

(12) DEMANDE INTERNATIONALE PUBLIÉE EN VERTU DU TRAITÉ DE COOPÉRATION
EN MATIÈRE DE BREVETS (PCT)

(19) Organisation Mondiale de la Propriété
Intellectuelle
Bureau international



(43) Date de la publication internationale
13 février 2003 (13.02.2003)

PCT

(10) Numéro de publication internationale
WO 03/012108 A2

(51) Classification internationale des brevets⁷ :
C12N 15/54, 15/11, 9/10, C12Q 1/48, 1/68, C12P 19/38

LIPOWSKI, Gérard [PL/FR]; 15 Résidence Le Bosquet,
Chambre A22, F-91940 Les Ulis (FR). **MARTIN,**
Jean-Louis [FR/FR]; 11 avenue de la Hacquinière,
F-91440 Bures / Yvette (FR).

(21) Numéro de la demande internationale :
PCT/FR02/02795

(74) Mandataires : **CATHERINE, Alain** etc.; Cabinet Harle
et Phelip, 7 rue de Madrid, F-75008 Paris (FR).

(22) Date de dépôt international : 2 août 2002 (02.08.2002)

(25) Langue de dépôt : français

(81) États désignés (*national*) : AE, AG, AL, AM, AT, AU, AZ,
BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ,
DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM,
HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK,
LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX,
MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI,
SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN,
YU, ZA, ZM, ZW.

(26) Langue de publication : français

(30) Données relatives à la priorité :
0110401 2 août 2001 (02.08.2001) FR
0205585 3 mai 2002 (03.05.2002) FR

(71) Déposants (*pour tous les États désignés sauf US*) :
INSTITUT NATIONAL DE LA SANTE ET DE LA
RECHERCHE MEDICALE (INSERM) [FR/FR]; 101
rue de Tolbiac, F-75654 Paris Cedex 13 (FR). **CENTRE**
NATIONAL DE LA RECHERCHE SCIENTIFIQUE
(CNRS) [FR/FR]; 3 rue Michel Ange, F-75794 Paris
cedex 16 (FR). **L'UNIVERSITE PARIS-SUD PARIS XI**
[FR/FR]; 15, rue Georges Clemenceau, F-91405 Orsay
Cedex (FR).

(84) États désignés (*régional*) : brevet ARIPO (GH, GM, KE,
LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), brevet
eurasien (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), brevet
européen (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), brevet
OAPI (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML,
MR, NE, SN, TD, TG).

(72) Inventeurs; et

(75) Inventeurs/Déposants (*pour US seulement*) : **MYL-**
LYKALLIO, Hannu [FI/FR]; 51 Résidence La Coc-
quetière, F-91470 Les Molières (FR). **LIEBL, Ursula**
[DE/FR]; 51 Résidence La Cocquetière, F-91470 Les
Molières (FR). **LEDUC, Damien** [FR/FR]; 11 rue de
la Ferme de Belleville, F-91190 Gif Sur Yvette (FR).

Publiée :

— sans rapport de recherche internationale, sera republiée
dès réception de ce rapport

En ce qui concerne les codes à deux lettres et autres abrégia-
tions, se référer aux "Notes explicatives relatives aux codes et
abréviations" figurant au début de chaque numéro ordinaire de
la Gazette du PCT.

(54) Title: USES OF A THYX POLYPEPTIDE OR A NUCLEIC ACID ENCODING SUCH A POLYPEPTIDE, IN PARTICULAR
FOR SCREENING ANTI-BACTERIAL OR ANTI-VIRAL COMPOUNDS

(54) Titre : UTILISATIONS D'UN POLYPEPTIDE THYX D'UN ACIDE NUCLEIQUE CODANT UN TEL POLYPEPTIDE,
NOTAMMENT POUR LE CRIBLAGE DE COMPOSES ANTI-BACTERIENS OU ANTI-VIRAUX.

(57) Abstract: The invention concerns various uses of a novel family of enzymes, called THYX, capable of catalyzing the synthe-
sis of thymidine 5'-monophosphate in the absence of an active thymidylate synthase enzyme (ThyA). The invention also concerns
reaction media for the thymidylate synthase activity of a THYX polypeptide and screening methods using said reaction media, as
well as kits for implementing said methods.

(57) Abrégé : L'invention est relative aux diverses utilisations d'une nouvelle famille d'enzymes, désignée THYX, capables de ca-
talyser la synthèse de la thymidine 5'-monophosphate en l'absence d'une enzyme thymidylate synthase (ThyA) active. L'invention
concerne aussi des milieux de réaction pour l'activité thymidylate synthase d'un polypeptide THYX ainsi que des procédés de cri-
blage mettant en uvre lesdits milieux réactionnels, ainsi que des kits pour la mise en uvre de ces procédés.



WO 03/012108 A2

Utilisations d'un polypeptide THYX ou d'un acide nucléique codant un tel polypeptide, notamment pour le criblage de composés anti-bactériens ou anti-viraux

5 **DOMAINE DE L'INVENTION**

La présente invention se rapporte au domaine des enzymes participant à la synthèse de l'ADN, et plus spécifiquement à la synthèse d'un composé intermédiaire, la thymidine 5'-monophosphate (dTMP), requis pour la production de la thymidine 5'-triphosphate (dTTP)
10 constitutive de la molécule d'ADN.

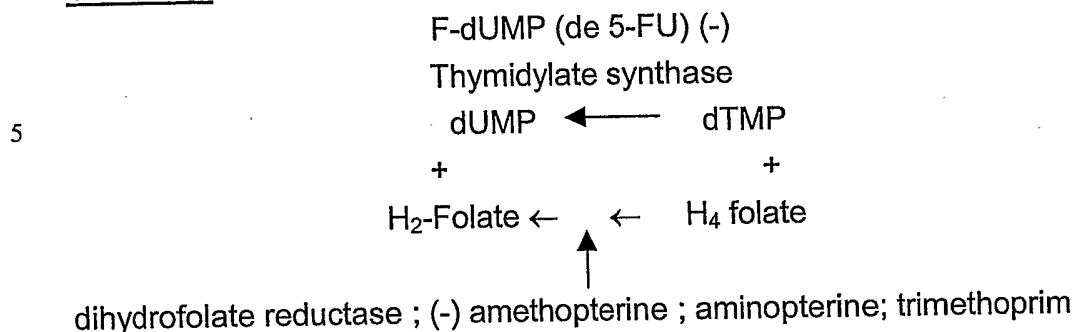
Elle est également relative aux diverses utilisations d'une nouvelle famille d'enzymes, désignée THYX, capables de catalyser la synthèse de la thymidine 5'-monophosphate en l'absence d'une enzyme thymidylate synthase (ThyA) active.

15 L'invention concerne aussi des milieux de réaction pour l'activité thymidylate synthase d'un polypeptide THYX ainsi que des procédés de criblage mettant en œuvre lesdits milieux réactionnels, ainsi que des kits pour la mise en œuvre de ces procédés.

20 **ART ANTERIEUR**

Le déoxythymidylate, au contraire des autres déoxynucléotides, n'est pas produit directement par une ribonucléotide réductase. Le thymidine 5'-monophosphate (dTMP) est décrit dans l'état de la technique comme étant le produit final de la méthylation de l'uridine 5'-
25 monophosphate (dUMP), méthylation qui est catalysée par une thymidylate synthase de type THYA, aussi bien chez les eucaryotes que chez les bactéries (Carreras et Santi, 1995). La thymidylate synthase THYA, qui est également exprimée chez les mammifères, a été envisagée comme cible potentielle de composés inhibiteurs de la
30 synthèse d'ADN, utilisables notamment dans le traitement du cancer colorectal (Papamichael, 2000).

Une illustration simplifiée de la production dTMP dans la cellule par méthylation du dUMP est représentée dans le schéma 1 ci-après.

Schéma 1

10 Dans le Schéma 1, on représente la synthèse de dTMP chez les bactéries et l'ensemble des eucaryotes par méthylation réductrice du dUMP sous l'action de la thymidylate synthase THYA.

15 Un autre composant, essentiel pour la synthèse de dTMP, est l'enzyme dihydrofolate réductase (DHFR), dont l'activité catalytique est nécessaire pour le recyclage du dihydrofolate via la formation d'un cofacteur actif, le CH₂H₄ folate, qui agit comme un donneur de méthyle. La thymidylate synthase peut être inactivée par un fluoro-dUMP et la DHFR peut être inactivée notamment par l'améthoptérine et l'aminoptérine et le triméthoprim, qui sont des dérivés du folate. (voir Schéma 1).

25 Les nombreux travaux de recherche réalisés sur la voie de synthèse des composés pyrimidine ont tous montré que la synthèse intracellulaire de dTMP ne pouvait être catalysée que par une seule enzyme, la thymidylate synthase codée par le gène *thyA* (voir Annual Reviews on Biochemistry, 1995, vol. 64:721-762; Nature Reviews Mol. Cell. Biol., 2001, vol.2, 147-151).

30 De plus, des travaux réalisés sur les inhibiteurs de la thymidylate synthase pour le traitement du cancer colorectal avancé précisent que « une caractéristique importante du métabolisme des nucléotides est la duplication des voies métaboliques ; l'inhibition de n'importe quelle enzyme peut être contournée par une ou plusieurs voies alternatives. Toutefois, une exception notable à cette règle est la thymidylate synthase *thyA*, qui est une enzyme incontournable

35 constituant le seul moyen d'ajouter un groupe méthyle en position 5 du

cycle pyrimidine dans la synthèse de *novo* de la thymidine (Papamichael, 1999).

SOMMAIRE DE L'INVENTION

5 De manière surprenante, il a été montré selon l'invention qu'une famille de polypeptides, désignés THYX, qui sont d'une structure complètement distincte des polypeptides codés par les gènes *thyA*, possèdent la capacité de synthétiser le dTMP dans les cellules.

10 En particulier, il a été montré selon l'invention que des cellules possédant un gène codant pour un polypeptide THYX possédait une activité thymidylate synthase, en l'absence de gène *thyA*. Il a aussi été montré que l'introduction d'une copie d'un gène *thyX*, codant pour un polypeptide THYX dans une cellule auxotrophe pour la thymidine permettait de restaurer, dans cette cellule, la capacité de synthétiser de
15 *novo* le dTMP et finalement le thymidylate.

Il a aussi été montré selon l'invention que les gènes *thyX* sont retrouvés dans le génome de nombreuses bactéries, bactériophages et virus bactériens ou eucaryotes, alors qu'ils sont absents dans le génome des mammifères, en particulier dans le génome humain.

20 La caractérisation, selon l'invention, d'une nouvelle voie de synthèse du dTMP, grâce à la famille des protéines THYX a rendu pour la première fois accessible à l'homme du métier les nombreuses applications qui en découlent directement et qui sont exposées dans la présente description. Les différentes utilisations d'un polypeptide THYX
25 ou d'un acide nucléique codant un polypeptide THYX exposées dans la présente demande de brevet sont liées techniquement du fait des caractéristiques fonctionnelles et structurales communes des polypeptides THYX définis ci-dessous.

L'invention a pour objet l'utilisation d'un polypeptide THYX
30 comprenant la séquence d'acides aminés suivante :
X₁HR(X)₇S, dans laquelle :

- X₁ représente l'acide aminé R (Arginine ou Arg) ou K (Lysine ou Lys), et

- (X)₇ est une chaîne de sept acides aminés consécutifs dans laquelle chaque X représente, indépendamment l'un de l'autre, l'un quelconque des 20 acides aminés naturels, dans un procédé de synthèse *in vitro* de la thymidine 5'-monophosphate (dTMP).

- S représente l'acide aminé Sérine ou Ser, selon le code à une lettre conforme à la nomenclature internationale.

De préférence, X₁ représente l'acide aminé R.

De préférence, le polypeptide THYX est choisi parmi les polypeptides comprenant les séquences d'acides aminés SEQ ID N°1 à SEQ ID N°37.

L'invention est également relative à l'utilisation d'un acide nucléique codant pour un polypeptide THYX tel que défini ci-dessus pour la production dudit polypeptide THYX.

De préférence, l'acide nucléique est choisi parmi les acides nucléiques comprenant les séquences nucléotidiques SEQ ID N°44 à SEQ ID N°64.

L'invention concerne aussi l'utilisation d'un polypeptide THYX tel que défini ci-dessus dans un procédé de criblage de composés inhibiteurs de thymidylate synthase, notamment de composés antibactériens ou antiviraux.

Elle est également relative à l'utilisation d'un acide nucléique codant pour un polypeptide THYX tel que défini ci-dessus dans un procédé de criblage de composés inhibiteurs de thymidylate synthase, notamment de composés anti-bactériens ou anti-viraux.

Elle a aussi pour objet l'utilisation d'un oligonucléotide antisens hybridant spécifiquement avec l'ARN messenger codant pour un polypeptide THYX pour inhiber *in vitro* la synthèse d'ADN dans une bactérie ou un virus.

Elle est également relative à l'utilisation d'un acide nucléique codant pour un polypeptide THYX tel que défini ci-dessus comme marqueur de sélection d'un événement de recombinaison génétique.

Elle concerne aussi l'utilisation d'une sonde ou d'une amorce nucléotidique spécifique d'un acide nucléique codant pour un polypeptide THYX pour détecter une bactérie ou un virus, particulièrement une

bactérie ou un virus pathogène pour les mammifères, et plus spécifiquement pour l'homme, ainsi qu'un nécessaire ou kit pour détecter cette bactérie ou ce virus, comprenant une sonde ou une amorce nucléotidique spécifique d'un acide nucléique codant pour un polypeptide
5 THYX.

Egalement, les travaux des inventeurs sur l'activité de ThyX ont permis d'identifier le mécanisme enzymatique et ainsi d'optimiser les conditions de la réaction en mettant en évidence l'importance de l'addition de certains composés dans le milieu réactionnel.

10 L'invention a donc pour but l'utilisation de tels milieux et leur application.

DESCRIPTION DES FIGURES

La figure 1 illustre un alignement des séquences d'acides aminés de divers polypeptides THYX.
15

Dans la colonne de gauche sont indiqués le nom des organismes desquels proviennent chacune des séquences d'acides aminés.

Les numéros identifient le numéro d'ordre de l'acide aminé du début de la lignée correspondante, dans la séquence du polypeptide
20 THYX de l'organisme.

La figure 2A représente des gels d'immunoempreintes réalisés sur des extraits cellulaires obtenus à partir de bactéries *E.coli* X2913(Δ thyA) transfectées avec le gène *ThyX* de *P.abysssi* (pistes numéro 1 et numéro 2) ou de bactéries *E. coli* X 2913 (Δ thyA) transfectées avec le gène *thyX* de *H.pylori* (pistes numéro 3 et numéro 4).
25

Les bactéries *E.coli* transformantes ont été cultivées en l'absence d'*Arabinose* (pistes numéro 1 et 3) ou en présence de 0,2% d'*Arabinose* (pistes numéro 2 et 4). Les bandes visibles sur le gel d'immunoempreintes (3Western blot ») correspondent respectivement au polypeptide THYX de *P.abysssi* (piste numéro 2) et au polypeptide THYX de *H.pylori* (piste numéro 4).
30

La figure 2B illustre des clichés photographiques de boîtes de Pétri qui ont étéensemencées :

- sur la partie gauche, par une bactérie *E.coli* X2193 (Δ thyA) transformée avec le gène *thyX* de *H.pylori* ; et
35

- sur la partie droite, avec des bactéries *E.coli* X2193 ($\Delta thyA$) transformées avec le gène *thyX* de *P.abysyi*.

Le cliché de gauche montre les résultats de croissance de la bactérie *E.coli* transformée cultivée sur de la gélose (agar) en milieu minimum M9 additionné de 0,2% d'*Arabinose*. Seules les bactéries *E.coli* transformées avec le gène *thyX* de *H.pylori* se multiplient.

Le cliché de droite représente les mêmes bactéries cultivées dans la gélose en présence de milieu minimal M9 en l'absence d'*Arabinose*. Aucune croissance bactérienne n'est observée, quelle que soit la bactérie transfectée.

La figure 3 : Migration sur gel SDS PAGE coloré avec Bleu de Comassie. Le premier puits correspond au marqueur Mid-Range de PROMEGA. Les flèches montrent les poids moléculaires qui correspondent aux bandes du marqueur. Le deuxième puits contient l'échantillon de notre protéine, après l'avoir purifiée sur une colonne de Ni-NTA, sa bande se trouve à environ 28 kDa.

La figure 4 : Western Blot des fractions après la purification de la protéine sur une colonne de Ni-NTA démontrent purification de THYX.

M = marqueur Mid-Range PROMEGA,

TF = la première fraction après l'incubation de la protéine avec la résine,

L = fraction après le lavage

E1 = protéine éluée.

La figure 5 présente des analyses biochimiques de la ThyX de *H. Pylori*.

La figure 6 illustre l'activité de formation du dTMP de la protéine ThyX de *H. pylori*.

DESCRIPTION DETAILLEE DE L'INVENTION

Il a été montré pour la première fois selon l'invention qu'une famille de polypeptides, désignée THYX, distincte de la famille des polypeptides codés par les gènes *thyA*, possédait une activité catalytique du type thymidylate synthase.

Le demandeur a cloné le gène *thyX* de *H. Pylori* dans un vecteur d'expression fonctionnel dans *Escherichia coli*. Ce vecteur d'expression a été utilisé pour transformer une souche de *Escherichia coli* auxotrophe pour la thymidine, plus spécifiquement la souche de *E. coli* X2913 ($\Delta thyA572$), dont le fond génétique est précisément caractérisé et dans laquelle le gène *thyA* est délété. Les résultats présentés dans les exemples montrent que l'expression du gène *thyX* qui a été artificiellement introduit dans la souche X2913($\Delta thyA$) a permis de restaurer la capacité de *E.coli* à synthétiser *de novo* le dTMP. Du fait de la grande connaissance accumulée sur les caractéristiques du génome de *E.coli* ainsi que sur la voie de synthèse normale du dTMP via l'expression de la thymidylate synthase THYA, les résultats obtenus montrent que le produit d'expression du gène *thyX* de *H. Pylori* pallie directement le défaut de production du polypeptide THYA dans cet organisme bactérien.

Par une étude d'homologie de séquences, il est montré selon l'invention que des polypeptides possédant des séquences homologues aux polypeptides *THYX* de *Helicobacter pylori* (SEQ ID n°21) étaient également codés par le génome de nombreuses bactéries, *Archaeobacteries* bactériophages ainsi que dans certains virus.

De manière surprenante, il est aussi montré selon l'invention que toutes les bactéries, *Archae bacterie* et virus possédant dans leur génome une copie d'un gène *thyX* ne possèdent pas simultanément le gène *thyA*, antérieurement connu comme le seul gène capable d'assurer la synthèse du dTMP. Ces organismes possédant le gène *thyX* en l'absence du gène *thyA* ne comprennent pas non plus de copie du gène *tdk* codant la thymidine kinase requise pour l'incorporation intracellulaire de thymidine exogène qui est ensuite convertie dans la cellule en dTMP, à l'exception notable de la bactérie *Mycobacterium tuberculosis*. La présence simultanée des gènes *thyA* et *thyX* chez *M.tuberculosis* est probablement due à un événement de transfert de gène.

L'exclusion mutuelle des gènes *thyA* et *thyX* dans le génome des organismes mentionnés ci-dessus montre clairement que *thyX* supplée la fonction thymidylate synthase qui n'est plus assurée en l'absence de *thyA*. De plus, seul le gène *thyA* est systématiquement absent du

génomique des bactéries possédant le gène *thyX*, alors que d'autres gènes impliqués dans le métabolisme des nucléotides sont indifféremment présents ou absents dans le génome de ces organismes.

Il est aussi montré selon l'invention que les bactéries déficientes en *thyA* et dans lesquelles est présent le gène *thyX* sont, pour la plupart, des bactéries pathogènes pour les mammifères. On peut citer notamment *Campylobacter jejuni* qui provoque des intoxications alimentaires, *Helicobacter pylori* qui est un agent causal des ulcères, *Rickettsia prowazekii* qui est un agent causal du typhus, *Borrelia burgdorferi* qui est impliqué dans la maladie de Lyme, *Treponema pallidum* qui est l'agent causal de la syphilis, les bactéries du genre *Chlamydiae* qui sont des pathogènes intracellulaires obligatoires ainsi que des virus à ADN eucaryotes tels que le *Chorella virus*.

On a aussi montré selon l'invention que des bactéries dans lesquelles le gène *thyX* a été inactivé deviennent auxotrophes pour la thymidine, c'est à dire qu'elles ne se multiplient que si l'on ajoute de la thymidine exogène dans le milieu de culture..

Toutes les protéines appartenant à la famille THYX possédant une activité thymidylate synthase nouvellement identifiée selon l'invention possède en commun les caractéristiques structurales et fonctionnelles indiquées ci-dessous :

Caractéristiques structurales :

- a) Les gènes *thyX* codent, tous sans exception, pour un polypeptide THYX comprenant la séquence d'acides aminés suivante :
- $X_1H R(X)_7 S$, dans laquelle :
- X_1 représente l'acide aminé R (Arginine ou Arg) ou K (Lysine ou Lys), et
 - $(X)_7$ est une chaîne de sept acides aminés consécutifs dans laquelle chaque X représente, indépendamment l'un de l'autre, l'un quelconque des 20 acides aminés naturels.

De préférence, X_1 représente l'acide aminé R.

X_1 représente K, notamment pour le polypeptide THYX codé par le génome du Roseophage S101.

b) Comme le montre clairement l'alignement des séquences d'acides aminés de protéines THYX d'origines diverses représenté à la figure 1, les polypeptides THYX possèdent un résidu d'acide aminé sérine conservé. De plus, les résultats des expériences de mutagenèse réalisées sur le gène *thyX* de *Helicobacter pylori*, présentés dans les exemples, montrent que l'acide aminé conservé sérine est indispensable à l'activité catalytique du polypeptide THYX, puisque le remplacement du résidu sérine respectivement par un résidu cystéine ou par un résidu alanine aboutit à la production d'un polypeptide qui est incapable de compléter la déficience en polypeptide THYA dans la souche de *E.coli* X de 2913 ($\Delta thyA$).

c) Les polypeptides THYX ne possèdent pas de résidus d'acides aminés cystéines conservés dans leurs séquences, contrairement aux polypeptides codés par les gènes *thyA*, dans lesquels le résidu cystéine conservé a un rôle essentiel nucléophile dans la réaction de méthylation catalysée par les polypeptides THYA.

d) En utilisant le logiciel BLAST (version 2.0) avec les paramètres par défaut, puis le module Psi-BLAST pour réaliser des cycles itératifs de recherche d'homologie avec la séquence du polypeptide THYX de *H. Pylori* comme « appât », aucune séquence cible de thymidylate synthase *thyA* n'a été sélectionnée, ce qui montre l'absence d'homologie entre les gènes *thya* et *thyX*.

