

ABRF Amino Acid Analysis Survey: Identification of Proteins Electroblotted to PVDF

A. Mahrenholz¹, T.T. Andersen², Y. Bao³, S.A. Cohen⁴, N.D. Denslow⁵, J. Hulmes⁶, P.E. Hunziker⁷, K. Mann⁸, K.M. Schegg⁹, and K. West⁶

¹Purdue University, West Lafayette, IN., ²Albany Medical College, Albany, NY, ³University of Virginia School of Medicine, Charlottesville, VA, ⁴Waters Corporation, Milford MA, ⁵University of Florida, Gainesville, FL, W. ⁶Alton Jones Cell Science Center, Lake Placid, NY, ⁷Universität Zürich, Zürich, Switzerland, ⁸Max-Planck-Institute, Martinsried, Germany, ⁹University of Nevada, Reno, Reno, NV,

Introduction

Amino Acid Analysis is currently enjoying renewed interest due to its ability to help identify proteins. Amino acid analysis (AAA) has previously been examined as a means for protein identification (1 [1]), however, more recently, widespread access to protein/DNA databases and network servers providing computer programs tailored to amino acid composition searches has multiplied the power of this tool. Thus, AAA has recently been shown to be practical and economical for protein identification, whether used as a stand-alone technique (2,3 [2]) or, especially, when combined with other information such as: pI, MW, and species (4 [3]), NH₂-terminal sequence tags (5 [4]), or mass fragment information (6,7 [5]).

One of the more promising applications of AAA in protein identification is to screen large numbers of proteins which have been separated by 2D-SDS-PAGE and blotted to PVDF. Such samples may become quite common as examination of the proteome ensues, since 2D-gels are perhaps the most powerful means of separating complex mixtures of proteins. With this in view, the ABRF amino acid analysis samples have recently focused on elements of this approach. ABRF-95 included a protein simply adsorbed from solution to PVDF (8 [6]), whereas ABRF-96 was a pure soluble protein which was to be identified using composition data alone (9 [7]). In view of the past difficulty with PVDF-adsorbed samples, namely, high background and high compositional errors (8,10 [8]), this year the AAA committee prepared electroblotted samples of both unknown and control protein in order to more closely approximate the preparations anticipated in laboratory settings and to, hopefully, reduce background. It was also hoped that additional practice with Internet searches would be beneficial to laboratories that do not routinely submit AAA data. The results reported here summarize the performance of 38 participating laboratories, the success of the identification routines, and usefulness of several of the options of the search utilities.

PVDF; polyvinylidene difluoride. EMBL: European Molecular Biology Laboratory, Heidelberg, Germany. ExPASy: Central Clinical Chemistry Laboratory and Medical Computing Center, Univ. Geneva, Switzerland. AQC; 6-aminoquinolyl-N-hydroxysuccinimidyl carbamate. PTC; phenylthiocarbonyl-

Materials and Methods

Chick ovalbumin and chick egg white lysozyme were purchased from Sigma; bovine serum albumin was obtained from the National Bureau of Standards. Solutions were prepared containing lysozyme and BSA, both at 0.05 mg/ml, 0.25 mg/ml or 0.5 mg/ml. Each solution also contained 0.25 mg/ml ovalbumin. Twenty μ l aliquots of these solutions were applied to BioRad Ready Gels (12% Tris-Glycine gels with 4% stacking gels).

Samples were electrophoresed using buffer composed of 25 mM Tris, 192 mM Glycine, and 0.1% SDS, pH 8.3. The samples were run for 15 minutes at 50V, then for 1 hour at 120V. Samples were electrotransferred to PVDF (Immobilon PSQ, Millipore) in a buffer system of 10 mM MES, pH 6, 20% methanol, 0.01% SDS for 4 h at 50 V. The resulting blots were stained with Coomassie blue, destained, and dried.

The dried blots were cut so that each participant received a piece of PVDF containing the unknown protein (either BSA or lysozyme) at 1 μ g, 5 μ g, and 10 μ g levels plus 5 μ g ovalbumin, the calibration protein. A photocopy of a 2D gel of the proteins with indicated molecular weight markers and pI ranges was included so that participants could estimate the molecular weight and pI of their unknown sample. Each participant was asked to analyze all 4 bands plus an unstained section of the blot as a blank. Data was to be submitted to the ExPASy (<http://expasy.hcuge.ch/ch2d/aacompi.html>) and/or Propsearch (<http://www.embl-heidelberg.de/aaa.html>) sites to identify the unknown protein.

Results and Discussion

The data received from the participating laboratories was evaluated using a “best fit” procedure previously described (8 [@]) to calculate the experimental mol % composition. By comparing with the theoretical compositions of the mature proteins, based on a 16 amino acid constellation excluding Trp and Cys, average mol % errors were determined for each data set, shown in **Table 1** [@]. The Average % Error overall was: 31% for the 1 μ g level [@], 20% for the 5 μ g level [@], 19% for the 10 μ g level [@] of unknown, and 17% for the known (Ovalbumin at 5 μ g, [@]). The level of Blanks ranged from 0.012 to 0.4 μ g total amino acids (n=20; one data set excluded), with an average of 0.13 μ g - over 10% of the predicted sample mass at the 1 μ g level. Low background levels (0.012 - 0.1 μ g) were common among the best performing sites at the 1 μ g level. However, low background was not a sufficient condition for high accuracy; several sites with relatively high % error reported similar low backgrounds.

The calculated amino acid compositions were submitted to the Geneva ExPASy site (**Table 2A** [@]), and to the Propsearch site in Heidelberg (**Table 2B** [@]). The ExPASy site allows inclusion of a calibration protein, which should be processed in parallel with the unknowns. This feature is designed to normalize for systematic errors at a particular site. Table 2A [@] lists identifications obtained using composition data submitted both without (no calibration), and with accompanying calibration data (with calib). Table 2A [@] & 2B [@] list the rank and distance score of the first correct protein family member identified (e.g., albumin or) without regard to species identification. When the correct species as well as protein family (e.g., albu_bovin) was identified as rank 1, the result is highlighted with shading.

