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

Scott Buckel[1], Richard Cook[2], Myron Crawford[3], Dave Dupont[4], Ben Madden[5], John Neveu[6], Laurey Steinke[7], and Joseph Fernandez[8]

[1]Amgen, Boulder, CO, [2]Baylor College of Medicine, Houston, TX, [3]Yale University, New Haven, CT, [4]Applied Biosystems, Foster City, CA, [5]Mayo Clinic, Rochester, MN, [6]Harvard, Cambridge, MA, [7]Nebraska Medical Center, Omaha, NE, [8]Rockefeller 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 um) 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 were 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 Minamino (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	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.

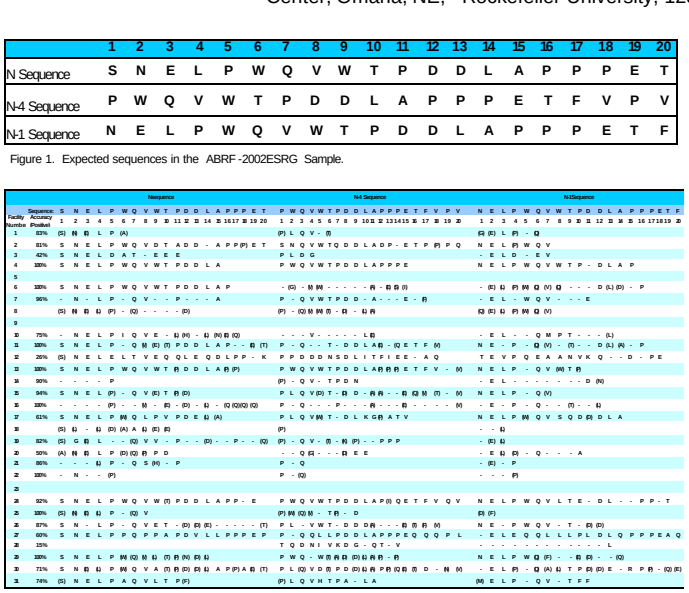

Figure 2. Sorted Tabulation of Submitted Results from ABRF-2002ESRG.

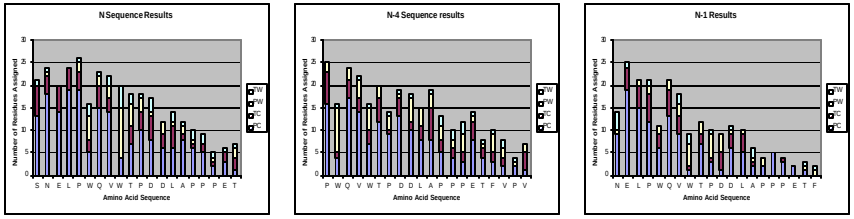

Figure 3. Summary of the Results from ABRF-2002ESRG. PC – Positive Correct, PW-Positive Wrong, TC – Tentative Correct and TW – Tentative Wrong

Sample	Sample type	Study purpose	Amount (pmol)	Positive Accuracy
STD-1	Peptide	Sequence analysis of a peptide	100	96
ABRF-89SEQ	2 peptides	Differentiation of two sequences	240/48	96
ABRF-90SEQ	Peptide conjugated to acetylated protein (PVDF)	Evaluate of PVDF-bound sample	30	83
ABRF-91SEQ	Peptide (rearrangement of STD-1)	Compare sequence results to STD-1	50	93
ABRF-92SEQ	Peptide	Determine 2 post translational modifications	500	84
ABRF-93SEQ	Peptide	Rearrangement of STD-1 sequence to determine improvement	50	91
ABRF-94SEQ	Protein	Cys and TRP determination	50	96
ABRF-95SEQ	Protein	4 cycles of microheterogeneity and length of sequence read	40	78
ABRF-96SEQA	Dataset	Sequence calling ability of a single sequence	40	100
ABRF-96SEQMmajor	Dataset	Sequence calling ability of a mixture	10	96
ABRF-96SEQMminor	Dataset	Sequence calling ability of a mixture	2	86
ABRF-97SEQmajor	Peptide (rearranged ABRF-96SEQmajor)	Compare lab generated data to previous year's dataset	10	92
ABRF-97SEQminor	Peptide (rearranged ABRF-96SEQminor)	Compare lab generated data to previous year's dataset	2	72
ABRF-98SEQ	Peptide	Ability to read a low level sequence and use of MS/MS data if possible	2,3	91
ABRF-99SEQprotein	Protein/peptide mixture	Differentiation of two sequences - Protein identification using BLAST search	10	99
ABRF-99SEQpeptide	Protein/peptide mixture	Differentiation of two sequences.	5	86
ABRF-00SEQ	Peptide (rearrangement of ABRF-92SEQ)	Determine 2 post translational modifications	5	86
ABRF-2002ESRG	Protein	Heterogeneous N-terminus. Protein identification using BLAST	35	76

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

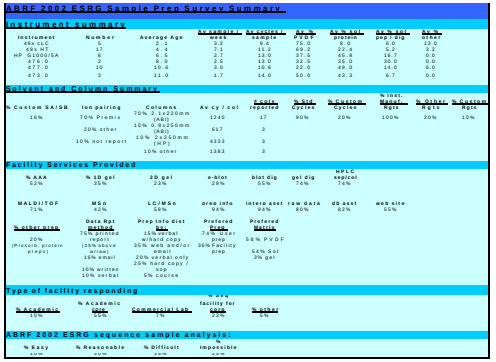

Figure 6. ESRG2002 Survey Results.

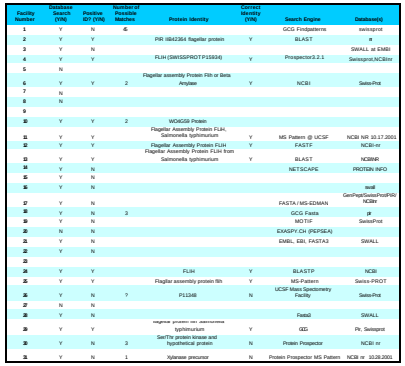

Figure 5. Database Search Results.

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.