

## Annales partie problème

### Problème :

- 1) Vous avez séquencé un fragment d'ADN. L'analyse informatique de cette séquence révèle la présence d'un gène et vous permet d'obtenir sa séquence protéique. Expliquer la stratégie que vous adopteriez pour avoir une idée de la fonction de la protéine codée par ce gène. Justifier chacune des étapes.

Le résultat de la démarche que vous avez suivie au 1) indique que votre protéine présente des similitudes avec les ATP-binding cassettes des transporteurs ABC. Un transporteur ABC permet le transport actif à travers une membrane. Ces transporteurs sont trouvés aussi bien chez les eucaryotes que chez les procaryotes. Un transporteur ABC typique est constitué de deux domaines transmembranaires et de deux domaines ATP-binding. Les domaines ATP-binding lient l'ATP et couplent l'hydrolyse de l'ATP au processus de transport.

Les différents domaines d'un transporteur ABC peuvent être exprimés :

- comme des peptides séparés,
- comme un peptide multifonctionnel dans lequel les domaines sont fusionnés.

Les analyses expérimentales réalisées ont montré que différents types de fusion étaient rencontrés *in vivo*:

- deux domaines ATP-binding fusionnés,
- un domaine transmembranaire et un domaine ATP-binding fusionnés,
- deux domaines transmembranaires fusionnés,
- les quatre domaines fusionnés.

- 2) Les ATP-binding cassettes des transporteurs ABC se caractérisent par la présence de trois motifs. Deux motifs, le motif Walker A et le motif Walker B, sont présents dans une majorité d'ATPases. Le troisième motif est caractéristique des ATP-binding cassettes des transporteurs ABC et est appelé signature ABC. En Annexe I, l'alignement multiple sur ces trois motifs d'un ensemble d'ATP-binding protéines connues vous est donné. Déterminer à partir de cet alignement une signature de type PROSITE pour la signature ABC et le motif Walker A. Pensez-vous que le motif Walker B permette de bien caractériser cette famille ?
- 3) A la suite de la recherche que vous avez faite en 3), vous détectez la présence de 11 ATP-binding cassettes dans un génome appelé Bsub. Vous voulez approfondir l'étude de ces séquences. Pour cela vous recherchez leurs structures en domaines à l'aide de PRODOM (Annexe II). PRODOM est une banque de domaines protéiques comme Pfam. De plus, l'utilisation d'un programme de recherche de fragments transmembranaires vous révèle la présence d'un domaine transmembranaire dans les régions N-terminales des séquences YGAD, YWJA et YVCC. L'analyse des résultats de PRODOM vous a permis de déterminer les bornes des domaines ATP-binding dans vos protéines. Vous avez ensuite réalisé un alignement multiple uniquement sur ces domaines (Annexe III).  
A l'aide de ces résultats et des différents types de fusion de domaines décrits ci-dessus, combien de classes différentes d'organisation pouvez-vous dégager ? Donnez les caractéristiques de ces classes et le nom des séquences y appartenant.
- 4) Replacez sur l'alignement de l'Annexe III (directement sur le sujet) les trois motifs caractéristiques des ATP-binding cassettes.