Inspection of the errors for individual amino acids, **Table 3** [@], showed Glycine, Methionine, and Proline to have the highest errors in this study. These amino acids were omitted, both individually

and as a group, from the composition and the data re-submitted to the ExPASy algorithm (using a batch process). The resulting changes in the distance scores, either lower (improved), unchanged, or increased (further away from theory) are reflected in **Table 4** [a] as a percentage of the number of data sets available. Only in two cases was this procedure seen to improve the result; omitting Gly in the 1 μ g sample, and omitting Met in the ovalbumin data. These two amino acids had the highest errors noted for individual residues (Table 3 [a]). Even though dropping the most error-prone residues improved the overall average % errors, Table 4 [a] suggests that this approach is ineffective in improving protein identification. This is probably due to loss of specificity incurred by omitting residues from the search.

In contrast, filtering results that are returned by the search program produces clear benefits. **Table 5** [a] shows that restricting the species displayed produces a dramatic enrichment of successful identifications. Apparently, armed with knowledge of the species, MW, and pI, (as one would usually have following 2D-SDS-PAGE) one can greatly reduce false positives by screening out irrelevant identifications. Of course, this presumes the protein data base contains the protein sequences of the species of interest.

Evaluating the ability of the search programs to correctly identify proteins using “real world” data was a primary goal of this project. **Table 6** [a] summarizes the performance of the two search sites, and reveals that they gave similar results when just the unknown composition data is used. The ability to normalize the data with a control protein using the ExPASy site affords a modest improvement in success rate. Note that success is defined as simply identifying the correct protein family without regard to species (see Table 2A [a] & 2B [a] for species-specific “hits”). From Tables 6 [a], 2A [a] & 2B [a], one can see a strong correlation between the mol % error and successful identification - within the unknowns. Careful examination of Table 2A [a] & 2B [a] shows that the cutoff between success and failure falls at roughly 15% average error for the 16 amino acid constellation. When the results were sorted for the method used, mol % accuracy did not depend on any particular chemistry, as excellent results could be found within all method categories, and the average % errors were similar at each level (data not shown, but see Table 1 [a]).

Conclusions:

1. The average % error for samples blotted to PVDF is still about two-fold higher compared to solution samples. The average errors for each sample level were 31%, 20%, and 19% for the 1 μ g, 5 μ g, and 10 μ g levels, respectively. Ovalbumin (5 μ g) had slightly lower error (17%).

The 1 μ g level analysis was difficult for most laboratories. Many analyses were probably adversely affected by the presence of high background levels of amino acids, in particular Gly, Glx, and Ser. The sum of background amino acids (μ g) averaged 10% and 7% relative to predicted sample mass for the 5 and 10 μ g levels, respectively, and accuracy improved substantially at these levels. This correlation suggests that increased attention be given to laboratory and/or integration techniques which may contribute to background amino acid levels.

2. In this study, problematic amino acids were Glycine, Methionine, and Proline (Table 3 [a]). Glycine is often high in sensitive analyses, and Met and Pro, especially in the case of lysozyme, are rare residues which are prone to exaggerated errors. Also, it appears that Gly contamination and Met oxidation with the use of Tris-Glycine gels is a concern. Omitting these problem amino acids from the data was generally not helpful for searches (Table 4 [a]).

3. In contrast, filtering the results for relevant species, estimated pI, and MW appears to producedramatic improvements in the success of identification of these proteins (Table 5 [@]). This is effected by screening out all intervening ‘hits’; the distance scores are unaffected. Of course, this presumes the target protein is archived in the data base. Lacking this, similarity with homologs in the data base can be convincing if several occupy the top-ranked identifications and have good scores. This is a major advantage of this method.
4. There is no clearly superior chemistry; these data show that all commonly used chemistries can beused to achieve excellent results. Similarly, there is no clearly superior search algorithm using ‘relaxed’ identification criteria (i.e., protein family without regard to species (Table 6 [@]). In this study, the ExPASy site seemed somewhat better at providing correct species-restricted identifications at the 5 and 10 µg levels (highlights on Table 2A [@] & 2B [@]). Also, use of the ovalbumin control did produce a small improvement in the number of non-restricted identifications (Table 6 [@]). One may speculate that some laboratories had systematic errors which are amenable to this treatment.
5. These data show a direct relationship between the accuracy of amino acid analysis and the ability to correctly identify a protein using the database. As an approximation, an average mol % error of 15% is adequate to achieve correct protein identifications as defined in this study. This is accomplished by a majority of the sites when given adequate material. However, good results at the 1 µg level are, for the most part, only seen at the “best sites”- those consistently achieving the lowest average errors.

