

TD 3 – Psi-blast et PSSM

Première partie

Le mutant VTC2 d'*Arabidopsis thaliana* est caractérisé par des niveaux très faibles en acide ascorbique (vitamine C). Le gène muté a été identifié comme étant le gène At4g26850 qui code pour la protéine référencée AAL07213.

On aimerait savoir de quel type de protéine il s'agit.

1) Recherchez la séquence protéique par Entrez. Avez-vous des informations sur la fonction moléculaire de cette protéine ?

Effectuez une recherche **psi-blast** (page blast du NCBI) à partir de la séquence d'*A. thaliana* dans la banque NR (avec en paramètre, un nombre maximum de séquences de 10 000).

2) Existe-t-il un domaine conservé d'après la recherche dans CDD ?

3) 1^{ère} itération :

- Combien de hits obtenez-vous avec $E < 10$?

Remarque : par défaut, le nombre maximum de hits affichés (graphique et alignements) est 100. Pour avoir l'ensemble des hits, modifiez les options d'affichage en haut de page.

- Dans quel(s) type(s) d'organismes ?

- Avez-vous des informations sur la fonction moléculaire de la protéine ? Quelle est l'annotation la plus fréquente ? Pouvez-vous savoir à quelle grande famille appartient la séquence requête ?

4) 2^{ème} itération

- Combien de hits obtenez-vous avec $E < 10$?

- Dans quel(s) type(s) d'organismes ?

- Quel type de protéines apparaissent avec $E < 0.005$?

- Quel motif ces nouvelles séquences possèdent-elles ?

5) 3^{ème} itération

- Combien de hits obtenez-vous avec $E < 10$?

- A quelle grande famille protéique appartient la séquence d'*A. thaliana* d'après les résultats obtenus ?

- D'après Interpro, il existe 3 sous-types à l'intérieur de cette superfamille. A quel sous-type appartient la séquence d'*A. thaliana* ?

6) KEGG pathway

Recherchez la voie de l'ascorbate dans les voies de la banque KEGG :

<http://www.kegg.jp/>

Plusieurs enzymes sont impliquées dans la voie allant du GDP-D-mannose au L-ascorbate. D'après les réactions indiquées, auriez-vous pu déduire la fonction précise du gène VTC2 ?

Deuxième partie : DNA Polymerase II de *Pyrococcus abyssi*

Il a été montré expérimentalement que l'ADN polymérase II de *Pyrococcus abyssi*, une enzyme dimérique, possède, en plus de son activité polymérase, une activité 3'-5' exonucléase.

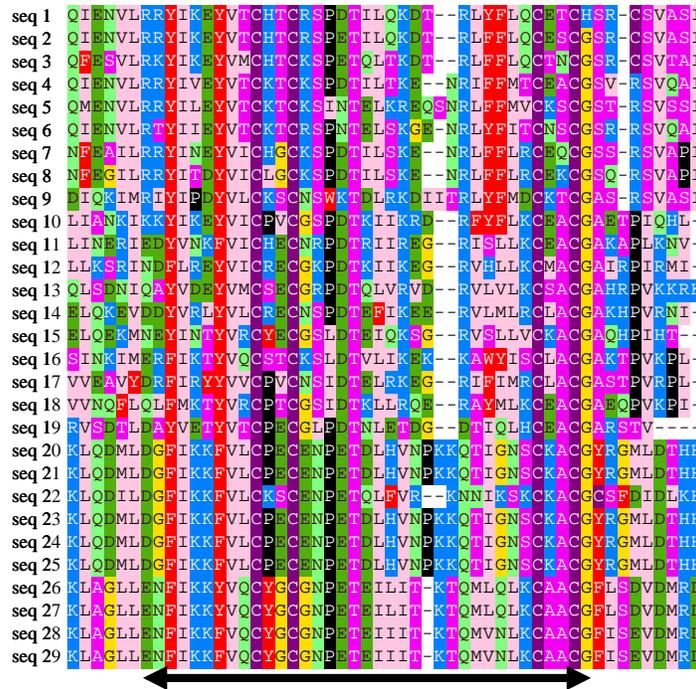
On déterminera la sous-unité responsable de l'activité exonucléase grâce à des comparaisons PSI-Blast

1) Rechercher les séquences similaires à la grande sous-unité par PSI-Blast. Que se passe-t-il au fur et à mesure des itérations ?

2) Rechercher les séquences similaires à la petite sous-unité par PSI-Blast. A quelles familles de polymérases eucaryotes est-elle apparentée ? Quel domaine est présent dans la région Cter ? Dans quelles autres familles est-il trouvé ? Citez 2 petits motifs conservés dans cette famille.

Troisième partie

Une région d'un alignement multiple est présentée ci-dessous.



Un profil (=PSSM) a été construit à partir de la région indiquée par une flèche (page suivante).

- Quelles sont les 4 positions les plus importantes dans ce profil ?
- Dans la ligne indiquée par la flèche (1), pourquoi la thréonine a une valeur très élevée ? Pourquoi la sérine a une valeur élevée dans cette même ligne ?
- Pourquoi toutes les valeurs de la ligne indiquée par la flèche (2) sont-elles proches de 0 ?

