample Prep Information

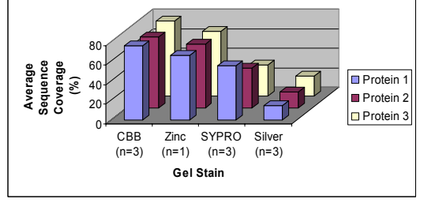
						Protein 1					Protein 2					Protein 3				
ID	Solvent	Volume/ul	Separation Method	Gel or Column Type	Staining Method	Protein 1 ID	Confidence	Total % Seq Coverage	N-terminus	C-terminus	Protein 2 ID	Confidence	Total % Seq Coverage	N-terminus	C-terminus	Protein 3 ID	Confidence	Total % Seq Coverage	N-terminus	C-terminus
715	SDS SB, pH 6.5	10	SDS-PAGE	10% Tris-Gly Standard	CBB	CAH2bov-R	P	79	Y	Y	CAH2bov-Q	P	79	Y	Y	CAH2hum	P	77	Y	
68166	SDS SB BME	20	SDS-PAGE	10-20% Tris-Gly Mini	Zinc	CAH2hum	P	66	Y		CAH2bov-R	P	65	Y		CAH2bov-Q	P	66	Y	
20702	1% FA	20				CAH2hum	P	63	Y		CAH2bov-R	P	59	Y		CAH2bov-Q	P	60	Y	
27974	water	20	SDS-PAGE	12% Bis-Tris NuPAGE	SYPRO Orange	CAH2bov-R	P	57	Y		CAH2bov-Q	P	58	Y		CAH2hum	P	47	Y	
65213	100 mM ABC	60				CAH2bov-R	P	58	Y		CAH2bov-Q	P	58	Y		CAH2hum	P	51	Y	
4343	25 mM ABC	20				CAH2bov-R	P	52	Y		CAH2bov-Q	P	52	Y		CAH2hum	P	52	Y	
29103	100 mM ABC, 0.5 M GuHCl	100				CAH2bov-R	P	43	Y		CAH2hum	P	40	Y		CAH2bov-Q	P	39	Y	
31113	50 mM ABC, pH 8	15	SDS-PAGE (33% of sample)	12% Tris-Gly Mini	MS Friendly Silver	CAH2hum	P	22	Y		CAH2bov-R	P	20	Y		CAH2bov-Q	P	20	Y	
110107	water	15				CAH2hum	P	55	Y		CAH2bov-R	P	49	Y		CAH2horse	T	31	Y	
1066352	0.1% TFA/10% ACN (9:1)	6				CAH2bov	P	42	Y		CAH2hum	P	45.9	Y		CAH2bov-Q	P	42	Y	
21562	0.1% FA, 0.20% N-octylglucopyranosi de	15	2DE (50% of sample), RP-HPLC (33% of sample)	C8 Vydac	Colloidal CBB	CAH2bov	P	79	Y		CAH2hum	P	69	Y						
25519	water	30				CAH2bov-R	P	73	Y		CAH2hum	P	64	Y						
11111	8 M Urea, 0.2 M Tris-HCl	20				CAH2bov-R	P	63	Y		CAH2hum	P	64	Y						
90894	8 M urea in 0.4 M ABC	10				CAH2hum	P	62	Y		CAH2bov-R	P	66	Y						
22626	6% ACN	5.3				CAH2bov	P	57	Y		CAH2hum	P	50	Y						
94591	water	10				CAH2hum	P	56	Y		CAH2bov-R	P	55	Y						
23312	ABC	20				CAH2bov	P	55	Y		CAH2hum	P	46	Y		keratin	P			
2115	50 mM ABC	30	SDS-PAGE	12% Tris-Gly Mini	CBB	CAHIhum	P	70			CAH2bov	P	66	Y						
24389	ABC	5	SDS-PAGE	12% MES Mini	SYPRO	CAH2bov	P	46	Y		CAH2hum	P	41	Y						
32569		30				CAH2hum	P	60	Y		CAH2bov	P	50	Y						
11787	25 mM ABC, 10% ACN	50				CAH2bov-R	P	54	Y		CAH2hum	P	44							
73108	100 mM ABC	60				CAH2hum	P	38			CAH2bov	P	30.4							
93743	ABC	10	SDS-PAGE	15% Tris-Gly Mini	Silver	CAH2hum	P	11			CAH2bov-R	P	23							
10567	Water	25	SDS-PAGE	4-12% NuPAGE	Silver	CAH2bov	P	11			CAHIhum	T	7							
24770		150	RP-HPLC	C18, 75 um x 15 cm, Picotip 20 um		CAH2bov	P	46			CAH2hum	T	20							
92711	8 M Urea	8				CAH2bov	P	54			CAH2bov	P	46			CAH2hum	P	47		
11735	5% ACN, 100 mM ABC	19				CAH2hum	P	42	Y		CAH2bov-R	P	77	Y		CAH2hum	P	16		
48583	50 mM ABC	10				CAH2bov	P	66	Y		CAH2hum	P	59	Y		CAH1thum	T	63		
69186	Tris-HCl, 6 M urea	5				CAH2bov	P	47.9			CAH2hum	P	48			CAH2sheep	P	32.4		
87050	SDS SB	60	SDS-PAGE	5-20% Tris-Gly Mini	SYPRO Ruby	CAH2bov	P	62.5	Y		CAH2hum	P	47.9			CAH1thum	T	63		
31815	0.1% TFA, 5% ACN	20				CAH2hum	P	38	Y		CAH2sheep	P	32			CAH2bov	P	15.8		
21068	5% ACN, 0.1% FA	20				CAH2bov	P	72		Y	CAH2hum	P	22	Y		CAH2sheep	P	17		
640921	2% ACN, 0.1% FA	50				CAH2sheep	P	50			CAH2hum	P	40			CAH2bov-R	P	67	Y	
uicmstw	50 mM ABC	10				no results														
80053	40% ACN	50				no results														
69117	5% FA	10				no results														
106369						no results														
13791	50 mM ABC	20				no results														
11747	SDS SB	100	SDS-PAGE	10% Tris-Gly	CBB	no results														
13053	30% ACN, 0.1% TFA	25	SDS-PAGE	12% Tris-Gly Mini	CBB	no results														
11596	3% TFA	30				calsequestrin	P	28			calsequestrin(dog)	P	25			calsequestrin (hur	P	4		
11128	25 mM ABC	10				BSA	P	17												

