Evolution et biodiversité des microorganismes : travaux dirigés

Recherche de fonction(s) pour des gènes de fonction inconnue chez *Thermococcus kodakarensis*

Objectif : Savoir se servir des outils d'analyse de séquences afin de prédire la fonction d'un gène hypothétique.

Consignes : Donnez vos réponses en couleur pour mieux les repérer, utilisez la police Courier New (taille de lettre constante) pour écrire les séquences. Pour les informations concernant les logiciels utilisés rapportez-vous aux **TUTORIELS**.

Lors de ce TD nous vous proposons d'annoter les séquences génomiques issues de *Thermococcus kodakarensis* KOD1. Il s'agit d'une Euryarchée hyperthermophile (température optimale de croissance 85°C) et anaérobie stricte. Cette archée est l'un des modèles d'étude le plus utilisé par des chercheurs qui s'intéressent aux mécanismes moléculaires et leur évolution chez les Archées. *T. kodakarensis* possède un petit génome d'environ 2 Mpb et code environ 2300 protéines. Dans les conditions du laboratoire *T. kodakarensis* KOD1 se divise toutes les 30 min, aussi vite qu'*E. coli*. Actuellement, il existe plusieurs outils génétiques permettant de modifier son génome.

Au cours de ce TD vous allez utiliser plusieurs outils d'analyse de séquences et de structures protéiques afin de prédire la fonction d'un gène de fonction inconnue. Ces prédictions pourront par la suite être testées au laboratoire grâce à des approches génétiques, biochimiques, ou encore des approches utilisant le séquençage à haut débit (RNA-seq, ChIP-seq etc.). Dans le cadre de ce TD vous allez tester l'utilisation et la qualité de ces outils sur l'exemple d'un gène codant la protéine Pcc1 qui fait partie d'un complexe conservé chez les eucaryotes et les archées.

1. **Utilisez le navigateur Chrome**. Trouvez dans la base de données NCBI la séquence du génome de *T. kodakarensis* KOD1 dont le numéro d'accession est NC_006624.1 Ouvrez la fiche GenBank, faites apparaître toutes les caractéristiques (« customize view » → »customize »→ »all features ») et cherchez dedans (ctrl + F) le mot « CDS » : combien de résultats avez-vous obtenus ? Faites la même recherche pour le mot « hypothetical » et notez le nombre de résultats. Que vous apprennent ces données ?

CDS → 2318

Hypothetical → 530

Cela montre que parmi environ 2000 gènes environ un quart n'a pas de fonction prédite.

2. Cherchez (toujours ctrl + F) maintenant dans la fiche GenBank le gène *TK1253*. Il s'agit du gène codant la protéine Pcc1. Quelles sont ses coordonnées génomiques ? Quelle est la taille de la protéine prédite en acides aminés ?

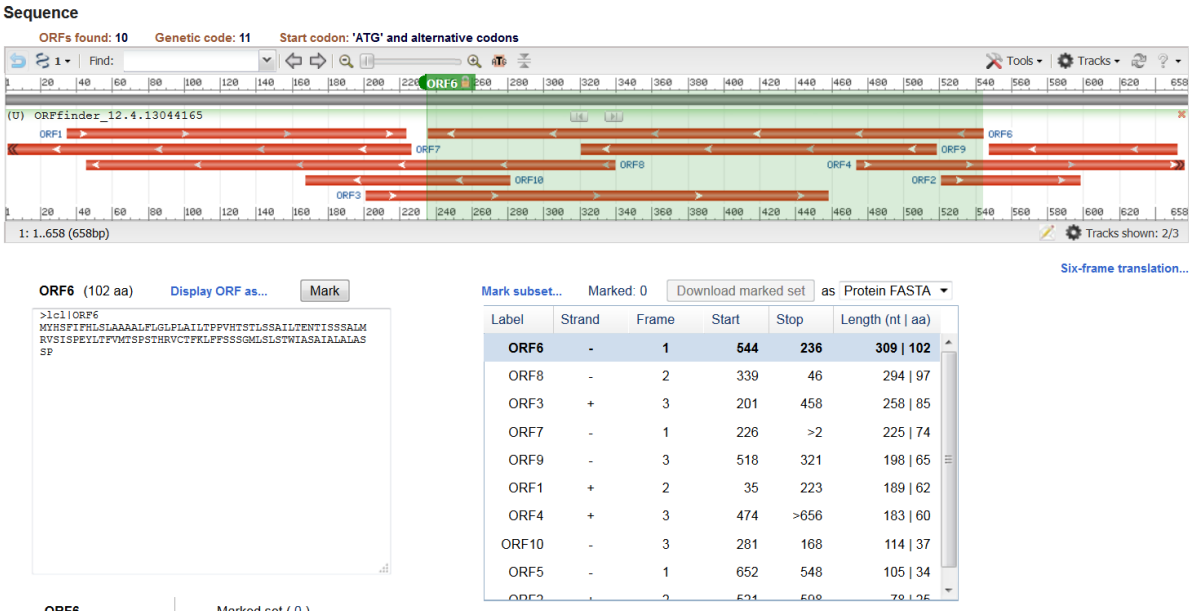
1098226..1098483
On peut cliquer sur protein_id : WP_011250204.1 cela ouvre la fiche GenBank pour TK1253 et on apprend qu'elle est constituée de 85 AA.

3. En utilisant maintenant l'option « change region shown » (voir TUTORIELS) récupérez la séquence contenant *TK1253* ainsi que 200 pb en amont et en aval. Copiez-collez la séquence sous format FASTA ci-dessous.