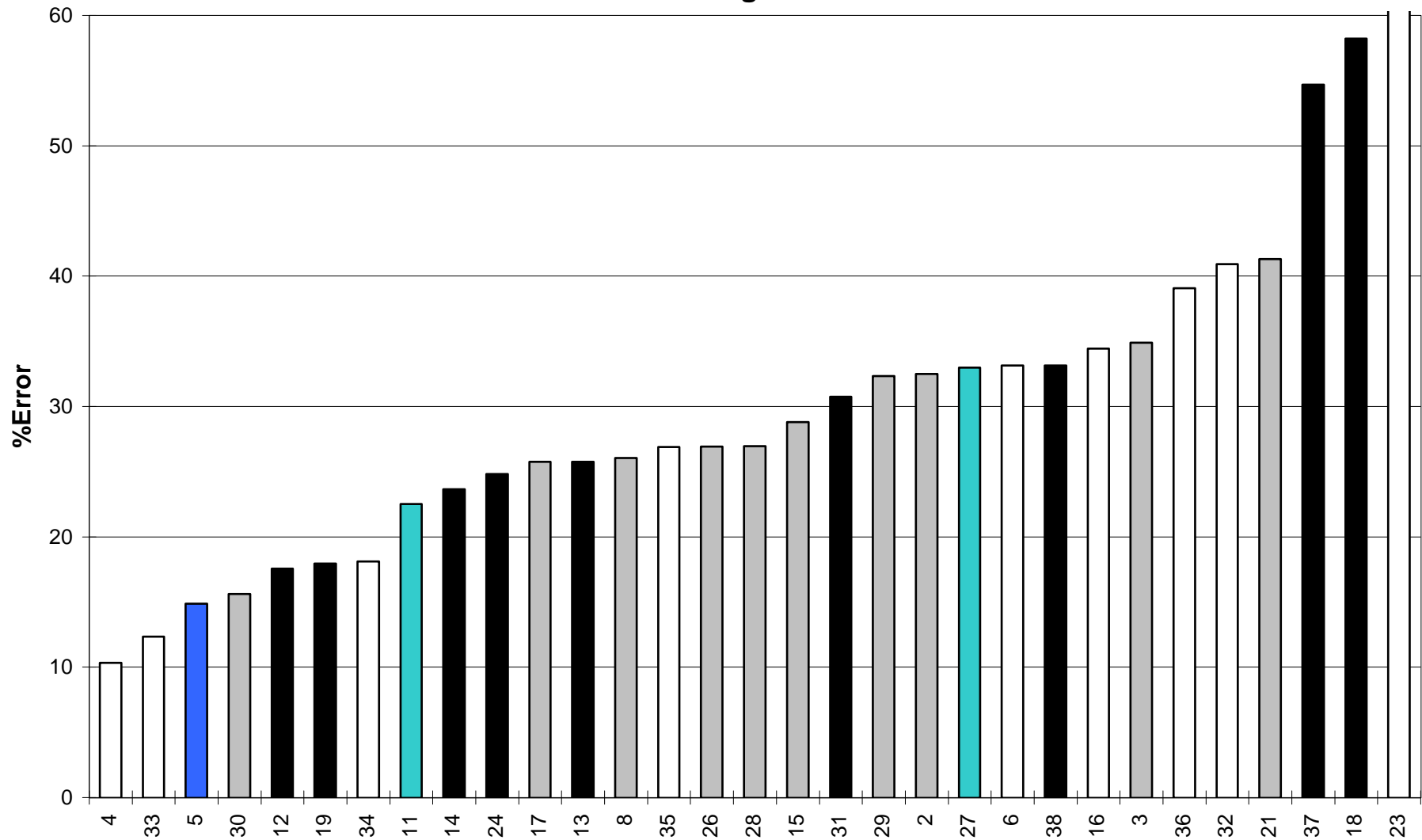
Acknowledgments: The authors are grateful to Dr. M. R. Wilkins for assistance with batch data submission for searches.

References:

1. Galat, A. et al., (1995) *Electrophoresis* **16**, 1095-1103 and references therein for a discussion of earlier efforts to apply this approach.
2. Hobohm, U., Houthaeve, T., and Sander, C. (1994) *Anal. Biochem.* **222**, 202-209. Amino acid analysis and protein database compositional search as a rapid and inexpensive method to identify proteins.
3. Touzet, et al., *Electrophoresis* (1996) **17**, 1393-1401. Amino Acid Analysis of Proteins Separated by Two-Dimensional Electrophoresis in maize: Isoform Detection and Function Identification.
4. Wilkins, M.R., et al. (1996) *Bio/Technology* **14**, 61-65. From Proteins to Proteomes: Large-Scale Protein Identification by Two-Dimensional Electrophoresis and Amino Acid Analysis.
5. Wilkins, M.R., et al. (1996) *BBRC* **221**, 609-613. Rapid Protein Identification Using N-Terminal “Sequence Tag” and Amino Acid Analysis.

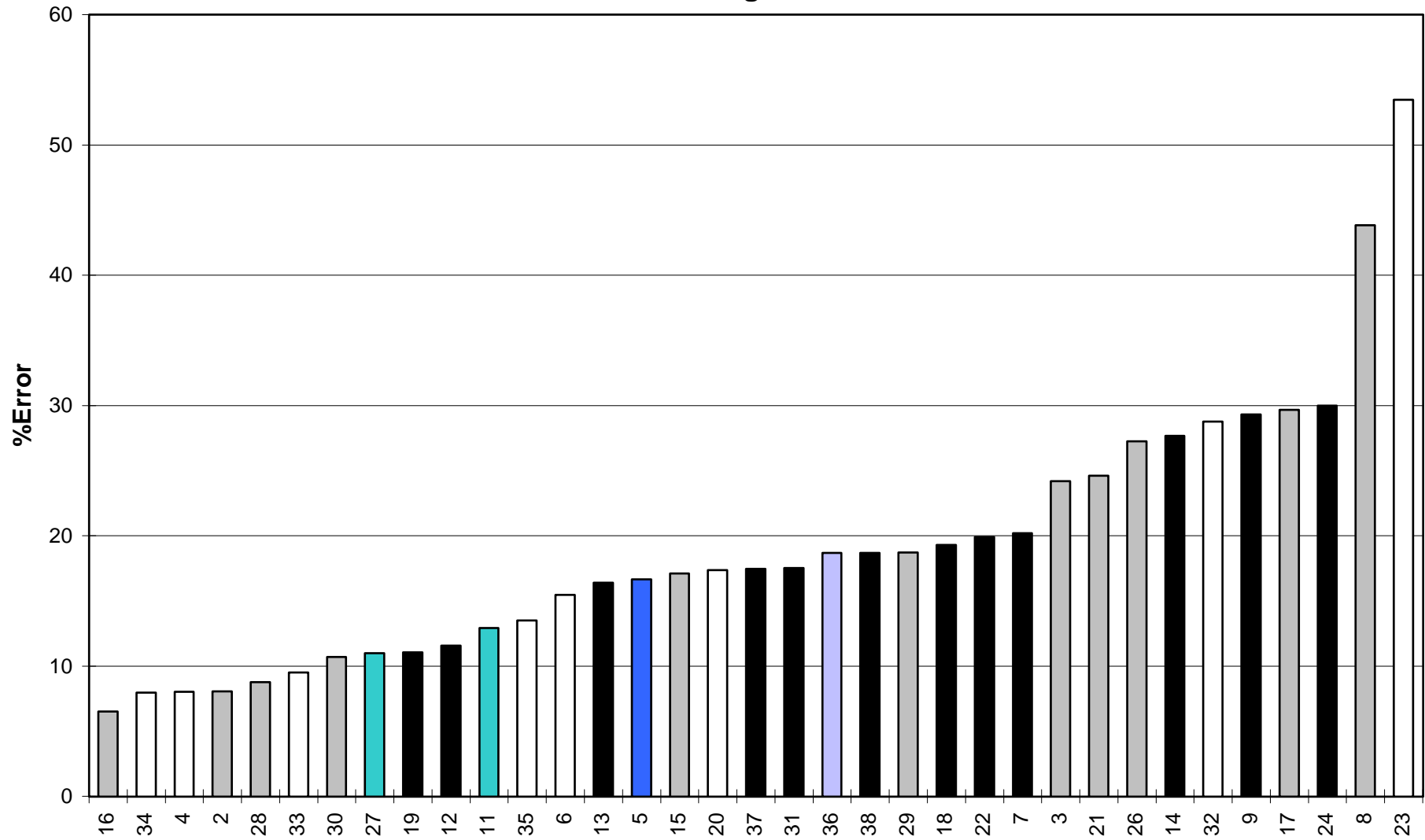
6. Cordwell et al. (1995) *Electrophoresis* 16, 438-443. Cross-species identification of proteins separated by two-dimensional electrophoresis using matrix-assisted laser desorption mass spectrometry and amino acid composition.
7. Wheeler et al (1996) *Electrophoresis* 17, 580-587. Characterisation of proteins from two-dimensional electrophoresis gels by matrix-assisted laser desorption mass spectrometry and amino acid compositional analysis.
8. Mahrenholz, A.M., et al., (1996) in *Techniques in Protein Chemistry VII*, Academic Press, San Diego, 323-330. Amino Acid Analysis - Recovery from PVDF membranes: ABRF-95AAA collaborative trial.
9. Schegg, K.M., et al. (1997) in *Techniques in Protein Chemistry VIII*, (Dan Marshak, ed.) Academic Press, San Diego, 207-216. Quantitation and Identification of Proteins by Amino Acid Analysis: ABRF-96AAA Collaborative Trial.
10. Tarr, G.E., et al. (1991) in *Techniques in Protein Chemistry II*, Academic Press, San Diego, 139-150. Amino Acid Analysis 1990: The Third Collaborative Study From the Association of Biomolecular Resource Facilities.

