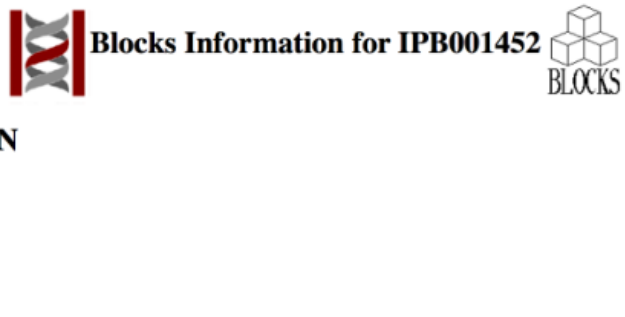
Project 2

Blosum

STEPS

1. retrieve blocks (see énonces)



Blocks Information for IPB001452

BLOCKS

IPB001452: SH3DOMAIN

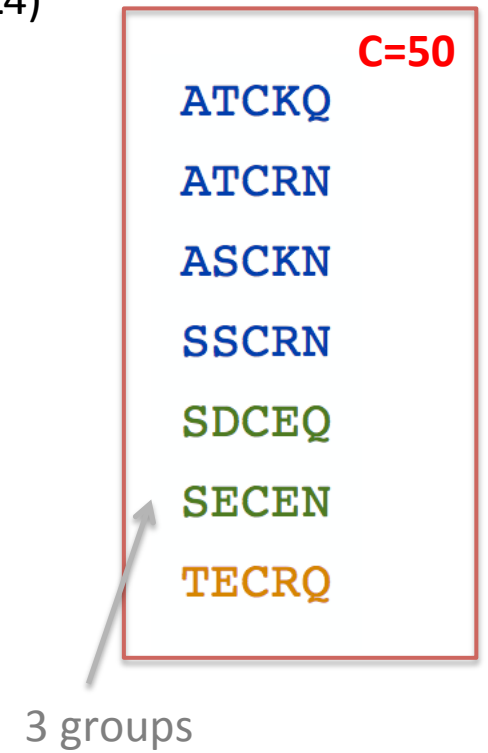
SH3 domain signature

- Introduction
- Block number IPB001452A
- Block number IPB001452B
- Block number IPB001452C
- Block number IPB001452D

STEPS

1. retrieve blocks (see énonces)

2. split blocks into groups with identity greater or equal given % C
   - first sequence -> group 1
   - for each following sequence:
     - if similarity with any of the sequences in group i >= C -> add to group I
   - grouping depends on ordering (slides 36-1 to 37-8 in L4)



STEPS

1. retrieve blocks (see énonces)
2. split blocks into groups with identity greater or equal given %
3. compute weighted frequencies f_{ab}
   - count # times AA a is present in group i and AA b is present in group j
   - example f_{QN}:

1st AA – 2nd AA	group 1st AA	group 2nd AA	#occurrences 1st AA	#occurrences 2nd AA
Q - N	1	2		
Q - N	1	3		
Q – N	2	3		
N – Q	1	2		
N – Q	1	3		
N - Q	2	3		

STEPS

1. retrieve blocks (see énonces)
2. split blocks into groups with identity greater or equal given %
3. compute weighted frequencies f_{ab}
   - count # times AA a is present in group i and AA b is present in group j
   - example f_{QN}:

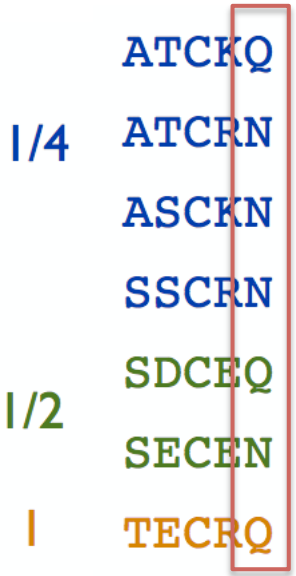
1st AA – 2nd AA	group 1st AA	group 2nd AA	#occurrences 1st AA	#occurrences 2nd AA
Q - N	1	2	1	1
Q - N	1	3	1	0
Q – N	2	3	1	0
N – Q	1	2	3	1
N – Q	1	3	3	1
N - Q	2	3	1	1

STEPS

1. retrieve blocks (see énonces)
2. split blocks into groups with identity greater or equal given %
3. compute weighted frequencies f_{ab}
   - count # times AA a is present in group i and AA b is present in group j
   - example f_{QN}:

1st AA – 2nd AA	group 1st AA	group 2nd AA	#occurrences 1st AA	#occurrences 2nd AA
Q - N	1	2	1	1
Q - N	1	3	1	0
Q – N	2	3	1	0
N – Q	1	2	3	1
N – Q	1	3	3	1
N - Q	2	3	1	1

   - multiply each with the corresponding group's weight
     1/ #sequences in group



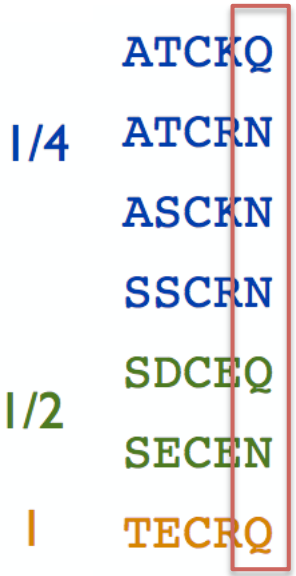
STEPS

1. retrieve blocks (see énonces)
2. split blocks into groups with identity greater or equal given %
3. compute weighted frequencies f_{ab}
   - count # times AA a is present in group i and AA b is present in group j
   - example f_{QN}:

