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

Members of Committee

- Scott Buckel
- Richard Cook
- Myron Crawford
- Dave Dupont
- Ben Madden
- John Neveu
- Laurey Steinke, EB Laison
- Joseph Fernandez, Chair

ABRF 2002 ESRG

- Information on actual sample
- Report form for ABRF2002ESRG
- Actual report forms to users

Rationale for the ABRF 2002ESRG

- To determine participating members ability to analyze a heterogeneous N-termini.
- To determine participating members ability to analyze an electroblotted sample.
- To determine participating members ability to identify the protein.

Choice of Sample

- Sample was a real life sample, which was submitted to a committee member's lab for analysis.
- The sequence is in a number of databases.

Instructions to Participants

- General Instructions:
  - Enter one letter codes for the common amino acids in the AA column.
  - Any residue that for any reason has a possibility of being incorrect should be listed as tentative. Enclose tentative calls in parentheses ( ).
  - Use a – if no residue is observed or if no identification can be made at that position. Please, do not use an “X”.
  - Use an “X” only if an unknown amino acid is observed.

Instructions to Participants

Positive Identification of a Residue:

- Previous ABRF Sequence Research Studies have shown that a significant percentage of incorrect calls were listed as positively identified. Only those residues that can be called with high confidence and no ambiguity should be listed as a positive identification.

Database Searches:

- Do not use database searching to influence the information reported on the data result sheet. Apply your normal interpretation processes, including database searching, only to the requested user report. Include only the experimentally determined sequence data on the data result sheet.

Sample Preparation

- Sample was first tested on PVDF by ESRG members.
- Sample was run using Novex precast gels from same lot to minimize gel variability.
- Proteins were blotted 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.
- Samples were destained with 40% methanol/10% acetic acid.
- Samples were washed with water overnight .
- Bands were excised and placed into 0.5 ml microcentrifuge tubes.
- Each ESRG member's lab tested a band from the final batch.
- Bands were then shipped to participating members.