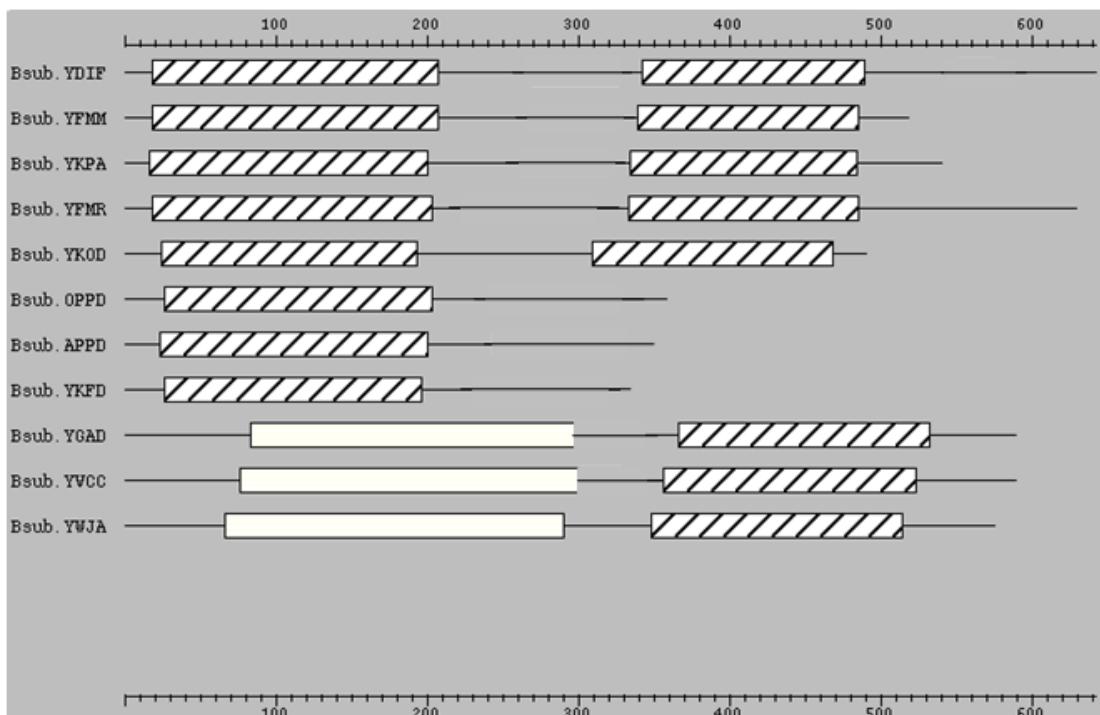
## ANNEXE I

	signature	
Walker A	ABC	Walker B
dppD	GESGCGKSV	... FSGGMR QRIVIAMALI CEPDILIADE PTTALD
yurJ	GPGCGKST	... LSGGQR QRVALGRSIV REPKVFLMDE PLSNLD
hisP	GPGSGKTT	... LSGGQK QRVGIARALA IHPDVLLFDE PTAALD
yqgK	GPGCGKST	... LSGGQQ QRLCIARALA TNPDILLMDE PTSALD
yxdl	GPGSGKTT	... VSGGQR QRAAIARAVI HKPSLILADE PTGNLD
yknY	GPGSGKST	... LSGGQK QRVAIARAIV NEPKLILADE PTGALD
fhuC	GPNCGKST	... LSGGQR QRVWIAMALA QGTELLLDE PTTYLD
ydbJ	GPNCGKTT	... YSLGMR QRLGIAQAIL HRPKLLILDE PTNGLD
yfiL	GPNAGAKST	... FSGGMK RRINIGAALM HKPELLIMDE PTVGID
yfiC	GPTGAGKTT	... ISQGQK QLISIARAVL ADPVLLLILDE ATSNID
yheH	GHTGSGKSS	... LSSGER QLISFARALA FDPAILILDE ATAHID
ftsE	GPGAGKST	... LSGGEQ QRVSIARSIV NNPDVVIADE PTGNLD
natA	GENGAGKTT	... FSKGMR QKVAIARALI HDPEIILFDE PTTGLD
ybxA	GHNGSGKST	... LSGGQK QRVAIAGVIA ARPDIIILDE ATSMLD
ycdI	GPNGASKST	... LSGGQK QKICIARMLA SNPDLMLMDE PTFAVD
ybdA	GPNAGAKTT	... LSRGQK QRVVLAGAIV QDALLYILDE PTVGLD
cydC	GASGAGVST	... LSGGQA QRVALARAFL GNRPILLLDE PTAHLD
ytrB	GRNGSGKTT	... LSTGLK KQLSLVLSFA ARPALILLDE PTDGID
ylmA	GLNGAGKTA	... LSQGEK QRALIARALM ADPELLILDE PITGLD

L'alignement en N-terminal précédant le motif Walker A, l'alignement en C-terminal suivant le motif Walker B et l'alignement entre le motif Walker A et la signature ABC ne sont pas représentés. Les ... remplacent N résidus alignés entre les 2 motifs.

