

Susan Ludwigsen, Proteome Software, Inc

Scaffold uses a protein grouping algorithm to reduce the number of proteins under consideration. Generally, this algorithm works by creating a table like that shown in the similarity view of Scaffold. This table shows each protein to which the peptides could potentially be assigned, and in that protein's column shows the probabilities of all of the peptides that match it. The sum of the probabilities is calculated for each protein. Only the "valid" peptides are considered (Fig. 1).

Index	Peptide	Exclusive To	Valid	E	F	G	H	I	J	K	L	M	N	O
2	1 AKWYPEVR		FALSE	9%	9%	9%	9%	9%			9%	9%	9%	
3	2 CVVVG DGAVGK		FALSE	28%	28%	28%			28%		28%			
4	3 DDKDTIEK		TRUE	73%	73%	73%		73%					73%	73%
5	4 GSPQAIK	Chain A, Small G-Protein	TRUE	75%										
6	5 IISAMQTIKCVVVG DGAVGK		TRUE											
7	6 KLTPIYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%			95%	95%	95%	95%	95%	95%	95%	
8	7 LIPITYPQGLAMAK	Ras-related C3 botulinum tox	TRUE		95%	95%								
9	8 LTPITYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%			95%	95%	95%	95%	95%	95%	95%	
10	9 LVPITYPQGLAMAK		TRUE											
11	10 TVFDEAIR		TRUE	95%	95%	95%	95%	95%	95%	95%	95%			95%
12	11 VDSKPVNLGLWDTAGQEDYDR		TRUE											92%
13				433%	263%	263%	358%	358%	285%	285%	285%	263%	263%	260%

Figure 1

Each peptide is then assigned to the protein with the highest total probability in whose column it appears (Fig. 2). If two or more proteins have equal total probabilities and that is the highest for that peptide, it is assigned to all of them.

Index	Peptide	Exclusive To	Valid	E	F	G	H	I	J	K	L	M	N	O
2	1 AKWYPEVR		FALSE	9%	9%	9%			9%	9%	9%	9%	9%	
3	2 CVVVG DGAVGK		FALSE	28%			28%		28%		28%			
4	3 DDKDTIEK		TRUE	73%	73%	73%				73%	73%	73%	73%	73%
5	4 GSPQAIK	Chain A, Small G-Protein	TRUE	75%										
6	5 IISAMQTIKCVVVG DGAVGK		TRUE											
7	6 KLTPIYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%	95%	95%	95%	95%	95%			95%	95%	
8	7 LIPITYPQGLAMAK	Ras-related C3 botulinum tox	TRUE		95%	95%				95%	95%			
9	8 LTPITYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%	95%	95%	95%	95%	95%			95%	95%	
10	9 LVPITYPQGLAMAK		TRUE											
11	10 TVFDEAIR		TRUE	95%	95%	95%	95%	95%	95%	95%	95%			95%
12	11 VDSKPVNLGLWDTAGQEDYDR		TRUE											92%
13				433%	358%	358%	285%	285%	285%	263%	263%	263%	263%	260%

Figure 2 Assigned peptides in green, unassigned in grey.

Now the grouping begins. Proteins with no peptides assigned are eliminated from consideration, as all of the evidence for those proteins has already been accounted for in proteins which are more likely. Proteins with the same peptides assigned to them are combined into a group (Fig. 3).

Index	Peptide	Exclusive To	Valid	Group 1	Group 2	Group 3
3	AKWYPEVR		FALSE	9%	9%	9%
4	CVVVG DGAVGK		FALSE	28%	28%	28%
5	DDKDTIEK		TRUE	73%	73%	73%
6	GSPQAIK	Chain A, Small G-Protein	TRUE	75%		
7	IISAMQTIKCVVVG DGAVGK		TRUE			
8	KLTPITYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%		
9	LIPITYPQGLAMAK	Ras-related C3 botulinum tox	TRUE		95%	95%
10	LTPITYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%		
11	LVPITYPQGLAMAK		TRUE			
12	TVFDEAIR		TRUE	95%	95%	95%
13	VDSKPVNLGLWDTAGQEDYDR		TRUE			92%
14				433%	263%	263%
15						260%

Figure 3

There is one further complication, however. If the only evidence for a group is a single protein with probability less than 95%, Scaffold disregards this group. This is based on a heuristic rule built into the algorithm which cuts down on the number of false protein matches displayed. In this case it would eliminate Group 3 (Fig. 4).

Index	Peptide	Exclusive To	Valid	Group 1	Group 2	Group 3
3	1 AKWYPEVR		FALSE	9%	9%	9%
4	2 CVVVG DGAVGK		FALSE	28%	28%	28%
5	3 DDKDTIEK		TRUE	73%	73%	73%
6	4 GSPQAIK	Chain A, Small G-Protein	TRUE	75%		
7	5 IISAMQTIKCVVVG DGAVGK		TRUE			
8	6 KLTPITYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%		
9	7 LIPITYPQGLAMAK	Ras-related C3 botulinum tox	TRUE		95%	95%
10	8 LTPITYPQGLAMAK	Chain A, Small G-Protein	TRUE	95%		
11	9 LVPITYPQGLAMAK		TRUE			
12	10 TVFDEAIR		TRUE	95%	95%	95%
13	11 VDSKPVNLGLWDTAGQEDYDR		TRUE			
14				433%	263%	263%

Figure 4

Generally, this works well to eliminate false assignments, however in certain rare instances, it can result in a protein that may actually be found in the sample being eliminated from consideration, and thus not seen in Scaffold's other views. Unfortunately, changing the filter settings has no effect upon the grouping algorithm.

Back to... www.proteomesoftware.com << OR >> [the newsletter](#)