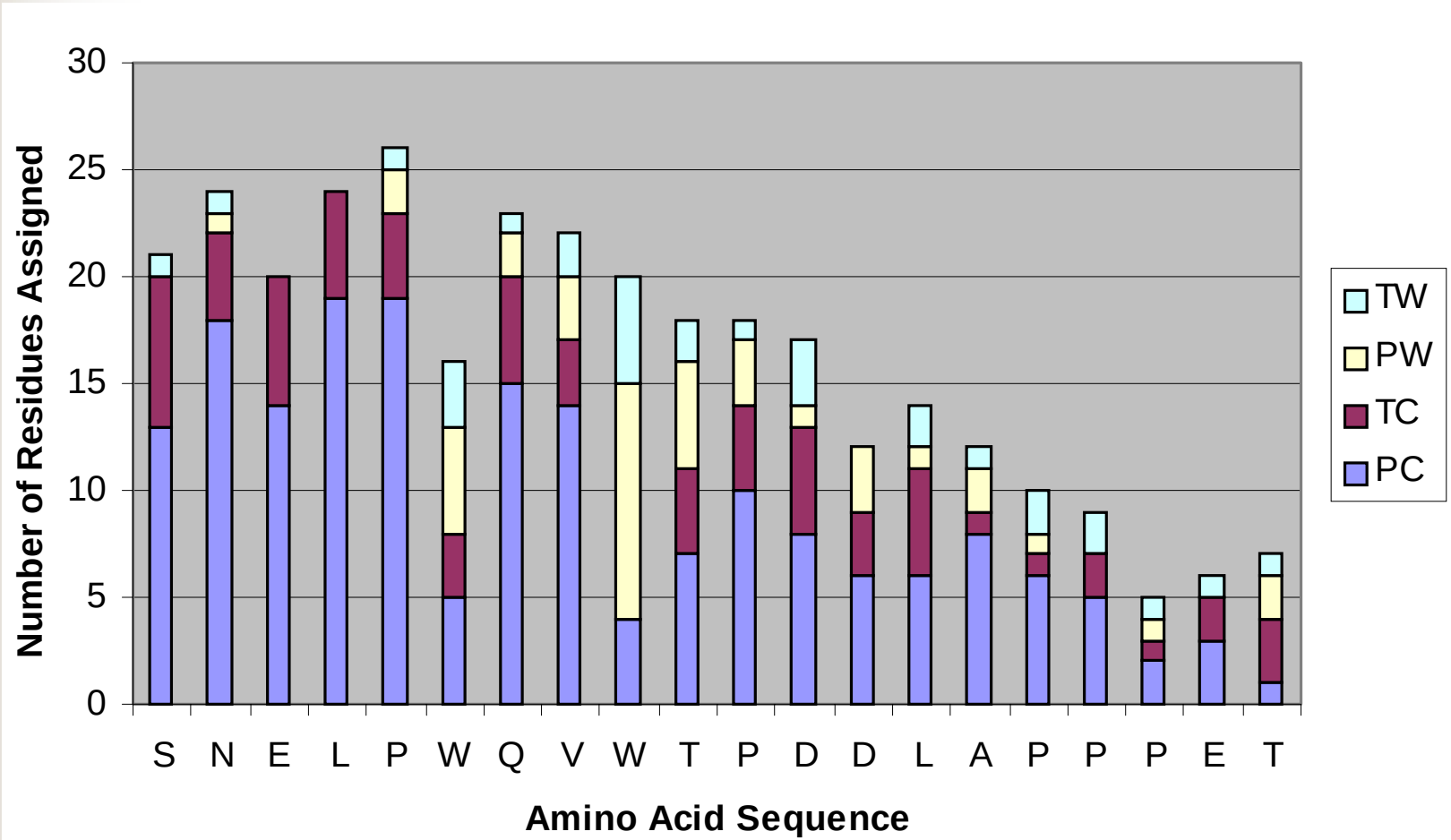
Data Sorting

- 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.
  - 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.