Table 1: 1 Microgram Level



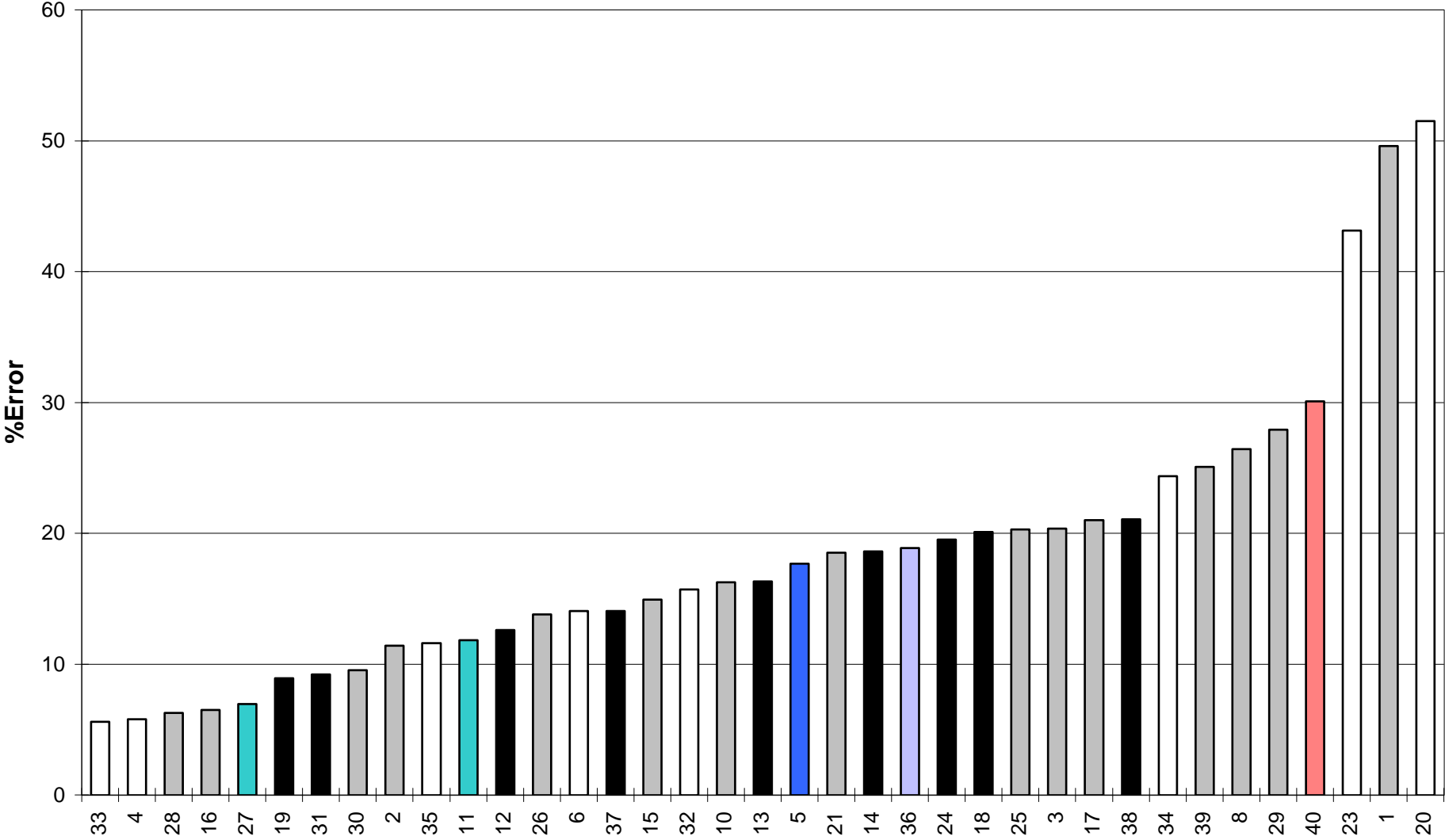
5 Microgram Level; 10 Microgram Level; Control (Ovalbumin); Identification Key