Il faut donc soustraire 200 bp de la première coordonnée 1098226 et rajouter 200 bp à la deuxième coordonnée 1098483 :

>AP006878.1:1098026-1098683 Thermococcus kodakarensis KOD1 DNA, complete genome
GGTAAAAGGCTCTGCCAGAACCACGGAGAAGGCACTGGAGAAGGGCTACCACCTTGGGGAAGCGTTGAAA
GAGGTTGCCGAAAAACTTGGCGGCGAGGGCGGTGGACACGCCATAGCCGCGGGCATACGCTTCCCGAAGA
ACAGGATAGACGAGTTCATTAAGCTCTTCAACGAAGCCCTCGGGAGGCAGGTGAAGGGCGGTGGAAGTGA
AGGCGAGGGTTGAGATAGTCTGGCACTACGGCGATGATGCCAAGGCGAGGGCAATAGCTGAAGCAATCCA
GGTGGACAACGAGAGCATGCCCGAGGAACTAAAGAAAAGTTTAAATGTGCAAACCCGATGGGTTGATGGA
GACGTCATAACAAAGGTTAAATACTCGGGTGAGATTGACACCCTCATCAAAGCGCTCGATGATATTGTGT
TTTCGGTCAAAATCGCCGAAGATAGCGTAGAGGTGTGAACTGGAGGTGTTAAGATGGCAAGAGGTAACCC
AAGGAAGAGGGCCGCTGCGGCTAAGGATAAGTGGAAGATGAAAGAGTGGTACATCGTTTACGCTCCCGAC
TTCTTTGGGAGCAAGGAGATAGGCCTCACTCCCGCTGACGACCCTGAGAAGGTTATAGGCAGGGTTATTG
AAACGACTCTGAAGGACCTCACCGGCGA

4. Nous allons maintenant utiliser l'outil ORFfinder du NCBI (voir TUTORIELS) pour traduire cette séquence en 6 cadres de lecture. Il s'agit ici de vérifier si la partie codante a été correctement annotée ou si des codons start alternatifs peuvent être trouvés. Choisissez l'option "ATG and alternative initiation codons" et le code génétique 11 (Bacteria, Archaea, plant plastids). Copiez-collez ci-dessous le résultat de votre recherche (capture d'écran).



5. Quelle ORF correspond à la protéine codée par *TK1253* ? Justifiez. Sur quel brin d'ADN est codé *TK1253* et quel est le cadre de lecture utilisé ?

Il s'agit d'ORF3, on s'attend à ce que notre ORF commence au nt. 201 (nous avons rajouté 200 nt en amont de TK1253 pour faire cette analyse) et la taille de la protéine prédite correspond à celle attendue (85 aa). l'ORF est codée par le brin + et le cadre de lecture est le numéro 3.

6. A votre avis est-ce que l'hypothèse d'un codon start alternatif est plausible ? Pouvez-vous en conclure que *TK1253* est correctement annoté ? Justifiez votre réponse.

Non, l'hypothèse n'est pas plausible. Si un start alternatif existaient, le programme aurait prédit une protéine >85 aa avec le même cadre de lecture (3) et sur le même brin (+). Cela n'est pas le cas, donc nous pouvons conclure que l'annotation du gène est probablement correcte.

7. Récupérez la séquence protéique de la protéine TK1253 (Pcc1) sous format FASTA et copiez-collez la ci-dessous.

```
>lcl|ORF3
MEVKARVEIVWHYGDDAKARAIAEAIQVDNESMPEELKKSLNVQTRWVDGDVITKVKYSGEIDTLIKA
LDDIVFSVKIAEDSVEV
```

8. Nous allons maintenant utiliser BLASTp pour chercher les orthologues de Pcc1 chez les Thermococcales. Au préalable, trouvez dans NCBI Taxonomy les noms des trois genres appartenant à l'ordre des Thermococcales et notez les ci-dessous. Vous pouvez ignorer les catégories « unclassified » ou « environmental samples »

Thermococcus, Paleococcus, Pyrococcus

9. Faites maintenant votre recherche BLASTp contre les Thermococcales. Trouvez-vous des orthologues chez tous les genres de cet ordre ? Que signifie cela en termes d'histoire évolutive de Pcc1 au sein de ce groupe d'organismes ? Gardez la page des résultats BLAST pour plus tard.

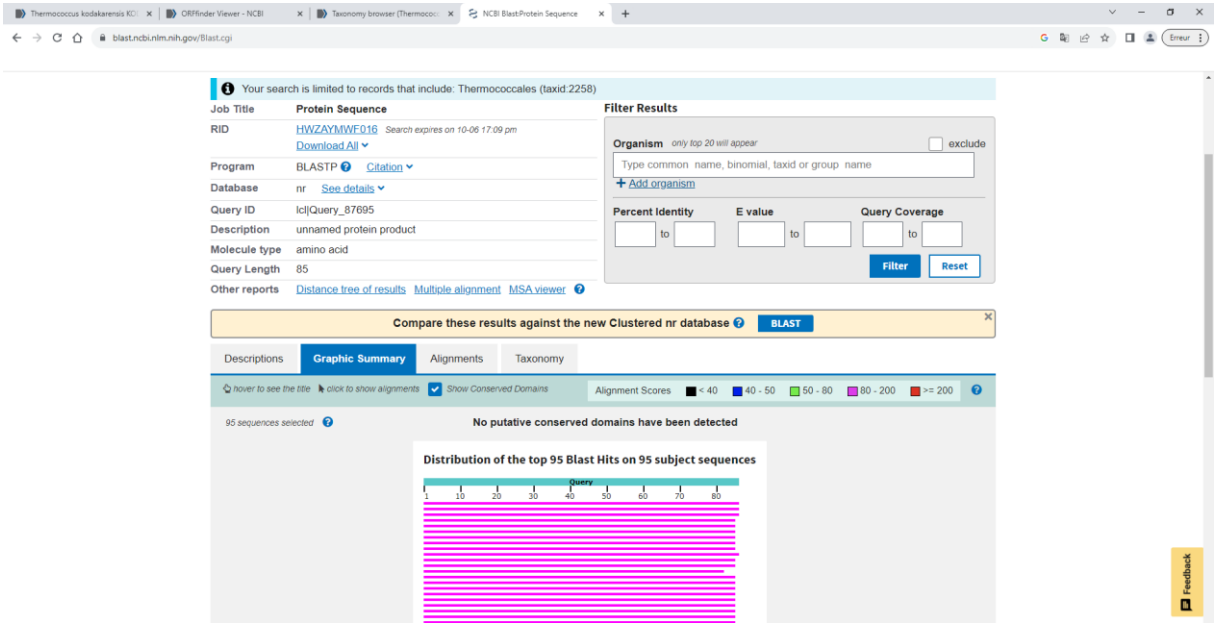
Oui, on trouve un orthologue chez plusieurs représentants des trois genres (onglet Taxonomy), cela signifie que la protéine est conservée chez les Thermococcales, nous pouvons donc émettre l'hypothèse que l'ancêtre commun de ce groupe codait pour le Pcc1.