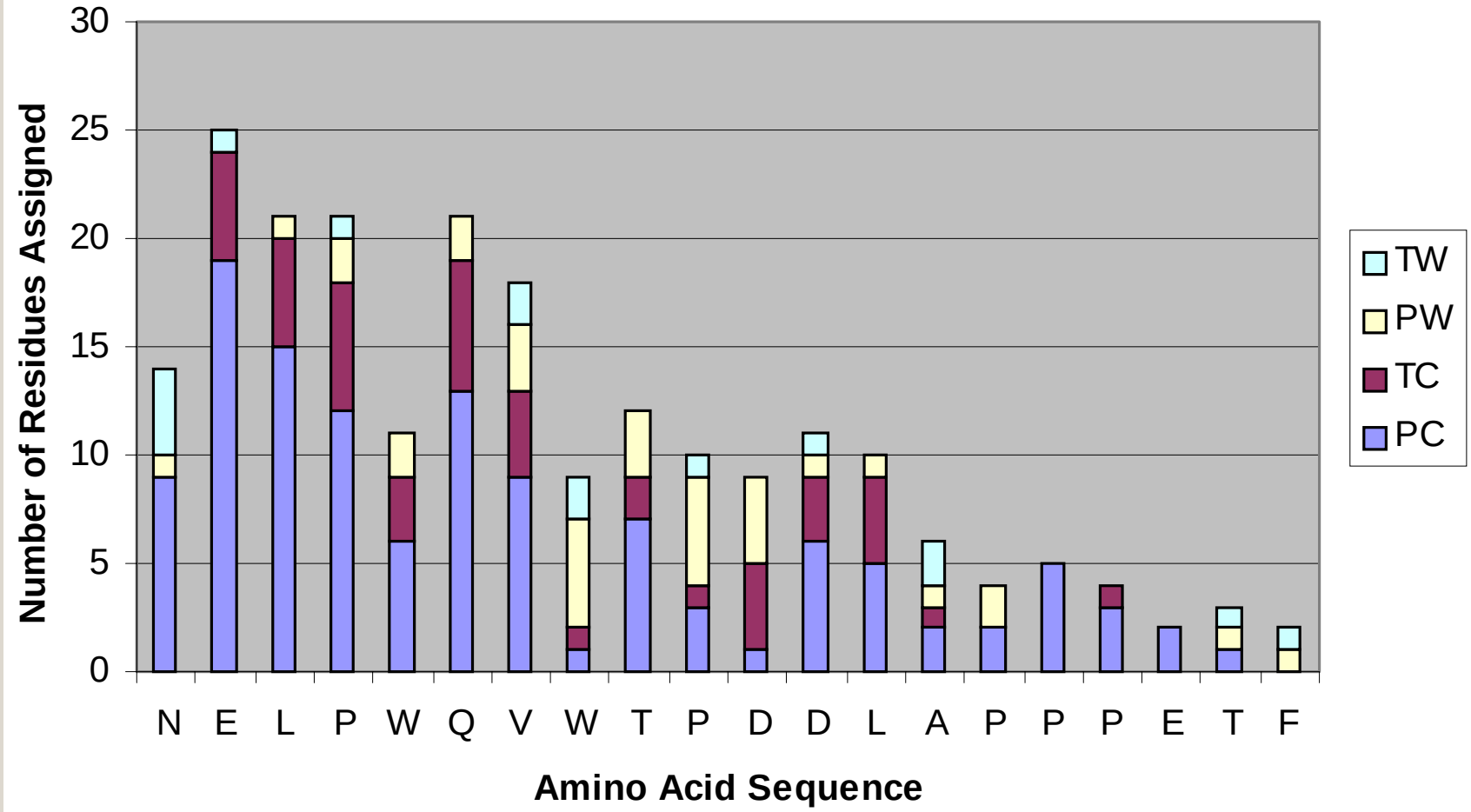
Sequence of Sample

	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

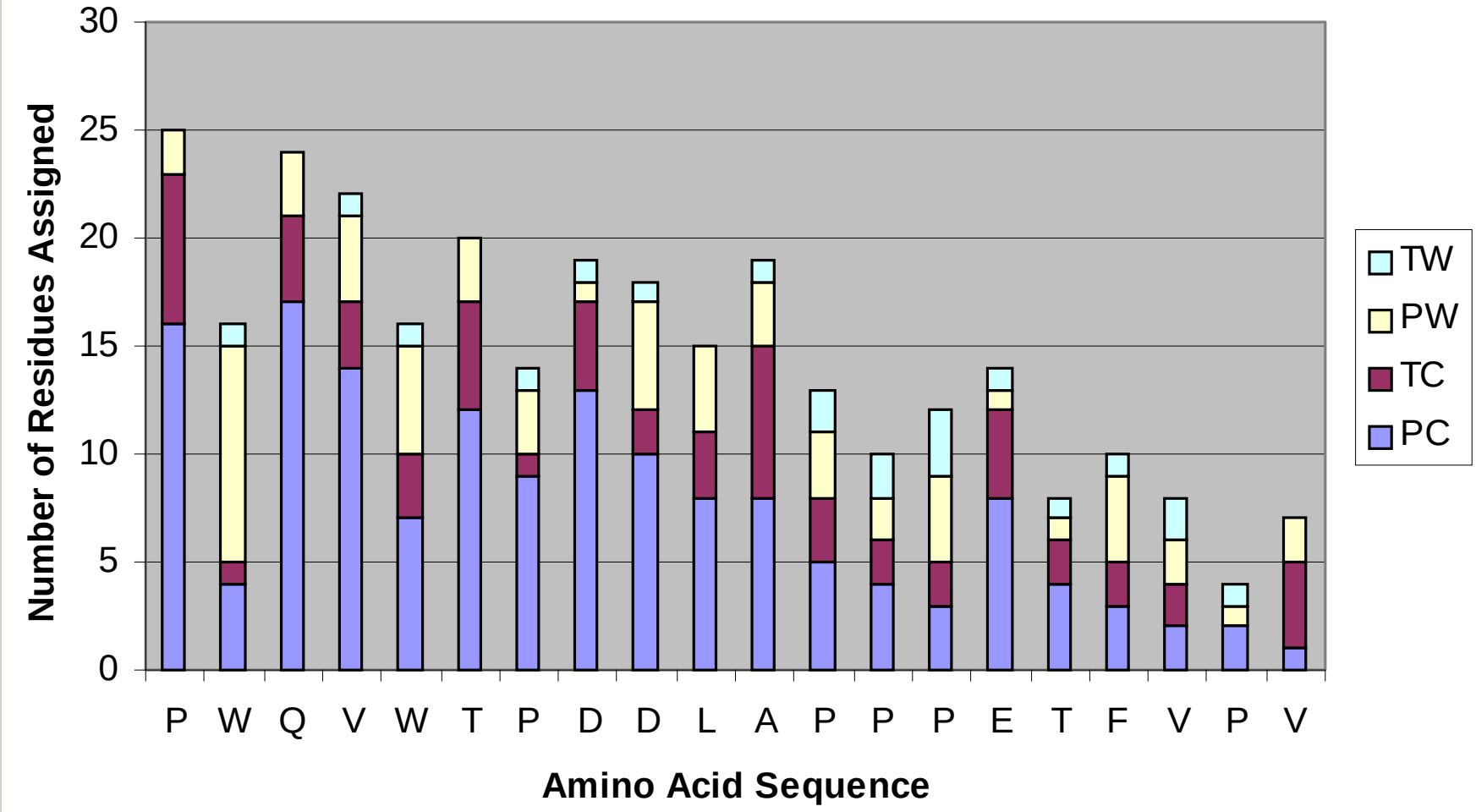
Summary of N Sequence Results



Summary of N-1 Sequence Results



Summary of N-4 Sequence Results



Accuracy Results from Past Samples

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 conjugated to	Compare solution sample to PVDF	50	83
ABRF-92SEQ	Peptide	Determine 2 post translational modifications	500	94
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	45	78
ABRF-96SEQA	Dataset	Sequence calling ability of a single sequence	40	100
ABRF-96SEQBmajor	Dataset	Sequence calling ability of a mixture	10	96
ABRF-96SEQBminor	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	Heterogenous N-terminus. Protein identification using BLAST	35	76

Database Search Results

Database Search (Y/N)	Positive ID? (Y/N)	# possible matches	Protein ID	Correct ID (Y/N)	Search Engine	Database(s)
Y	Y		PIR IIB42364 flagellar protein	Y	BLAST	nr
Y	Y		FLAGELLAR ASSEMBLY PROTEIN FLIH (SWISSPROT P15934)	Y	MS Edman from Protein Prospector 3.2.1	Swissprot,NCBI nr
Y	Y	2	Flagellar assembly Protein Flih or Beta Amylase	Y	NCBI	Swiss-Prot
Y	Y	2	WO4G59 Protein			
Y	Y		Flagellar Assembly Protein FLiH, Salmonella typhimurium	Y	MS Pattern @ UCSF	NCBI NR 10.17.2001
Y	Y		Flagellar Assembly Protein FLIH	Y	FASTF	NCBI-nr
Y	Y		Flagellar Assembly Protein FLIH from Salmonella typhimurium	Y	BLAST	NCBI\NR
Y	Y		FLAGELLAR ASSEMBLY PROTEIN FLIH	Y	BLASTP	NCBI
Y	Y		Flagllar assembly protein flih	Y	MS-Pattern	Swiss-PROT
Y	N	?	PM348	N	UCSF Mass Spectometry Facility	Swiss-Prot
Y	N	3	Ser/Thr protein kinase and hypothetical protein	N	Protein Prospector	NCBI nr
Y	N	1	Xylanase precursor	N	Protein Prospector MS Pattern	NCBI nr 10.28.2001

Sample Summary

- The sample was a very complex and difficult to analyze.
- The participants did a great job with this sample.