ABC: ammonium bicarbonate
ACN: acetonitrile
FA: formic acid
GuHCl: guanidinium hydrochloride
SDS SB BME: SDS sample buffer plus beta mercaptoethanol
TFA: trifluoroacetic acid
CBB: Coomassie Brilliant Blue

Sample Preparation Summary:

- 106 samples were sent to requesting laboratories, and 42 labs returned data
- 11/42 labs separated the proteins by 1-D SDS-PAGE (see corresponding protein ID results in Figure 2), 1 lab used 2-D SDS-PAGE
  - 5/11 labs used Coomassie Brilliant Blue to stain the gels
  - 1/11 labs used zinc staining
  - 3/11 labs used SYPRO orange or ruby staining
  - 3/11 labs used silver staining
- 2/42 labs separated the intact proteins by reverse phase HPLC
- 2/11 labs did not see any proteins by Coomassie staining of 1-D SDS-PAGE gels

Figure 2. Effect of Gel Stain on Protein Sequence Coverage




```
CAH2 hum          1   Ac-SHHWGYGKHN GPEHWHKDFP IAKGERQSPV DIDTHTAKYD PSLKPLSVSY
CAH2 bov 56R      1   Ac-SHHWGYGKHN GPEHWHKDFP IANGERQSPV DIDTKAVVQD PALKPLALVY
CAH2 bov 56Q      1   Ac-SHHWGYGKHN GPEHWHKDFP IANGERQSPV DIDTKAVVQD PALKPLALVY

CAH2 hum         51     DQSTSLRILN NGHAFNVEFD DSQDKAVLKG GPLDGTYRLI QFHFHWGSLD
CAH2 bov 56R     51     GEATSRRMVN NGHSFNVEYD DSQDKAVLKD GPLTGTYRLV QFHFHWGSSD
CAH2 bov 56Q     51     GEATSQRMVN NGHSFNVEYD DSQDKAVLKD GPLTGTYRLV QFHFHWGSSD

CAH2 hum        101     GQGSEHTVDK KKYAAELHLV HWNTKYGDFG KAVQQPDGLA VLGIFLKVGS
CAH2 bov 56R    101     DQGSEHTVDR KKYAAELHLV HWNTKYGDFG TAAQQPDGLA VVGVFLKVGD
CAH2 bov 56Q    101     DQGSEHTVDR KKYAAELHLV HWNTKYGDFG TAAQQPDGLA VVGVFLKVGD

CAH2 hum        151     AKPGLQKVVD VLDSIKTKGK SADFTNFDPR GLLPESLDYW TYPGSLTTPP
CAH2 bov 56R    151     ANPALQKVLD ALDSIKTKGK STDFPNFDPG SLLPNVLDYW TYPGSLTTPP
CAH2 bov 56Q    151     ANPALQKVLD ALDSIKTKGK STDFPNFDPG SLLPNVLDYW TYPGSLTTPP

CAH2 hum        201     LLECVTWIVL KEPISVSSEQ VLKFRKLNFN GEGEPEELMV DNWRPAQPLK
CAH2 bov 56R    201     LLESVTWIVL KEPISVSSQQ MLKFRTLNFN AEGEPELLML ANWRPAQPLK
CAH2 bov 56Q    201     LLESVTWIVL KEPISVSSQQ MLKFRTLNFN AEGEPELLML ANWRPAQPLK

CAH2 hum        251     NRQIKASFK
CAH2 bov 56R    251     NRQVRGFPK
CAH2 bov 56Q    251     NRQVRGFPK
```

Figure 3. Amino acid sequences of the two bovine and one human CAH2 isoforms present in the sample mixture. Tryptic peptides unique to human CAH2 are shown in red or green, peptides unique to the two bovine CAH2 isoforms are shown in blue. All isoforms were acetylated at the N-terminal serine (in green).

Protein Identification Summary:

- 8/42 responding labs correctly identified all three CAH2 isoforms with supporting evidence to discriminate between the isoforms as well as identifying N-terminal acetylation of the proteins (dark blue section in Tables).
- 16/42 labs correctly identified two of the CAH isoforms with supporting evidence (light blue section in Tables). Eleven of these labs also identified N-terminal acetylation. One of these labs (11010) also provided MS/MS evidence for a peptide unique to horse CAH2, but which most likely arose from tryptic cleavage of an asparagine residue (the horse sequence has a lysine in this position).
- 9/42 labs correctly identified two CAH isoforms, but also made one or more incorrect protein or posttranslational modification identification (white section in Tables).
- 7/42 labs did not identify any proteins (yellow section in Tables). Reasons for this varied, but several groups did not observe any proteins in the sample
- 2/42 labs did not make any correct protein identifications, but made one or more incorrect identifications (violet/pink section in Tables). Since the incorrectly identified proteins are not similar in sequence to any CAH, it is most likely that these proteins were contaminants not present in the sample.
- A total of 33/42 labs identified human and bovine isoforms of CAH2.
- 27/42 labs reported identification of the N-terminus of at least one isoform, while only 2/42 labs reported seeing the C-terminus.
- One laboratory attempted to identify the intact proteins by Edman sequencing, but did not obtain any results.

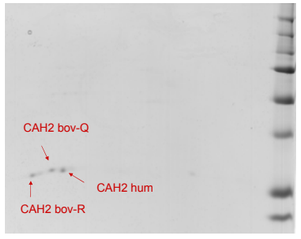


Figure 1. 2-D PAGE of 3 pmol carbonic anhydrase protein mixture. The first dimension pH gradient was 3-10 nonlinear, the second dimension SDS-PAGE gel contained 12% acrylamide. Proteins were visualized by Colloidal Coomassie Blue staining.