10. Question préliminaire : comment peut-on définir un domaine protéique ?

Un domaine protéique correspond à une séquence et une structure tertiaire protéiques conservées au sein d'un polypeptide. Un domaine peut exister et fonctionner indépendamment du reste du polypeptide. Souvent les protéines sont des assemblages de deux ou plusieurs domaines.

11. Est-ce que BLAST trouve des domaines conservés dans votre séquence ? Comment peut on expliquer cela ?

Lorsque BLAST identifie un domaine conservé une image avec le nom du domaine apparait au-dessus de la représentation « graphic summary » Dans notre cas, il semblerait que BLAST n'a pas détecté de domaine particulier au sein de la séquence de la protéine Pcc1. Cela peut être due au fait que ce domaine n'est pas encore répertorié dans les bases de données ou du fait que les séquences de Pcc1 de Thermococcales sont très divergentes par rapport aux Pcc1 d'autres organismes.



12. D'autres méthodes (autre que les recherches par BLAST) peuvent être utilisés pour détecter les protéines contenant des domaines conservés. C'est le cas de PFAM qui utilise HMM (hidden markov model) pour détecter des résidus signature au sein d'un groupe de séquences homologues. PFAM (https://www.ebi.ac.uk/interpro) répertorie toutes les superfamilles, familles et domaines protéiques connues ainsi que des alignements de séquences pour chaque famille ou domaine. Faites une recherche avec le mot-clé Pcc1 dans la fenêtre prévue à cet effet (cliquez sur « Search » ensuite « by text »). Quel est le numéro d'accession de la base de données PFAM pour la famille Pcc1 ?

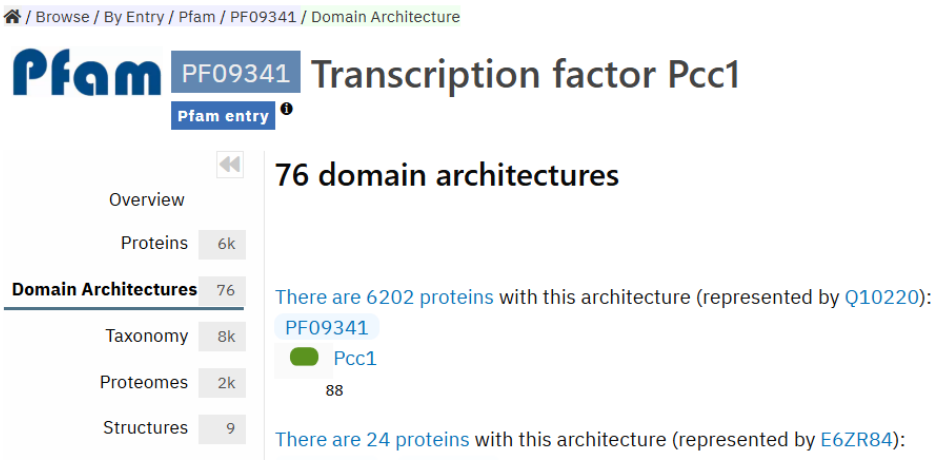
PF09341

1 - 20 of **28** entries — Export

ACCESSION	NAME	SOURCE DATABASE	DESCRIPTION
PF09341	Transcription factor Pcc1	PFAM	Pcc1 is a transcription factor that functions in regulating genes involved in cell cycle progression and polarised growth [PMID:16874308].
IPR015419	CTAG/Pcc1 family	INTERPRO	This entry represents the CTAG/Pcc1 family. Its members include yeast EKC/KEOPS complex subunit Pcc1, mammalian EKC/KEOPS complex subunit Lage3 and human cancer/testis antigen (CTAG) 1/2. In Sacch...
PTHR31283	EKC/KEOPS COMPLEX SUBUNIT PCC1 FAMILY MEMBER	PANTHER	
IPR027893	EKC/KEOPS complex subunit GON7, metazoa	INTERPRO	...threonylcarbamoyladenosine (t6A). In eukaryotes, KEOPS is composed of OSGEP/Kae1, PRPK/Bud32, TPRKB/Cgi121, LAGE3/Pcc1 and GON7 [[cite:PUB00084171]]. This family consists of subunit GON7 from Metazoa.

13. Regardez maintenant les informations contenues dans l'onglet « domain architectures » (à gauche). Ici le PFAM répertorie toutes les protéines contenant la séquence étudiée en tant que domaine protéique isolé ou fusionné à d'autres domaines. Sous quelle forme trouve-t-on Pcc1 majoritairement ?

Sous forme isolé, en tant qu'un seul peptide, PFAM trouve presque 6000 protéines de ce type.