Table 1: 5 Microgram Level



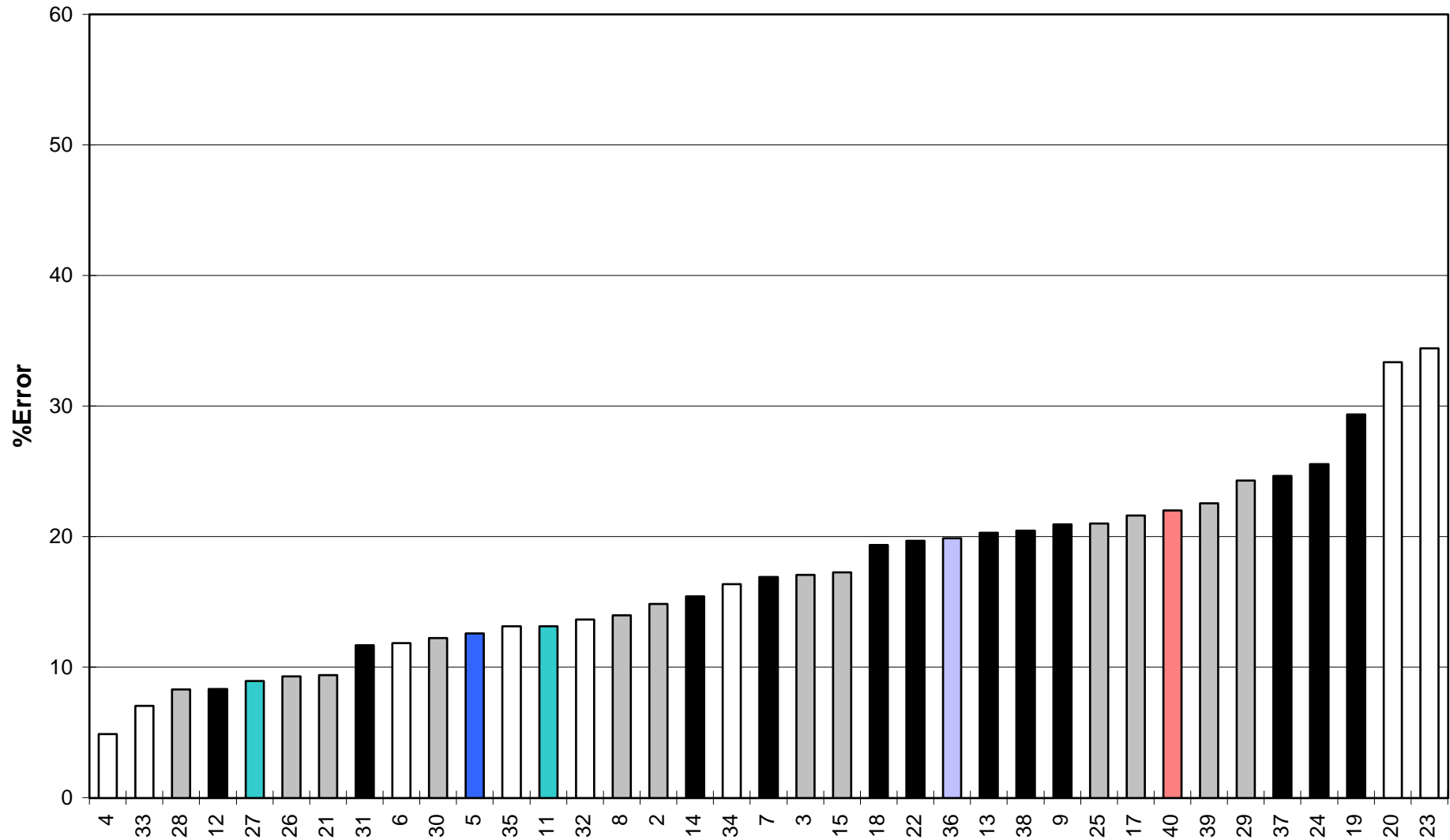
1 Microgram Level; 10 Microgram Level; Control (Ovalbumin); Identification Key

Table 1: 10 Microgram Level










1 Microgram Level; 5 Microgram Level; Control (Ovalbumin); Identification Key

Table 1: Control (Ovalbumin)



1 Microgram Level; 5 Microgram Level; 10 Microgram Level; Identification Key

Key to Derivatization Chemistry

Post-Column		Pre-Column	
Ninhydrin		AQC	
OPA		PTC	
		OPA/FMOC	
		OPA	
		Floram	

Note: Site 16 used AQC for the 1µg unknown and ninhydrin for the other samples.

Key to Site Identification

Site	Identifier	Site	Identifier
1	3931	21	9559
2	4951	22	1340
3	5368	23	3935
4	9327	24	1733
5	8284	25	4057
6	8931	26	6254
7	1812	27	2601
8	7106	28	4306
9	466A	29	1242
10	466B	30	1919
11	2274	31	8395
12	2868	32	5130
13	7609	33	1715
14	775	34	5439
15	6085	35	5439
16	715	36	1547
17	2454	37	9511
18	1568	38	9665
19	3407	39	8817
20	7826	40	8817