Caractéristiques fonctionnelles

a) Un polypeptide THYX selon l'invention est capable de catalyser l'oxydation du méthylène tétrahydrofolate en tétrahydrofolate, ce qui est une caractéristique de l'activité catalytique du type thymidylate synthase, et montre que les polypeptides THYX appartiennent à la classe des enzymes de type thymidylate synthase.

b) la croissance de microorganismes transformés par un gène *thyX* peut être inhibée par une forte concentration de triméthoprim, qui est un inhibiteur spécifique de la dihydrofolate réductase ; le fait que seules de fortes concentrations de triméthoprim inhibent THYX suggère qu'il existe des différences fonctionnelles entre les voies métaboliques de formation

du thymidylate dans lesquelles sont impliquées respectivement *thyA* et *thyX*.

c) des bactéries dans lesquelles le gène *thyX* a été inactivé deviennent auxotrophes pour la thymidine ;

5 d) un polypeptide THYX selon l'invention n'est catalytiquement actif qu'en présence d'un co-facteur du type flavine. On peut noter que les polypeptides codés par les gènes *thyA*, antérieurement connus comme les seuls gènes codant pour des enzymes du type thymidylate synthase, sont actifs sans requérir la présence de flavine ;

10 Sont inclus dans les polypeptides appartenant à la famille de polypeptides THYX possédant des caractéristiques structurales et fonctionnelles définies ci-dessus les polypeptides THYX comprenant les séquences d'acides aminés SEQ ID N°1 à SEQ ID N°37.

15 Les polypeptides THYX de séquences SEQ ID n°1 à SEQ ID n°37 ont été rendus accessibles au public notamment par leur publication dans les bases de données de séquences d'acides aminés.

L'invention a pour objet l'utilisation d'un polypeptide THYX comprenant la séquence d'acides aminés suivante :

X_1 HR(X)₇ S, dans laquelle

20 - X_1 représente l'acide aminé R (Arginine ou Arg) ou K (Lysine ou Lys), et

- (X)₇ est une chaîne de sept acides aminés consécutifs dans laquelle chaque X représente, indépendamment l'un de l'autre, l'un quelconque des 20 acides aminés naturels,

25 dans un procédé de synthèse *in vitro* de la thymidine 5'-monophosphate (dTMP).

De préférence, l'acide aminé X_1 représente l'acide aminé R.

30 Afin de réaliser la synthèse *in vitro* de dTMP à l'aide d'un polypeptide THYX selon l'invention, l'homme du métier peut notamment se référer aux exemples de la présente demande de brevet, dans lesquels l'activité thymidylate synthase du polypeptide THYX de *Helicobacter pylori* (SEQ ID N°21) est mise en évidence dans des extraits cellulaires par la détection de la réaction d'oxydation du composé méthylène tétrahydrofolate.

De préférence, l'utilisation ci-dessus est caractérisée en ce que le polypeptide THYX est choisi parmi les polypeptides comprenant les séquences d'acides aminés SEQ ID N°1 à SEQ ID N°37.

Selon un autre aspect, l'utilisation ci-dessus est caractérisée en ce que le polypeptide THYX est choisi parmi les polypeptides consistant en les séquences d'acides aminés SEQ ID N°1 à seq id N°37.

Le polypeptide THYX de séquences en acides aminés SEQ ID N°5 est codé par un gène provenant de l'organisme *Dictyostelium discoideum* qui a été décrit en 1989 par Dynes et Firtel, et désigné « Thy1 » par ces auteurs.

Cependant, DYNES et FIRTEL avaient explicitement exclu que le gène Thy1 de *Dictyostelium discoideum*, puisse coder une thymidylate synthase. La voie de biosynthèse de la thymidine par *Dictyostelium discoideum*, qui n'est toujours pas connue à ce jour, était a fortiori inconnue en 1989, comme étaient inconnues aussi les bases moléculaires causales de l'autotrophie de la thymidine. De plus, l'organisme *Dictyostelium discoideum* n'a pas encore fait l'objet de travaux de séquençage systématique de son génome. A fortiori, en 1989, on ne disposait d'aucune donnée concernant la caractérisation du fond génétique de cet organisme, ce qui constitue encore aujourd'hui une barrière technique insurmontable pour identifier le rôle fonctionnel direct d'une mutation, en particulier d'une mutation provoquant une altération d'une voie métabolique aussi complexe que celle des nucléotides, en particulier du nucléotide thymidine. De fait, DYNES et FIRTEL n'avaient pas caractérisé la nature de la mutation chez *Dictyostelium discoideum*. La fonction de l'insert d'ADN du clone permettant de compléter l'organisme afin de restaurer l'autotrophie par la thymidine était totalement inconnue. De plus, depuis 1989, les nombreuses équipes de recherches étudiant la voie de biosynthèse cellulaire du dTMP ont continué d'accumuler des résultats expérimentaux montrant que la thymidilate synthase codée par les gènes *thyA* serait l'unique voie de synthèse du dTMP (D. PAPAMICHAEL The Oncologist, 1989, précité).

L'invention est également relative à l'utilisation d'un acide nucléique codant pour un polypeptide THYX tel que défini ci-dessus en vue de produire ledit polypeptide THYX pour sa mise en œuvre dans les

différentes utilisations d'un polypeptide THYX décrites dans la présente description.

A partir des séquences en acides aminés SEQ ID N°1, à SEQ ID N°37 et/ou des séquences nucléotidiques SEQ ID N°44 à SEQ ID N°64, l'homme du métier est capable de détecter, isoler, cloner et caractériser tout acide nucléique codant pour un polypeptide THYX tel que défini ci-dessus, par exemple en synthétisant des sondes nucléotidiques spécifiques d'un acide nucléique codant le peptide X₁HR(X)₇S ou RHR(X)₇ S.

Pour construire de telles sondes, l'homme du métier peut adapter leur séquence en fonction de l'usage du codon pour un organisme déterminé. La détection d'un gène *thyX* peut être réalisée par hybridation sur un gel d'ADN (« Southern Bot ») ou encore par amplification PCR, par exemple en utilisant la sonde définie ci-dessus comme amorce nucléotidique.

Par exemple, le gène *thyX* de *H. Pylori* peut être isolé à l'aide des amorces nucléotidiques de séquences SEQ ID N°38 et SEQ ID N°39 et le gène *thyX* de *P.abyssi* peut être isolé à l'aide des amorces nucléotidiques de séquences SEQ ID N°40 et SEQ ID N°41. Le gène *thyX* de *Campylobacter jejuni* peut être isolé à l'aide des amorces nucléotidiques de séquences SEQ ID N°42 et SEQ ID N°43.

De préférence, l'acide nucléique codant pour un polypeptide THYX est choisi parmi les acides nucléiques codant pour un polypeptide THYX comprenant l'une des séquences d'acides aminés SEQ ID N°1 à SEQ ID N°37.

Selon un autre aspect, l'acide nucléique est choisi parmi les acides nucléiques codant pour un polypeptide THYX consistant en une des séquences d'acides aminés SEQ ID N°1 à SEQ ID N°37.

Avantageusement, l'acide nucléique est choisi parmi les acides nucléiques comprenant les séquences nucléotidiques SEQ ID N°44 à SEQ ID N°64.

Font partie des polypeptides THYX susceptibles d'être mis en œuvre selon l'invention les polypeptides THYX possédant au moins 95%

d'identité en acides aminés avec un polypeptide THYX choisi parmi les séquences SEQ ID N°1 à SEQ ID N°37.

Font partie des acides nucléiques codant pour un polypeptide THYX susceptible d'être mis en œuvre selon l'invention un acide nucléique possédant au moins 95% d'identité en nucléotides avec un acide nucléique choisi parmi les séquences nucléotidiques SEQ ID N°44 à SEQ ID N°64.

Identité entre deux acides nucléiques ou entre deux polypeptides

Aux fins de la présente description, l'expression « séquence nucléotidique » est employée pour désigner indifféremment un polynucléotide ou un acide nucléique. L'expression « séquence nucléotidique » englobe le matériel génétique lui-même et n'est donc pas restreinte à l'information concernant sa séquence.

Selon l'invention, un premier acide nucléique ayant au moins 95% d'identité avec un second acide nucléique de référence, possédera au moins 95%, de préférence au moins 96%, 97%, 98%, 98,5%, 99%, 99,1%, 99,2%, 99,3%, 99,4%, 99,5%, 99,6%, 99,7%, 99,8% ou 99,9% d'identité en nucléotides avec ce second polynucléotide de référence, le pourcentage d'identité entre deux séquences étant déterminé comme décrit ci-dessous.

Selon l'invention, un premier polypeptide ayant au moins 95% d'identité avec un second polypeptide de référence, possédera au moins 95%, de préférence au moins 96%, 97%, 98%, 98,5%, 99%, 99,1%, 99,2%, 99,3%, 99,4%, 99,5%, 99,6%, 99,7%, 99,8% ou 99,9% d'identité en acides aminés avec ce second polypeptide de référence, le pourcentage d'identité entre deux séquences étant déterminé comme décrit ci-dessous.

Le " pourcentage d'identité " entre deux séquences de nucléotides ou d'acides aminés, au sens de la présente invention, peut être déterminé en comparant deux séquences alignées de manière optimale, à travers une fenêtre de comparaison.

La partie de la séquence nucléotidique ou polypeptidique dans la fenêtre de comparaison peut ainsi comprendre des additions ou des

délétions (par exemple des "gaps") par rapport à la séquence de référence (qui ne comprend pas ces additions ou ces délétions) de manière à obtenir un alignement optimal des deux séquences.

Le pourcentage est calculé en déterminant le nombre de positions auxquelles une base nucléique ou un résidu d'acide aminé identique est observé pour les deux séquences (nucléique ou peptidique) comparées, puis en divisant le nombre de positions auxquelles il y a identité entre les deux bases ou résidus d'acides aminés comparés, par le nombre total de positions dans la fenêtre de comparaison, puis en multipliant le résultat par cent afin d'obtenir le pourcentage d'identité de séquence.

L'alignement optimal des séquences pour la comparaison peut être réalisé de manière informatique à l'aide d'algorithmes connus.

De manière préférée, le pourcentage d'identité de séquences est déterminé à l'aide du logiciel BLAST (version BLAST 2.06 de Septembre 1998), en utilisant exclusivement les paramètres par défaut.

Un acide nucléique possédant au moins 95% d'identité en nucléotides avec un acide nucléique selon l'invention englobe les "variants" d'un acide nucléique selon l'invention.

Par "variant" d'un acide nucléique selon l'invention, on entend un acide nucléique qui diffère de l'acide nucléique de référence par une ou plusieurs substitutions, additions ou délétions d'un nucléotide, par rapport à l'acide nucléique de référence. Un variant d'un acide nucléique selon l'invention peut être d'origine naturelle, tel qu'un variant allélique qui existe naturellement. Un tel acide nucléique variant peut être également un acide nucléique non naturel obtenu, par exemple, par des techniques de mutagenèse.

En général, les différences entre l'acide nucléique de référence et l'acide nucléique "variant" sont réduites de telle sorte que l'acide nucléique de référence et l'acide nucléique variant ont des séquences nucléotidiques très similaires et, dans de nombreuses régions, identiques. Les modifications nucléotidiques présentes dans un acide nucléique variant peuvent être silencieuses, ce qui signifie qu'elles n'affectent pas la séquence d'acides aminés qui peut être codée par cet acide nucléique variant.

Les modifications de nucléotides dans l'acide nucléique variant peuvent aussi résulter en des substitutions, additions ou délétions d'un ou plusieurs acides aminés dans la séquence du polypeptide qui peut être codé par cet acide nucléique variant.

5 De manière tout à fait préférée, un acide nucléique variant selon l'invention comportant une phase de lecture ouverte, code pour un polypeptide qui conserve la même fonction ou la même activité biologique que le polypeptide codé par l'acide nucléique de référence.

De manière tout à fait préférée, un acide nucléique variant selon
10 l'invention et qui comporte une phase de lecture ouverte, code pour un polypeptide THYX qui conserve l'activité catalytique d'une thymidylate synthase, qui peut notamment être détectée par la capacité du polypeptide THYX à oxyder le méthylène-tétrahydrofolate *in vitro* ou encore à restaurer la capacité d'une bactérie ou d'une cellule eucaryote
15 déficiente en thymidylate synthase THYA à synthétiser l'ADN, comme cela est décrit dans les exemples.

De manière tout à fait préférée, un polypeptide THYX variant ne comprendra aucune modification d'acide aminé sur le motif $X_1HR(X)_7S$, et l'acide aminé Sérine conservé est présent.

20 Font partie des polypeptides THYX susceptibles d'être mis en œuvre selon l'invention, les polypeptides THYX codés par un acide nucléique tel que défini ci-dessus.

Sont également inclus dans la définition d'un polypeptide THYX selon l'invention les polypeptides THYX comprenant une ou plusieurs
25 substitutions d'acides aminés dans l'une des séquences SEQ ID N°1 à SEQ ID N°37 par un acide aminé « équivalent ». Sont compris dans la définition d'acides aminés "équivalents" les acides aminés appartenant à la même classe, tels que les acides aminés acides (D et E), basiques (K, R et H), non polaires (A, V, L, I, P, M, F et W) ou encore polaires non
30 chargés (G, S, T, C, Y, N et Q).

Font également partie de l'invention des polypeptides dits "homologues" à l'un quelconque des polypeptides THYX de séquences d'acides aminés SEQ ID N°1 à SEQ ID N°37, ou de leurs variants.

De tels polypeptides homologues ont des séquences d'acides aminés possédant une ou plusieurs substitutions d'un acide aminé par un acide aminé équivalent, par rapport aux polypeptides de référence.

On entendra par acide aminé équivalent selon la présente invention, par exemple remplacement d'un résidu sous la forme L par un résidu sous la forme D ou encore le remplacement d'un acide glutamique (E) par un acide pyro-glutamique selon des techniques bien connues de l'homme du métier. A titre illustratif, la synthèse de peptide contenant au moins un résidu sous la forme D est décrite par KOCH (1977).

Selon un autre aspect, sont également considérés comme des acides aminés équivalents deux acides aminés appartenant à la même classe, c'est-à-dire deux acides aminés acide, basique, non polaire ou encore polaire non chargé.

De préférence, les polypeptides selon l'invention comprenant une ou plusieurs additions, délétions, substitutions d'au moins un acide aminé conserveront leur capacité à être reconnus par des anticorps dirigés contre les polypeptides non modifiés. Ces polypeptides conserveront également leur activité catalytique thymidylate synthase.

De préférence, un polypeptide THYX ou un acide nucléique codant pour un polypeptide THYX susceptible d'être mis en œuvre selon l'invention se présente sous la forme isolée ou purifiée.

Le terme "isolé" au sens de la présente invention désigne un matériel biologique qui a été soustrait à son environnement originel (l'environnement dans lequel il est localisé naturellement). Par exemple, un polypeptide ou un polynucléotide présent à l'état naturel dans un animal ou une plante n'est pas isolé. Le même polypeptide séparé de son environnement naturel ou le même polynucléotide séparé des acides nucléiques adjacents au sein desquels il est naturellement inséré dans le génome de l'animal ou de la plante est isolé.

Un tel polynucléotide peut être inclus dans un vecteur et/ou un tel polynucléotide peut être inclus dans une composition et demeurer néanmoins à l'état isolé, du fait que le vecteur ou la composition ne constituent pas son environnement naturel.

Le terme "purifié" ne nécessite pas que le matériel soit présent sous une forme de pureté absolue, exclusive de la présence d'autres

composés. Il s'agit plutôt d'une définition relative. Un polypeptide ou un polynucléotide est à l'état purifié après purification du matériel de départ d'au moins un ordre de grandeur, de préférence 2 ou 3 et préférentiellement 4 ou 5 ordres de grandeur.

5 Font également partie de l'invention les polypeptides THYX de séquence SEQ ID N°1 à SEQ ID N° 37 rendus résistants à la protéolyse par l'introduction d'une ou plusieurs liaisons non peptidiques, telles qu'une liaison réduite (CH_2NH), une liaison rétro-inverso (NHCO), une liaison méthylène-oxy ($\text{CH}_2\text{-O}$), une liaison thiométhylène ($\text{CH}_2\text{-S}$), une
10 liaison carba ($\text{CH}_2\text{-CH}_2$), une liaison cétométhylène (CO-CH_2), une liaison hydroxyéthylène (CHOH-CH_2) ou encore une liaison CH=CH .

Dans tous les cas, un polypeptide THYX susceptible d'être mis en œuvre selon l'invention conserve l'activité catalytique d'une thymidylate synthase, qui peut notamment être détectée par la capacité du
15 polypeptide THYX à oxyder le méthylène-tétrahydrofolate *in vitro* ou encore à restaurer la capacité d'une bactérie ou d'une cellule eucaryote déficiente en thymidylate synthase THYA à synthétiser l'ADN, comme cela est décrit dans les exemples.

20 **Production d'un polypeptide THYX susceptible d'être mis en œuvre selon l'invention.**

L'invention est également relative à un procédé pour la production de l'un des polypeptides THYX tels que définis ci-dessus, en particulier d'un polypeptide choisi parmi les polypeptides THYX de séquences
25 d'acides aminés SEQ ID N°1 à SEQ ID N°37 ou d'un variant de ce dernier, ladite méthode comprenant les étapes de :

- a) insérer un acide nucléique codant pour ledit polypeptide dans un vecteur approprié ;
- b) cultiver, dans un milieu de culture approprié, une cellule hôte
30 préalablement transformée ou transfecter avec le vecteur recombinant de l'étape a) ;
- c) récupérer le milieu de culture conditionné ou lyser la cellule hôte, par exemple par sonication ou par choc osmotique ;
- d) séparer et purifier à partir dudit milieu de culture ou encore à
35 partir des lysats cellulaires obtenus à l'étape c), ledit polypeptide ;

e) le cas échéant, caractériser le polypeptide recombinant produit.

Les polypeptides THYX selon l'invention peuvent être caractérisés par fixation sur une colonne de chromatographie d'immunoaffinité sur laquelle les anticorps dirigés contre ce polypeptide ou contre un fragment
5 ou un variant de ce dernier ont été préalablement immobilisés.

Selon un autre aspect, un polypeptide THYX recombinant selon l'invention peut être purifié par passage sur une série appropriée de colonnes de chromatographie, selon les méthodes connues de l'homme de l'art.

10 Un polypeptide THYX selon l'invention peut être également préparé par les techniques classiques de synthèse chimique indifféremment en solution homogène ou phase solide.

A titre illustratif, un polypeptide THYX selon l'invention pourra être préparé par la technique ou en solution homogène décrite par HOUBEN
15 WEYL (1974) ou encore la technique de synthèse en phase solide décrite par MERRIFIELD (1965a; 1965b).

PROCEDES DE CRIBLAGE DE COMPOSES INHIBITEURS DE THYMIDYLATE SYNTHASE

20 Comme cela a déjà été exposé précédemment, les gènes *thyX* sont retrouvés dans de nombreuses bactéries pathogènes pour les mammifères, en particulier pathogènes pour l'homme et dans certains virus. De plus, au contraire des gènes *thyA*, les gènes *thyX* ne sont pas retrouvés dans le génome des mammifères. En particulier, le génome
25 humain est dépourvu d'un gène *thyX*.

En conséquence, les polypeptides THYX et les acides nucléiques codant pour les polypeptides THYX constituent des cibles privilégiées pour des composés inhibant spécifiquement l'expression des gènes *thyx* ou inhibant spécifiquement l'activité thymidylate synthase des
30 polypeptides THYX. De tels composés inhibiteurs sont susceptibles d'inhiber la synthèse d'ADN dans de nombreuses bactéries, bactériophages et virus, en particulier ceux qui sont pathogènes pour les mammifères, y compris l'homme, ainsi que la multiplication de ces bactéries, ces bactéries et ces virus, tout en ne provoquant pas d'effets
35 indésirables chez les mammifères, ou tout au moins en provoquant des

effets indésirables très réduits chez les mammifères, y compris l'homme, dont le génome ne comprend aucune copie d'un gène *thyX*.

Procédé de criblage de composés inhibiteurs du polypeptide *thyX*

5 L'invention a aussi pour objet l'utilisation d'un polypeptide THYX tel que défini dans la description dans un procédé de criblage de composés antibactériens ou anti-viraux.

10 Selon un premier mode de réalisation d'un procédé de criblage d'un composé antibactérien ou antiviral selon l'invention un polypeptide THYX tel que défini dans la description peut être utilisé pour cribler des molécules se fixant sur celui-ci.

La fixation du polypeptide avec la molécule ou substance peut ou inhiber (molécule antagoniste) l'activité thymidylate synthase dudit polypeptide.

15 De telles molécules capables de se fixer sur l'un quelconque des polypeptides selon l'invention comprennent des anticorps, des oligonucléotides, d'autres protéines et de manière générale des petites molécules de toute nature.

20 Dans un tel test de criblage, on peut simplement mettre en évidence la fixation de la molécule candidate aux polypeptides, l'un des deux partenaires étant marqué par un composé détectable (polypeptide d'intérêt ou molécule candidate), la visualisation du complexe polypeptide THYX/molécule candidate étant alors visualisée par détection du marqueur détectable, après élimination des molécules candidates non
25 liées spécifiquement.

A titre d'exemple, un test de criblage d'une molécule candidate capable de se fixer sur un polypeptide selon l'invention pourra comprendre avantageusement une première étape au cours de laquelle le polypeptide d'intérêt ou la molécule candidate est immobilisé sur un
30 support, une seconde étape au cours de laquelle le second partenaire (molécule candidate ou polypeptide d'intérêt) est mis en présence du premier composé préalablement immobilisé sur le support, une troisième étape au cours de laquelle un ou plusieurs lavages sont effectués dans des conditions appropriées à l'élimination des composés n'étant pas lié
35 spécifiquement, et enfin une quatrième étape au cours de laquelle le

complexe éventuellement formé entre le polypeptide d'intérêt et la molécule candidate est détecté.

Dans le mode de réalisation du test de criblage selon lequel la molécule candidate est préalablement immobilisée sur un support, puis
5 mise en présence du polypeptide d'intérêt selon l'invention, la détection du complexe formé par la molécule candidate et le polypeptide d'intérêt selon l'invention pourra être avantageusement réalisée à l'aide d'un anticorps tel que décrit ci-dessus.

Dans un autre mode de réalisation du test de criblage selon lequel
10 c'est le polypeptide d'intérêt selon l'invention qui est préalablement immobilisé sur un support, la molécule candidate sera avantageusement marquée à l'aide d'un marqueur détectable préalablement à sa mise en contact avec le polypeptide d'intérêt immobilisé.

Un tel marqueur détectable peut être radioactif ou non radioactif,
15 par exemple fluorescent ou correspondre à un ligand pour un troisième partenaire utilisé pour la détection comme une molécule de biotine.