14. Il peut être informatif de s'intéresser à des protéines « fusion » ou la protéine de fonction inconnue fait partie d'un polypeptide plus grand contenant un ou plusieurs domaines de fonction connue. Une telle fusion peut indiquer que la protéine d'intérêt est associée avec les domaines de fonction connue dans une même voie métabolique. Il y a-t-il des protéines fusion contenant le domaine Pcc1 qui vous semblent intéressantes à mentionner ? Cliquez sur les domaines qui vous intéressent pour apprendre leur fonction. Que pouvez-vous déduire de ces données ?

Oui, on peut identifier 21 séquence chez les mouches/moustiques/mites dans lesquelles le domaine Pcc1 est fusionné avec un domaine PIG-P impliqué dans la synthèse de GPI (ancrage des protéines à la membrane)

Chez le riz (Oryza) on trouve 8 séquences dans lesquelles le domaine Pcc1 est fusionné avec un domaine impliqué dans la fusion de membrane.

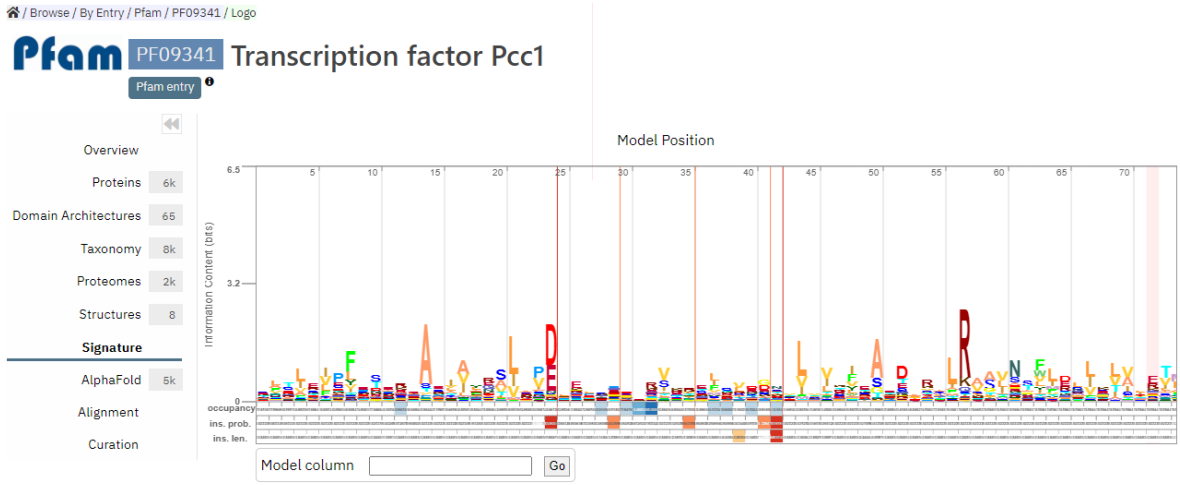
Chez les champignons on trouve 11 séquences dans lesquelles Pcc1 est fusionné avec une lyase (Une lyase est une enzyme capable de casser des liaisons covalentes (par des moyens autre que l'hydrolyse ou oxydation), créant souvent de nouvelles doubles liaisons).

Chez les euryarchées une fusion avec une Asparagine tRNA synthétase peut être trouvé (20-taine de séquences).

Il existe une multitude d'autres séquences associées à d'autres domaines fonctionnels.

L'ensemble des données montre que Pcc1 est le plus souvent présent en tant que protéine isolée. Cependant chez certains eucaryotes il peut faire partie d'une plus grande protéine contenant également un domaine impliqué dans des fonctions variées dont le métabolisme membranaire ou une lyase. Il serait par exemple intéressant de tester si Pcc1 est ancré dans la membrane plasmique chez les Eucaryotes. Chez les archées, on trouve Pcc1 fusionné avec une Asn-tRNA synthétase ce qui implique potentiellement une fonction liée avec les ARNt et/ou la traduction.

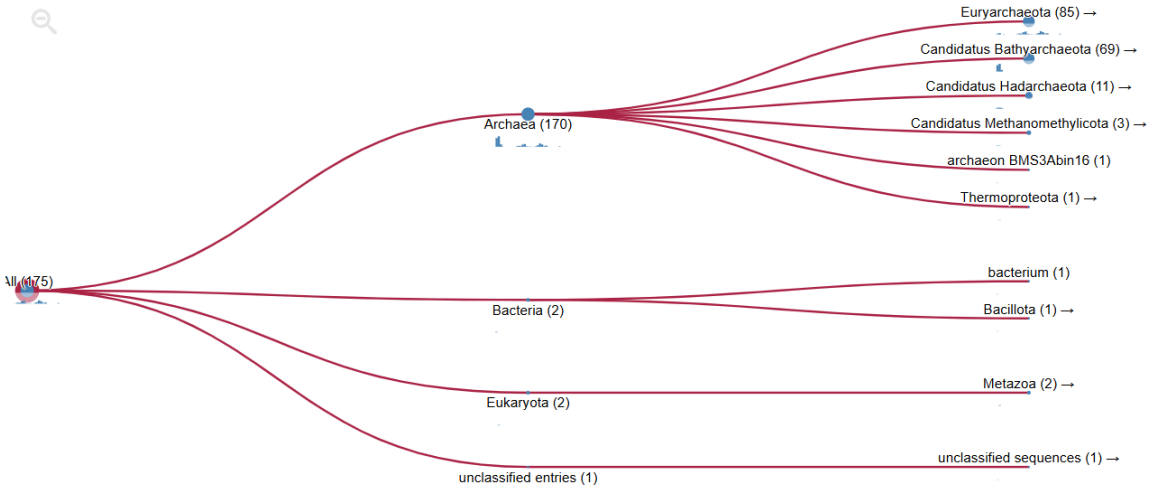
15. Regardez maintenant les informations contenues dans l'onglet « Profile HMM ». Copiez-collez l'image ci-dessous. A partir de l'alignement d'un grand nombre de séquences PFAM va construire un profil de séquence typique pour la famille Pcc1 en utilisant le HMM. Le graphique montre la fréquence d'acides aminés pour chaque position. Répertoriez ci-dessous les résidus conservés. Ces résidus sont potentiellement importants pour la fonction du Pcc1.