1st AA – 2nd AA	group 1st AA	group 2nd AA	#occurrences 1st AA	#occurrences 2nd AA
Q - N	1	2	¼*1	½*1
Q - N	1	3	¼*1	1*0
Q – N	2	3	½*1	1*0
N – Q	1	2	¼*3	½*1
N – Q	1	3	¼*3	1*1
N - Q	2	3	½*1	1*1

   - multiply each with the corresponding group's weight

     1/ #sequences in group



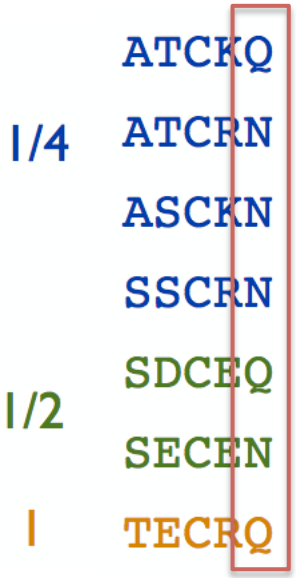
STEPS

1. retrieve blocks (see énonces)
2. split blocks into groups with identity greater or equal given %
3. compute weighted frequencies f_{ab}
   - count # times AA a is present in group i and AA b is present in group j
   - example f_{QN}:

1st AA – 2nd AA	group 1st AA	group 2nd AA	#occurrences 1st AA	#occurrences 2nd AA
Q - N	1	2	¼*1	½*1
Q - N	1	3	¼*1	1*0
Q – N	2	3	½*1	1*0
N – Q	1	2	¼*3	½*1
N – Q	1	3	¼*3	1*1
N - Q	2	3	½*1	1*1

   - multiply each with the corresponding group's weight

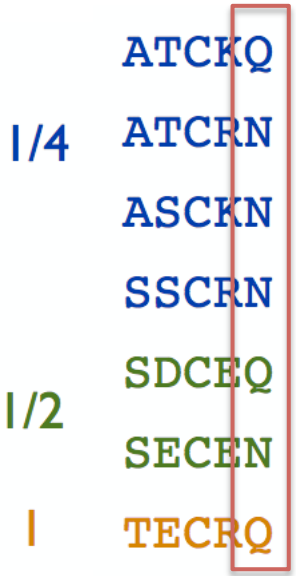
     1/ #sequences in group



STEPS

1. retrieve blocks (see énonces)
2. split blocks into groups with identity greater or equal given %
3. compute weighted frequencies f_{ab}
   - count # times AA a is present in group i and AA b is present in group j
   - example f_{QN}:

1st AA – 2nd AA	group 1st AA	group 2nd AA	#occurrences 1st AA	#occurrences 2nd AA
Q - N	1	2	¼*1	½*1
Q - N	1	3	¼*1	1*0
Q – N	2	3	½*1	1*0
N – Q	1	2	¼*3	½*1
N – Q	1	3	¼*3	1*1
N - Q	2	3	½*1	1*1



   - multiply each count with the corresponding group's weight
     1/ #sequences in group

example: (¼*1* ½*1)+(¼*1*1*0)+(1/2*1*1*0)+(¼*3*½*1) +(¼*3*1*1)+(½*1*1*1)= 7/4

   - repeat for each column and for each block

STEPS

1. retrieve blocks (see énonces)
2. split blocks into groups with identity greater or equal given %
3. compute weighted frequencies f
4. compute observed probabilities q
   - divide each f_{ab} by the the sum of the frequencies

$$q_{a,b} = \frac{f_{a,b}}{\sum_{1 \le b \le a} f_{a,b}}$$

	A	C	D	E	K	N	Q	R	S	T
A	0	0	0	0	0	0	0	0	3/4	3/4
C	0	3	0	0	0	0	0	0	0	0
D	0	0	0	1/2	0	0	0	0	1/4	1/4
E	0	0	1/2	1/2	1/2	0	0	6/4	3/4	3/4
K	0	0	0	1/2	0	0	0	1/2	0	0
N	0	0	0	0	0	3/8	14/8	0	0	0
Q	0	0	0	0	0	14/8	7/8	0	0	0
R	0	0	0	6/4	1/2	0	0	1/2	0	0
S	3/4	0	1/4	1/2	0	0	0	0	1/4	5/4
T	3/4	0	1/4	1/2	0	0	0	0	5/4	0

STEPS

1. retrieve blocks (see énonces)
2. split blocks into groups with identity greater or equal given %
3. compute weighted frequencies f
4. compute observed probabilities q
5. compute BLOSUM score s

$$s_{a,b} = 2\log_2\left(\frac{q_{a,b}}{e_{a,b}}\right)$$

where

$e_{aa} = p_a^2$ the same AA

$e_{ab} = 2p_a p_b$ different AA

and $p_Q = q_{Q,Q} + (1/2)\sum_{b \neq Q} q_{Q,b}$

	A	C	D	E	K	N	Q	R	S	T
A	0	0	0	0	0	0	0	0	0,05	0,05
C	0	5	0	0	0	0	0	0	0	0
D	0	0	0	0,033	0	0	0	0	0,0167	0,0167
E	0	0	0,033	0,033	0,033	0	0	0,1	0,05	0,05
K	0	0	0	0,033	0	0	0	0,033	0	0
N	0	0	0	0	0	0,025	0,1167	0	0	0
Q	0	0	0	0	0	0,1167	0,0583	0	0	0
R	0	0	0	0,1	0,033	0	0	0,033	0	0
S	0,05	0	0,0167	0,05	0	0	0	0	0,0167	0,0833
T	0,05	0	0,017	0,05	0	0	0	0	0,083	0