1 Microgram Level; 5 Microgram Level;
10 Microgram Level; Control (Ovalbumin);

Table 2A [\(cont. next page\)](#). ExPasy Results Using All Amino Acids

		1 ug unknown			5 ug unknown			10 ug unknown			Ovalbumin								
		no calibration		with calib	error	no calibration		with calib	error	no calibration		with calib	error						
Site	sample	rank	score	1ug			5ug			10 ug			rank	score	oval error				
				rank	score	Error	rank	score	Error	rank	score	Error							
1	bsa																		
2	bsa	>20	87	>20	89	32.5	1	3	1	13	8.1a	1	4	1	14	11.4	7	18	14.9
3	bsa	>20	74	>20	27	34.9	>20	34	1	8	24.2	15	24	1	5	20.3	>20	32	17.1
4	bsa	1	6	1	9	10.3	1	2	1	5	8.0	1	2	1	5	5.8	1	3	4.9
5	bsa	3	16	3	20	14.9	1	11	1	24	16.7	1	19	1	16	17.7	1	10	12.6
6	lyc	8	84	10	76	33.1	1	32	1	30	15.5	1	32	1	30	14.0	3	34	11.8
7	lyc						>20	116	1	7	20.2						>20	43	16.9
8	lyc	>20	86	7	42	26.0	>20	560	>20	417	43.8	1	33	1	16	26.4	1	17	14.0
9	bsa						7	18	>20	53	29.3						>20	151	20.9
10	bsa											3	10			16.3			
11	lyc	3	29	1	7	22.5	1	7	1	10	12.9	1	8	1	8	11.8	8	25	13.1
12	lyc	1	8	1	7	17.6	1	3	1	5	11.6	1	6	1	4	12.6	1	12	8.3
13	lyc	>20	97	1	24	25.7	1	45	1	14	16.4	1	36	1	14	16.3	>20	39	20.3
14	lyc	>20	197	>20	97	23.7	>20	174	>20	149	27.7	1	18	1	12	18.6	1	20	15.4
15	lyc	1	21	1	19	28.8	1	11	1	7	17.1	1	9	1	13	14.9	9	30	17.3
16	bsa	>20	76	>20	76	34.4	1	13	1	13	6.5	1	14	1	14	6.5	1	0*	
17	bsa	>20	72	>20	72	25.7	>20	64	>20	55	29.7	>20	58	>20	49	21.0	2	41	21.6
18	bsa	>20	>257	>20	119	58.2	1	17	1	8	19.3	>20	31	2	21	20.1	>20	36	19.4
19	bsa	1	11	1	22	17.9	1	3	1	25	11.1	1	2	1	51	8.9	>20	79	29.4
20	lyc	>20	813	>20	371		1	38	9	121	17.4	>20	167	1	11	51.5	>20	109	33.4
21	lyc	>20	94	>20	93	41.3	1	16	1	20	24.6	1	18	1	20	18.5	2	10	9.4
22	bsa						7	21	1	7	19.9						>20	25	19.7
23	lyc	>20	571	5	65	77.9	>20	291	>20	73	53.5	>20	166	1	11	43.1	>20	156	34.4
24	bsa	>20	41	17	30	24.8	>20	77	>20	44	30.0	29	31	>20	40	19.5	>20	41	25.5
25	lyc											1	21	1	20	20.3	29	24	21.0

Table 2A, Cont.

		1 ug unknown			5 ug unknown			10 ug unknown			Ovalbumin								
		no calibration	with calib		error	no calibration	with calib		error	no calibration	with calib		error						
Site	sample	rank score		rank score		1ug Error	rank score		rank score		5ug Error	rank score		rank score		10 ug Error	rank score		oval error
		26	lyc	22	51	>20	58	26.9	4	23	6	20	27.2	1	11	1	19	13.8	1
27	bsa	>20	62	>20	36	33.0	1	2	1	2	11.0	1	3	1	2	7.0	1	8	8.9
28	lyc	2	28	1	31	26.9	1	3	1	6	8.8	1	2	1	6	6.3	1	4	8.3
29	bsa	>20	63	>20	39	32.3	3	25	1	11	18.7	>20	97	>20	63	27.9	19	45	24.3
30	bsa	1	16	1	6	15.6	1	5	1	4	10.7	1	5	1	3	9.5	1	9	12.2
31	bsa	>20	39	>20	41	30.7	1	11	1	8	17.5	1	4	1	2	9.2	1	6	11.7
32	lyc	>20	169	>20	90	40.9	1	29	1	17	28.8	1	11	1	11	15.7	5	17	13.6
33	bsa	1	7	1	5	12.4	1	6	1	5	9.5	1	1	1	1	5.6	1	4	7.0
34	bsa	5	20	>20	25	18.1	1	4	3	12	7.9	>20	58	>20	76	24.4	5	19	12.9
35	lyc	>20	146	>20	165	26.9	1	20	1	28	13.5	1	8	1	9	11.6	2	19	13.1
36	lyc	>20	123	27	39	39.1	1	9	1	21	18.7	1	15	1	16	18.9	5	34	19.9
37	bsa	>20	212	>20	1362	54.7	1	10	>20	2296	17.5	1	9	>20	2268	14.1	>20	75	24.6
38	lyc	1	54	1	26	33.2	1	37	1	8	18.7	1	42	1	17	21.1	>20	62	20.5
39	lyc											1	32	1	18	25.1	>20	55	22.5
40	lyc											1	16	1	16	30.1	8	31	22.0

Shading denotes identification of correct family and species as rank 1, OTHERWISE, highest ranked protein family member reported. **No Calibration:** composition submitted without known as calibrant, **With Calib:** unknown submitted with ovalbumin as calibration standard. *Site used theoretical ovalbumin composition for control; data not used in calculations.

