

ABRF-2002 ESRG, A Difficult Sequence: Analysis of a PVDF-Bound Known Protein with a Heterogenous Amino Terminus.

Scott Buckel¹, Richard Cook², Myron Crawford³, Dave Dupont⁴, Ben Madden⁵, John Neveu⁶, Laurey Steinke⁷, and Joseph Fernandez⁸

¹Amgen, Boulder, CO, ²Baylor College of Medicine, Houston, TX, ³Yale University, New Haven, CT, ⁴Applied Biosystems, Foster City, CA, ⁵Mayo Clinic, Rochester, MN, ⁶Harvard, Cambridge, MA, ⁷Nebraska Medical Center, Omaha, NE, ⁸Rockefeller University, 1230 York Ave., New York, NY 10021.

ABSTRACT:

The ABRF-2002ESRG sample is the 14th study designed as an education and self-evaluation tool for laboratories that perform Edman Sequence Analysis. This year's study, one of the more challenging protein samples distributed by an ABRF research group, is a known protein with a staggered amino terminus which was a "real-life" sample submitted to a ESRG member lab. The protein was purified using commercially available, precast SDS-PAGE gels and transferred to PVDF. Protein bands were distributed to 72 members who requested ABRF-2002ESRG, along with a data instruction sheet and a brief survey. Participating members were requested to report observed raw data, interpret the data as they normally would for an investigator, and identify the protein using a BLAST search. Results from the study will be presented from 31 responses to show how labs fare with a difficult, but sequencable sample.

Materials and Methods

Sample Preparation:

The protein in this study was submitted to one of the committee members as part of a protein complex in 1.5 ml of 10 mM ammonium bicarbonate. A test SDS-PAGE gel transferred to PVDF and sequenced by the members showed the components of the complex could be readily separated with the lower MW band containing the protein of interest which has a staggered amino-terminus. The remainder of the sample was mixed with an equal volume of 2X Laemli resolubilization buffer plus B-mercaptoethanol and boiled for 3 minutes. The protein was then run onto ten 15-lane precast 10-20% Tris glycine 1.0 mm gels (NOVEX). After electrophoresis, the proteins were electrophoretically transferred to PVDF (NOVEX, 0.22 µm) at 250 mAmps for 2 hours in a full immersion transfer tank system. The PVDF membranes were then stained with 0.05% Coomassie blue G-250 in 50% methanol/10% acetic acid for five minutes, followed by destaining with 40% methanol/10% acetic acid, and finally washing with water overnight. Bands were excised and placed into 0.5 ml microcentrifuge tubes.

Scoring of Data:

Amino acids assigned in each cycle by the submitters were sorted by ESRG members. In the cases where more than three amino acids were listed in a cycle (three sequences were expected), priority was given to amino acids that matched the expected sequence. (Ex: Q,P,A,V,E assigned where Q,P,V expected was scored as three positive correct assignments. Q,P,A,E would have been scored as two positive correct and one positive wrong assignments.) Due to the complexity of the sample, unassigned positions were not counted.

*Our thanks to Dr. Robert Macnab (Yale University, New Haven, CT, USA) and Dr. Tohru Minamoto (Protonic Nanomachine ERATO, Kyoto, Japan) for the use of their protein complex in this study.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
N Sequence	S	N	E	L	P	W	Q	V	W	T	P	D	D	L	A	P	P	P	E	T
N-4 Sequence	P	W	Q	V	W	T	P	D	D	L	A	P	P	E	T	F	V	P	V	
N-1 Sequence	N	E	L	P	W	Q	V	W	T	P	D	D	L	A	P	P	P	E	T	F

Figure 1. Expected sequences in the ABRF-2002ESRG Sample.

Entry	Sequence	Match	Score
1	S N E L P W Q V W T P D D L A P P P E T	100%	100
2	S N E L P W Q V W T P D D L A P P P E T	100%	100
3	S N E L P W Q V W T P D D L A P P P E T	100%	100
4	S N E L P W Q V W T P D D L A P P P E T	100%	100
5	S N E L P W Q V W T P D D L A P P P E T	100%	100
6	S N E L P W Q V W T P D D L A P P P E T	100%	100
7	S N E L P W Q V W T P D D L A P P P E T	100%	100
8	S N E L P W Q V W T P D D L A P P P E T	100%	100
9	S N E L P W Q V W T P D D L A P P P E T	100%	100
10	S N E L P W Q V W T P D D L A P P P E T	100%	100
11	S N E L P W Q V W T P D D L A P P P E T	100%	100
12	S N E L P W Q V W T P D D L A P P P E T	100%	100
13	S N E L P W Q V W T P D D L A P P P E T	100%	100
14	S N E L P W Q V W T P D D L A P P P E T	100%	100
15	S N E L P W Q V W T P D D L A P P P E T	100%	100
16	S N E L P W Q V W T P D D L A P P P E T	100%	100
17	S N E L P W Q V W T P D D L A P P P E T	100%	100
18	S N E L P W Q V W T P D D L A P P P E T	100%	100
19	S N E L P W Q V W T P D D L A P P P E T	100%	100
20	S N E L P W Q V W T P D D L A P P P E T	100%	100

Figure 2. Sorted Tabulation of Submitted Results from ABRF-2002ESRG.