En conséquence, l'invention a également pour objet un procédé de criblage d'une molécule ou d'une substance candidate interagissant avec
20 un polypeptide selon l'invention, ladite méthode comprenant les étapes de :

- a) mettre en contact un polypeptide conforme à l'invention avec la substance ou molécule candidate à tester ;
- b) détecter les complexes éventuellement formés entre ledit
25 polypeptide et ladite substance ou molécule candidate.

L'invention concerne également un nécessaire ou kit pour le criblage d'une molécule ou d'une substance candidate interagissant avec un polypeptide selon l'invention, ledit nécessaire comprenant :

- 30 a) un polypeptide conforme à l'invention ;
- b) le cas échéant, des moyens nécessaires à la détection du complexe formé entre ledit polypeptide et la molécule ou substance candidate.

Procédé de criblage de composés antibactériens ou antiviraux dans un système acellulaire.

Selon cet autre procédé de criblage de composés antibactériens ou antiviraux selon l'invention, le test d'inhibition de l'activité thymidylate synthase d'un polypeptide THYX est réalisé dans un système acellulaire, par exemple dans un lysat de culture de cellules exprimant le polypeptide THYX et qui n'exprime pas simultanément de polypeptide codé par un gène *thyA*.

Les cellules à partir desquelles est obtenu le lysat cellulaire sont de préférence des cellules possédant les mêmes caractéristiques que celles qui sont mises en œuvre dans le procédé de criblage dans un système acellulaire qui a été décrit en détail précédemment dans la description.

Brièvement, les cellules à partir desquelles est obtenu le lysat cellulaire sont respectivement :

- soit des cellules exprimant naturellement THYX en l'absence de THYA ;
- soit des cellules dont le génome desquelles le gène *thyA* a été inactivé et qui sont transfectées avec un vecteur recombinant exprimant un gène *thyX*.

Le lysat cellulaire peut être obtenu à partir d'une culture des cellules décrites ci-dessus, par exemple par sonication ou par choc osmotique, selon des techniques bien connues de l'homme du métier.

Après l'étape de lyse cellulaire proprement dite, les débris cellulaires peuvent être éliminés par une étape de centrifugation à l'issue de laquelle ces débris cellulaires sont retrouvés en culot, le surnageant de centrifugation comprenant notamment l'ensemble des protéines y compris la protéine THYX, et qui est récupérée pour la mise en œuvre du procédé de criblage.

Selon le procédé de criblage dans un système acellulaire, l'activité thymidylate synthase est quantifiée respectivement dans des échantillons témoins contenant seulement le lysat cellulaire et dans des échantillons de tests contenant un composé inhibiteur candidat, le cas échéant pour une série de concentrations croissantes du composé inhibiteur candidat.

De préférence, on mettra en œuvre une série d'échantillons de tests comprenant un composé inhibiteur candidat donné, à des concentrations croissantes.

5 L'activité thymidylate synthase peut être quantifiée notamment par la détection de l'oxydation du méthylène-tétrahydrofolate, comme cela est décrit dans les exemples.

L'invention a encore pour objet un procédé de criblage d'un composé antibactérien ou antiviral *in vitro* dans un système acellulaire, caractérisé en ce que ledit procédé comprend les étapes suivantes :

10 a) préparer un lysat cellulaire à partir d'une culture de cellules exprimant un polypeptide THYX en l'absence d'un polypeptide codé par un gène thyA.

b) ajouter au lysat cellulaire obtenu à l'étape a) le composé inhibiteur à tester ;

15 c) comparer l'activité thymidylate synthase respectivement dans le lysat cellulaire obtenu à l'étape a) et dans le lysat cellulaire obtenu à l'étape b) ; et

d) sélectionner les composés candidats pour lesquels une inhibition de l'activité thymidylate synthase a été détectée.

20 L'invention a également pour objet un kit ou nécessaire pour le criblage d'un composé inhibiteur de la thymidylate synthase caractérisé en ce qu'il comprend :

a) une composition comprenant un polypeptide THYX en solution ou sous forme lyophilisée ;

25 b) facultativement, un ou plusieurs réactifs nécessaires à la quantification de l'activité thymidylate synthase.

Selon un premier mode de réalisation, la composition contenant le polypeptide THYX consiste en un lysat cellulaire préparé comme décrit ci-dessus.

30 Selon un second mode de réalisation, la composition comprenant le polypeptide THYX comprend une quantité du polypeptide THYX sous forme purifiée adaptée à l'obtention d'un échantillon d'essai en solution comprenant une concentration du polypeptide THYX comprise entre 10^{-10} et 10^{-2} M, de préférence 10^{-8} à 10^{-3} M et de manière tout à fait préférée
35 entre 10^{-7} M et 10^{-5} M..

Procédé de criblage de composés antibactériens ou antiviraux dans un système cellulaire

L'invention concerne aussi l'utilisation d'un acide nucléique codant pour un polypeptide THYX tel que défini dans la description dans un
5 procédé de criblage de composés antibactériens ou antiviraux.

De préférence l'acide nucléique code pour un polypeptide THYX choisi parmi les polypeptides comprenant les séquences d'acides aminés SEQ ID N°1 à SEQ ID N°37.

De manière tout à fait préférée, l'acide nucléique est choisi parmi
10 les acides nucléiques comprenant les séquences nucléotidiques SEQ ID N°44 à SEQ ID N°64.

L'invention a également trait à un kit ou nécessaire pour le criblage d'un composé antibactérien ou antiviral, caractérisé en ce qu'il
15 comprend :

a) un vecteur d'expression recombinant comprenant un acide nucléique codant un polypeptide THYX tel que défini dans la présente description, sous le contrôle d'un promoteur fonctionnel dans une cellule hôte dans laquelle son expression est recherchée ou une cellule hôte
20 transfectée avec un tel vecteur recombinant ;

b) facultativement, un ou plusieurs réactifs nécessaire à la quantification de l'activité thymidylate synthase.

Selon un second mode de réalisation d'un procédé de criblage de composés antibactériens ou antiviraux selon l'invention, l'activité des
25 composés inhibiteurs de l'activité thymidylate synthase de THYX peut être testée dans des cultures de cellules exprimant THYX, en l'absence d'expression de *thyA*.

Selon un premier aspect, un tel procédé de criblage peut être réalisé sur des cultures de cellules pour le génome possède une copie
30 d'un gène THYX mais pas de gène *thyA*, comme par exemple des cultures de *Campylobacter jejuni*, *Helicobacter pylori*, *Rickettsia prowazekii*, *Borrelia burgdorferi* ou *Chlamydia*.

Selon un second aspect, un tel procédé de criblage de composés antibactériens ou antiviraux dans un système cellulaire peut être réalisé à
35 l'aide de cultures de cellules dans lesquelles le gène *thyA* a été inactivé

et qui ont été transfectés par un acide nucléique ou un vecteur recombinant exprimant un gène *thyX* dans ces cellules, comme par exemple la souche de *E.coli* n°X 2913 ($\Delta thyA$) qui a été transfectée avec un vecteur d'expression codant un polypeptide THYX.

5

L'invention a aussi pour objet un procédé de criblage d'un composé antibactérien ou antiviral caractérisé en ce qu'il comprend les étapes suivantes :

a) cultiver des cellules exprimant le gène *thyX* en l'absence d'expression du gène *thyA* dans un milieu de culture approprié ;

b) mettre en contact les cellules avec un composé candidat à tester ;

et

c) sélectionner les composés candidats inhibant l'activité thymidylate synthase du polypeptide codé par le gène *thyX*.

15

Les composés sélectionnés selon le procédé de criblage ci-dessus sont ceux pour lesquels est observée une inhibition de l'activité thymidylate synthase dans les cellules, par rapport à l'activité thymidylate synthase observée dans des cultures de cellules témoins qui ne sont pas mises en présence des composés candidats à tester.

20

La quantification de l'activité thymidylate synthase peut être réalisée par exemple par incorporation d'uracile radiomarquée dans l'ADN des cellules en culture, selon les techniques bien connues de l'homme du métier, la quantité de dTTP radiomarquée dans l'ADN reflétant le niveau d'activité thymidylate synthase dans la culture cellulaire. Dans ce cas, les cellules cultivées à l'étape a) du procédé sont incubées en présence d'uracil radiomarquée, par exemple de (^3H)-uracile ou (^{14}C)-uracile.

25

La radioactivité de dTTP dans l'ADN est mesurée après hydrolyse d'ADN purifié à l'aide des techniques standard (compteur de radioactivité).

30

Après lyse des cellules, par exemple par sonication ou par choc osmotique, les lysats cellulaires sont filtrés sur une membrane de nitrocellulose retenant l'ADN, puis la radioactivité contenue sur le filtre est mesurée à l'aide d'un compteur de radioactivité adapté.

35

Dans le mode de réalisation du procédé de criblage ci-dessus dans lequel les cellules cultivées consistent en des cellules dont le génome ne comprend aucune copie active d'un gène *thyA* et qui ont été
5 transfectées avec un vecteur recombinant comprenant un insert d'ADN codant un polypeptide *thyX*, le vecteur recombinant sera choisi de telle manière à permettre l'expression du polypeptide THYX dans la cellule hôte qui est cultivée durant le procédé.

Des exemples de vecteurs recombinants utilisables pour la mise
10 en œuvre du procédé de criblage ci-dessus sont détaillés ci-après.

Vecteurs recombinants susceptibles d'être utilisés selon l'invention

L'invention concerne aussi l'utilisation d'un vecteur recombinant comprenant un acide nucléique codant pour un polypeptide THYX tel que
15 défini ci-dessus ou un variant de ce polypeptide.

Avantageusement, un tel vecteur recombinant comprendra un acide nucléique choisi parmi les acides nucléiques suivants :

a) un acide nucléique codant pour un polypeptide ayant une
séquence en acides aminés choisie dans le groupe des séquences SEQ
20 ID N°1 à SEQ ID N°37 ou un variant de ce polypeptide, éventuellement fusionné à un polypeptide hétérologue ;

b) un acide nucléique comprenant un polynucléotide choisi parmi
les séquences SEQ ID N°44 à SEQ ID N°64, ou un variant de ce dernier
;

25 c) un acide nucléique codant pour un polypeptide THYX ayant au moins 95% d'identité en acides aminés avec un polypeptide choisi dans le groupe constitué des séquences SEQ ID N°1 à SEQ ID N°37 ou un variant de ce dernier ;

d) un acide nucléique ayant au moins 95% d'identité en
30 nucléotides avec un acide nucléique choisi dans le groupe constitué des séquences SEQ ID N°44 à SEQ ID N°64 ou un variant de ce dernier ;

Par " vecteur " au sens de la présente invention on entendra une molécule d'ADN ou d'ARN circulaire ou linéaire qui est indifféremment sous forme de simple brin ou double brin.

Sont préférés les vecteurs d'expression comprenant, outre un acide nucléique codant un polypeptide THYX conforme à l'invention, des séquences régulatrices permettant d'en diriger sa transcription et/ou sa traduction.

5 Selon un mode de réalisation avantageux, un vecteur recombinant selon l'invention comprendra notamment les éléments suivants :

(1) des éléments de régulation de l'expression de l'acide nucléique à insérer, tels que des promoteurs et des enhanceurs ;

10 (2) la séquence codante comprise dans l'acide nucléique conforme à l'invention à insérer dans un tel vecteur, ladite séquence codante étant placée en phase avec les signaux de régulation décrits aux (1) ; et

(3) des séquences d'initiation et d'arrêt de la transcription appropriées.

15 En outre, les vecteurs recombinants selon l'invention pourront inclure une ou plusieurs origines de répllication chez les hôtes cellulaires dans lesquels leur expression est recherchée, un ou plusieurs marqueurs de sélection.

20 A titre d'exemples, les promoteurs bactériens pourront être les promoteurs LacI, LacZ, les promoteurs de l'ARN polymérase du bactériophage T3 ou T7, les promoteurs PR, ou PL du phage lambda.

Les promoteurs pour cellules eucaryotes comprendront le promoteur de la thymidine kinase du virus HSV ou encore le promoteur de la métallothionéine-L de souris.

25 De manière générale, pour le choix d'un promoteur adapté, l'homme du métier pourra avantageusement se référer à l'ouvrage de SAMBROOK et al. (1989) précité ou encore aux techniques décrites par FULLER et al. (1996).

30 Des vecteurs particulièrement adaptés pour une expression des acides nucléiques selon l'invention dans des bactéries sont par exemple les vecteurs pQE70, pQE60 ou pQE-9 (commercialisés par la société QIAGEN), les vecteurs pBluescript, Page script, pNH8A ; pNH16a, pNH18a, pNH46A (commercialisés par la société Stratagene), les vecteurs pKK223-3, pKK233-3, pDR540 et pRIT5 (commercialisés par la
35 société Pharmacia) .

Des vecteurs particulièrement adaptés pour une expression dans cellules eucaryotes sont par exemple les vecteurs pWLNEO, pSV2CAT, pOG44, pXT1 et pSG (commercialisés par la société Stratagene), les vecteurs pSVK3, pBPV, pMSG et pSVL (commercialisés par la société Pharmacia)

Un premier vecteur préférentiellement mis en œuvre dans le cadre de l'invention est le vecteur pcDNA3 commercialisé par la société Invitrogen .

Un second vecteur particulièrement préféré est le vecteur pBluescript SK (-), commercialisé par la société Stratagene.

Les vecteurs bactériens préférés selon l'invention sont par exemple les vecteurs pBR322(ATCC37017) ou encore des vecteurs tels que pAA223-3 (Pharmacia, Uppsala, Suède), et pGEM1 (Promega Biotech, Madison, WI, ETATS-UNIS).

On peut encore citer d'autres vecteurs commercialisés tels que les vecteurs psiX174, pBluescript SA, pNH8A, pNH16A, pNH18A, pNH46A, pWLNEO, pSV2CAT, pOG44, pXTI, pSG(Stratagene).

Il peut s'agir également de vecteurs de type *baculovirus* tel que le vecteur pVL1392/1393 (Pharmingen) utilisé pour transfecter les cellules de la lignée Sf9 (ATCC N°CRL 1711) dérivées de *Spodoptera frugiperda*.

Il peut encore s'agir de vecteurs adénoviraux tels que l'adénovirus humain de type 2 ou 5.

Un vecteur recombinant selon l'invention peut aussi être un vecteur rétroviral ou encore un vecteur adéno-associé (AAV). De tels vecteurs adéno-associés sont par exemple décrits par FLOTTE et al. (1992), SAMULSKI et al. (1989), ou encore McLAUGHLIN BA et al. (1996).

De manière tout à fait préférée, on met en œuvre le vecteur pBAD TOPO commercialisé par la Société Invitrogen, qui permet une expression du gène *thyX* chez *E. Coli* qui est précisément régulée par la présence ou l'absence d'arabinose dans le milieu de culture des cellules transfectées avec un tel vecteur recombinant, du fait que le vecteur pBAD TOPO comprend le promoteur P_{BAD} qui est inductible par l'arabinose.

Exemples de composés inhibiteurs de l'activité thymidylate synthase d'un polypeptide THYX.

Ces composés inhibiteurs de l'activité thymidylate synthase d'un polypeptide THYX sont potentiellement des composés anti-bactériens et/ou antiviraux sans effets indésirables ou avec des effets indésirables réduits pour les mammifères, y compris l'homme.

La croissance de transformants contenant *thyX* dans la souche *E.coli* X2913($\Delta thyA$) peut être inhibée par le triméthoprim, qui est un inhibiteur spécifique de la dihydrofolate réductase. Ce fait indique que la croissance est dépendante des folates.

De même, il est montré selon l'invention que l'activité thymidylate synthase du polypeptide codé par le gène THYX de *Pyrococcus abyssi* est inhibé par un dérivé métabolique du 5-fluorouracile, ce qui suggère que le polypeptide THYX interagit avec le composé fluoro-dUMP.

D'autres composés susceptibles d'inhiber l'activité thymidylate synthase d'un polypeptide THYX sont respectivement les acides nucléiques antisens hybridant spécifiquement avec l'ARN messager codant pour un polypeptide THYX et les anticorps dirigés contre un polypeptide THYX.

Acides nucléiques antisens

Afin d'inhiber ou de bloquer l'expression d'un acide nucléique codant un polypeptide THYX, l'homme du métier pourra recourir à l'utilisation de polynucléotides antisens.

Ainsi, l'invention concerne aussi l'utilisation d'un polynucléotide ou oligonucléotide antisens, capable de s'hybrider spécifiquement avec l'ARN messager codant pour un polypeptide THYX et capable d'inhiber ou de bloquer sa transcription et/ou sa traduction. Un tel polynucléotide a la structure générale qui est définie dans la présente description pour les sondes et les amorces selon l'invention.

De préférence, un polynucléotide antisens susceptible d'être utilisé selon l'invention comprend une séquence correspondant à une séquence localisée dans la région de l'extrémité 5' de l'ARN messager ,

et de manière tout à fait préférée à proximité du codon d'initiation de la traduction (ATG) de l'acide nucléique codant le polypeptide THYX.

Selon un second mode de réalisation préférentiel, un polynucléotide antisens selon l'invention comprend une séquence
5 correspondant à l'une des séquences localisées au niveau des jonctions exon/intron d'un gène codant le polypeptide THYX et de manière tout à fait préférée des séquences correspondant à un site d'épissage.

Un polynucléotide antisens selon l'invention peut être préparé à partir d'un acide nucléique codant un polypeptide THYX choisi
10 parmi les polypeptides de séquences SEQ ID N°1 à SEQ ID N°37.

Un polynucléotide antisens selon l'invention peut être préparé à partir d'un acide nucléique choisi parmi les séquences nucléotidiques SEQ ID N°44 à SEQ ID N°64.

De manière générale, les polynucléotides antisens doivent
15 avoir une longueur et une température de fusion suffisante pour permettre la formation d'un hybride duplex intracellulaire ayant une stabilité suffisante pour inhiber l'expression de l'ARNm codant le polypeptide THYX considéré. Des stratégies pour construire des polynucléotides antisens sont notamment décrites par Green et al. (1986)
20 et Izant et Weintraub (1984).

Des méthodes de construction de polynucléotides antisens sont également décrites par Rossi et al. (1991) ainsi que dans les demandes PCT N° WO 94/23026, WO 95/04141, WO 9218522 et dans la demande de brevet européen n° EP 0 572 287.

25 Avantageusement, un polynucléotide antisens selon l'invention a une longueur de 15 à 200 nucléotides. Un polynucléotide sens de l'invention a ainsi une longueur allant de 15, 20, 25, 30, 35, 40, 45 ou 50 à 75, 100, 150 ou 200 nucléotides.

Afin d'inhiber ou de bloquer l'expression d'un acide nucléique
30 codant un polypeptide THYX tel que défini dans la description, on peut aussi avoir recours simultanément à une pluralité de polynucléotides antisens tels que définis ci-dessus, chacun des polynucléotides antisens hybridant avec une région distincte du gène ou de son ARN messenger.

D'autres méthodes de mise en œuvre des polynucléotides antisens sont par exemple celles décrites par Sczakiel et al. (1995) ou encore celles décrites dans la demande PCT n° WO 95/24223.

5 **Anticorps**

Les polypeptides THYX tels que définis selon l'invention, en particulier les polypeptides de séquences en acides aminés SEQ ID N°1 à SEQ ID N°37, ou les variants de ces derniers ainsi que les peptides homologues, peuvent être utilisés pour la préparation d'anticorps qui
10 peuvent être sélectionner pour leur capacité à inhiber ou à bloquer leur activité thymidylate synthase.

L'invention a aussi pour objet l'utilisation d'anticorps dirigés contre un polypeptide THYX pour inhiber ou bloquer l'activité thymidylate syn
15 thase de ce polypeptide.

De tels anticorps dirigés spécifiquement contre un polypeptide THYX constituent un nouvel exemple illustratif d'un composé inhibiteur de l'activité thymidylate synthase d'un polypeptide THYX selon l'invention.

20 De tels anticorps représentent potentiellement des composés antibactériens ou antiviraux.

Par " anticorps " au sens de la présente invention, on entendra notamment des anticorps polyclonaux ou monoclonaux ou des fragments (par exemple les fragments F (ab)'₂, Fab) ou encore tout polypeptide
25 comprenant un domaine de l'anticorps initial reconnaissant le polypeptide ou le fragment de polypeptide cible selon l'invention .

Des anticorps monoclonaux peuvent être préparés à partir d'hybridomes selon la technique décrite par KOHLER et MILSTEIN (1975).

30 La présente invention concerne également des anticorps dirigés contre un polypeptide tel que décrit ci-dessus ou un fragment ou un variant de ce dernier, tels que produits dans la technique du trioma ou encore la technique d'hybridome décrite par KOZBOR et al. (1983).

L'invention a également trait à des fragments d'anticorps simple chaîne Fv (ScFv) tels que décrits dans le brevet US N° 4,946,778 ou encore par MARTINEAU et al. (1998).

Les anticorps selon l'invention comprennent également des fragments d'anticorps obtenus à l'aide de banques de phages RIDDER et al., (1995) ou encore des anticorps humanisés (REIMANN et al., 1997; LEGER et al., 1997).

COMPOSITIONS PHARMACEUTIQUES

L'invention a également pour objet une composition pharmaceutique antibactérienne ou antivirale comprenant, à titre de principe actif un oligonucléotide antisens hybridant spécifiquement avec un ARN messager codant pour un polypeptide THYX tel que défini dans la présente description, en association avec un ou plusieurs excipients physiologiquement compatibles.

Elle est également relative à l'utilisation d'un oligonucléotide antisens hybridant spécifiquement avec l'ARN messager codant pour un polypeptide THYX tel que défini dans la présente description pour la fabrication d'un médicament antibactérien ou anti-viral.

Une telle composition pharmaceutique comprendra de préférence des concentrations d'oligonucléotides antisens qui sont au moins équimolaires à celles de l'ARN messager correspondant dans la cellule.

Parmi les excipients utilisables en association avec un oligonucléotide antisens tel que défini ci-dessus, on peut citer les molécules cationiques synthétiques se liant aux sites anioniques de l'oligonucléotide antisens, qui facilite le passage de l'oligonucléotide antisens à travers la membrane cellulaire par endocytose non spécifique, tels que ceux décrits par Schofield en 1995 ou ceux décrits par BEHR EN 1994.

Un autre excipient utilisable en association avec un oligonucléotide antisens selon l'invention est le composé lipofectinTM, qui est composé d'une formulation 1 :1 du composé d'ammonium quaternaire DOTMA et de dioleoylphosphatidylethanolamine, soniqués sous la forme de petites vésicules unilamellaires dans l'eau.

L'invention a aussi pour objet une composition pharmaceutique antibactérienne ou antivirale comprenant, à titre de principe actif, un anticorps dirigé spécifiquement contre un polypeptide THYX tel que défini dans la description, en association avec un ou plusieurs excipients physiologiquement compatibles.