Table 2. Mass Spectrometry Results

			Peptides Masses Measured Without PSD or ms/ms						Peptide Masses Measured with PSD or ms/ms						Protein 1							Protein 2							Protein 3						
			Protein 1		Protein 2		Protein 3		Protein 1		Protein 2		Protein 3				Total %							Total %							Total %			
ID	Proteins Digested	Enzyme(s) used	#Peptides Matched	%Sequence Coverage	#Peptides Matched	%Sequence Coverage	#Peptides Matched	%Sequence Coverage	#Peptides Matched	%Sequence Coverage	#Peptides Matched	%Sequence Coverage	#Peptides Matched	%Sequence Coverage	Protein 1 ID	Confidence	Seq Coverage	N-terminus	C-terminus	PTM	Protein 2 ID	Confidence	Seq Coverage	N-terminus	C-terminus	PTM	Protein 3 ID	Confidence	Seq Coverage	N-terminus	C-terminus	PTM
715	Y	T	19	79	19	79	21	77	18	78	18	78	16	67	CAH2bov-R	P	79	Y	Y	Acetyl	CAH2bov-Q	P	79	Y	Y	Acetyl	CAH2hum	P	77	Y		Acetyl
68166	Y	T, Ch							14	66	13	65	13	66	CAH2hum	P	66	Y		Acetyl	CAH2bov-R	P	65	Y		Acetyl	CAH2bov-Q	P	66	Y		Acetyl
20702	Y	T							15	63	15	59.2	15	59.6	CAH2hum	P	63	Y		Acetyl	CAH2bov-R	P	59	Y		Acetyl	CAH2bov-Q	P	60	Y		Acetyl
27974	Y	T	12	57	12	58	12	43	2	11	2	12	2	20	CAH2bov-R	P	57	Y		Acetyl	CAH2bov-Q	P	58	Y		Acetyl	CAH2hum	P	47	Y		Acetyl
65213	Y	T, Glu-C, Ch	11	58	11	58	12	51	6	28	6	28	8	30	CAH2bov-R	P	58	Y		Acetyl	CAH2bov-Q	P	58	Y		Acetyl	CAH2hum	P	51	Y		Acetyl
4343	Y	T							14	52	14	53	13	52	CAH2bov-R	P	52	Y		Acetyl	CAH2bov-Q	P	52	Y		Acetyl	CAH2hum	P	52	Y		Acetyl
29103	Y	T	7	35	10	40	6	32	3	8	3	8	1	8	CAH2hum	P	43	Y		Acetyl	CAH2bov-R	P	40	Y		Acetyl	CAH2bov-Q	P	39	Y		Acetyl
31113	Y	T							4	22	3	20	3	20	CAH2hum	P	22	Y		Acetyl	CAH2bov-R	P	20	Y		Acetyl	CAH2bov-Q	P	20	Y		Acetyl
110107	Y	T	8	30	7	34	0	0	6	28	5	20	4	15	CAH2hum	P	55	Y		Acetyl	CAH2bov-R	P	49	Y		Acetyl	CAH2horse	T	31	Y		
1066352	Y	Lys-C	3	13.9	0	0			18	42	13	45.9	18	42	CAH2bov	P	42	Y		Acetyl	CAH2hum	P	45.9	Y		Acetyl	CAH2bov-Q	P	42	Y		
21562	Y	T	18	70	14	58			17	63	5	25			CAH2bov	P	79	Y		Acetyl	CAH2hum	P	69	Y		Acetyl						
25519	Y	T							15	73	16	64			CAH2bov-R	P	73	Y		Acetyl	CAH2hum	P	64	Y		Acetyl						Acetyl
11111	Y	T	12	59	13	56			1						CAH2bov-R	P	63	Y		Acetyl	CAH2hum	P	64	Y		Acetyl						
90894	Y	T							18	61.8	20	66			CAH2hum	P	62	Y		Acetyl	CAH2bov-R	P	66	Y		Acetyl						