On peut repérer : Ala à la position 14, un résidu hydrophobe (Ile, Leu, Val) à la position 21, un résidu acide (Asp ou Glu) à la position 24, un résidu hydrophobe (Ile, Leu, Val) à la position 44, une Ala/Gly/Ser/Thr à la position 49, une Arg/Lys à la position 57.

16. Vous allez maintenant utiliser le serveur HMMER (https://www.ebi.ac.uk/Tools/hmmer/) pour trouver les orthologues de la protéine TK1253 et étudier leur distribution au sein du vivant. Ce serveur utilise HMM (hidden markov model) pour trouver les orthologues éloignés. Si vous souhaitez connaitre le principe de fonctionnement de ce site vous pouvez consulter la page Wikipedia dédiée (https://en.wikipedia.org/wiki/HMMER). Entrez la séquence protéique de TK1253 (Pcc1) dans la fenêtre prévue à cet effet et cochez la base de séquences « UniProtKB ». Combien de résultats avez-vous obtenu ? Comment sont distribuées ces séquences au sein du vivant ? Pour répondre à cette question utilisez l'onglet « Taxonomy » en haut de la page. Copiez-collez l'image montrant la distribution de la TK1253 au sein du vivant.

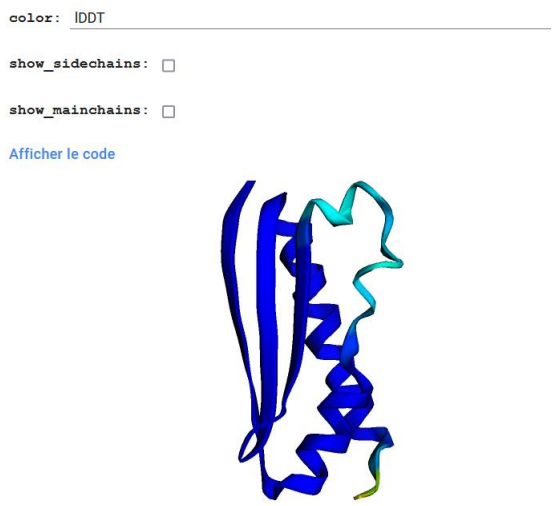
On obtient 183 résultats significatifs.



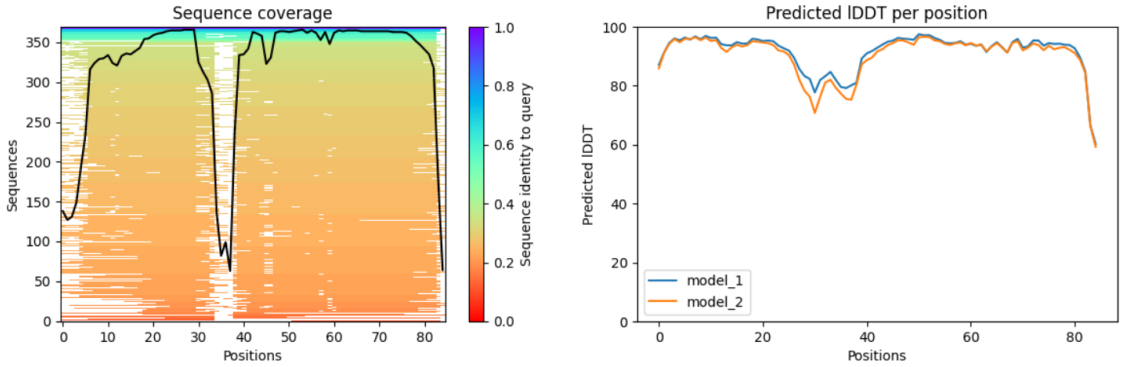
On remarque que selon cette analyse le TK1253 est présent en grande majorité chez les Archées. La majorité des séquences est issue d'Euryarchaeota le superphylum auquel appartiennent les Thermococcales. On retrouve également cette protéine chez les groupes appartenant au superphylum TACK (Bathyarchaeota). Cela montre que cette protéine est distribuée au minima dans deux des quatre superphyla d'Archées. Par ailleurs, le HMMER peut effectuer des recherches itératives : à partir de la première série de séquences trouvées le programme va établir un profile HMM qu'il utilisera ensuite pour trouver la seconde série de séquences etc.

17. Vous avez sans doute entendu parler du logiciel AlphaFold2. Il s'agit d'un algorithme de « machine learning » (intelligence artificielle) qui utilise les réseaux neuronaux pour prédire grâce à un échantillon de structures résolus la structure tridimensionnelle des protéines uniquement à partir d'une séquence protéique. Dernièrement, une nouvelle version de ce programme a été publié (DOI: 10.1038/s41586-021-03819-2) avec des résultats de qualité de prédiction époustouflants à tel point qu'on parle d'une révolution dans le domaine de la biologie structurale ! Si vous possédez un compte gmail vous pouvez utiliser l'interface Colab qui permet d'utiliser facilement AlphaFold2 : **https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=kOblAo-xetgx**

Reportez-vous au TUTORIELS pour quelques consignes d'utilisation et pour savoir comment interpréter les résultats. Faites une capture d'écran de principaux résultats (le modèle 3D, le graphique montrant l'alignement des séquences et le graphique montrant la distribution de l'indicateur IDDT au sein de la séquence). Télécharger le fichier .zip contenant tous les résultats. Vous allez en avoir besoin pour l'exercice suivant.