Elle est également relative à l'utilisation d'un anticorps dirigé spécifiquement contre un polypeptide THYX tel que défini dans la description pour la fabrication d'un médicament antibactérien ou antiviral.

L'invention est également relative à un procédé pour la prévention ou le traitement d'une affection bactérienne ou virale, ledit procédé comprenant une étape d'administration d'une quantité thérapeutiquement efficace d'un oligonucléotide antisens ou d'un anticorps spécifique d'un polypeptide THYX tel que décrit ci-dessus.

Une composition pharmaceutique selon l'invention peut être administrée par n'importe quelle voie, par exemple par voie intraveineuse, intramusculaire, orale ou mucausale, en association avec un véhicule et/ou un adjuvant ou excipient physiologiquement compatible.

Un anticorps dirigé spécifiquement contre un polypeptide THYX est présent dans une composition pharmaceutique selon l'invention en des quantités adaptées pour une administration quotidienne de 10 nanogrammes à 10 mg d'anticorps, de préférence de 100 nanogrammes à 1 mg et de manière tout à fait préférée de 1 µg à 100 µg d'anticorps.

Des techniques de formulation et d'administration de composés inhibiteurs de la thymidylate synthase d'un polypeptide THYX peuvent être retrouvées par l'homme du métier dans l'ouvrage suivant : « REMINGTON'S PHARMACEUTICAL SCIENCES-MACK publication co., Easton, PA », dans sa dernière édition.

UTILISATION DE SONDAS ET D'AMORCES NUCLEOTIDIQUES HYBRIDANT AVEC UN ACIDE NUCLEIQUE CODANT UN POLYPEPTIDE THYX

Comme cela a déjà été exposé précédemment, les gènes *thyX* ont été retrouvés selon l'invention dans diverses bactéries et virus

pathogènes pour les mammifères, en particulier des bactéries pathogènes pour l'homme, et des virus.

En conséquence, des sondes ou des amorces dérivées des acides nucléiques génomique ou de l'ARN messager codant pour un polypeptide THYX constituent des moyens de détection de la présence
5 d'une bactérie ou d'un virus pathogène dans un échantillon, et plus spécifiquement d'un échantillon biologique prélevé sur l'homme ou l'animal, par exemple un échantillon de salive, de larmes, de sang, de plasma ou de sérum ou encore un prélèvement de biopsie ou un frottis.

10 Les acides nucléiques dérivés de l'une quelconque des séquences nucléotidiques codant un polypeptide THYX tel que défini dans la description, en particulier les acides nucléiques de séquences SEQ ID N°44 à SEQ ID N°64 sont utiles pour la détection de la présence d'au moins une copie d'une séquence nucléotidique choisie parmi les
15 séquences SEQ ID N°44 à SEQ ID N°64 ou encore d'un fragment ou d'un variant de cette dernière dans un échantillon.

De préférence, des sondes ou amorces nucléotidiques selon l'invention auront une longueur de 10, 12, 15, 18 ou 20 à 25, 35, 40, 50, 70, 80, 100, 200, 500, 1000, 1500 nucléotides consécutifs d'un acide
20 nucléique codant un polypeptide THYX ou d'un acide nucléique de séquence complémentaire.

Alternativement, une sonde ou une amorce nucléotidique selon l'invention consistera et/ou comprendra les fragments d'une longueur de 12, 15, 18, 20, 25, 35, 40, 50, 100, 200, 500, 1000, 1500 nucléotides
25 consécutifs d'un acide nucléique codant un polypeptide THYX selon l'invention, plus particulièrement d'un acide nucléique choisi parmi les séquences SEQ ID N°44 à SEQ ID N°64, ou d'un acide nucléique de séquence complémentaire.

La définition d'une sonde et d'une amorce nucléotidique selon
30 l'invention englobe des oligonucléotides qui hybrident, dans les conditions d'hybridation de forte stringence définies ci-après, avec un acide nucléique codant un polypeptide THYX, en particulier un acide nucléique choisi parmi les séquences SEQ ID N°44 à SEQ ID N°64 ou avec une séquence complémentaire de ces derniers.

Définition des conditions d'hybridation

Par conditions d'hybridation de forte stringence, au sens de l'invention, on entend les conditions d'hybridation suivantes:

5

Préhybridation:

mêmes conditions que pour l'hybridation
durée: 1 nuit.

10 **Hybridation:**

5 x SSPE (0.9 M NaCl, 50 mM phosphate de sodium pH 7.7, 5 mM EDTA)

5 x Denhardt's (0.2% PVP, 0.2% Ficoll, 0.2% SAB)

100 µg/ml ADN de sperme de saumon

15

0.1% SDS

durée: 1 nuit.

Lavages:

2 x SSC, 0.1% SDS 10 min 65°C

1 x SSC, 0.1 % SDS 10 min 65°C

20

0.5 x SSC, 0.1 % SDS 10 min 65°C

0.1 x SSC, 0.1% SDS 10 min 65°C.

Les paramètres définissant les conditions de stringence dépendent de la température à laquelle 50% des brins appariés se séparent (T_m).

25

Pour les séquences comprenant plus de 360 bases, T_m est définie par la relation:

$$T_m = 81,5 + 0,41 (\%G+C) + 16,6 \text{ Log}(\text{concentration en cations}) - 0,63 (\% \text{ formamide}) - (600/\text{nombre de bases})$$
 (SAMBROOK et al., (1989), pages 9.54-9.62).

30

Pour les séquences de longueur inférieure à 30 bases, T_m est définie par la relation: $T_m = 4(G+C) + 2(A+T)$.

Dans des conditions de stringence appropriées, dans lesquelles les séquences aspécifiques n'hybrident pas, la température d'hybridation est approximativement de 5 à 30°C, de préférence de 5 à 10°C en-
dessous de T_m .

35

Les conditions d'hybridation ci-dessus décrites sont mises en œuvre pour l'hybridation d'un acide nucléique de 200 bases de longueur et peuvent être adaptées en fonction de la longueur de l'acide nucléique dont l'hybridation est recherchée ou du type de marquage choisi, selon les techniques connues de l'homme du métier.

Les conditions convenables d'hybridation peuvent par exemple être adaptées selon l'enseignement contenu dans l'ouvrage de HAMES et HIGGINS (1985) ou encore dans l'ouvrage de AUSUBEL et al. (1989).

En particulier, il faut prendre en compte que le niveau et la spécificité d'hybridation dépend de différents paramètres, tels que :

- a) la pureté de la préparation de l'acide nucléique sur lequel la sonde ou l'amorce doit s'hybrider ;
- b) la composition en base de la sonde ou de l'amorce, les paires de bases G-C possédant une plus grande stabilité thermique que les paires de bases A-T ou A-U ;
- c) la longueur de la séquence de bases homologue entre la sonde ou l'amorce et l'acide nucléique ;
- d) la force ionique : le taux d'hybridation augmente avec l'accroissement de la force ionique et la durée du temps d'incubation ;
- e) la température d'incubation ;
- f) la concentration de l'acide nucléique sur lequel la sonde ou l'amorce doit s'hybrider ;
- g) la présence d'agents dénaturants, tels que des agents favorisant la rupture de liaisons hydrogène, comme le formamide ou l'urée, qui accroissent la stringence de l'hybridation ;
- h) le temps d'incubation, le taux d'incubation augmentant avec la durée de l'incubation ;
- i) la présence d'agents d'exclusion de volume, tels que le dextran ou le sulfate de dextran, qui augmentent le taux d'hybridation du fait qu'ils accroissent les concentrations effectives de la sonde ou de l'amorce et de l'acide nucléique qui doit s'hybrider, au sein de la préparation.

Une amorce ou une sonde nucléotidique selon l'invention peut être préparée par toute méthode adaptée bien connue de l'homme du métier, y compris par clonage et action d'enzymes de restriction ou encore par synthèse chimique directe selon des techniques telles que la méthode au

phosphodiester de NARANG et al. (1979) ou de BROWN et al. (1979), la méthode aux diéthylphosphoramidites de BEAUCAGE et al. (1980) ou encore la technique sur support solide décrite dans le brevet EU N°EP 0 707 592.

5 Chacun des acides nucléiques selon l'invention, y compris les sondes et amorces oligonucléotidiques décrites ci-dessus, peuvent être marqués, si désiré, en incorporant un marqueur détectable par des moyens spectroscopiques, photochimiques, biochimiques, immunochimiques ou encore chimiques.

10 Par exemple, de tels marqueurs peuvent consister en des isotopes radioactifs (^{32}P , ^{33}P , ^3H , ^{35}S ,), des molécules fluorescentes (5-bromodeoxyuridine, fluorescéine, acétylamino fluorène, digoxigénine) ou encore des ligands tels que la biotine.

Le marquage des sondes est fait de préférence par incorporation
15 de molécules marquées au sein des polynucléotides par extension d'amorces, ou bien par rajout sur les extrémités 5' ou 3'

Les sondes oligonucléotides selon l'invention peuvent être utilisées notamment dans des hybridations de type Southern à l'ADN génomique ou encore dans des hybridations à l'ARN messager
20 correspondant lorsque l'expression du transcrit correspondant est recherchée dans un échantillon.

Les sondes selon l'invention peuvent aussi être utilisées pour la détection de produits d'amplification PCR ou encore pour la détection de mésappariements.

25 Des sondes ou amorces nucléotidiques selon l'invention peuvent être immobilisées sur un support solide. De tels supports solides sont bien connus de l'homme du métier et comprennent des surfaces des puits de plaques de microtitration, des lits de polystyrène, des lits magnétiques, des bandes de nitrocellulose, ou encore des
30 microparticules telles que des particules de latex.

En conséquence, l'invention a aussi pour objet l'utilisation d'une sonde ou d'une amorce nucléique hybridant avec un acide nucléique codant pour un polypeptide THYX, tel que défini dans la description, dans un procédé de détection d'une bactérie ou d'un virus, en particulier d'une

bactérie ou d'un virus pathogène pour les mammifères, y compris l'homme.

La présente invention concerne également un procédé de détection de la présence d'un acide nucléique tel que décrit ci-avant dans un échantillon, ladite méthode comprenant les étapes de :

1) mettre en contact une ou plusieurs sondes nucléotidiques selon l'invention avec l'échantillon à tester ;

2) détecter le complexe éventuellement formé entre la ou les sondes et l'acide nucléique présent dans l'échantillon.

10

Selon un mode de réalisation particulier du procédé de détection selon l'invention, la ou les sondes oligonucléotidiques sont immobilisées sur un support.

Selon un autre aspect, les sondes oligonucléotidiques comprennent un marqueur détectable.

L'invention concerne en outre un nécessaire ou kit pour la détection de la présence d'un acide nucléique selon l'invention dans un échantillon, ledit nécessaire comprenant :

a) une ou plusieurs sondes nucléotidiques telles que décrites ci-dessus ;

b) le cas échéant, les réactifs nécessaires à la réaction d'hybridation.

Selon un premier aspect, le nécessaire ou kit de détection est caractérisé en ce que la ou les sondes sont immobilisées sur un support.

Selon un second aspect, le nécessaire ou kit de détection est caractérisé en ce que les sondes oligonucléotidiques comprennent un marqueur détectable.

Selon un mode de réalisation particulier du kit de détection décrit ci-dessus, un tel kit comprendra une pluralité de sondes oligonucléotidiques conformes à l'invention qui pourront être utilisées pour détecter des séquences cibles d'intérêt ou alternativement détecter des mutations dans les régions codantes ou les régions non codantes des acides nucléiques selon l'invention, plus particulièrement des acides

nucléiques de séquences SEQ ID N°44 à SEQ ID N°64 ou les acides nucléiques de séquence complémentaire.

Ainsi, les sondes selon l'invention immobilisées sur un support peuvent être ordonnées en matrices telles que les "puces à ADN". De
5 telles matrices ordonnées ont été en particulier décrites dans le brevet US N° 5,143,854, dans les demandes PCT N° WO 90/150 70 et 92/10092.

Des matrices supports sur lesquelles des sondes oligonucléotidiques ont été immobilisées à une haute densité sont par
10 exemple décrites dans les brevets US N°5,412,087 et dans la demande PCT N°WO 95/11995.

Les amorces nucléotidiques selon l'invention peuvent être utilisées pour amplifier l'un quelconque des acides nucléiques selon l'invention, et plus particulièrement tout ou partie d'un acide nucléique de séquences
15 SEQ ID N°44 à SEQ ID N°64, ou encore un variant de celui-ci.

Un autre objet de l'invention concerne un procédé pour l'amplification d'un acide nucléique selon l'invention, et plus particulièrement un acide nucléique de séquences SEQ ID N°44 à SEQ
ID N°64 ou un fragment ou un variant de celui-ci contenu dans un
20 échantillon, ledit procédé comprenant les étapes de :

a) mettre en contact l'échantillon dans lequel la présence de l'acide nucléique cible est suspectée avec une paire d'amorces nucléotidiques dont la position d'hybridation est localisée respectivement
du côté 5' et du côté 3' de la région de l'acide nucléique cible dont
25 l'amplification est recherchée, en présence des réactifs nécessaires à la réaction d'amplification ; et

b) détection des acides nucléiques amplifiés.

Pour mettre en oeuvre le procédé d'amplification tel que défini ci-dessus, on aura avantageusement recours à l'une quelconque des
30 amorces nucléotidiques décrites ci-avant.

L'invention a en outre pour objet un nécessaire ou kit pour l'amplification d'un acide nucléique selon l'invention, et plus particulièrement tout ou partie d'un acide nucléique de séquences SEQ
35 ID N°44 à SEQ ID N°64, ledit nécessaire ou kit comprenant :

a) un couple d'amorces nucléotidiques conformes à l'invention, dont la position d'hybridation est localisée respectivement du côté 5' et du côté 3' de l'acide nucléique cible dont l'amplification est recherchée ;

5 b) le cas échéant, les réactifs nécessaires à la réaction d'amplification.

Un tel nécessaire ou kit d'amplification comprendra avantageusement au moins une paire d'amorces nucléotidiques telles que décrites ci-dessus.

10

UTILISATION D'UN ACIDE NUCLEIQUE CODANT UN POLYPEPTIDE THYX COMME MARQUEUR DE SELECTION.

Il existe un besoin constant dans l'état de la technique pour de nouveaux gènes marqueurs de sélection, particulièrement dans des
15 procédés visant à introduire un ou plusieurs gènes d'intérêt dans un organisme hôte, par exemple une cellule hôte.

Les gènes marqueurs de sélection, qui sont portés par la molécule d'ADN codant le ou les gènes d'intérêt que l'on souhaite introduire dans l'organisme hôte ou la cellule hôte, et qui sont en conséquence introduits
20 dans la cellule hôte simultanément aux gènes d'intérêt, permettent de sélectionner les cellules hôtes recombinantes.

Un acide nucléique codant un polypeptide THYX, lorsqu'il est utilisé pour la transfection de cellules hôtes auxotrophes pour la synthèse de la thymidine, constitue un excellent marqueur de sélection d'un
25 événement de transformation, de transfection ou de recombinaison de la cellule hôte. En effet, l'introduction d'un acide nucléique codant un polypeptide THYX simultanément à un ou plusieurs gènes d'intérêt dans la cellule hôte restaure l'autotrophie de la cellule hôte ayant subi avec succès la transfection, la transformation ou la recombinaison.

30 De cette manière, on peut sélectionner des cellules hôtes ayant subi l'événement de transfection, de transformation ou de recombinaison par pression de sélection du milieu, notamment en cultivant les cellules hôtes en l'absence de thymidine. Dans ce cas, seules les cellules hôtes ayant été recombinées survivent en l'absence de thymidine dans le
35 milieu de culture.

De plus, un acide nucléique codant un polypeptide THYX est d'une totale innocuité pour l'environnement.

Des vecteurs comprenant un acide nucléique codant pour un polypeptide THYX comme gène marqueur de sélection peuvent être
5 préparés à partir de vecteurs conventionnels selon des techniques bien connues de l'homme du métier, par exemple à partir des vecteurs préférés selon l'invention.

L'invention a encore pour objet l'utilisation d'un acide nucléique codant pour un polypeptide THYX tel que défini dans la description
10 comme marqueur de sélection d'un événement de transfection, de transformation ou de recombinaison génétique d'une cellule hôte ou d'un organisme hôte.

Elle est également relative à un vecteur de clonage et/ou d'expression comprenant un acide nucléique codant pour un polypeptide
15 THYX tel que défini dans la description à titre de gènes marqueurs de sélection d'un événement de transfection, de transformation ou de recombinaison génétique.

MILIEUX DE REACTION POUR L'ACTIVITE THYMIDYLATE 20 SYNTHASE DE THYX ET LEURS UTILISATIONS, NOTAMMENT DANS DES PROCEDES DE CRIBLAGE

Les travaux des inventeurs sur l'activité de ThyX ont permis d'identifier le mécanisme enzymatique et ainsi d'optimiser les conditions de la réaction en mettant en évidence l'importance de l'addition de
25 certains composés dans le milieu réactionnel.

L'invention a donc aussi pour but l'utilisation de tels milieux et leur application.

Elle vise aussi un milieu de réaction pour l'activité thymidylate synthase de la ThyX caractérisé en ce qu'il comporte des flavines réduites et du CH₂H₄folate.
30

Plus particulièrement, les flavines réduites de ce milieu sont obtenues par réduction *in situ* de flavines oxydées. La concentration en flavines oxydées est alors de 50µM à 1mM, de préférence, 0,5mM. De préférence, les flavines oxydées sont des flavines mononucléotides
35 (FMN) et/ou de dinucléotides flavine adénine (FAD). La réduction des

flavines peut être effectuée par voie chimique, enzymatique, photochimique ou électrochimique et plus particulièrement, avec du NADH (β -nicotinamide adenine dinucléotide) et/ou du NADPH (β -nicotinamide adenine dinucléotide phosphate).

5 Par ailleurs, la concentration en $\text{CH}_2\text{H}_4\text{folate}$ est de $50\mu\text{M}$ à 2mM , de préférence, 1mM .

De manière préférentielle, le milieu peut comporter en outre du dUMP (uridine 5'-monophosphate), à une concentration allant de $1\mu\text{M}$ à $800\mu\text{M}$, de préférence, $500\mu\text{M}$.

10 Lorsque la réduction des flavines est effectuée par du NADH et/ou du NADPH, la concentration en NADH est de $0,1$ à 1mM , et celle en NADPH, de $0,5$ à 5mM .

L'invention vise d'autre part des procédés de criblage. En particulier, elle a pour objet de fournir un procédé de criblage d'un composé antibactérien ou antiviral *in vitro* dans un système acellulaire caractérisé en ce que ledit procédé comprend les étapes suivantes :

- a) préparer un lysat cellulaire à partir d'une culture de cellules exprimant un polypeptide THYX en l'absence d'un polypeptide codé par un gène *thyA* et comportant un milieu selon l'invention;
- b) ajouter au lysat cellulaire obtenu à l'étape a) le composé inhibiteur à tester ;
- c) comparer l'activité thymidylate synthase ThyX respectivement dans le lysat cellulaire obtenu à l'étape a) et dans le lysat cellulaire obtenu à l'étape b) ; et
- d) sélectionner les composés candidats pour lesquels une inhibition de l'activité thymidylate synthase ThyX a été détectée.

30 Dans ce procédé de criblage, on utilise de préférence le dTMP et ^3H comme marqueurs de l'activité thymidylate synthase ThyX.

L'invention couvre également des kits ou nécessaires pour le criblage d'un composé inhibiteur de la thymidylate synthase ThyX caractérisé en ce qu'il comprend :

- 5
- a) une composition comprenant un polypeptide THYX ainsi qu'un milieu selon invention, en solution ou sous forme lyophilisée ;
 - b) facultativement, un ou plusieurs réactifs nécessaires à la quantification de l'activité thymidylate synthase ThyX.

De manière alternative, le kit ou nécessaire pour le criblage d'un composé antibactérien ou antiviral, peut également comprendre :

- 10
- a) un vecteur d'expression recombinant comprenant un acide nucléique codant un polypeptide ThyX sous le contrôle d'un promoteur fonctionnel dans une cellule hôte dans laquelle son expression est recherchée ou une cellule hôte transfectée avec un tel vecteur recombinant ;
 - 15 b) un milieu selon l'invention;
 - c) facultativement, un ou plusieurs réactifs nécessaires à la quantification de l'activité thymidylate synthase ThyX.

D'autres caractéristiques et avantages de l'invention sont donnés, à titre illustratif dans les exemples suivants avec références aux figures.

20 La figure 5 présente des analyses biochimiques de la ThyX de *H. Pylori*.

- 25
- (A) 12% SDS-PAGE et analyses immunoblot de protéine isolée ThyX de *H. pylori*. 1,5 (Coomassie) et 0,3 (anti-V5) µg de protéine pure ont été détectés respectivement par coloration au Bleu de Coomassie ou utilisation d'anticorps monoclonaux contre un épitope anti-V5 (Invitrogen), par détection chimioluminescente. Le poids moléculaire attendu de la ThyX de *H. pylori* est de 31,5 kDa
- 30
- (B) Les analyses spectroscopiques de la ThyX de *H. pylori* indiquent que ThyX est une flavoprotéine. Le spectre absolu de 10 µM d'enzyme isolée (ligne continue) et de co-facteur après sa libération de la protéine (ligne
- 35

pointillée). La fenêtre montre le spectre pour l'enzyme oxydée et la protéine dithionique réduite.

(C) L'activité de libération du Tritium de la protéine ThyX purifiée de *H. pylori* a été enregistrée en présence (+MTHF) et en l'absence (-MTHF) de CH₂H₄tétrahydrofolate .

La figure 6 illustre l'activité de formation du dTMP de la protéine ThyX de *H. pylori*. Les réactions enzymatiques ont été réalisées comme décrit dans le tableau 1, en utilisant du dUMP marqué en position 6. Les produits de réaction ont été analysés en utilisant une colonne à phase inversée C 18 par élution isocratique, en utilisant 10 mM de tampon de phosphate. Les durées d'élution pour dUMP et dTMP ont été déterminées en utilisant les références authentiques.

L'invention est en outre illustrée, sans pour autant être limitée, aux exemples suivants :

EXEMPLES

EXEMPLE 1 : Expression du polypeptide THYX de *Helicobacter pylori* dans *E.coli*.

Les ADNs comprenant le cadre de lecture ouvert codant pour le polypeptide THYX de *Helicobacter pylori* (souche 26.695) (séquence SEQ ID n°21) à été obtenue par amplification PCR à l'aide des amorces spécifiques de séquences SEQ ID N°38 et SEQ ID N°39 à partir du clone GHPEH26 publiquement accessible auprès de l'AMERICAN TYPE CULTURE COLLECTION sous le numéro d'accès n°628.507.

L'ADN comprenant le cadre de lecture ouvert codant pour le polypeptide THYX de *Pyrococcus. abyssi* (souche ORSAY) (SEQ ID n°12) a été préparé par amplification PCR à l'aide d'amorces spécifiques de séquences SEQ ID N°40 et SEQ ID N°41 à partir de l'ADN chromosomique de *H. pylori* (HP 1533).