22626	Y	T	10	56	12	50			5	27	8	30			CAH2bov-R	P	57	Y		Acetyl	CAH2hum	P	50	Y		Acetyl						
94591	Y	T	9	53.1	9	51.9			1	2.6	1	2.6			CAH2hum	P	56	Y		Acetyl	CAH2bov-R	P	55	Y		Acetyl						
23312	Y	T	12	55	8	46			4	23	6	25			CAH2bov	P	55	Y		Acetyl	CAH2hum	P	46	Y		Acetyl	keratin	P				
2115	Y	T							16	70	14	68			CAH2bov	P	70				CAH2bov	P	66	Y		Acetyl						
24389	Y	T, Glu-C	10	45	11	41			10	45	11	41			CAH2bov	P	46	Y		Acetyl	CAH2hum	P	41	Y		Acetyl						
32569	Y	T			10	50					5	28			CAH2hum	P	60	Y			CAH2bov	P	50	Y								
11787	Y	T	11	53	9	44									CAH2bov-R	P	54	Y			CAH2hum	P	44									
73108	Y	T	10	44	9	30			6	38	5	30.40			CAH2hum	P	38				CAH2bov	P	30.4									
93743	Y	T	2	11	6	24			2	11	6	24			CAH2hum	P	11				CAH2bov-R	P	23									
10567	Y	T							3	11	2	7			CAH2bov	P	11				CAHIhum	T	7									
24770	Y	Sub, ProK (pH11)							17	45.8	2	7.7			CAH2bov	P	46			phosphorylation	CAH2hum	T	20									
92711	Y	T							8	54	6	46	8	47	CAH2bov	P	54			Acetyl	CAH2bov	P	46			Acetyl	CAH2hum	P	47			Acetyl
11735	Y	T	11	42			16		7	34	7	30	6	34	CAH2hum	P	42	Y			CAH2bov-R	P	77	Y			CAH2hum	P	16			
48583	Y	T, Glu-C, Elas							15	66	12	59	11	63	CAH2bov	P	66	Y		Acetyl	CAH2hum	P	59	Y		Acetyl	CAH1hum	T	63			
69186	Y	T	11		10		5		14	47.9	11	48.3	5	32.4	CAH2bov	P	47.9				CAH2hum	P	48				CAH2sheep	P	32.4			
87050	Y	T, Lys-C, Asp-N							11	62.5	4	22.5	4	15.80	CAH2hum	P	62.5	Y		Acetyl	CAH2bov	P	22.7				CAH2sheep	P	15.8			
31815	Y	T	9	37	8	32	8	31	9	37	5	24	6	23	CAH2hum	P	38	Y		Acetyl	CAH2sheep	P	32			Acetyl	CAH2bov-R	P	24	Y		
21068	N	T							30	73	10	22	5	17	CAH2hum	P	72		Y		CAH2hum	P	22	Y			CAH2sheep	P	17			
640921	Y	T	9	50	9	40	14	67	9	50	9	40	14	67	CAH2sheep	P	50				CAH2hum	P	40				CAH2bov-R	P	67	Y		Acetyl
uicmstw															no results																	
80053	Y	T													no results																	
69117	Y	T													no results																	
106369															no results																	
13791	Y														no results																	
11747		T													no results																	
13053	Y														no results																	
11596	Y	T							15	28	15	25	1	4	calsequestrin	P	28				calsequestrin(dog)	P	25				calsequestrin (hum)	P	4			
11128	Y	T	14	18					14	17					BSA	P	17															