	Cons	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
	R	-47	-109	34	46	-105	-55	-8	-105	34	-95	-57	25	-52	38	61	-12	-32	-96	-114	-76
	R	-20	-85	-19	-3	-88	-26	-30	-80	17	-69	-46	-1	-57	-1	42	-13	-25	-70	-99	-72
	Y	-87	-87	-131	-100	169	-131	49	-31	-100	-31	-31	-100	-144	-69	-100	-87	-87	-44	74	254
	I	-41	-47	-143	-134	-6	-182	-140	176	-132	91	54	-140	-135	-130	-138	-93	-41	142	-138	-47
	K	-38	-95	-27	13	-94	-66	-32	-77	81	-60	-35	12	-39	12	37	-8	-21	-58	-108	-68
	E	-34	-108	28	85	-89	-72	-23	-86	40	-78	-54	-4	-41	31	-2	0	-2	-57	-95	-60
	Y	-110	-110	-165	-116	182	-165	93	-49	-116	-49	-49	-116	-171	-67	-116	-110	-110	-55	104	363
	V	0	-100	-300	-200	-100	-300	-300	300	-200	100	100	-300	-200	-200	-300	-200	0	400	-300	-100
	L	-23	-44	-78	-56	-36	-100	-72	25	-43	31	25	-54	-70	-32	-36	-29	22	24	-75	-45
	C	0	900	-300	-400	-200	-300	-300	-100	-300	-100	-100	-300	-300	-300	-300	-100	-100	-100	-200	-200
	H	-38	-89	-43	-11	-71	-67	32	-85	18	-72	-46	-18	21	-4	-1	-16	-38	-70	-87	-20
	T	-10	-82	0	37	-85	-34	-45	-65	-14	-71	-51	-6	-42	-1	-32	18	45	-33	-89	-72
	C	0	900	-300	-400	-200	-300	-300	-100	-300	-100	-100	-300	-300	-300	-300	-100	-100	-100	-200	-200
	K	-29	-106	-14	8	-102	30	-25	-113	36	-100	-66	34	-55	0	17	2	-39	-89	-97	-79
	S	14	-62	-6	-3	-91	-14	-31	-89	9	-82	-44	61	-53	1	-17	124	25	-89	-130	-83
	P	-50	-126	-68	-64	-154	-108	-101	-98	-61	-100	-69	-104	245	-60	-100	-57	-48	-66	-158	-121
	D	-71	-131	185	102	-122	-48	-26	-122	-12	-148	-108	44	-43	19	-52	3	-38	-112	-152	-109
(1) →	T	0	-100	-100	-100	-200	-200	-200	-100	-100	-100	-100	0	-100	-100	-100	100	500	0	-200	-200
	E	-37	-91	12	42	-74	-75	-33	-40	8	-51	-30	-14	-52	20	-15	-17	-30	-28	-100	-57
	L	-49	-49	-181	-142	8	-189	-140	107	-104	170	85	-142	-144	-104	-104	-95	-49	59	-98	-42
	M	-24	-59	-51	-23	-39	-79	-25	-9	-22	-13	-5	-31	-63	-9	-22	-11	-10	-10	-78	-31
	K	-35	-107	-57	14	-107	-87	-47	-83	135	-61	-29	-17	-53	20	66	-15	-28	-52	-119	-75
	E	-38	-110	64	85	-98	-44	-15	-92	8	-98	-65	21	-43	32	-9	4	-21	-72	-109	-75
	G	-16	-69	5	5	-70	8	-37	-65	-16	-68	-48	-7	-13	-9	-30	-1	-7	-46	-68	-61
(2) →	K	-4	-15	-8	0	-15	-14	-10	-10	17	-8	-4	-3	-10	0	3	3	-4	-7	-21	-12
	N	-19	-37	0	1	-40	-14	-2	-37	18	-33	-22	44	-22	1	6	10	9	-30	-48	-29
	R	-50	-141	-70	12	-141	-90	-1	-141	87	-98	-47	9	-87	61	198	-37	-46	-135	-139	-91
	L	-13	-35	-97	-76	-16	-102	-85	50	-62	57	32	-72	-80	-61	-68	-41	0	44	-77	-37
	F	-50	-56	-95	-81	82	-100	-21	10	-78	9	0	-78	-106	-68	-80	-51	-49	-10	4	74
	F	-49	-67	-92	-77	84	-81	-47	-4	-67	6	14	-80	-106	-58	-73	-53	-54	-18	-15	37
	L	-40	-46	-133	-102	-11	-138	-101	61	-69	116	80	-85	-109	-63	-70	-58	-35	31	-84	-44
	K	-22	-80	-27	14	-84	-59	-19	-73	59	-59	-28	-1	-43	29	37	7	-7	-53	-93	-55
	C	0	900	-300	-400	-200	-300	-300	-100	-300	-100	-100	-300	-300	-300	-300	-100	-100	-100	-200	-200
	E	-25	-99	-3	65	-84	-71	-30	-71	48	-52	-30	-9	-46	28	4	0	-21	-49	-97	-64
	A	122	-19	-64	-30	-88	-11	-71	-54	-26	-51	-42	-50	-43	-26	-34	52	15	-18	-123	-82
	C	0	900	-300	-400	-200	-300	-300	-100	-300	-100	-100	-300	-300	-300	-300	-100	-100	-100	-200	-200
	G	-3	-239	-80	-156	-236	466	-146	-317	-158	-317	-237	1	-159	-156	-156	-1	-159	-239	-159	-232