## ANNEXE II

### Résultats de PRODOM



**Nom et prénom :**

**ANNEXE III**  
**Alignment multiple des domaines ATP-binding**

		10	20	30	40	50	60	70	80	90	
Bsub . APPD	23	200	AVDGVDHFISKGETVALVGE SGSKSITSLSIMGLVQSSGGKIMD-----GSIKLEDKDLTSFTENDYCKIRGNEVSMIFQEPMTSLN P-----VLT I								
Bsub . OPPD	26	203	AIRGVNFHLDKGETLAIVGE SGSKSVTSQAIMKLIPMPGGYFKR-----GEILFEGKDVLVPLSEKEMQNVRGKEIGMI FQDPMTSLNP-----TMKV								
Bsub . YKFD	26	196	AVDGVTQIREGETFGLVGE SGCGKSTLGRVLMLR YQPT-----E-----GSVTYRGTNLHALSEKEQFAFN-RKLQMFQDPYASLN P-----RMTV								
Bsub . YKOD	24	193	VFQDISFELQKGE CVLLLGP SGCGKSSLALCLNGLYPEACDGIQS-----GHVFLFQKPVTD AETSETIT---OHAGVVFQDPDQQF-C-----MLTV								
Bsub . YKOD	309	468	IFKD ISFSLREGSLTA L VGPNGTGKSTLLSVLASLMKP-----QS-----GKILLYDQPLQKYKEKE-LR---KRMGFVFQNP EHQF-V-----TDTV								
Bsub . YGAD	366	532	I LHDVSLKVNRGETVALVGMSGGKSTLVS LIPRFYDVT-----S-----GRLLIDGTDIRDYEARS LR---NQVMVLQDTFL-FSE-----TIRE								
Bsub . YWJA	348	514	VLDINL SIQAGETVA FVGPGSAGKSTLCSLLP RFYEA S-----E-----GDITIDGISIKDMTLSSLR---GQIGVVQ QDVFL-FSG-----TLRE								
Bsub . YVCC	356	523	I LKEVSAVIEAGKVT AIVGP SGGGKTTLFKLLERFY SPT-----A-----GTIRLGDEPVDTYSLESWR---EHIGYVSQESPL-MSG-----TIRE								
Bsub . YDIF	18	207	I LNNIKLEVNRDR DIAIVGRNGAGKSTLLKIIAGQLSYEKGEI IKPKDITMGYLAQHTGLDSKLTIKEELLTVFDHLKAMEKEMRAMEEKMAA ADP GEL								
Bsub . YFMM	18	207	I FNNSVFSRLLKGEHVG LIGANGEGKSTFMNIITGKLEPDEGKVEWSKNVRVG YLDQHTVLEKGKSIRDVLKDAHFYL FAMEEE MNEIYNKMGEADPDEL								
Bsub . YFMR	18	203	LFDHISFHIEENERIGLIGPNGTGKSTLLKVIAGLESIEEGEITKSGSVQVFELHQDPELPAGQTVLEHIYSGESAVMKT LREYEKALYELGK-DP-EN								
Bsub . YKPA	16	200	L FEDVNIFKFTPGNCYGLIGANGAGKSTFLKVL SGEIEPQTGDVHMSPGERLAVLKQNHFEYEEYEVLKVVIMGHKRLYEV MQE KDAI YM K-----PDFS								
Bsub . YDIF	342	489	L LTEVSKFMLTRGE SAALVGPN GIGKSTLLKTLID TLKPDQ-----GTISYGSNVSVGYYDQE Q-----AELT								
Bsub . YFMM	339	485	L SRLP LNRLMERGQKIALY GANGIGKTTLLKSL GEIQPLE-----GSVERGEHIYTGYFEQEV-----KE-T								
Bsub . YKPA	334	484	VLDNVSFIMNREDKIAFTGRNELAVTTLFKIISGEMEADS-----GTFKWGVTTSQAYFPKDN-----SEYF								
Bsub . YFMR	333	485	L VDRF NELVIPGERIGIIGPNGIGKTTLLNALAGRHTPDG-----GDI TIGQTVRIGYYTQDH-----SEM N								
		100	110	120	130	140	150	160	170	180	190