T: trypsin
Ch: chymotrypsin
Lys-C: endoproteinase Lys-C
Glu-C: endoproteinase Glu-C
Sub: subtilisin
ProK: proteinase K
Elas: elastase
Asp-N: endoproteinase Asp-N
CAH: carbonic anhydrase
Q: 56R isoform (bovine)
R: 56R isoform (bovine)
BSA: bovine serum albumin
Acetyl: acetylated N-terminal
P: positive
T: tentative

ACKNOWLEDGMENTS

We thank Isabel Birg of the Max Planck Institute of Psychiatry for 2D-PAGE analysis of the carbonic anhydrase mixture, and Lora Goodrich of Columbia University for mailing and receiving correspondence and for ensuring that the participants remain anonymous.

CONCLUSIONS

•Almost all of the labs employed tryptic digestion of proteins followed by tandem mass spectrometry to identify the proteins.

•Separating the proteins before analysis did not influence the success rate of the analyses of this simple protein mixture.

•For this study, use of enzymes other than trypsin was not useful for most analyses.

•In almost all cases relying on protein identification software (database search engines) alone was not sufficient to make accurate discriminations between very closely related isoforms (i.e. the two bovine isoforms that differed by a single amino acid). Careful inspection of raw MS data was usually required. Examination of notation in the database entry could also be helpful. However, in most cases database search engines were able to identify the human and bovine forms of carbonic anhydrase.

•Though the sample number was small, in the case of gel separation the type of gel stain seemed to have an effect on the average sequence coverage that was obtained by mass spectrometry.

After March 15 a PDF version of this poster can be found at http://www.abrf.org/index.cfm/group.show/Proteomics.34.htm