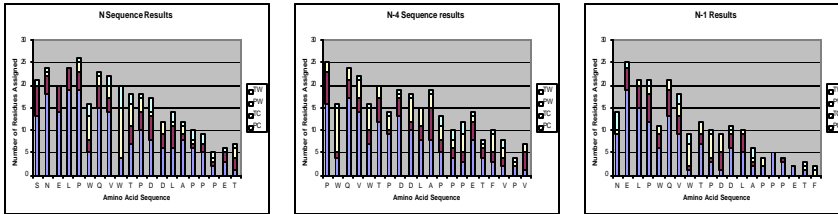


Figure 3. Summary of the Results from ABRF-2002ESRG. PC – Positive Correct, PW-Positive Wrong, TC – Tentative Correct and TW – Tentative Wrong

Temp	Temp type	Study process	Amount (µmol)	Positive Accuracy
STD-1	Peptide	Sequence analysis of a peptide	50	96
ABRF-ESRG	2 samples	Determination of the sequence	2400	96
ABRF-ESRG	Peptide	Evaluation of PVDF bound sample	31	83
ABRF-ESRG	Peptide	Determination of 2 post-translational modifications	100	94
ABRF-ESRG	Peptide	Determination of 2 post-translational modifications	50	91
ABRF-ESRG	Protein	Cy5 and TRP derivatization	50	96
ABRF-ESRG	Protein	Sequence calling ability of a single sequence	45	78
ABRF-ESRG	Protein	Sequence calling ability of a single sequence	40	100
ABRF-ESRG	Protein	Sequence calling ability of a mixture	10	96
ABRF-ESRG	Protein	Sequence calling ability of a mixture	2	86
ABRF-ESRG	Protein	Sequence calling ability of a mixture	10	92
ABRF-ESRG	Protein	Sequence calling ability of a mixture	2	72
ABRF-ESRG	Protein	Sequence calling ability of a mixture	23	91
ABRF-ESRG	Protein	Sequence calling ability of a mixture	10	99
ABRF-ESRG	Protein	Sequence calling ability of a mixture	5	86
ABRF-ESRG	Protein	Sequence calling ability of a mixture	5	86
ABRF-ESRG	Protein	Sequence calling ability of a mixture	22	70

Figure 4. Tabulation of past test samples and accuracy of calls and quantity of sample.

Facility Name	Number	Average Age	Academic	Research	Govt	Other	Other	Other
ABR-1	10	4.4	7.1	10.0	10.0	10.0	10.0	10.0
ABR-2	10	4.4	7.1	10.0	10.0	10.0	10.0	10.0
ABR-3	10	4.4	7.1	10.0	10.0	10.0	10.0	10.0
ABR-4	10	4.4	7.1	10.0	10.0	10.0	10.0	10.0
ABR-5	10	4.4	7.1	10.0	10.0	10.0	10.0	10.0
ABR-6	10	4.4	7.1	10.0	10.0	10.0	10.0	10.0
ABR-7	10	4.4	7.1	10.0	10.0	10.0	10.0	10.0
ABR-8	10	4.4	7.1	10.0	10.0	10.0	10.0	10.0
ABR-9	10	4.4	7.1	10.0	10.0	10.0	10.0	10.0
ABR-10	10	4.4	7.1	10.0	10.0	10.0	10.0	10.0

Figure 6. ESRG2002 Survey Results.

Facility	Search Type	Positive Results	Number of Positive Results	Protein Identity	Search Engine	Database
1	Y	Y	4	Protein Identity	BLAST	SwissProt
2	Y	Y	4	Protein Identity	BLAST	SwissProt
3	Y	Y	4	Protein Identity	BLAST	SwissProt
4	Y	Y	4	Protein Identity	BLAST	SwissProt
5	N	N	0	Protein Identity	BLAST	SwissProt
6	Y	Y	2	Protein Identity	BLAST	SwissProt
7	N	N	0	Protein Identity	BLAST	SwissProt
8	N	N	0	Protein Identity	BLAST	SwissProt
9	Y	Y	2	Protein Identity	BLAST	SwissProt
10	Y	Y	2	Protein Identity	BLAST	SwissProt
11	Y	Y	2	Protein Identity	BLAST	SwissProt
12	Y	Y	2	Protein Identity	BLAST	SwissProt
13	Y	Y	2	Protein Identity	BLAST	SwissProt
14	Y	Y	2	Protein Identity	BLAST	SwissProt
15	Y	Y	2	Protein Identity	BLAST	SwissProt
16	Y	Y	2	Protein Identity	BLAST	SwissProt
17	Y	Y	2	Protein Identity	BLAST	SwissProt
18	Y	Y	2	Protein Identity	BLAST	SwissProt
19	Y	Y	2	Protein Identity	BLAST	SwissProt
20	Y	Y	2	Protein Identity	BLAST	SwissProt

Figure 5. Database Search Results.

Conclusions:

As expected, this was a difficult sample. The early Trp, Ser, and Pro residues and several overlapping residues between the sequences made the pattern of the heterogeneous N-termini less apparent. Only 24 of the 31 respondents searched the sequence against a database. Of these, 9 correctly identified the protein. All of the databases were able to correctly determine the identity of the protein.