Structure prédite : couleur bleu indique que cette prédiction est globalement très fiable mis à part une boucle qui est affichée en vert. Cette boucle correspond aux résidus 30-40 AA qui sont sous-représentés dans l'alignement des séquences (voir le graphe ci-dessous)



A gauche :

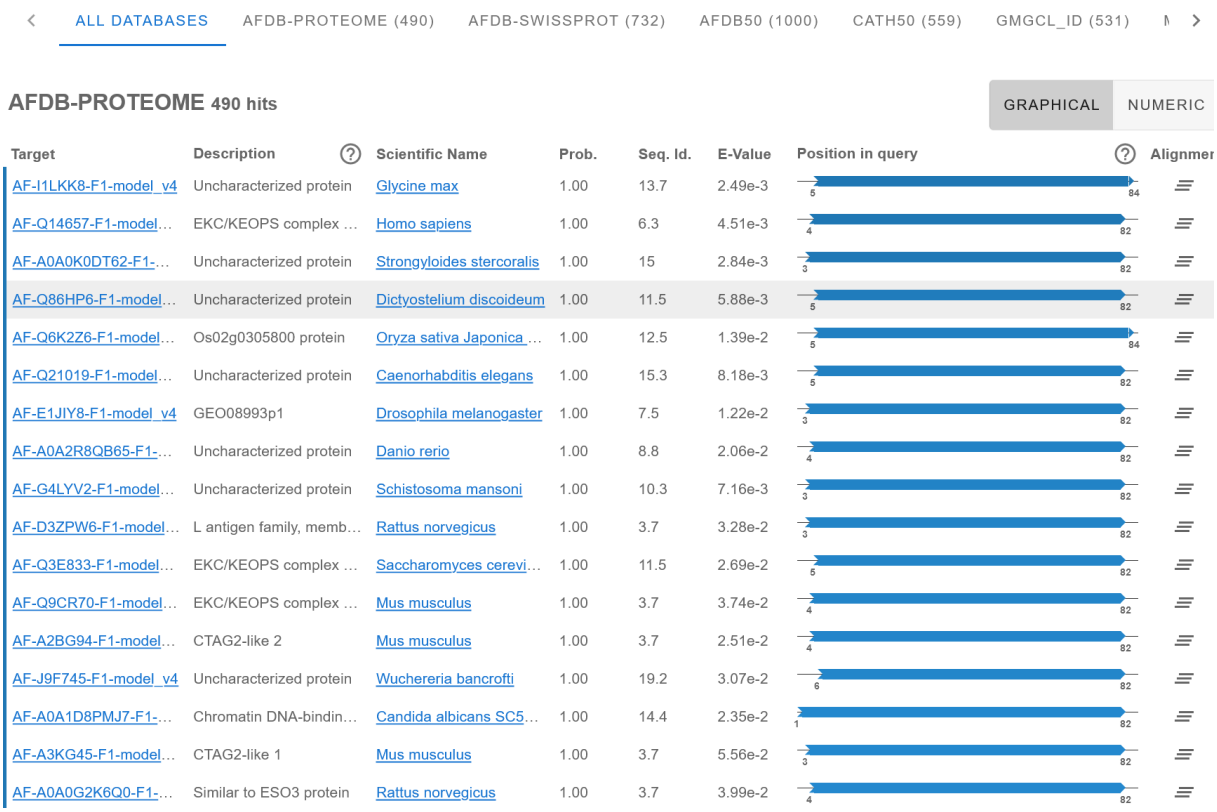
Cette image montre qu'environ 350 séquences ont été alignées pour calculer le modèle de la structure. On note que les résidus 30-40 sont présents chez une minorité des séquences y compris la séquence qui a été analysée. On note également la faible conservation de séquence en particulier en N-ter et C-ter.

A droite :

L'indicateur lDDT (Local Distance Difference Test) montre le niveau de confiance (allant de 0 à 100) de prédiction. On voit que ce niveau est globalement très élevé sauf pour la région 30-40 AA et les deux extrémités de la séquence.

18. Récupérez le fichier pdb qui correspond au modèle le plus fiable « unrelaxed_rank_1_model». Vous allez utiliser ce fichier pour effectuer une recherche avec l'outil Foldseek (https://search.foldseek.com/search). Ce serveur vous permettra de chercher parmi des centaines de millions de structures prédites par AlphaFold2 et parmi les structures déposées dans la PDB (donc des structures réelles déterminées expérimentalement. Faites une capture d'écran avec la liste des meilleurs résultats trouvées dans toutes les bases de données (« ALL DATABASES »). Quelles

informations vous donne cette liste ? Affichez la superposition de structure pour le premier résultat (en cliquant sur l'icône « Alignment »), et commentez-la notamment en terme de sa robustesse.

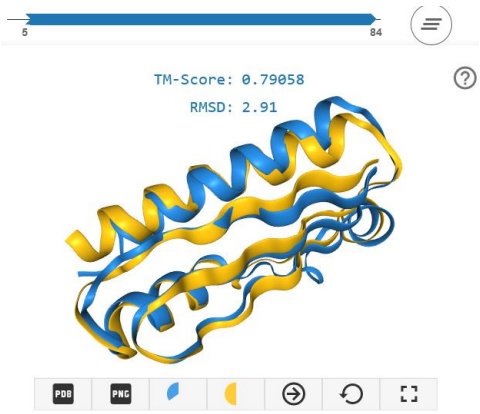


ALL DATABASES | AFDB-PROTEOME (490) | AFDB-SWISSPROT (732) | AFDB50 (1000) | CATH50 (559) | GMGCL_ID (531)