Table 2B. Propsearch Results Using All Amino Acids

		1 ug unknown			5 ug unknown			10 ug unknown			Ovalbumin		
Site	sample	rank	score	error	rank	score	error	rank	score	error	rank	score	error
1	bsa							>50		49.6			
2	bsa	>50		32.5	1	1.01	8.1	1	1.02	11.4	>50		14.9
3	bsa	>50		34.9	>50		24.2	>50		20.3	1	1.41	17.1
4	bsa	1	0.93	10.3	1	0.8	8.0	1	0.86	5.8	1	0.54	4.9
5	bsa	3	1.35	14.9	3	1.62	16.7	1	1.97	17.7	1	1.11	12.6
6	lyc	8	2.77	33.1	1	1.62	15.5	1	1.06	14.0	1	0.88	11.8
7	lyc				>50		20.2				>50		16.9
8	lyc	>50		26.0	>50		43.8	1		2.3	>50		14.0
9	bsa				>50		29.3				>50		20.9
10	bsa							1	1.3	16.3			
11	lyc	2	1.73	22.5	1	1.5	12.9	1	1.37	11.8	1	1.55	13.1
12	lyc	1	1.51	17.6	1	1.35	11.6	1	1.57	12.6	>50		8.3
13	lyc	39	2.66	25.7	2	2.1	16.4	1	2.05	16.3	>50		20.3
14	lyc	>50		23.7	>50		27.7	1	1.52	18.6	>50		15.4
15	lyc	1	2.23	28.8	1	2.01	17.1	1	2.01	14.9	>50		17.3
16	bsa	>50		34.4	1	0.94	6.5	1	0.94	6.5	1	0.21*	
17	bsa	>50		25.7	>50		29.7	44	2.57	21.0	6.00	2.17	21.6
18	bsa	>50		58.2	1	1.6	19.3	12	1.8	20.1	>50		19.4
19	bsa	1	1.34	17.9	1	0.92	11.1	1	1.01	8.9	>50		29.4
20	lyc	>50			2	2.08	17.4	>50		51.5	>50		33.4
21	lyc	>50		41.3	1	1.61	24.6	1	1.52	18.5	>50		9.4
22	bsa				1	1.4	19.9				>50		19.7
23	lyc	>50		77.9	>50		53.5	>50		43.1	>50		34.4
24	bsa	6	1.81	24.8	>50		30.0	1	1.38	19.5	>50		25.5
25	lyc							1	1.9	20.3	>50		21.0
26	lyc	18	2.25	26.9	38	1.97	27.2	1	1.68	13.8	>50		9.3
27	bsa	>50		33.0	1	0.96	11.0	1	0.83	7.0	1	1.16	8.9
28	lyc	2	1.93	26.9	1	1.5	8.8	1	1.29	6.3	1	0.76	8.3
29	bsa	>50		32.3	8	1.85	18.7	>50		27.9	>50		24.3
30	bsa	1	1.42	15.6	1	1.22	10.7	1	1.05	9.5	1	1.06	12.2
31	bsa	>50		30.7	1	1.3	17.5	1	1.03	9.2	1	1.07	11.7
32	lyc	>50		40.9	1	1.8	28.8	1	1.74	15.7	7	1.38	13.6
33	bsa	1	1.01	12.4	1	1.14	9.5	1	0.91	5.6	1	0.88	7.0
34	bsa	4	1.52	18.1	1	0.89	7.9	>50		24.4	>50		12.9
35	lyc	27	2.39	26.9	1	1.65	13.5	1	1.52	11.6	>50		13.1
36	lyc	>50		39.1	1	1.44	18.7	1	1.68	18.9	12	1.86	19.9
37	bsa	>50		54.7	1	1.3	17.5	1	1.28	14.1	>50		24.6
38	lyc	5	2.62	33.2	1	2.23	18.7	1	2.12	21.1	>50		20.5
39	lyc							2	2.21	25.1	7	2.19	22.5
40	lyc							1	1.79	30.1	2	1.7	22.0

Rank and score for highest placed homolog is given. Where the identification is true for exact species, (eg., oval_chick, albu_bovin, or lyc_chick) the score is shaded.

* theoretical composition was reported by site. This data was not used in calculations.

Table 3. Individual Amino Acid Errors (%)

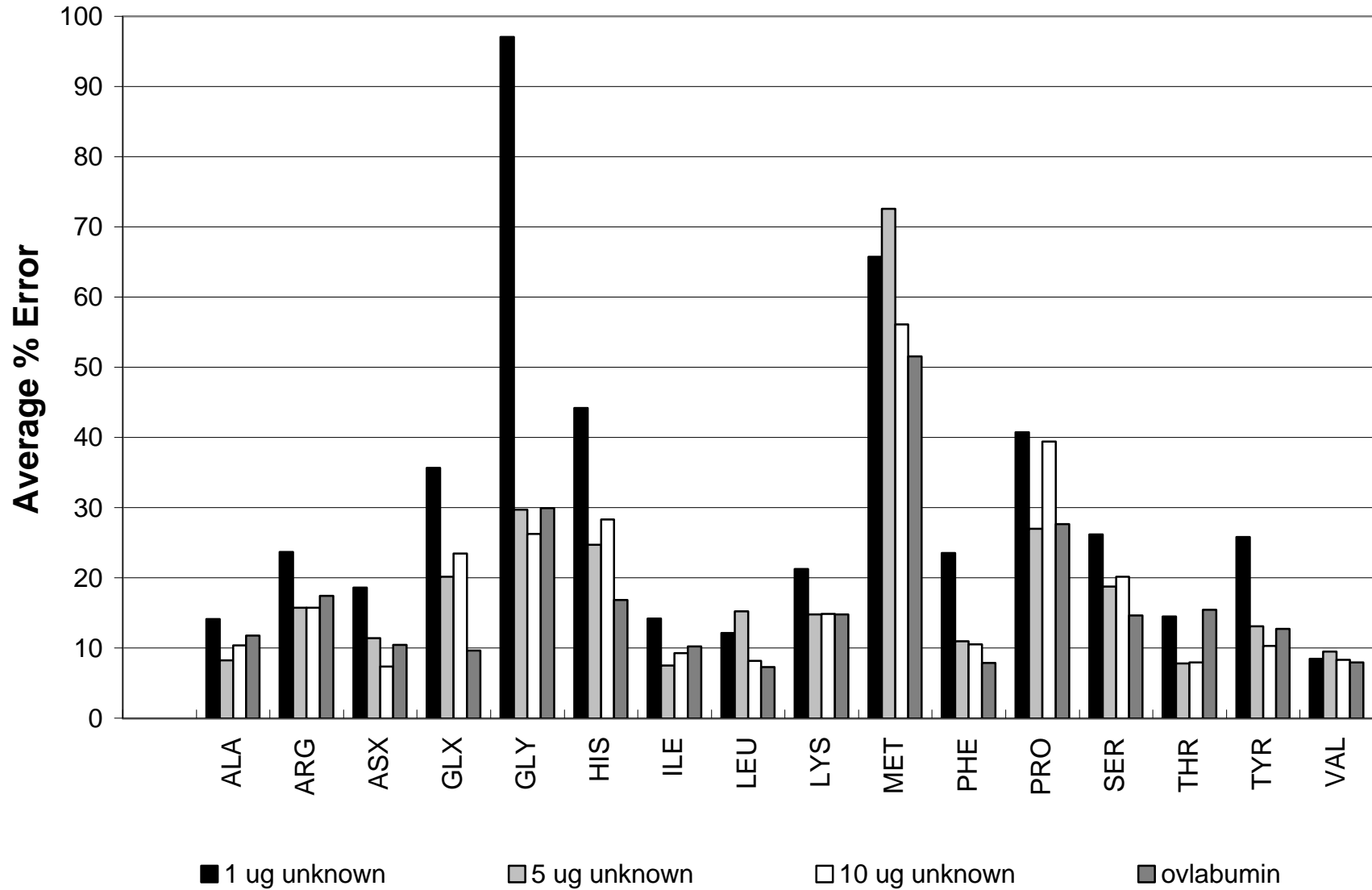


Table 4. ExPASy score changes upon omitting high-error amino acids from composition data. Results expressed as % of total in group.