L'ADN comprenant le cadre de lecture ouvert codant pour le polypeptide THYX de *Campylobacter jejuni* (souche NCTC 11168) (SEQ ID N°27) a été préparé par amplification PCR à l'aide d'amorces spécifiques de séquences SEQ ID N°42 et SEQ ID N°43 à partir de l'ADN génomique publiquement accessible auprès de l'ATCC (Aerican Culture Collection) sous le n° d'accès 7008199.

Les produits d'amplification PCR ont été clonés dans le vecteur pBAD TOPO TA activé par la topoisomérase-I, commercialisée par la Société Invitrogen, qui permet une expression strictement régulée du gène d'intérêt artificiellement inséré dans le vecteur, dans la bactérie *E.coli*.

Les clones de *E.coli* ont été caractérisés par séquençage des inserts d'ADN contenus dans le vecteur pBADTOPOTA.

Les poids moléculaires théoriques des polypeptides THYX de *H. pylori* (HP1533), de *P.abyssii* (PAB 0861) et de *C. jejuni* (NCTC 11168) y compris un activateur de traduction amino-terminal et des épitopes carboxy-terminaux V5 et histidine, sont respectivement de 31,5 kDa, de 33,7 kDa et de 28 kDa.

Les résultats présentés dans la figure **2A** montrent que l'expression du polypeptide THYX est induite en présence de 0,2% d'Arabinose dans le milieu de culture, comme cela a été détecté en utilisant des anticorps monoclonaux anti-V5 commercialisés par la Société Invitrogen et utilisés selon les recommandations du fabricant.

Puis on a testé la capacité des polypeptides THYX de *H.pylori* et de *P.abyssii* à permettre la croissance de la souche de *E.coli* X2913 ($\Delta thya572$), auxotrophe pour la thymidine, à se multiplier en l'absence de thymidine.

La capacité de croissance des souches de *E.coli* exprimant les polypeptides THYX de *H.pylori* et de *P.abyssii* a été déterminée après culture des cellules de *E.coli* recombinantes pendant 3 à 4 jours en présence ou en l'absence d'une concentration de 0,2% d'Arabinose sur un milieu agar minimal M9 dépourvu de thymidine (Michaels et al., 1990).

Les résultats sont présentés dans la figure **2B**.

Comme le montrent les résultats de la figure **2B**, l'expression du gène *thyX* de *H.pylori* dans la souche de *E.coli* X2193 qui a été induite

par l'Arabinose, a permis de compléter cette souche de *E.coli* initialement déficiente en activité thymidylate synthase et restaurer la prototrophie de cette souche de *E.coli* pour la thymidine.

En revanche, le gène *thyX* de *P.abysyi*, après induction par l'Arabinose, n'a pas permis de compléter la souche de *E.coli* X 2193, ce qui reflète l'incapacité d'une protéine hyperthermophile d'être fonctionnelle dans un hôte mésophile.

Les deux souches de *E.coli* recombinantes étaient néanmoins capables de se multiplier dans un milieu agar minimal en présence de thymidine à la concentration finale de 50µg/ml.

Le clone d'*Helicobacter pylori* GHPEH26 contient un insert d'ADN de 1,5421 kb correspondant aux nucléotides 1613133-1611.613 de l'ORF référence AE00511, locus 10, HP 1533, référencé dans les bases de données comme DNA seq. Acc : AE000 511.

Le clone GHPEH26 est commercialisé par la Société TIGR/ATCC Microbial Genome Special Collection. Le gène *thyX* de *H.pylori* obtenu par PCR à partir de ce clone représente un fragment d'ADN d'une longueur de 693 pb.

Les résultats de gel d'électrophorèse SDS PAGE illustrés à la figure 3 montrent que la souche de *E. coli* transformée avec l'ADN de *Campylobacter jejuni* produit un polypeptide THYX du poids moléculaire attendu.

De plus, les résultats d'immunoempreintes illustrés à la figure 4 montrent que la protéine THYX de *Campylobacter jejuni* peut être efficacement purifiée sur une colonne de Nickel Ni-NTA.

Les résultats décrits ci-dessus représentent la première démonstration expérimentale que la réaction de conversion du dUMP en dTMP peut être réalisée dans la cellule par une thymidylate synthase autre qu'une thymidylate synthase codée par un gène *thyA*, c'est-à-dire par la thymidylate synthase codée par un gène *thyX*.

EXEMPLE 2 :**Identification d'une famille de polypeptides THYX**

Par l'analyse des séquences d'environ 50.000 gènes référencée
5 dans la base de données des « clusters » de protéines orthologues
(Tatuson et al., 2000) par une méthode itérative de recherche de
similarité (Altschul et al., 1997) il a été montré que les séquences *THYX*,
similaires à la séquence *THY1* de *H. pylori* (HP 1533) étaient la seule
famille de gènes présentant une distribution mutuellement exclusive avec
10 *thyA*, à l'exception unique de *Mycobacterium tuberculosis* qui comprend
simultanément un gène *thyX* et un gène *thyA*.

La recherche itérative de similarité a été réalisée à l'aide du
programme itératif PSI-BLAST (Version 2.0) en utilisant diverses
15 séquences *THYA* comme séquences « appâts ». Des sélections
(« hits ») ayant une valeur attendue inférieure à 1.10^{-5} ont été
considérées comme statistiquement significatives. Une valeur de seuil
pour le recrutement des alignements dans les itérations successives était
de 0,02.

20 Des résultats non exhaustifs de l'analyse de similarité protéique ci-
dessus sont représentés dans les tableaux 1 et 2 ci-après.

Tableau 1

	Comparaison	
	comparaison (%)	Similarité (%)
<i>A. aeolicus</i>	26	47
<i>A. pernix</i>	25	38
<i>B. anthracis</i>	30	47
<i>B. burgdorferi</i>	23	38
<i>Bacteriophage phi-C31</i>	25	40
<i>B. lactis</i>	23	42
<i>C. difficile (partial)</i>	40	59
<i>C. diptherie</i>	25	45
<i>C. glutamicum</i>	23	42
<i>C. jejuni</i>	56	72
<i>Chlamydia sp.</i>	22	36
<i>Chlorella virus</i>	20	32
<i>D. discoideum</i>	23	41
<i>D. vulgaris</i>	23	46
<i>G. sulfuureducens</i>	28	48
<i>Gp16 (bacteriophage)</i>	24	39
<i>Gp48 (bacteriophage)</i>	25	44
<i>H. pylori (HP1533)</i>	100	100
<i>H. pylori (jhp1421)</i>	96	97
<i>H. salinarium</i>	22	43
<i>Halobacterium sp. NRC-1</i>	22	43

- 5 % Les pourcentages d'identité et de similarité représentés dans le tableau 1 ci-dessus ont été obtenus après recherches itératives en utilisant les logiciels BLAST ou Psi-BLAST, en utilisant exclusivement les paramètres par défaut. Les résultats non exhaustives du Tableau 1 montrant que de nombreuses bactéries bactériophages et virus
- 10 possèdent une copie d'un gène codant pour un polypeptide THYX..

Tableau 1(suite)

	Comparaison	
	comparai son (%)	Similarité (%)
<i>M. avium</i>	24	41
<i>M. bovis</i>	24	41
<i>M. leprae</i>	21	40
<i>M. tuberculosis</i>	23	39
<i>P. abyssi</i>	23	47
<i>P. furiosus</i>	27	52
<i>P. horikoshii</i>	25	46
<i>R. capsulatus</i>	18	38
<i>R. prowazekii</i>	22	41
<i>Roseophage S101</i>	18	36
<i>S. coelicolor</i>	17	33
<i>S. solfaraticus</i>	26	45
<i>Synechocystis sp.</i>	19	34
<i>T. acidophilum</i>	21	42
<i>T. denticola</i>	28	48
<i>T. maritime</i>	28	46
<i>T. pallidum</i>	22	37
<i>T. volcanium</i>	20	34

Tableau 2

Espèces	thyA ¹	thyX	DHFR	tdk	upp	Commentaires
Bactéries :						
Campylobacter jejuni	-	+	-	-	+	Intoxications alimentaires
Helicobacter pylori	-	+	-	-	-	Formation des ulcères stomacaux
Rickettsia prowazekii	-	+	-	-	-	agent causal du typhus
Borrelia burgdorferi	-	+	-	-	-	impliqué dans la maladie de Lyme
Treponema pallidum	-	+	-	-	+	agent causal de la Syphilis
Chlamydia (3 espèces)	-	+	+	-	+	Pathogène intracellulaires obligatoires
Mycobacterium tuberculosis	+	+	+	-	+	tuberculose
Thermotoga maritima	-	+	+	+	+	thermophile
Archaeobactéries						
Pyrococcus abyssi	-	+	-	-	+	hyperthermophile
Pyrococcus horikoshii	-	+	-	-	-	hyperthermophile
Eucaryotes:						
Dictyostelium discoideum	N.A	+	N.A	N.A	N.A	pas de séquence génomique complète accessible
Virus :						
Virus à ADN bactériens et eucaryotes (5 espèces – Lesquelles ?	-	+	N.A	N.A	N.A	

- ¹thyA, gène codant la thymidylate synthase nécessaire pour la synthèse de NOVO du dTMP ; thyX, nouvelle famille de gène impliquée dans la biosynthèse des pyrimidines ; DHFR, dihydrofolate reductase essentielle pour le recyclage d'un donneur de méthyl essentiel dans la synthèse du dTMP ; tdk thymidine kinase requise pour le sauvetage de la thymidine exogène qui est ensuite convertie en dTMP ; UPP, uracylphosphoribosyl transferase (UPRTase) requise pour le sauvetage de l'uracyle.

Dans le tableau 2 ci-dessus, on présente les résultats des tests de la présence notamment des gènes *thyX* et *thyA* dans certains des organismes listés dans le tableau 1.

Les résultats présentés dans le tableau 1 ci-dessus montrent que
5 les gènes *thyX* sont présents à la fois dans les bactéries, les bactériophages et les virus. Parmi les eucaryotes, l'organisme *D.dyscoidum* est le seul organisme a porté une copie d'un gène *thyX*.

La présence unique d'un gène *thyX* dans de nombreuses bactéries et le virus à ADN eucaryote ainsi que dans de nombreuses
10 bactéries pathogènes pour l'homme en fait une cible privilégiée pour des composés antibactériens ou antiviraux n'interférant pas avec la voie métabolique de la thymidylate synthase THYA présente chez l'homme.

De plus, parmi les organismes listés dans le tableau 1 ci-dessus, nombre d'entre eux ne comprennent pas de gènes DHFR codant une
15 dihydrofolate reductase, qui est requise pour le recyclage d'un cofacteur essentiel du métabolisme du thymidylate, le CH₂H₄-folate.

EXEMPLE 3 :

Mutagenèse dirigée du gène THYX de *Helicobacter pylori*

20

A. MATERIELS ET METHODES

Pour l'obtention des mutants 1 à 6 décrits dans la Section « Résultats », la mutagenèse dirigée a été réalisée avec le kit « Quick
25 Change™ » commercialisé par la Société Stratagène, selon les recommandations du fabricant.

Pour l'obtention des mutants 7 et 8 décrits dans la Section « Résultats », la mutagenèse dirigée a été réalisée avec le kit « QuickChange™ Multi Site-Directed Mutagenesis Kit » commercialisé
30 par la Société Stratagène, selon les recommandations du fabricant.

L'ADN de départ (« Template ») utilisé est le plasmide pBAD TOPO, commercialisé par la Société InVitrogen, dans lequel a été inséré le gène *thyX* de *Helicobacter pylori* 26695.

Le résidu sérine 107 (codon « AGT ») conservé a été remplacé
35 respectivement par les résidus cystéine (codon « TGC ») ou alanine

(codon « GCT »), en utilisant les oligonucléotides mutagéniques appropriés.

Le résidu Tyrosine 110 (codon « TAC ») a été remplacé respectivement par les résidus thréonine (codon « ACT ») et
5 phénylalanine (codon « TTC »), en utilisant les oligonucléotides mutagéniques appropriés.

Le résidu Glutamate 142 (codon « GAA ») a été remplacé respectivement par les résidus Alanine (codon « GCT ») et Aspartate (codon « GAT »), en utilisant les oligonucléotides mutagéniques
10 appropriés.

Le résidu Histidine 71 (codon « CAT ») a été remplacé par le résidu Glutamine (codon « CAA »), en utilisant les oligonucléotides mutagéniques appropriés.

Le résidu Glutamate 205 (codon « GAA ») a été remplacé par le
15 résidu Leucine (codon « TTA »), en utilisant les oligonucléotides mutagéniques appropriés.

B. Résultats

20 Les expériences de mutagenèse ont montré que les bactéries de *E.coli* Chi2193 transformées par les vecteurs pBAD TOTO TA contenant respectivement des inserts d'ADN codant pour la protéine THYX dans lequel le résidu d'acide aminé sérine conservé a été remplacé respectivement par un résidu de cystéine ou d'alanine exprimait une
25 protéine THYX mutée contenant une flavine.

En revanche, les deux bactéries transformées mutantes ont perdu leur capacité de restauration de la prototrophie pour la thymidine.

Ces résultats montrent que le résidu d'acide aminé sérine conservé est essentiel pour l'activité catalytique de la protéine THYX.

30 De plus, la capacité du gène *thyX* de compléter les bactéries pour l'activité thymidylate synthase est perdue lorsque les résidus Tyrosine 110 et Glutamate 142 sont mutés.

Une mutation par remplacement de l'histidine 71 par la Glutamine conduit à la production d'un polypeptide capable d'interagir avec le co-

facteur flavine, mais qui ne complémente pas les bactéries pour l'activité thymidylate synthase.

Le remplacement du résidu Glutamate 205 par un résidu Leucine bloque la production du polypeptide THYX.

5 Dans des expériences additionnelles, le gène *thyX* de la bactérie *Rhodobacter capsulatus* a été inactivé (par « knock out »). Les bactéries déficientes en protéine THYX ont exprimé un caractère phénotypique d'auxotrophie pour la thymidine, c'est à dire qu'elles ne se multiplient qu'en présence de thymidine. Ces derniers résultats montrent clairement
10 que les protéines THYX sont essentielles pour le métabolisme de la thymidine. Ces expériences montrent que les protéines *thyX* sont essentielles pour la croissance bactérienne en l'absence de thymidine, ce qui prouve l'utilité du gène *thyX* ou du polypeptide THYX comme cible thérapeutique.

15

EXEMPLE 5 : Identification du mécanisme de réaction de ThyX

Pour identifier les réactions biochimiques catalysées par les protéines ThyX, une THYX de *H. pylori* portant un marqueur d'histidine sur sa terminaison carboxy a été purifiée à partir d'extraits acellulaires de
20 souche χ 2913 d'*E. coli* soumis à une induction par de l'arabinose en chromatographie d'affinité de nickel immobilisé .

Des cadres de lecture ouverts codant pour des protéines hypothétiques, apparaissant maintenant correspondre aux gènes *thyX* de *P. abyssi* (PAB0861) et *H. pylori* (HP1533), ont été obtenus par PCR, en
25 utilisant des amorces spécifiques et de l'ADN chromosomique de *P. abyssi* et le clone GHPEH26 (American Type Culture Collection n° 628507), respectivement, comme matrices. Les produits PCR ont été clonés dans le vecteur I-activé pBAD TOPO® TA de topoisomérase (Invitrogen), permettant l'expression strictement contrôlée du gène dans
30 *E. coli*. Tous les clones plasmidiques ont été confirmés par séquençage d'ADN. Les masses moléculaires attendues de PAB0861 et HP1533, comprenant une séquence activatrice de traduction amino-terminale et une région V5 carboxy-terminale et des épitopes d'hexahistidine, sont respectivement de 33,7 et de 31,5 kDa. L'expression de la protéine a été
35 induite avec 0,2% de L-arabinose. Les protéines exprimées ont été

détectées par utilisation d'anticorps monoclonaux V5-spécifiques (Invitrogen) selon les recommandations du fabricant.

La protéine ThyX à site antigénique marqué de *H. pylori* et biologiquement active a été purifiée à partir de 200ml de cultures de pGL2/*E. coli* X2913 [Δ thyA (tableau 3)] après 2 heures d'induction par 0,2% de L-arabinose. Un kit d'expression QIA (Qiagen) sous conditions endogènes standard a été utilisé pour la purification comme indiqué par le fabricant, comprenant 10% (volume/volume) de glycérol dans tous les tampons. Les échantillons de protéine obtenus ont été dialysés contre 50mM d'un tampon de phosphate, à pH 7,4 et de glycérol à 10% (volume/volume) après élution pour éliminer l'imidazole. La concentration de protéine dans des échantillons purs a été estimée par lecture A_{280} , justifiant l'absorbance A_{280} du cofacteur de flavine à liaison non covalente, à $35560 \text{ M}^{-1}\text{cm}^{-1}$, calculée par rapport à la séquence connue d'acides aminés, et a été utilisée pour l'apoprotéine ThyX de *H. pylori*.

Les préparations de protéines obtenues (pureté > 95%) contenaient habituellement 1 à 2 mg/ml de protéine, avec une masse moléculaire d'environ 31 kDa sur gels SDS-PAGE (la masse moléculaire attendue de ThyX de *H. pylori* est de 31,5 kDa) (Figure 5A), et avaient une couleur jaune clair. La chromatographie par exclusion de taille au Superdex 200 utilisant des marqueurs de poids moléculaire standards, a révélé une masse moléculaire endogène de 111 kDa ($r=0,9874$) pour cette protéine, suggérant que sa forme active pouvait correspondre à un homotétramère. Des analyses spectroscopiques de la protéine isolée (oxydée) ont montré des caractéristiques d'absorbance typiques d'une flavoprotéine (figure 5B), avec de larges pics à 447,5 et 375 nm. Ces pics d'absorption se sont révélés absents dans l'enzyme réduite au dithionite. Des caractéristiques d'absorption similaires ont été trouvées pour le cofacteur après sa libération de la protéine par dénaturation à la chaleur à 80°C pendant 5 minutes. Par chromatographie HPLC, le cofacteur associé à la ThyX de *H. pylori* a été identifié comme un FAD (flavine-adénine dinucléotide). Il a été estimé que les différentes préparations d'enzyme de ThyX de *H. pylori* contiennent de 0,4 à 0,5 molécules de FAD par monomère. En tout, ces propriétés spectroscopiques indiquent

que la ThyX de *H. pylori* est une flavoprotéine et/ou utilise les cofacteurs de flavine dans la catalyse.

Dans la formation de dTMP, la perte de tritium du [5-H]dUMP dans le solvant est un intermédiaire obligatoire, permettant de quantifier l'activité de la thymidylate-synthétase après élimination des nucléotides radioactifs des mélanges réactionnels (ROBERTS, 1966). Pour aborder le mécanisme biochimique, par lequel la ThyX peut circonvenir l'exigence pour la ThyA dans la synthèse *de novo* du thymidilate, la protéine ThyX purifiée a été utilisée dans la même mesure.

10

EXEMPLE 6: Optimisation des conditions de réaction de ThyX

Le N⁵,N¹⁰-CH₂H₄folate a été formé de manière non-enzymatique, par incubation de 2mM d'acide tétrahydrofolique (Sigma®) avec 100 mM de β-mercaptoéthanol et 20 mM de formol pendant 30 minutes dans l'obscurité et à température ambiante. Les dosages de libération de tritium avec une enzyme purifiée ont été réalisés dans 50 mM de Tris-Cl, à pH 7,9, comprenant une préparation de 1 mM de CH₂H₄folate obtenue comme décrit ci-dessus. Les réactions de contrôle ont été réalisées sous des conditions analogues, sans acide tétrahydrofolique. Les réactions à 50 µl ont été démarrées par addition de 6 µM de [5-³H]dUMP, l'activité spécifique à 16,2 Ci/mmol (Amersham) et arrêtées après 60 minutes à 37°C par deux extractions avec 250µl de charbon actif [10% (poids/volume) de Norit A] dans 2% d'acide trichloroacétique pour éliminer les nucléotides des mélanges réactionnels. La radioactivité rémanente dans le surnageant a été déterminée selon MYLLYKALLIO (2000).

Comme l'indique la figure 5C, la ThyX catalyse *in vitro* la libération de tritium du (5-³H)dUMP en fonction de la concentration en protéine et de manière CH₂H₄folate dépendante, ce qui démontre l'activité biochimique des protéines ThyX. Des conditions optimisées de réaction sont rapportées dans le tableau 3.

30

Tableau 3. Optimisation des conditions de réaction pour la protéine ThyX d'*H. pylori*

Conditions du test	nmol de ^3H libéré / mg de protéine (incubation de 60 minutes)
ThyX de <i>H. pylori</i> :	
Complet	63.0 (100%)
Complet, - protéine	0.7 (1.1%)
Complet, - H_4 folate	0.8 (1.2%)
Complet, - NADH, - NADPH, -FMN	2.2 (3.4%)
Complet, -FMN	2.70 (4.3%)
Complet, - NADPH	37.3 (59.2%)
Complet, - NADH	19.0 (30.1%)
Complet, + 500 μM dUMP	9.3 (14.8%)
Complet, + 500 μM UMP	64.4 (102.2%)

- 5 ¹ Le test complet contient 50 mM de Tris-HCl, pH 7.0, 1 mM de
préparation de CH_2H_4 folate, 10 mM de MgCl_2 , 2 mM de NADPH, 1 mM
de NADH, 0.5 mM de FMN et 9 μM de ^3H -dUMP (activité spécifique
1,7357 Ci/mmol). La préparation de CH_2H_4 folate est obtenue par
10 incubation 30 minutes dans l'obscurité de 2 mM de H_4 -folate, de 96 mM
de 2-mercaptoethanol et de 42 mM de formaldéhyde et de 50 mM de
Tris-HCl. Dans la réaction, le FMN peut être remplacée efficacement par
du FAD (flavine adénine di-nucléotide). Dans ces conditions de réactions,
20 20 μM de dUMP et 100 μM de CH_2H_4 folate sont suffisant pour saturer
l'activité de libération de tritium de la protéine ThyX de *H. pylori* lors d'une
15 incubation de 60 minutes.

De manière surprenante, il a été trouvé que l'addition de
nucléotides réduits de flavine augmente drastiquement l'activité de
libération de tritium de la ThyX de *H. pylori* ($\approx 0,01 \mu\text{mol}$ de H_2O formée
20 par min et par mg de protéine, tel que mesuré lors d'expériences
supplémentaires de temps d'absorption).

De même, il a été démontré par de simples expériences de
compétition, que l'activité de la ThyX n'utilise pas UMP comme substrat
(tableau 3) et que cette activité est inhibée par des concentrations

micromoléculaires de dTMP. L'activité libératoire de tritium de la ThyX de *H. pylori* est directement liée à la formation de dTMP (figure 6).