- We hope that everyone participates next year.

Data Reporting

- Participants were asked to report the data without using a database search to refine the data set.
- Some laboratories reported only the amino acids they saw from their analysis.
- Some laboratories reported a single protein sequence.
- Some laboratories reported three sequences and commented that they were truncated.
- Committee also asked for report forms that would be used to report data.

User Reports

- The committee was interested in the reports that members used to report data to clients
- Committee asked members to submit reports in the same format they would submit to clients

User Reports

- Six members sent user reports to the committee
- These reports were all different
- All gave the same basic information
- Few gave very extensive evaluation parameters
- Few gave search information as well as the sequence data

User Reports

- Reports ranged from being very formal format while others were informal
- They ranged from being handwritten to a typed memo style report
- Reports varied from having yield data to having histograms, to not including any yield data

Invitation

The committee welcomes any comments from the membership about this year's sample or about next year's sample or survey. Contact the committee with any ideas they may have for next years survey or sample for next year.

Conclusions

- As expected, this was a difficult but doable 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.

Conclusions

- Only 24 of the 31 respondents searched the sequence against a database.
- Of these, 9 correctly identified the protein.
- A variety of search engine and database combinations were able to correctly identify the protein.

Acknowledgements

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.

Thanks go out to all of the participants for taking the time to analyze the sample and send in their results.