10 ug Sample	Lower Score	No Change	Higher Score
without Gly	35	19	46
without Met	46	46	8
without Pro	27	30	43
without Gly, Met and Pro	41	18	41

5 ug Sample	Lower Score	No Change	Higher Score
without Gly	40	20	40
without Met	46	37	17
without Pro	17	26	57
without Gly, Met and Pro	46	14	40

1 ug Sample	Lower Score	No Change	Higher Score
without Gly	75	6	19
without Met	41	34	25
without Pro	38	14	50
without Gly, Met and Pro	78	3	19

Ovalbumin Sample	Lower Score	No Change	Higher Score
without Gly	47	11	42
without Met	84	8	8
without Pro	29	16	55
without Gly, Met and Pro	84	8	8

Note: Lowering the distance score indicates a better fit and is an improvement while a higher score indicates increased distance from theory.

Table 5.

When the results file is filtered to show only proteins of relevant species and pI, there is a dramatic improvement in identification of the species-specific protein as rank 1. (Shown for ExPASy data).

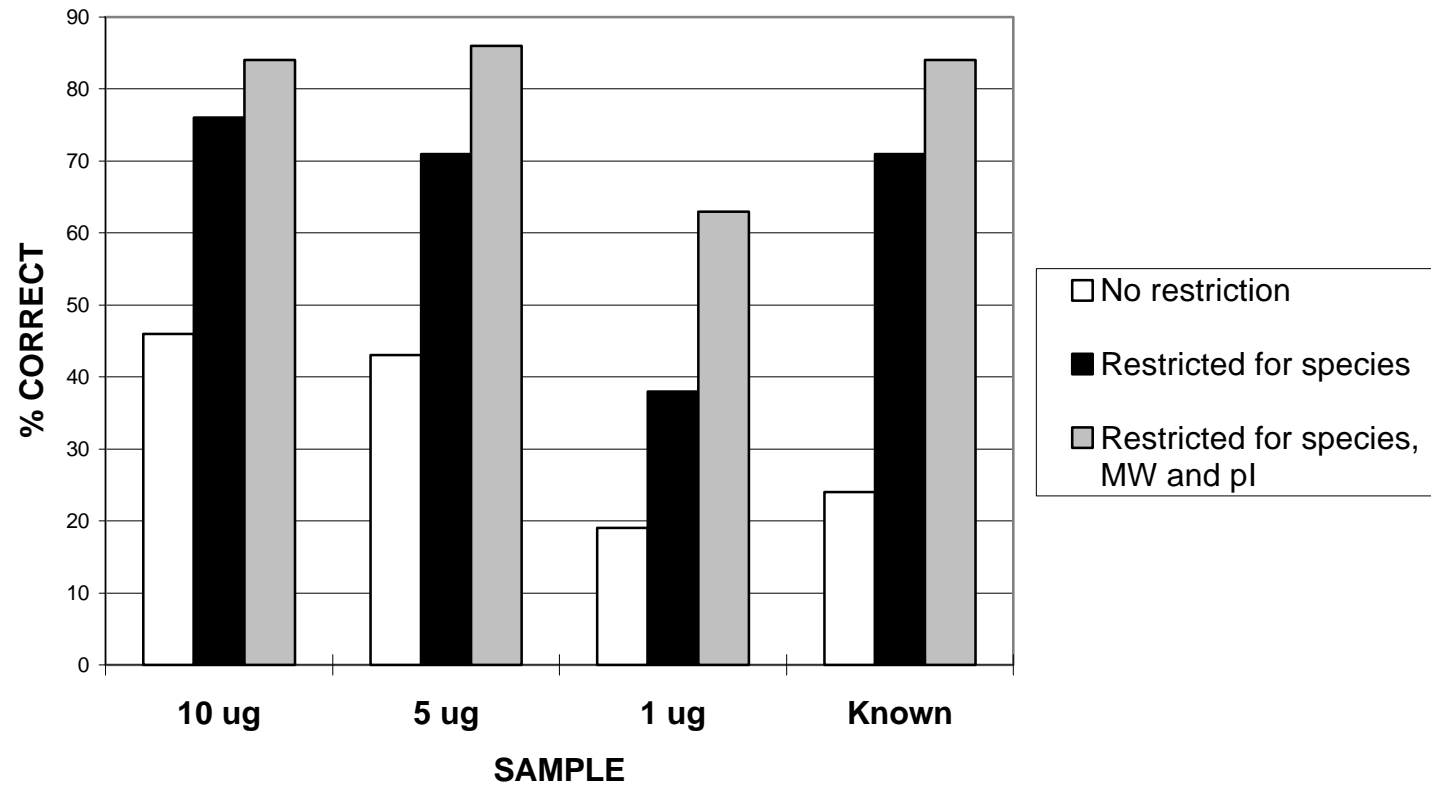


Table 6. Correlation of Identification with % Error

Sample Level	Avg % Error	N	Propsearch % Correct	ExPASy % Correct	
				no calib	with calib
1 µg	30	32	19	21	31
5 µg	19	35	63	68	71
10µg	19	37	76	72	80
Ovalbumin (5µg)	17	38	26	31	-

Results are expressed as a % of the total number of data sets (N) scoring Rank=1 (not species restricted). Avg % Error = average % composition error. ExPASy results given for data submitted without calibration protein (no calib) and with calibration protein (with calib).