Les résultats expérimentaux ont donc clairement établi que les protéines ThyX fonctionnent comme une thymidilate-synthétase dUMP dépendante (figures 5, 6).

De plus, on a aussi montré que le fluoro-dUMP agit en tant qu'inhibiteur des protéines THYX. Ces résultats montrent que les déoxynucléotides monophosphates peuvent être utilisés pour identifier de nouveaux inhibiteurs des THYX.

Pour une application pratique, ces résultats montrent aussi que les tests sont aussi utiles pour le ciblage des protéines THYX.

Par conséquent, alors que la catalyse par ThyX est dépendante des nucléotides de flavine réduite, la ThyA utilise des électrons du H₄folate pour la formation du groupe fonctionnel méthyle. Le CH₂H₄folate fonctionne dans la réaction catalysée par ThyX uniquement comme donneur d'un carbone, provoquant ainsi la formation de H₄folate comme produit de la réaction. Ce mécanisme de réaction explique clairement pourquoi la dihydrofolate-réductase n'est pas indispensable à la formation de thymidylate par les protéines ThyX (tableau 4).

Tableau 4

Espèces	thyA	DHFR	tdk	thyX	Commentaires
Bactéries :					
Campylobacter Jejuni	-	-	-	+	Intoxications Alimentaires
Helicobacter pylori	-	-	-	+	Formation des ulcères stomacaux
Rickettsia prowazekii	-	-	-	+	Agent causal du typhus
Borrelia burgdorferi	-	-	-	+	Impliqué dans la maladie de Lyme
Treponema pallidum	-	-	-	+	Agent causal de la Syphilis
Chlamydia (3 espèces)	-	+	-	+	Pathogène intracellulaire s obligatoires
Mycobacterium tuberculosis	+	+	-	+	Tuberculose
Archaeobactéries :					
Pyrococcus abyssi	-	-	-	+	Hyperthermophil e
Pyrococcus horikoshii	-	-	-	+	Hyperthermophil e
Sulfolobus solfaraticus	-	-	-	+	Hyperthermophil e
Eucaryotes :					
Dictyostellium	N.A		N.A.	+	Pas de séquence génomique complete
Virus :					
Virus à ADN bactériens et eucaryotes (5 espèces)	-		N.A	+	

5 Tdk : thymidine kinase, nécessaire à la récupération de thymidine exogène.

DHFR : dihydrofolate reductase.

On notera que, dans la séquence de Chlamydia, le résidu Sérine est soit absent, soit localisé à l'extrémité aminoterminal de la protéine.

10

Les différences dans le mécanisme enzymatique des deux différentes classes de thymidylate-synthétases sont aussi dues à

l'absence de motifs de séquence essentiels pour la catalyse dans les protéines ThyA et ThyX. Ces données ont permis d'identifier des analogues de dUMP, de dTMP, de folate et de nucléotides de flavine comme candidats idéaux pour des composés clés pour identifier des
5 composés nouveaux inhibant l'activité de la ThyX.

REFERENCES

- ALSTCHUL SF, Madden TL, SCHAFFER AA, ZHANG J, ZHANG Z, Miller W. & LIPMAN DJ *Nucleic Acids Res.* 25, 3389-3402 (1997).
- 5 Beaucage et al., *Tetrahedron Lett* 1981, **22**: 1859-1862
- Behr, *Bioconjugate Chem.*, 5, 382-389 (1994)
- Brown EL, Belagaje R, Ryan MJ, Khorana HG, *Methods Enzymol* 1979;**68**:109-151
- CARRERAS, C.W. and SANTI, D.V., 1995, *Annu. Rev. Biochem.* vol.64: 721-763.
- 10 DYNES J.L. and FIRTEL R.A., 1989, *Proc. Natl. Acad. Sci. USA*, vol.86: 7966-7970.
- Fuller S.A. et al., 1996, *Immunology in Current Protocols in Molecular Biology*, Ausubel et al. Eds, John Wiley & Sons, Inc., USA
- 15 Green et al., *Ann. Rev. Biochem.* **55**:569-597 (1986)
- Houben Weyl, 1974, in *Meuthode der Organischen Chemie*, E. Wunsch Ed., Volume 15-I et 15-II,
- Izant JG, Weintraub H, *Cell* 1984 Apr;**36**(4):1007-15
- Koch Y., 1977, *Biochem. Biophys. Res. Commun.*, **74**:488-491
- 20 Kohler G. and Milstein C., 1975, *Nature*, **256** : 495.
- Kozbor et al., 1983, *Hybridoma*, **2**(1):7-16.
- Leger OJ, et al., 1997, *Hum Antibodies*, **8**(1): 3-16
- Martineau P, Jones P, Winter G, 1998, *J Mol Biol*, **280**(1):117-127
- Merrifield RB, 1965a, *Nature*, **207**(996): 522-523.
- 25 Merrifield RB., 1965b, *Science*, **150**(693): 178-185.
- Michaels, M.L. Kim, C.W. Matthews D.A. & Miller, J.H. *Proc. Natl. Acad. Sci. USA* 87, 3957-3961 (1990).
- Myllykallio, H., et al., *Science* **288**, 2212 (2000).
- Narang SA, Hsiung HM, Brousseau R, *Methods Enzymol* 1979;**68**:90-98

- PAPAMICKAEL D. 2000 , STEM. Cells, vol.18. 166-175.
- Reimann KA, et al., 1997, AIDS Res Hum Retroviruses. **13**(11): 933-943
- Ridder R, Schmitz R, Legay F, Gram H, 1995, Biotechnology (N Y), **13**(3):255-260
- 5 Roberts, D., *Biochemistry* **5**, 3546 (1996).
Rossi et al., *Pharmacol. Ther.* **50**:245-254, (1991)
- Sambrook, J. Fritsch, E. F., and T. Maniatis. 1989. Molecular cloning: a laboratory manual. 2ed. Cold Spring Harbor Laboratory, Cold spring Harbor, New York.
- 10 Schofield, Brit. Microencapsulated. Bull., 51(1):56-71 (1995)
- Sczakiel G. et al., 1995, *Trends Microbiol.*, 1995, **3**(6):213-217
- Carreras C.W. and Santi D.V., ANNUAL Review on Biochemistry, 1995, 64: 721-762.
- 15 Poole A., Penny D. and Sjöberg B.M.____Nature Reviews, Mol. Cell. Biol. 2001, 2: 147-151
- Papamichael D., 1999, The Oncologist, 4(6): 478-487.
- TATUSOV R.L. GALPERIN, M.Y. Natale, D.A. & KOONIN, E.V. Nucleic Acids Res. 28, 33-36 (2000).

REVENDICATIONS

1. Utilisation d'un polypeptide THYX comprenant la séquence d'acides aminés suivante :
X₁HR(X)₇S, dans laquelle :
- 5 - X₁ est l'acide aminé R (Arginine ou Arg) ou K (Lysine ou Lys), et
 - (X)₇ est une chaîne de sept acides aminés consécutifs dans laquelle chaque X représente, indépendamment l'un de l'autre, l'un quelconque des 20 acides aminés naturels,
dans un procédé de synthèse *in vitro* de la thymidine 5'-monophosphate
10 (dTMP).
2. Utilisation selon la revendication 1, caractérisée en ce que le polypeptide THYX est choisi parmi les polypeptides comprenant les séquences d'acides aminés SEQ ID N°1 à SEQ ID N°37.
- 15
3. Utilisation d'un acide nucléique codant pour un polypeptide THYX selon l'une des revendications 1 ou 2, pour produire ledit polypeptide THYX.
- 20
4. Utilisation selon la revendication 3 , caractérisée en ce que l'acide nucléique est choisi parmi les acides nucléiques comprenant les séquences nucléotidiques SEQ ID n°44 à SEQ ID n°64 .
5. Procédé de criblage d'une molécule ou d'une substance
25 candidate interagissant avec un polypeptide selon l'invention, ladite méthode comprenant les étapes de :
- a) mettre en contact un polypeptide selon l'une des revendications 1 ou 2 avec la substance ou molécule candidate à tester ;
 b) détecter les complexes éventuellement formés entre ledit
30 polypeptide et ladite substance ou molécule candidate.
6. Nécessaire ou kit pour le criblage d'une molécule ou d'une substance candidate interagissant avec un polypeptide tel que défini dans l'une des revendications 1 ou 2, ledit nécessaire comprenant :

a) un polypeptide THYX tel que défini dans l'une des revendications 1 ou 2 ;

b) le cas échéant, des moyens nécessaires à la détection du complexe formé entre ledit polypeptide et la molécule ou substance candidate.

7. Procédé de criblage d'un composé antibactérien ou antiviral *in vitro* dans un système acellulaire caractérisé en ce que ledit procédé comprend les étapes suivantes :

10 a) préparer un lysat cellulaire à partir d'une culture de cellules exprimant un polypeptide THYX tel que défini dans l'une des revendications 1 ou 2 en l'absence d'un polypeptide codé par un gène *thyA* ;

b) ajouter au lysat cellulaire obtenu à l'étape a) le composé inhibiteur à tester ;

15 c) comparer l'activité thymidylate synthase respectivement dans le lysat cellulaire obtenu à l'étape a) et dans le lysat cellulaire obtenu à l'étape b) ; et

d) sélectionner les composés candidats pour lesquels une inhibition de l'activité thymidylate synthase a été détectée.

20

8. Kit ou nécessaire pour le criblage d'un composé inhibiteur de la thymidylate synthase caractérisé en ce qu'il comprend :

25 a) une composition comprenant un polypeptide THYX tel que défini dans l'une des revendications 1 ou 2 en solution ou sous forme lyophilisée ;

b) facultativement, un ou plusieurs réactifs nécessaires à la quantification de l'activité thymidylate synthase.

9. Utilisation d'un acide nucléique codant pour un polypeptide THYX tel que défini dans l'une des revendications 1 ou 2 dans un procédé de criblage de composés antibactériens ou antiviraux.

30

10. Kit ou nécessaire pour le criblage d'un composé antibactérien ou antiviral, caractérisé en ce qu'il comprend :

a) un vecteur d'expression recombinant comprenant un acide nucléique codant un polypeptide THYX tel que défini dans l'une des revendications 1 ou 2 sous le contrôle d'un promoteur fonctionnel dans une cellule hôte dans laquelle son expression est recherchée ou une
5 cellule hôte transfectée avec un tel vecteur recombinant ;

b) facultativement, un ou plusieurs réactifs nécessaires à la quantification de l'activité thymidylate synthase.

11. Utilisation d'un oligonucléotide antisens s'hybridant
10 spécifiquement avec un acide nucléique codant pour un polypeptide THYX tel que défini selon l'une des revendications 1 ou 2 pour inhiber ou bloquer la transcription et/ou la traduction dudit acide nucléique.

12. Utilisation d'un anticorps dirigé contre un polypeptide THYX
15 tel que défini selon l'une des revendications 1 ou 2 pour inhiber ou bloquer l'activité thymidylate synthase dudit polypeptide.

13. Composition pharmaceutique antibactérienne ou antivirale comprenant, à titre de principe actif un oligonucléotide antisens hybridant
20 spécifiquement avec un ARN messager codant pour un polypeptide THYX tel que défini dans l'une des revendications 1 ou 2, en association avec un ou plusieurs excipients physiologiquement compatibles.

14. Utilisation d'un oligonucléotide antisens hybridant
25 spécifiquement avec l'ARN messager codant pour un polypeptide THYX tel que défini dans l'une des revendications 1 ou 2 pour la fabrication d'un médicament antibactérien ou anti-viral.

15. Composition pharmaceutique antibactérienne ou antivirale
30 comprenant, à titre de principe actif, un anticorps dirigé spécifiquement contre un polypeptide THYX tel que défini dans l'une des revendications 1 ou 2 en association avec un ou plusieurs excipients physiologiquement compatibles.

16. Utilisation d'un anticorps dirigé spécifiquement contre un polypeptide THYX tel que défini dans l'une des revendications 1 ou 2 pour la fabrication d'un médicament antibactérien ou antiviral.

5 17. Utilisation d'une sonde ou d'une amorce nucléique hybridant avec un acide nucléique codant pour un polypeptide THYX, tel que défini dans l'une des revendications 1 ou 2, dans un procédé de détection d'une bactérie ou d'un virus.

10 18. Utilisation d'un acide nucléique codant pour un polypeptide THYX, tel que défini dans l'une des revendications 1 ou 2 comme marqueur de sélection d'un événement de transfection, de transformation ou de recombinaison génétique d'une cellule hôte ou d'un organisme hôte.

15 19. Milieu de réaction pour l'activité thymidylate synthase de la ThyX caractérisé en ce qu'il comporte des flavines réduites et du $\text{CH}_2\text{H}_4\text{folate}$.

20 20. Milieu selon la revendication 19, caractérisé en ce que les flavines réduites sont obtenues par réduction *in situ* de flavines oxydées.

25 21. Milieu selon l'une quelconque des revendications 19 ou 20, caractérisé en ce que la concentration en flavines oxydées est de $50\mu\text{M}$ à 1mM , de préférence, $0,5\text{mM}$.

 22. Milieu selon l'une quelconque des revendications 19 à 21, caractérisé en ce que la concentration en $\text{CH}_2\text{H}_4\text{folate}$ est de $50\mu\text{M}$ à 2mM , de préférence, 1mM .

30 23. Milieu selon l'une quelconque des revendications 19 à 22, caractérisé en ce qu'il comporte en outre du dUMP, à une concentration allant de $1\mu\text{M}$ à $800\mu\text{M}$, de préférence, $500\mu\text{M}$.

24. Milieu selon l'une quelconque des revendications 19 à 23, caractérisé en ce que les flavines oxydées sont des flavines mononucléotides et/ou de dinucléotides flavine adénine.

5 25. Milieu selon l'une quelconque des revendications 19 à 24, caractérisé en ce que la réduction des flavines est effectuée par voie chimique, enzymatique, photochimique ou électrochimique.

10 26. Milieu selon l'une quelconque des revendications 19 à 25, caractérisé en ce que la réduction des flavines est effectuée avec du NADH et/ou du NADPH.

15 27. Milieu selon la revendication 26, caractérisé en ce que la concentration en NADH est de 0,1 à 1mM, et celle en NADPH, de 0,5 à 5mM.

28. Procédé de criblage d'un composé antibactérien ou antiviral *in vitro* dans un système acellulaire caractérisé en ce que ledit procédé comprend les étapes suivantes :

- 20 a) préparer un lysat cellulaire à partir d'une culture de cellules exprimant un polypeptide THYX en l'absence d'un polypeptide codé par un gène *thyA* et comportant un milieu selon l'une des revendications 19 à 27;
- 25 b) ajouter au lysat cellulaire obtenu à l'étape a) le composé inhibiteur à tester ;
- c) comparer l'activité thymidylate synthase ThyX respectivement dans le lysat cellulaire obtenu à l'étape a) et dans le lysat cellulaire obtenu à l'étape b) ; et
- 30 d) sélectionner les composés candidats pour lesquels une inhibition de l'activité thymidylate synthase ThyX a été détectée.

29. Procédé de criblage selon la revendication 28, caractérisé en ce que on utilise le dTMP et ³H comme marqueurs de l'activité thymidylate synthase ThyX.

35

30. Kit ou nécessaire pour le criblage d'un composé inhibiteur de la thymidylate synthase ThyX caractérisé en ce qu'il comprend :

- 5 a) une composition comprenant un polypeptide THYX ainsi qu'un milieu selon l'une quelconque des revendications 19 à 27, en solution ou sous forme lyophilisée ;
- b) facultativement, un ou plusieurs réactifs nécessaires à la quantification de l'activité thymidylate synthase ThyX.

10 31. Kit ou nécessaire pour le criblage d'un composé antibactérien ou antiviral, caractérisé en ce qu'il comprend :

- 15 a) un vecteur d'expression recombinant comprenant un acide nucléique codant un polypeptide ThyX sous le contrôle d'un promoteur fonctionnel dans une cellule hôte dans laquelle son expression est recherchée ou une cellule hôte transfectée avec un tel vecteur recombinant ;
- b) un milieu selon l'une quelconque des revendications 19 à 27 ;
- c) facultativement, un ou plusieurs réactifs nécessaires à la quantification de l'activité thymidylate synthase ThyX.

C. diphteriae	1	KVDS - A A T D S E A L V E F A G A C A C A T E T	1	F D K P N P R T A A N D A Y T R H I M E V G - H M A L L E	1	H M A L L E
Phage_D29	1	V G D D E T T S A D E L A E F A G R N C Y L S	1	F D R P N P K T R E N V D Y I N H I I D V G - H E S V L E	1	H E S V L E
P. abyssi	1	T R R P L E T T I T W A A L V S W D E W S T E S	1	F E K I M - - E D D V K A H L P R I I G Y G - H E S I L E	1	H E S I L E
H. salinarum	1	T E N P E L I C Q S A R N D Y M S D W C D T	1	P L D T A M A S V D G D T T D E K L S N L I A Q L L T R G - H Y C P F E	1	H Y C P F E
H. pylori	1	L E T P C K H Y T P L D A S Q A I R T C W Q S	1	- - - - - F E Y S D D G C K D R D L I H R V G N I F R H S S T L E	1	H S S T L E
Synechocystis_sp.	1	M T K P I V I D G E P L S P E G L I A Y C A R - - - -	1	- - - - - V S S P N Q E N P N Y T K L L Q F C I R E G - H T S P F E	1	H T S P F E
D. discoideum	1	H G K V A L V D T M P R L A P V G Q T A D S I C Q A A R V S M G A G T K K V T E D K G L I R Y L Y R H Q - H I S P F G	1	- - - - - S D G R D K G L S G R D T K L I K Y L A K H K - H I S P F G	1	H I S P F G
Roseophage_SIO1	1	H M G S D L S V N A A R V S C K K S E W V Y C G Q - - - -	1	- - - - - S D G R D K G L S G R D T K L I K Y L A K H K - H I S P F G	1	H I S P F G
C. diphteriae	52	H P A T V Y T I R G S R S A T H E L V R H R H F S F S Q L S Q R V H S D E T H - - - - - V V I P P - - - - - L I D N	53	H P A T V Y T I R G S R S A T H E L V R H R H F S F S Q L S Q R V H S D E T H - - - - - V V I P P - - - - - L I D N	54	H P A T V Y T I R G S R S A T H E L V R H R H F S F S Q L S Q R V H S D E T H - - - - - V V I P P - - - - - L I D N
Phage_D29	53	H S A T F Y T I E - A S R S V L T E L E R H R H I S F S V S Q R Y V D P T E L G - - - - - I H V P P A F T E L S G S	54	H S A T F Y T I E - A S R S V L T E L E R H R H I S F S V S Q R Y V D P T E L G - - - - - I H V P P A F T E L S G S	55	H S A T F Y T I E - A S R S V L T E L E R H R H I S F S V S Q R Y V D P T E L G - - - - - I H V P P A F T E L S G S
P. abyssi	51	H A T F T F S I E G C S R V C T H Q L V R H R I A S T Q S Q R Y V L N E E N - - - - - V E E T F V I P P S A T D P	52	H A T F T F S I E G C S R V C T H Q L V R H R I A S T Q S Q R Y V L N E E N - - - - - V E E T F V I P P S A T D P	53	H A T F T F S I E G C S R V C T H Q L V R H R I A S T Q S Q R Y V L N E E N - - - - - V E E T F V I P P S A T D P
H. salinarum	60	H P S A T F A I E G V S R S C M A O L T R H R H A S F D V Q S M R Y V A F D V D P A A V A E G E L V V I P P S A T D P	61	H P S A T F A I E G V S R S C M A O L T R H R H A S F D V Q S M R Y V A F D V D P A A V A E G E L V V I P P S A T D P	62	H P S A T F A I E G V S R S C M A O L T R H R H A S F D V Q S M R Y V A F D V D P A A V A E G E L V V I P P S A T D P
H. pylori	54	H L Y Y N F E I K G L S R G A L Q L S R H R I A S F S Q R Y T L R - - - - - E L K E V E S F L P L N E T	55	H L Y Y N F E I K G L S R G A L Q L S R H R I A S F S Q R Y T L R - - - - - E L K E V E S F L P L N E T	56	H L Y Y N F E I K G L S R G A L Q L S R H R I A S F S Q R Y T L R - - - - - E L K E V E S F L P L N E T
Synechocystis_sp.	53	M V D M T L E I T - T T R A L A P Q L R H R S T F S F Q E F S L R Y S C A T E Y E - - - - - C Y E A R R Q	54	M V D M T L E I T - T T R A L A P Q L R H R S T F S F Q E F S L R Y S C A T E Y E - - - - - C Y E A R R Q	55	M V D M T L E I T - T T R A L A P Q L R H R S T F S F Q E F S L R Y S C A T E Y E - - - - - C Y E A R R Q
D. discoideum	60	M V E F R E H C V - M P V F A R Q W L R H R T F A N V N E Y S A R Y S V L P D K F - - - - - Y H P S I E E V R K Q	61	M V E F R E H C V - M P V F A R Q W L R H R T F A N V N E Y S A R Y S V L P D K F - - - - - Y H P S I E E V R K Q	62	M V E F R E H C V - M P V F A R Q W L R H R T F A N V N E Y S A R Y S V L P D K F - - - - - Y H P S I E E V R K Q
Roseophage_SIO1	57	H A F A S E F H V K - A P I F V A R Q L V K K F L R N E I S R R Y V D D E P E F - - - - - Y T P D V W R G R	58	H A F A S E F H V K - A P I F V A R Q L V K K F L R N E I S R R Y V D D E P E F - - - - - Y T P D V W R G R	59	H A F A S E F H V K - A P I F V A R Q L V K K F L R N E I S R R Y V D D E P E F - - - - - Y T P D V W R G R
C. diphteriae	102	D P - Q L R E L F L S T V D E V R F A V S E L M T A L D N K K L A D E P N A I L R R K Q A R Q A A R S L P N A T E S R I	103	D P - Q L R E L F L S T V D E V R F A V S E L M T A L D N K K L A D E P N A I L R R K Q A R Q A A R S L P N A T E S R I	104	D P - Q L R E L F L S T V D E V R F A V S E L M T A L D N K K L A D E P N A I L R R K Q A R Q A A R S L P N A T E S R I
Phage_D29	106	D A D K A K E V L L D V Q S F A Q E A V E Y L V H I F S - - - - - D A G F P R K K A R E A A R A V L P N M F N S P M	107	D A D K A K E V L L D V Q S F A Q E A V E Y L V H I F S - - - - - D A G F P R K K A R E A A R A V L P N M F N S P M	108	D A D K A K E V L L D V Q S F A Q E A V E Y L V H I F S - - - - - D A G F P R K K A R E A A R A V L P N M F N S P M
P. abyssi	105	D R - E L Y E K W K K A M A E T I K L Y K - - - - - E S T K K G I H Q E D A R F A V L P Q A V R S K I T	106	D R - E L Y E K W K K A M A E T I K L Y K - - - - - E S T K K G I H Q E D A R F A V L P Q A V R S K I T	107	D R - E L Y E K W K K A M A E T I K L Y K - - - - - E S T K K G I H Q E D A R F A V L P Q A V R S K I T
H. salinarum	120	D W G R N O D A G D I D E E T M A E R E A V F Q A S V R R A V E D Y Q E L L G L G M P P E D A R F V L P I G T E V N V	121	D W G R N O D A G D I D E E T M A E R E A V F Q A S V R R A V E D Y Q E L L G L G M P P E D A R F V L P I G T E V N V	122	D W G R N O D A G D I D E E T M A E R E A V F Q A S V R R A V E D Y Q E L L G L G M P P E D A R F V L P I G T E V N V
H. pylori	105	N L E R A K E F L V F V D D E K V N E M S V L A L E N L R - - - - - V L L S E H N I K N D L A R A Y A M P E S Y K T H I	106	N L E R A K E F L V F V D D E K V N E M S V L A L E N L R - - - - - V L L S E H N I K N D L A R A Y A M P E S Y K T H I	107	N L E R A K E F L V F V D D E K V N E M S V L A L E N L R - - - - - V L L S E H N I K N D L A R A Y A M P E S Y K T H I
Synechocystis_sp.	100	D V K N R Q N S L D D F D E S T K K W F N Q A A A V W E K S H Q L Y E A L A K G I A K E C A R S L P L N T V T R L	101	D V K N R Q N S L D D F D E S T K K W F N Q A A A V W E K S H Q L Y E A L A K G I A K E C A R S L P L N T V T R L	102	D V K N R Q N S L D D F D E S T K K W F N Q A A A V W E K S H Q L Y E A L A K G I A K E C A R S L P L N T V T R L
D. discoideum	111	S T S N R O G G E E A E P K T A Q E F L D Y L D K V E E N - Y K T Y N E - - - - - L Q A L D I G I S P E Q A R M L P O S T M T E W	112	S T S N R O G G E E A E P K T A Q E F L D Y L D K V E E N - Y K T Y N E - - - - - L Q A L D I G I S P E Q A R M L P O S T M T E W	113	S T S N R O G G E E A E P K T A Q E F L D Y L D K V E E N - Y K T Y N E - - - - - L Q A L D I G I S P E Q A R M L P O S T M T E W
Roseophage_SIO1	106	S A D K K Q G S D G V V N P E Y N P Q Y L D N K I K F A Y - - - - - L Q A L D I G I S P E Q A R M L P O S T M T E W	107	S A D K K Q G S D G V V N P E Y N P Q Y L D N K I K F A Y - - - - - L Q A L D I G I S P E Q A R M L P O S T M T E W	108	S A D K K Q G S D G V V N P E Y N P Q Y L D N K I K F A Y - - - - - L Q A L D I G I S P E Q A R M L P O S T M T E W
C. diphteriae	161	V V T G N F R A W R H F F G M R A T E H A D V E I R S L A V R C L E I L K E K A P T	162	V V T G N F R A W R H F F G M R A T E H A D V E I R S L A V R C L E I L K E K A P T	163	V V T G N F R A W R H F F G M R A T E H A D V E I R S L A V R C L E I L K E K A P T
Phage_D29	159	V V T G N H R A W R Y V V K N R W H E A A D A D V E I R E L A G E L L R Q L R R E I A P N	160	V V T G N H R A W R Y V V K N R W H E A A D A D V E I R E L A G E L L R Q L R R E I A P N	161	V V T G N H R A W R Y V V K N R W H E A A D A D V E I R E L A G E L L R Q L R R E I A P N
P. abyssi	149	V V T M N H R E L K H F F G L R L C E R A Q W E I R E L A G E L L R Q L R R E I A P N	150	V V T M N H R E L K H F F G L R L C E R A Q W E I R E L A G E L L R Q L R R E I A P N	151	V V T M N H R E L K H F F G L R L C E R A Q W E I R E L A G E L L R Q L R R E I A P N
H. salinarum	180	V V T L N P R S L M H V A D M R A A A D A Q W E I R E L T E Q L L D A A A Q W C P E H	181	V V T L N P R S L M H V A D M R A A A D A Q W E I R E L T E Q L L D A A A Q W C P E H	182	V V T L N P R S L M H V A D M R A A A D A Q W E I R E L T E Q L L D A A A Q W C P E H
H. pylori	159	A Y S I N A R S L Q N L T L R S S Q A T Q K E H R E L A L A A R K I F P Y E H Q Y	160	A Y S I N A R S L Q N L T L R S S Q A T Q K E H R E L A L A A R K I F P Y E H Q Y	161	A Y S I N A R S L Q N L T L R S S Q A T Q K E H R E L A L A A R K I F P Y E H Q Y
Synechocystis_sp.	160	Y M K G S V R S W I H Y F S V R C D Q A T Q K E H R E L A L A A R K I F P Y E H Q Y	161	Y M K G S V R S W I H Y F S V R C D Q A T Q K E H R E L A L A A R K I F P Y E H Q Y	162	Y M K G S V R S W I H Y F S V R C D Q A T Q K E H R E L A L A A R K I F P Y E H Q Y
D. discoideum	170	Y W K I D L H N L F F R L R M D S H S Q K E I R D Y A N T T F A L I T R P I V P V	171	Y W K I D L H N L F F R L R M D S H S Q K E I R D Y A N T T F A L I T R P I V P V	172	Y W K I D L H N L F F R L R M D S H S Q K E I R D Y A N T T F A L I T R P I V P V
Roseophage_SIO1	160	Y W S G S L D A F A D M C R L R C K E D T Q Y E S R V V A D Q L S K K A D L Y P V	161	Y W S G S L D A F A D M C R L R C K E D T Q Y E S R V V A D Q L S K K A D L Y P V	162	Y W S G S L D A F A D M C R L R C K E D T Q Y E S R V V A D Q L S K K A D L Y P V