Bsub . APPD	23	200	GEQITEVLIYHKNMKKKEARQRAV-ELLQMVGFSRAEQIMKEYPHRLSGGMQRQVMIAIALSCNPKLLIADEPTTALDVTIQAQVLEILMKD								
Bsub . OPPD	26	203	GKQITEVLFKHEKISKEAKKRAV-ELLELVGIPMPEKRVNQFPHEFSGGMRQRVVIA M ALAANPKLLIADEPTTALDVTIQAQVLEILMKD								
Bsub . YKFD	26	196	REIILEPMEIHNLYNTHKARLSV D VDELLAEVGLH--PDFGSRYPHEFSGGQRQ RIGIARALS LNP EFIVADEPISAL DVSVQAQVVNLK R								
Bsub . YKOD	24	193	EDEIAFGLENLQI PKEEMTEK--INA VLGKLRIT--HLKEKMISTLSGGQKQKV VALACI LAMEPEL I DEPTSLLDPFSAREFVHLMKD								
Bsub . YKOD	309	468	Y DELLFG---QKANAETEKK--AQHLLQRFGLA--HLADHHPFAISQGQKRR LS VATMLMDVKVLLLDEPTFGQDARTAAECM EMI QR								
Bsub . YGAD	366	532	NIAIG-KPDATLEEII EAAKAANAHEFIMS PEGYETRVGERGV-KLSGGQKQRISIARVFLKNPPLL ILDEATSALD LESEHYI QEAMDK								
Bsub . YWJA	348	514	NIAYG-RLGASEE DIWQAVQAHLEELVHNMPDGLDTMIGERGV-KLSGGQKQR LSIARMFLKNPSI L DEATSALD TETEAAI QKALQE								
Bsub . YVCC	356	523	NICYGLERDVTDAEIEKAAEMAYALNFIKE LPNQFDTEVGERGI-MLSGGQRQRIA IAR ALLRN PNSI MLDEATSSLDSQSEKSVQ QALEV								
Bsub . YDIF	18	207	E SIMKTYDRLQQEFKD KGGYQYEAD VRSV LHGLGF SHFDD STQVQSLSGGQKTRI ALGK LLLT QPD L LDEPT NHLD I DLT WLEH YLQ G								
Bsub . YFMM	18	207	E KLLEEVGV I QDALT NND FV IDSK VEEIARGL GLSDIGLERDVT DLSGGQKRT KVLLA KLL LEK PEI LL DEPT NYL D EQHIEWL KRYL QE								
Bsub . YFMR	18	203	EQRQKHLLAAQAKMDANNAWDANTLAKTVLSKLG VNDV--TKPVN ELSGGQK KRVIAK NLIQPADL L DEPT NHLD NETI EWLEG YLSQ								
Bsub . YKPA	16	200	DEDGIRAAELEGEFAELNGWEAESEA II LKGLG ISED LHTK KMA DLGG SEK V KVLLA QAL FGKPD VLL DEPT NHLD LQAI QWLE EFLIN								
Bsub . YDIF	342	489	SSKRV LDELW-DEY PG LPEKE--IRT--CLGNFLFSGDDV LKP VHSLS GG EKARL A LAKLMLQKANFL I DEPT NHLD L DSKEV L ENAL ID								
Bsub . YFMM	339	485	NNNT CIEEVW-SEF PSY TQYE--IRA--APAKCGLTTKHIESRV SVLS GG EKAKV RLCK LINSET TNLL VL DEPT NHLD ADAKE ELK RAL KE								
Bsub . YKPA	334	484	EGSDLNLVDWLRQYSPHDQSESFLRG--FLGRMLFSGEEVHKKANVLS GG EKVR CMSKAM LSGA NI L DEPT NHLD LES ITA LN NGLIS								
Bsub . YFMR	333	485	GELKVIDYIKETAEVVKTAGDMITA EQMLERFLPRSMQ QT YIRKLSGG EKRR LYLLQVLMQEPNVLFLDEPT ND LDT ETLSVLEDYIDQ								

A la suite du nom de la protéine, les deux nombres indiquent les positions de début et de fin des domaines ATP-binding trouvés dans les protéines.