AFDB-PROTEOME 490 hits — GRAPHICAL | NUMERIC

Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment
AF-I1LKK8-F1-model_v4	Uncharacterized protein	Glycine max	1.00	13.7	2.49e-3	5–84	
AF-Q14657-F1-model…	EKC/KEOPS complex …	Homo sapiens	1.00	6.3	4.51e-3	4–82	
AF-A0A0K0DT62-F1-…	Uncharacterized protein	Strongyloides stercoralis	1.00	15	2.84e-3	3–82	
AF-Q86HP6-F1-model…	Uncharacterized protein	Dictyostelium discoideum	1.00	11.5	5.88e-3	5–82	
AF-Q6K2Z6-F1-model…	Os02g0305800 protein	Oryza sativa Japonica …	1.00	12.5	1.39e-2	5–84	
AF-Q21019-F1-model…	Uncharacterized protein	Caenorhabditis elegans	1.00	15.3	8.18e-3	5–82	
AF-E1JIY8-F1-model_v4	GEO08993p1	Drosophila melanogaster	1.00	7.5	1.22e-2	3–82	
AF-A0A2R8QB65-F1-…	Uncharacterized protein	Danio rerio	1.00	8.8	2.06e-2	4–82	
AF-G4LYV2-F1-model…	Uncharacterized protein	Schistosoma mansoni	1.00	10.3	7.16e-3	3–82	
AF-D3ZPW6-F1-model…	L antigen family, memb…	Rattus norvegicus	1.00	3.7	3.28e-2	3–82	
AF-Q3E833-F1-model…	EKC/KEOPS complex …	Saccharomyces cerevi…	1.00	11.5	2.69e-2	5–82	
AF-Q9CR70-F1-model…	EKC/KEOPS complex …	Mus musculus	1.00	3.7	3.74e-2	4–82	
AF-A2BG94-F1-model…	CTAG2-like 2	Mus musculus	1.00	3.7	2.51e-2	4–82	
AF-J9F745-F1-model_v4	Uncharacterized protein	Wuchereria bancrofti	1.00	19.2	3.07e-2	6–82	
AF-A0A1D8PMJ7-F1-…	Chromatin DNA-bindin…	Candida albicans SC5…	1.00	14.4	2.35e-2	1–82	
AF-A3KG45-F1-model…	CTAG2-like 1	Mus musculus	1.00	3.7	5.56e-2	3–82	
AF-A0A0G2K6Q0-F1-…	Similar to ESO3 protein	Rattus norvegicus	1.00	3.7	3.99e-2	4–82	

On trouve majoritairement des protéines avec une fonction inconnue mais aussi des résultats dont la description est EKC/KEOPS complex subunit PCC1. Ces résultats correspondent effectivement aux orthologues de notre protéine Pcc1 qui fait partie d'un complexe nommé KEOPS et impliqué dans la synthèse d'une modification universelle d'ARN de transfert, nommé t^6A. Nous pouvons noter par ailleurs que ce résultat pertinent est obtenu malgré des valeurs d'identité de séquences très faibles et la valeur E d'une robustesse modérée ce qui montre l'intérêt de l'alignement structurel en plus de l'alignement de séquence. Par ailleurs, vous remarquerez que la probabilité que l'algorithme donne à la justesse des résultats est maximale (Colonne Prob. = 1).



La superposition de deux structures montre une bonne correspondance et l'alignement comprend la quasi-totalité de notre structure, la valeur de TM-score est >0.5 et RMSD est assez proche de 2 Angström. Il s'agit donc d'un résultat fiable.

19. Affichez maintenant la liste qui correspond aux résultats trouvés dans la PDB (onglet « PDB100 »). Existe-t-il des structures de Pcc1 déterminées expérimentalement ?

PDB100 486 hits — GRAPHICAL | NUMERIC

Target	Description	Scientific Name	Prob.	Seq. Id.	E-Value	Position in query	Alignment
7a66_C	structure of Pcc2 from …	unclassified	1.00	52.4	1.17e-8	1 82	
7a66_A	structure of Pcc2 from …	unclassified	1.00	53.1	7.00e-8	3 81	
7a67_B	Pcc1Pcc2 complex	unclassified	1.00	52.5	2.30e-7	3 82	
7a67_A	Pcc1Pcc2 complex	Pyrococcus abyssi GE5	1.00	12.1	5.61e-4	4 85	
5jmv_F	Crystal structure of mj…	Pyrococcus furiosus D…	1.00	15	1.42e-3	5 83	
5jmv_D	Crystal structure of mj…	Pyrococcus furiosus D…	1.00	20.4	8.34e-4	3 85	
3eno_D	Crystal structure of Pyr…	Pyrococcus furiosus D…	1.00	15.1	3.13e-3	5 83	
3eno_F	Crystal structure of Pyr…	Pyrococcus furiosus D…	1.00	18	1.09e-3	2 84	
5jmv_G	Crystal structure of mj…	Pyrococcus furiosus D…	1.00	15	3.81e-3	5 84	

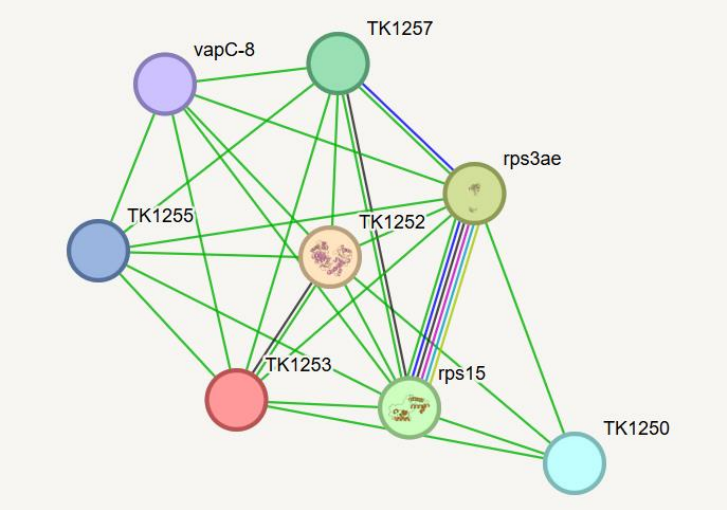
Oui, on voit qu'il y a plusieurs structures de Pcc1 provenant de l'archée Pyrococcus abyssi mais aussi qu'il existe un paralogue structuralement très proche et qu'on a nommé Pcc2.