FIGURE 1

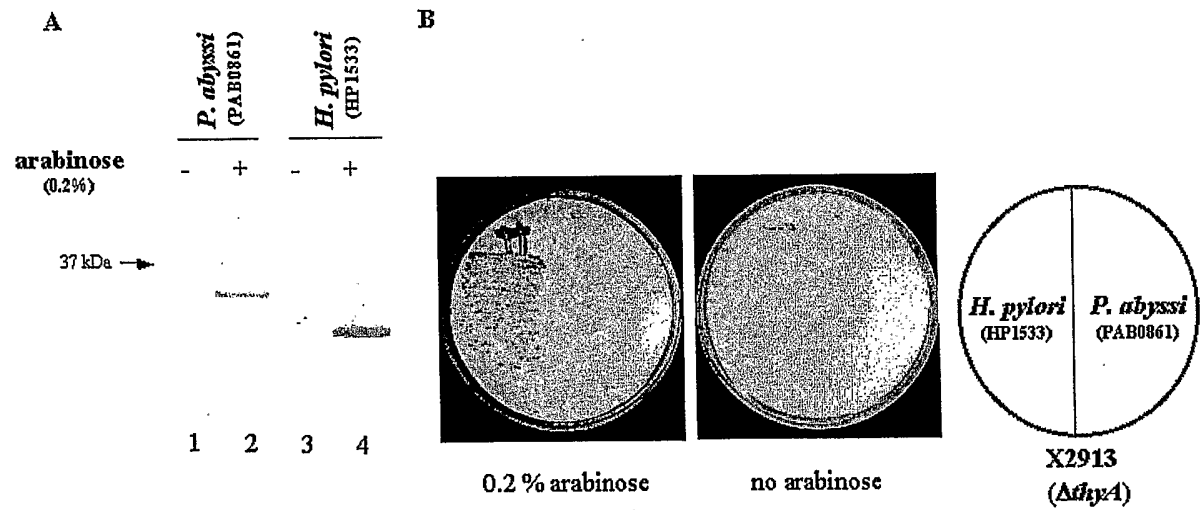


FIGURE 2

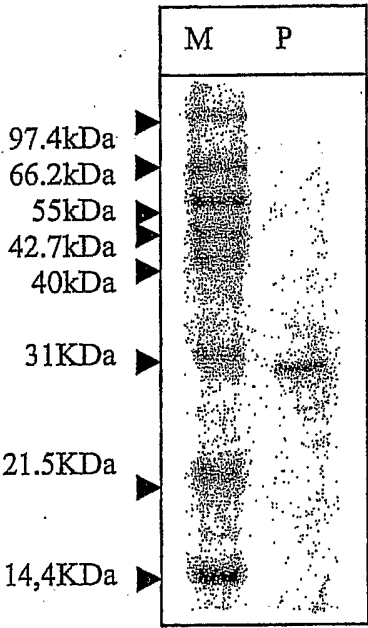


FIGURE 3

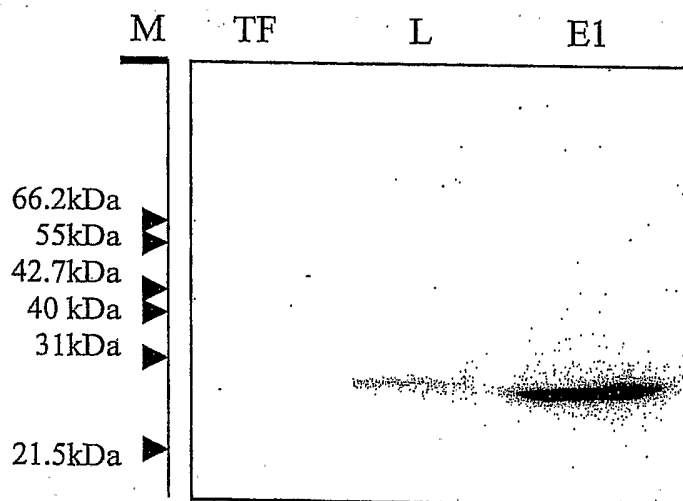


FIGURE 4

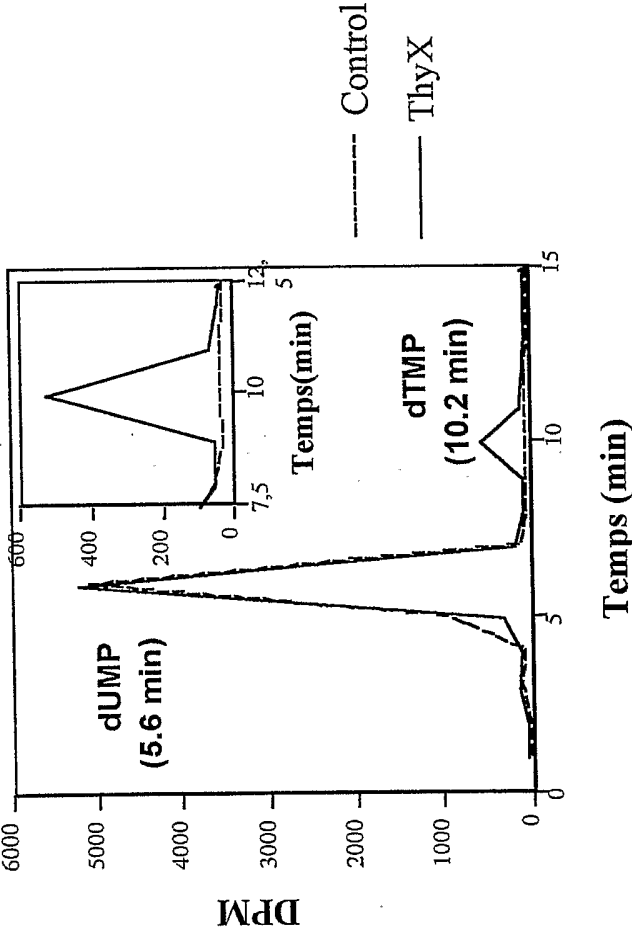


Figure 6

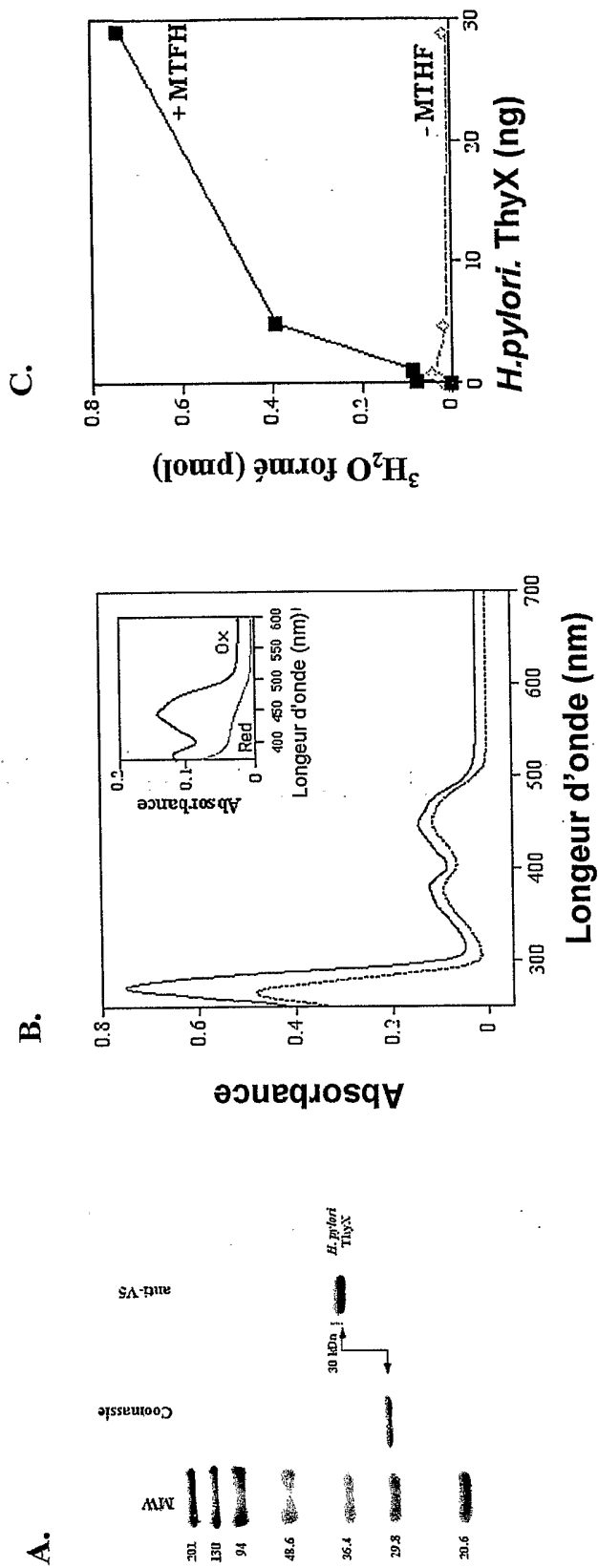


Figure 5

LISTE DE SEQUENCES

<110> Institut National de la Santé et de la Recherche M

<120> Utilisations d'un polypeptide THYX ou d'un acide
nucléique codant un tel polypeptide, notamment pour le
criblage de composés anti-bactériens ou anti-viraux

<130> THYX - INSERM

<140>

<141>

<160> 64

<170> PatentIn Ver. 2.1

<210> 1

<211> 265

<212> PRT

<213> Borrelia burgdorferi

<400> 1

Met Asn Lys Glu Tyr Lys Ile Leu Asp Asn Gly Phe Leu Lys Leu Ile
1 5 10 15

Asp Phe Met Gly Asp Asp Arg Arg Ile Val Lys Ala Ala Arg Ile Ser
20 25 30

Tyr Arg Glu Glu Ser Val Lys Arg Lys Asp Ala Glu Leu Ile Asp Tyr
35 40 45

Leu Ile Arg Asn Gly His Thr Ser Pro Leu Glu Gln Val Val Phe Thr
50 55 60

Phe His Val Lys Ala Pro Ile Phe Val Ala Arg Gln Trp Met Arg His
65 70 75 80

Arg Thr Ala Arg Ile Asn Glu Val Ser Gly Cys Tyr Ser Leu Ala Arg
85 90 95

Glu Glu Phe Tyr Val Pro Leu Glu Glu Asp Leu Lys Cys Gln Thr Ser
100 105 110

Ser Asn Ser Ser Glu Lys Glu Phe Lys Ser Leu Glu Lys Leu Ser Asp
115 120 125

Lys Ile Lys His His Gln Lys His Ser Tyr Glu Leu Tyr Gln Asp Met
130 135 140

Ile Asn Ala Asn Ile Pro Lys Glu Leu Ser Arg Ile Val Leu Pro Leu
145 150 155 160

Ser Leu Tyr Thr Glu Trp Tyr Trp Gln Ile Asp Leu Asn Asn Leu Phe
165 170 175

His Phe Ile Lys Leu Arg Leu Ala Leu Asp Ser Pro Lys Glu Ile Lys
180 185 190

Glu Asn Ser Pro Lys Glu Met Arg Glu Tyr Ala Lys Ala Leu Ile Ser
195 200 205

Ile Val Arg Glu Ile Val Pro Ile Ala Phe Asn Ser Phe Glu Asn His
 210 215 220

Phe Leu Arg Gly Lys Arg Phe Ser His Glu Glu Ile Ile Ala Ile Ile
 225 230 235 240

Asn Ala Leu Asp Leu Asn Lys Leu Ser Met Asp Ala Glu Lys Leu Asn
 245 250 255

Leu Leu Lys Asp Lys Leu Gly Ile Asp
 260 265

<210> 2

<211> 305

<212> PRT

<213> Treponema pallidum

<400> 2

Met Thr Leu Arg Thr Leu Gln Ala Gly Val Ala Val Ser Ile Ala Leu
 1 5 10 15

Asp Arg Val Cys Phe Phe Cys Tyr Asn Gly Ala Val Ala His Cys Val
 20 25 30

Val Glu Ala Ala Glu Asp Ile Leu Asp Arg Arg Phe Ser Val Leu Asp
 35 40 45

Lys Gly Phe Val Arg Leu Ile Asp Tyr Leu Gly Gly Asp Ala Arg Ile
 50 55 60

Val Gln Ala Ala Arg Val Ser Tyr Gly Ala Gly Thr Arg Thr Ala Arg
 65 70 75 80

Asp Asp Ala Ala Leu Ile Asp Phe Leu Leu Arg Asn Lys His Thr Ser
 85 90 95

Pro Phe Glu Gln Val Val Leu Thr Phe His Val Arg Ala Pro Ile Phe
 100 105 110

Val Ala Arg Gln Trp Met Arg His Arg Thr Ala Arg Ile Ser Glu Val
 115 120 125

Ser Ser Arg Tyr Ser Leu Leu Ser His Asp Cys Tyr Val Pro Gln Glu
 130 135 140

Thr Ser Val Ala Val Gln Ser Thr Arg Asn Lys Gln Gly Arg Ala Ser
 145 150 155 160

Glu Gly Ile Ser Pro Glu Gln Gln Gln Glu Val Arg Ala Ala Phe Glu
 165 170 175

Ala Gln Gln Lys Ala Ala Cys Ala Ala Tyr Asp Ala Leu Ile Gln Lys
 180 185 190

Asn Ile Ala Arg Glu Leu Ala Arg Ile Asn Val Pro Leu Ser Leu Tyr
 195 200 205

Thr Glu Trp Tyr Trp Gln Ile Asp Leu His Asn Leu Phe His Phe Leu

210 215 220
 Arg Leu Arg Ala Ser Ala His Ala Gln Ala Glu Ile Arg Ala Tyr Ala
 225 230 235 240
 Glu Val Ile Ile Glu Ile Thr Arg Ala Val Ala Pro Cys Ala Thr Ala
 245 250 255
 Ser Phe Glu Asn His Glu Lys Asp Gly Val Gln Phe Ser Gly Arg Glu
 260 265 270
 Phe Ala Ala Leu Lys Ala Leu Leu Ala Gly Glu Gly Leu Ser Leu Glu
 275 280 285
 Gly Lys Glu Arg Ala Arg Phe Glu Glu Lys Leu Arg Ser Gly Leu Gln
 290 295 300
 Gln
 305

 <210> 3
 <211> 294
 <212> PRT
 <213> Rickettsia prowazekii

 <400> 3
 Met His Asn Thr Thr Lys Arg Val Thr Val Pro Ala Leu Glu Ala Met
 1 5 10 15
 Leu Tyr Glu Thr Ile Lys Val Leu Asp His Gly Phe Ile Arg Val Ile
 20 25 30
 Asp Tyr Met Gly Asp Asp Ser Ser Ile Val Gln Ala Ala Arg Val Ser
 35 40 45
 Tyr Gly Lys Gly Thr Lys Gln Leu Asn Gln Asp Lys Gly Leu Ile Asn
 50 55 60
 Tyr Leu Leu Arg His Tyr His Thr Thr Pro Phe Glu Met Cys Asp Ile
 65 70 75 80
 Lys Phe His Ile Lys Leu Pro Ile Phe Ile Ala Arg Gln Trp Ile Arg
 85 90 95
 His Arg Thr Ala Ser Val Asn Glu Tyr Ser Ala Arg Tyr Ser Ile Leu
 100 105 110
 Gly Asn Glu Phe Tyr Leu Pro Asp Pro Ala Asn Ile Ala Ser Gln Ser
 115 120 125
 Val Val Asn Lys Gln Cys Arg Ala Gly Asp Ser Val Pro Lys Lys Val
 130 135 140
 Ser Glu Lys Val Leu Ala Ile Leu Glu Glu Asp Ala Arg Arg Cys Tyr
 145 150 155 160
 Arg His Tyr Lys Glu Leu Met Asn Ala Asp Glu Asp Gly Asn Ile Leu
 165 170 175