Pour terminer nous allons interroger la base de données STRING (https://string-db.org/) qui répertorie les réseaux d'interaction physique et/ou fonctionnelle pour de nombreuses protéines. Pour créer les réseaux protéiques STRING s'appuie sur les prédictions de contextes génomiques, sur les données expérimentales (comme par exemple les expériences de co-purification à grande échelle ou les données de transcriptome) et sur les recherches dans la littérature scientifique.

20. Question préliminaire : à quoi correspond la notion de contexte génomique ou de synthénie ? Que peut suggérer la conservation d'une synthénie particulière chez de nombreuses espèces et comment cela peut être intéressant pour l'annotation des gènes hypothétiques ?

Il s'agit de la conservation de l'ordre des gènes sur le génome de plusieurs espèces. Un tel phénomène peut suggérer que ces gènes sont fonctionnellement reliés (« guilty by association »). Cela peut s'avérer utile pour identifier la fonction des gènes hypothétiques.

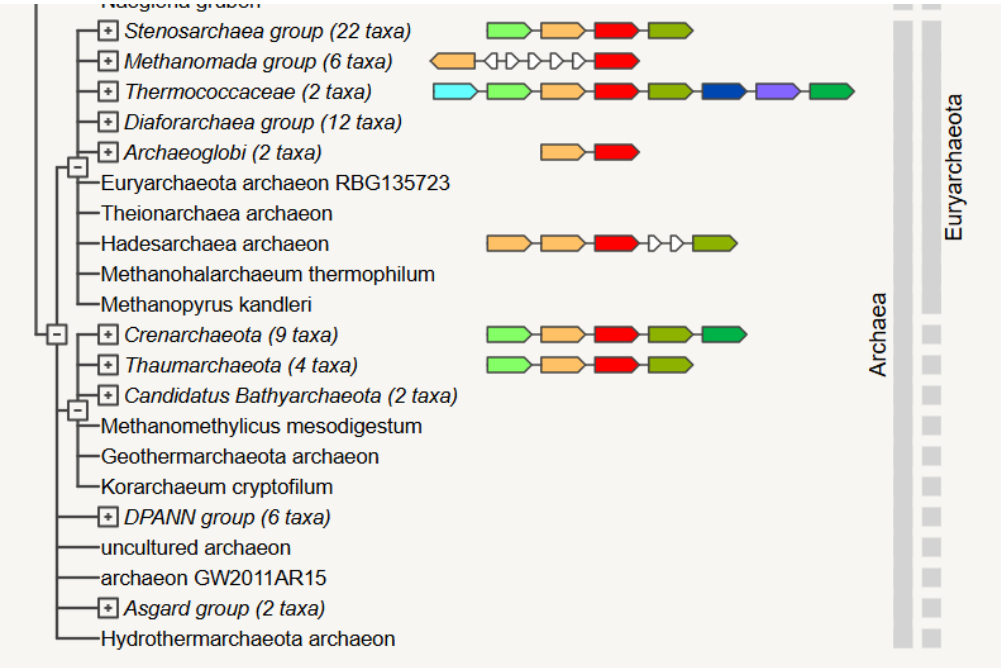
21. Sur la page d'accueil de STRING (https://string-db.org) cliquez sur « search » puis à gauche sur « protein by sequence » puis faites la recherche avec la séquence protéique de TK1253. Copiez-collez ci-dessous le résultat graphique montrant le réseau.



22. La couleur des lignes reliant les différentes protéines a une signification. Le code couleur est indiqué sous l'image. Quelle est la nature des interactions entre le TK1253 et les autres protéines selon ce code couleur ?

Il s'agit des prédictions d'interaction basées essentiellement sur la conservation des contextes génomiques et l'interrogation des articles scientifiques.

23. Cliquez sur « Viewers » pour afficher les différentes catégories d'informations, copiez-collez ci-dessous la catégorie qui vous semble la plus informative. Justifiez votre choix et commentez les données.



J'ai choisi de montrer la synthénie (gene neighborhood) car le réseau est principalement basé sur ces données. On voit que la synthénie est largement conservée chez les Archées et qu'il ne s'agit pas d'une spécificité des Thermococcales.

24. En conclusion, qu'indique ce réseau ? Dans quel processus moléculaire pourrait participer TK1253 ? Comment tester cette hypothèse ?

Dans le réseau on trouve deux protéines ribosomales et quatre protéines qui interagissent avec l'ARN. On peut donc proposer que TK1253 (Pcc1) participe à un processus impliquant l'ARN et le ribosome. C'est effectivement le cas. Le complexe KEOPS dont Pcc1 fait partie est impliqué dans la synthèse d'une modification d'ARNt, nommé t6A, essentielle pour la fidélité de la traduction (Perrochia et al., 2013, doi: 10.1093/nar/gks1287). On pourrait tester cette hypothèse au laboratoire en générant un mutant de délétion (si Pcc1 n'est pas essentiel) chez *T. kodakarensis* puis en mesurant le taux de la modification t6A au sein des cellules ainsi que la fidélité de la traduction grâce à un système rapporteur ou (encore mieux) en utilisant la technique de « ribosome profiling ».