Asp Glu Asn Val Ser Gly Ile Ala Arg Glu Leu Ala Arg Ile Asn Leu
 180 185 190

Thr Leu Asn Tyr Tyr Thr Glu Trp Tyr Trp Lys Ile Asn Leu His Asn
 195 200 205

Leu Leu His Phe Leu Arg Leu Arg Thr Asp Pro Lys Ala Gln Tyr Glu
 210 215 220

Ile Arg Val Tyr Ala Glu Lys Ile Leu Asp Ile Val Lys Ala Trp Val
 225 230 235 240

Pro Phe Thr Tyr Glu Ala Phe Glu Glu Tyr Arg Leu Gln Gly Ala Asn
 245 250 255

Ile Ser Arg Lys Gly Leu Glu Val Ile Lys Arg Met Ile Lys Gly Glu
 260 265 270

Lys Val Ile His Glu Thr Ser Gly Met Asn Lys Arg Glu Trp Glu Glu
 275 280 285

Leu Val Lys Ile Phe Arg
 290

<210> 4

<211> 220

<212> PRT

<213> Thermogata maritima

<400> 4

Met Lys Ile Asp Ile Leu Asp Lys Gly Phe Val Glu Leu Val Asp Val
 1 5 10 15

Met Gly Asn Asp Leu Ser Ala Val Arg Ala Ala Arg Val Ser Phe Asp
 20 25 30

Met Gly Leu Lys Asp Glu Glu Arg Asp Arg His Leu Ile Glu Tyr Leu
 35 40 45

Met Lys His Gly His Glu Thr Pro Phe Glu His Ile Val Phe Thr Phe
 50 55 60

His Val Lys Ala Pro Ile Phe Val Ala Arg Gln Trp Phe Arg His Arg
 65 70 75 80

Ile Ala Ser Tyr Asn Glu Leu Ser Gly Arg Tyr Ser Lys Leu Ser Tyr
 85 90 95

Glu Phe Tyr Ile Pro Ser Pro Glu Arg Leu Glu Gly Tyr Lys Thr Thr
 100 105 110

Ile Pro Pro Glu Arg Val Thr Glu Lys Ile Ser Glu Ile Val Asp Lys
 115 120 125

Ala Tyr Arg Thr Tyr Leu Glu Leu Ile Glu Ser Gly Val Pro Arg Glu
 130 135 140

Val Ala Arg Ile Val Leu Pro Leu Asn Leu Tyr Thr Arg Phe Phe Trp
 145 150 155 160

Thr Val Asn Ala Arg Ser Leu Met Asn Phe Leu Asn Leu Arg Ala Asp
 165 170 175

Ser His Ala Gln Trp Glu Ile Gln Gln Tyr Ala Leu Ala Ile Ala Arg
 180 185 190

Ile Phe Lys Glu Lys Cys Pro Trp Thr Phe Glu Ala Phe Leu Lys Tyr
 195 200 205

Ala Tyr Lys Gly Asp Ile Leu Lys Glu Val Gln Val
 210 215 220

<210> 5

<211> 260

<212> PRT

<213> Dictyostelium discoideum

<400> 5

Met Gly Leu Asp Ile Gln Thr Glu Ile Asp Lys Ile Val Ile Glu Lys
 1 5 10 15

Val Lys Pro Glu Val Glu Tyr Tyr Asp Val Met Gly Gly Ser His Arg
 20 25 30

Trp Glu Val Lys Val His Asp His Gly Lys Val Ala Leu Val Asp Thr
 35 40 45

Met Pro Arg Leu Ala Pro Val Gly Gln Thr Ala Asp Phe Ser Ile Cys
 50 55 60

Gln Ala Ala Arg Val Ser Tyr Gly Ala Gly Thr Lys Lys Val Thr Glu
 65 70 75 80

Asp Lys Gly Leu Ile Arg Tyr Leu Tyr Arg His Gln His Thr Ser Pro
 85 90 95

Phe Glu Met Val Glu Phe Lys Phe His Cys Val Met Pro Val Phe Ile
 100 105 110

Ala Arg Gln Trp Ile Arg His Arg Thr Ala Asn Val Asn Glu Tyr Ser
 115 120 125

Ala Arg Tyr Ser Val Leu Pro Asp Lys Phe Tyr His Pro Ser Ile Glu
 130 135 140

Glu Val Arg Lys Gln Ser Thr Ser Asn Arg Gln Gly Gly Glu Glu Ala
 145 150 155 160

Leu Glu Pro Lys Thr Ala Gln Glu Phe Leu Asp Tyr Leu Asp Lys Val
 165 170 175

Glu Glu Asn Tyr Lys Thr Tyr Asn Glu Leu Leu Glu Lys Gly Leu Ser
 180 185 190

Arg Glu Leu Gly Arg Ile Gly Leu Pro Val Ser Ile Tyr Thr Glu Trp
 195 200 205

Tyr Trp Lys Ile Asp Leu His Asn Leu Phe His Phe Leu Arg Leu Arg

210

215

220

Met Asp Ser His Ser Gln Lys Glu Ile Arg Asp Tyr Ala Asn Thr Ile
 225 230 235 240

Phe Ala Leu Ile Arg Pro Ile Val Pro Val Ala Cys Glu Gly Ile Tyr
 245 250 255

Arg Leu Cys Phe
 260

<210> 6

<211> 222

<212> PRT

<213> Roseophage

<400> 6

Met Thr Gln Ile Glu Ala Thr Tyr Ile Asp His Met Gly Ser Asp Leu
 1 5 10 15

Ser Val Val Asn Ala Ala Arg Val Ser Phe Gly Lys Lys Ser Glu Trp
 20 25 30

Val Tyr Cys Gly Gln Ser Asp Gly Arg Asp Lys Gly Leu Ser Gly Arg
 35 40 45

Asp Thr Lys Leu Ile Lys Tyr Leu Ala Lys His Lys His Ile Ser Pro
 50 55 60

Phe Gly His Ala Phe Ala Ser Phe His Val Lys Ala Pro Ile Phe Val
 65 70 75 80

Ala Arg Gln Leu Val Lys His Lys Phe Leu Arg Trp Asn Glu Ile Ser
 85 90 95

Arg Arg Tyr Val Asp Asp Glu Pro Glu Phe Tyr Thr Pro Asp Val Trp
 100 105 110

Arg Gly Arg Ser Ala Asp Lys Lys Gln Gly Ser Asp Gly Val Val Asn
 115 120 125

Pro Glu Tyr Asn Pro Gln Tyr Leu Asp Asn Lys Ile Lys Phe Ala Tyr
 130 135 140

Leu Gln Ala Leu Asp Ile Gly Ile Ser Pro Glu Gln Ala Arg Met Leu
 145 150 155 160

Leu Pro Gln Ser Thr Met Thr Glu Trp Tyr Trp Ser Gly Ser Leu Asp
 165 170 175

Ala Phe Ala Asp Met Cys Arg Leu Arg Cys Lys Glu Asp Thr Gln Tyr
 180 185 190

Glu Ser Arg Val Val Ala Asp Gln Ile Ser Glu Lys Met Ala Asp Leu
 195 200 205

Tyr Pro Val Ser Trp Ala Ala Leu Met Glu Gly Glu Lys Gln
 210 215 220

<210> 7

<211> 246

<212> PRT

<213> Streptomyces coelicolor

<400> 7

Met Thr Asp Ile Pro Ala Asp Asp Pro Lys Ile Glu Leu Arg Ser Asp
 1 5 10 15

Ile Thr Val Glu Leu Val Lys Ser Ala Ala Thr Asp Ser Asp Val Leu
 20 25 30

Phe Ala Ala Arg Val Ser Thr Ala Gly Glu Gln Ser Leu Asp Glu Leu
 35 40 45

Lys Lys Asp Pro Glu Arg Ser Lys Gly Leu Ile Asn Tyr Leu Met Arg
 50 55 60

Asp Arg His Gly Ser Pro Phe Glu His Asn Ser Met Thr Phe Phe Val
 65 70 75 80

Ser Ala Pro Ile Phe Val Phe Arg Glu Phe Met Arg His Arg Val Gly
 85 90 95

Trp Ser Tyr Asn Glu Glu Ser Gly Arg Tyr Arg Glu Leu Gln Pro Val
 100 105 110

Phe Tyr Ala Pro Asp Ala Ser Arg Lys Leu Val Gln Gln Gly Arg Pro
 115 120 125

Gly Lys Tyr Val Phe Val Glu Gly Thr Pro Glu Gln His Glu Leu Val
 130 135 140

Gly Ser Ala Met Glu Asp Ser Tyr Arg Gln Ala Tyr Ala Thr Tyr Gln
 145 150 155 160

Gln Met Leu Ala Ala Gly Val Ala Arg Glu Val Ala Arg Ala Val Leu
 165 170 175

Pro Val Gly Leu Tyr Ser Ser Met Tyr Ala Thr Cys Asn Ala Arg Ser
 180 185 190

Leu Met His Phe Leu Gly Leu Arg Thr Gln His Glu Leu Ala Lys Val
 195 200 205

Pro Ser Phe Pro Gln Arg Glu Ile Glu Met Ala Gly Glu Lys Met Glu
 210 215 220

Ala Glu Trp Ala Arg Leu Met Pro Leu Thr His Ala Ala Phe Asn Ala
 225 230 235 240

Asn Gly Arg Val Ala Pro
 245

<210> 8

<211> 317

<212> PRT

<213> Aquifex aeolicus

<400> 8

```

Met Met Lys Ile Tyr Leu Met Gly Ser Asp Gln Arg Ile Val Arg Cys
 1              5              10              15

Ala Arg Val Ser Phe Ala Lys Asp Ser Tyr Val Asp Glu Lys Arg Asp
      20              25              30

Lys Arg Leu Ile Arg Tyr Leu Phe Lys His Arg His Ala Ser Pro Phe
      35              40              45

Glu His Asn Ile Ile Ala Phe Glu Trp Lys Lys Glu Lys Trp Ile Glu
      50              55              60

Leu Leu Ser Lys Leu Glu Asn Pro Thr Val Gln Val Tyr Tyr Ser Asn
      65              70              75              80

Gly Phe Val Phe Leu Asn Leu Arg Asn Ala Ile Asn Val Trp Glu Leu
      85              90              95

Leu Pro Asp Ala Val Lys Glu Arg Ile Lys Glu Ala Phe Pro Thr Thr
      100             105             110

Tyr Gly Val Ile Gln Arg Arg Gly Glu Ile Glu Asp Glu Glu Leu Tyr
      115             120             125

Ser Leu Pro Tyr Thr Lys Asp Lys Ala Tyr Val Lys Glu Lys Ile Glu
      130             135             140

Thr Ser Ser Gly Trp Ile Gly Leu Val Asp Lys Leu Glu Leu Glu Thr
      145             150             155             160

Asp Met Asp Phe Tyr Thr Phe Val Val Glu Cys Pro Leu Phe Val Ala
      165             170             175

Arg Gln Trp Met Arg His Arg Phe Gly Ser Tyr Asn Glu Val Ser Lys
      180             185             190

Arg Tyr Val Gly Lys Glu Phe Leu Glu Phe Tyr Leu Pro Lys Tyr Ile
      195             200             205

Arg Lys Gln Ala Glu Lys Asn Lys Gln Ala Ser Val Asp Glu Pro Ile
      210             215             220

Ser Glu Ser Glu Val Phe Ile Lys Lys Ile Glu Asn Leu Ile Ser Lys
      225             230             235             240

Ser Val Lys Leu Tyr Glu Glu Ile Ile Glu Lys Gly Gly Ala Lys Glu
      245             250             255

Leu Ala Arg Gly Val Leu Pro Gln Phe Met Lys Thr Arg Phe Tyr Trp
      260             265             270

Thr Val Pro Arg Ile Ser Leu Asp Asn Phe Ile Thr Leu Arg Thr His
      275             280             285

Glu Gly Ala Gln Lys Glu Ile Arg Glu Phe Ala Glu Ala Ile Lys Glu
      290             295             300

Met Val Gly Tyr Arg Gly Thr Asp Lys Lys Asn Val Ile

```

305

310

315

<210> 9

<211> 216

<212> PRT

<213> Chorella virus

<400> 9

Met Ser Ala Lys Leu Ile Ser Val Thr Lys Pro Val Val Glu Gly Val
 1 5 10 15

Asn Thr Ala Glu Glu Leu Ile Ala Tyr Ala Ala Arg Val Ser Asn Pro
 20 25 30

Glu Asn Gln Ile Asn Asn Lys Thr Ala Ser Gly Leu Leu Lys Tyr Cys
 35 40 45

Ile Arg His Lys His Trp Ser Ile Phe Glu Thr Ala Phe Met Thr Leu
 50 55 60

Glu Leu Lys Thr Ser Arg Gly Ile Ala Ala Gln Val Leu Arg His Arg
 65 70 75 80

Ser Phe His Phe Gln Glu Phe Ser Gln Arg Tyr Ala Ser Val Met Glu
 85 90 95

Thr Pro Pro Pro His Gln Ala Arg Phe Gln Asp His Lys Asn Arg Gln
 100 105 110

Asn Ser Leu Asp Thr Val Pro Glu Asp Asp Gln Thr Trp Trp Ala Thr
 115 120 125

Glu Gln Glu Lys Leu Tyr Ala Gln Ser Met Glu Leu Tyr Asn Lys Ala
 130 135 140

Leu Glu Lys Gly Ile Ala Lys Glu Cys Ala Arg Phe Ile Leu Pro Leu
 145 150 155 160

Ser Thr Pro Thr Thr Ile Tyr Met Ser Gly Thr Ile Arg Asp Trp Ile
 165 170 175

His Tyr Ile Glu Leu Arg Thr Ser Asn Gly Thr Gln Arg Glu His Ile
 180 185 190

Asp Leu Ala Asn Ala Cys Lys Glu Ile Phe Ile Lys Glu Phe Pro Ser
 195 200 205

Ile Ala Lys Ala Leu Asp Trp Val
 210 215

<210> 10

<211> 215

<212> PRT

<213> Synechocystis sp.

<400> 10

Met Asp Val Arg Phe Ile Ser Leu Thr Lys Pro Glu Ile Val Ile Asp

1 5 10 15
 Gly Glu Pro Leu Ser Pro Glu Gly Leu Ile Ala Tyr Cys Ala Arg Val
 20 25 30
 Ser Ser Pro Asn Gln Glu Asn Pro Asn Tyr Thr Lys Leu Leu Gln Phe
 35 40 45
 Cys Ile Arg Glu Gly His Trp Ser Ile Phe Glu Met Val Asp Met Thr
 50 55 60
 Leu Glu Ile Thr Thr Thr Arg Ala Ile Ala Pro Gln Ile Leu Arg His
 65 70 75 80
 Arg Ser Phe Ser Phe Gln Glu Phe Ser Leu Arg Tyr Ser Cys Ala Thr
 85 90 95
 Glu Tyr Glu Cys Tyr Glu Ala Arg Arg Gln Asp Val Lys Asn Arg Gln
 100 105 110
 Asn Ser Leu Asp Asp Phe Asp Glu Ser Thr Lys Lys Trp Phe Asn Gln
 115 120 125
 Ala Gln Ala Ala Val Trp Glu Lys Ser His Gln Leu Tyr Glu Glu Ala
 130 135 140
 Leu Ala Lys Gly Ile Ala Lys Glu Cys Ala Arg Ser Ile Leu Pro Leu
 145 150 155 160
 Asn Thr Val Thr Arg Leu Tyr Met Lys Gly Ser Val Arg Ser Trp Ile
 165 170 175
 His Tyr Phe Ser Val Arg Cys Asp Gln Ala Thr Gln Lys Glu His Arg
 180 185 190
 Glu Ile Ala Leu Ala Ala Arg Lys Ile Phe Met Lys His Phe Pro Thr
 195 200 205
 Val Ala Ala Ala Leu Glu Trp
 210 215

<210> 11

<211> 243

<212> PRT

<213> *Pyrococcus horikoshii*

<400> 11

Met Val Lys Val Lys Leu Ile Asn Tyr Thr Pro Lys Pro Leu Glu Thr
 1 5 10 15
 Val Thr Trp Ala Ala Leu Ile Ser Tyr Trp Asp Gly Trp Ser Thr Glu
 20 25 30
 Ala Phe Glu Lys Ile Ser Pro Asn Asp Val Glu Ile His Leu Pro Arg
 35 40 45
 Ile Leu Ser Tyr Gly His Glu Ser Ile Leu Glu His Ala Thr Phe Thr
 50 55 60

Phe Ser Ile Glu Gly Cys Ser Arg Val Cys Thr His Gln Leu Val Arg
 65 70 75 80
 His Arg Ile Ala Ser Tyr Thr Gln Gln Ser Gln Arg Tyr Ile Lys Ile
 85 90 95
 Asn Pro Glu Asp Val Glu Glu Thr Phe Val Ile Pro Glu Ser Ile Lys
 100 105 110
 Lys Asp Ser Glu Leu Leu Lys Glu Trp Lys Glu Leu Leu Lys Arg Ser
 115 120 125
 Leu Glu Leu Tyr Glu Lys Ser Ile Glu Arg Gly Ile His Gln Glu Asp
 130 135 140
 Ala Arg Phe Ile Leu Pro Gln Ser Val Lys Thr Lys Ile Val Val Thr
 145 150 155 160
 Met Asn Leu Arg Glu Leu Lys His Phe Phe Gly Leu Arg Leu Cys Glu
 165 170 175
 Arg Ala Gln Trp Glu Ile Arg Glu Val Ala Trp Lys Met Leu Glu Glu
 180 185 190
 Ile Ala Lys Arg Lys Glu Leu Lys Pro Ile Ile Glu Trp Ala Lys Leu
 195 200 205
 Gly Pro Arg Cys Ile Gln Leu Gly Tyr Cys Pro Glu Arg Glu Leu Met
 210 215 220
 Pro Pro Gly Cys Leu Lys Arg Thr Arg Glu Arg Trp Lys Asn Leu Leu
 225 230 235 240
 Glu Lys Tyr

<210> 12
 <211> 244
 <212> PRT
 <213> Pyrococcus abyssi

<400> 12
 Met Val Arg Val Thr Leu Val Asn Tyr Thr Arg Arg Pro Leu Glu Thr
 1 5 10 15
 Ile Thr Trp Ala Ala Leu Val Ser Tyr Trp Asp Glu Trp Ser Thr Glu
 20 25 30
 Ser Phe Glu Lys Ile Asn Glu Asp Asp Val Lys Ala His Leu Pro Arg
 35 40 45
 Ile Leu Gly Tyr Gly His Glu Ser Ile Leu Glu His Ala Thr Phe Thr
 50 55 60
 Phe Ser Ile Glu Gly Cys Ser Arg Val Cys Thr His Gln Leu Val Arg
 65 70 75 80
 His Arg Ile Ala Ser Tyr Thr Gln Gln Ser Gln Arg Tyr Ile Val Leu
 85 90 95

Asn Glu Glu Asn Val Glu Glu Thr Phe Val Ile Pro Glu Ser Ile Lys
 100 105 110

Lys Asp Arg Glu Leu Tyr Glu Lys Trp Lys Lys Ala Met Ala Glu Thr
 115 120 125

Ile Lys Leu Tyr Lys Glu Ser Leu Lys Arg Gly Ile His Gln Glu Asp
 130 135 140

Ala Arg Phe Ile Leu Pro Gln Ala Val Arg Ser Lys Ile Val Val Thr
 145 150 155 160

Met Asn Leu Arg Glu Leu Lys His Phe Phe Gly Leu Arg Leu Cys Glu
 165 170 175

Arg Ala Gln Trp Glu Ile Arg Glu Val Ala Trp Lys Met Leu Glu Glu
 180 185 190

Ile Ala Lys Arg Glu Glu Leu Arg Pro Ile Ile Lys Trp Ala Lys Leu
 195 200 205

Gly Pro Arg Cys Ile Gln Leu Gly Tyr Cys Pro Glu Arg Glu Leu Met
 210 215 220

Pro Pro Gly Cys Phe Lys Arg Thr Arg Glu Arg Trp Met Lys Leu Leu
 225 230 235 240

Glu Lys Pro Leu

<210> 13
 <211> 266
 <212> PRT
 <213> Halobacterium sp.

<400> 13
 Met Val Pro Ala Arg Gly Phe Gly Val Phe Leu Pro Pro Ala Gly Thr
 1 5 10 15

Pro Ser Ser Met Arg Val Arg Leu Leu Glu Ala Thr Glu Asn Pro Glu
 20 25 30

Glu Leu Ile Cys Gln Ser Ala Arg Asn Asp Tyr Met Ser Asp Trp Val
 35 40 45

Gly Asp Thr Pro Leu Asp Thr Ala Met Ala Ser Val Asp Gly Asp Thr
 50 55 60

Thr Asp Glu Lys Leu Ser Asn Leu Ile Ala Gln Leu Leu Thr Arg Gly
 65 70 75 80

His Tyr Gly Pro Phe Glu His Pro Ser Ala Thr Phe Ala Ile Glu Gly
 85 90 95

Val Ser Arg Ser Cys Met Ala Gln Leu Thr Arg His Arg His Ala Ser
 100 105 110

Phe Asp Val Gln Ser Met Arg Tyr Val Ala Phe Asp Asp Val Asp Pro

115 120 125
 Ala Ala Val Ala Glu Gly Glu Leu Val Val Thr Pro Pro Ser Ala Thr
 130 135 140
 Asp Pro Asp Trp Val Gly Arg Asn Gln Asp Ala Gly Asp Ile Asp Glu
 145 150 155 160
 Glu Thr Met Ala Glu Arg Glu Ala Val Phe Gln Ala Ser Val Arg Arg
 165 170 175
 Ala Val Glu Asp Tyr Gln Glu Leu Leu Gly Leu Gly Met Pro Pro Glu
 180 185 190
 Asp Ala Arg Phe Val Leu Pro Ile Gly Thr Glu Val Asn Val Val Ile
 195 200 205
 Thr Leu Asn Pro Arg Ser Leu Met His Val Ala Asp Met Arg Ala Ala
 210 215 220
 Ala Asp Ala Gln Trp Glu Ile Arg Glu Leu Thr Glu Gln Leu Leu Asp
 225 230 235 240
 Ala Ala Ala Gln Trp Cys Pro His Thr Phe Glu Tyr Tyr Asp Ala Glu
 245 250 255
 Met Lys His Arg Lys Asn Arg Leu Ala Pro
 260 265

<210> 14
 <211> 235
 <212> PRT
 <213> Mycobacteriophage D29

<400> 14
 Met Lys Val Gln Leu Ile Ala Ser Thr Ile Leu Glu Asp Pro Ser Trp
 1 5 10 15
 Ala Gly Thr Asp Tyr Val Gly Asp Asp Glu Thr Val Thr Ser Ala Asp
 20 25 30
 Glu Leu Ala Glu Phe Ala Gly Arg Asn Cys Tyr Leu Ser Phe Asp Arg
 35 40 45
 Pro Asn Pro Lys Thr Arg Glu Asn Val Asp Tyr Leu Asn His Ile Leu
 50 55 60
 Asp Val Gly His Glu Ser Val Leu Glu His Ser Ser Ala Thr Phe Tyr
 65 70 75 80
 Ile Glu Ala Ser Arg Ser Val Leu Thr Glu Leu Glu Arg His Arg His
 85 90 95
 Leu Ser Phe Ser Val Val Ser Gln Arg Tyr Val Asp Pro Thr Glu Leu
 100 105 110
 Gly Ile His Val Pro Pro Ala Phe Thr Glu Leu Ser Gly Ser Asp Ala
 115 120 125

Asp Lys Ala Lys Glu Val Leu Leu Asp Val Gln Ser Phe Ala Gln Glu
130 135 140

Ala Tyr Glu Tyr Leu Val His Ile Phe Ser Asp Ala Gly Phe Pro Arg
145 150 155 160

Lys Lys Ala Arg Glu Ala Ala Arg Ala Val Leu Pro Asn Met Thr Asn
165 170 175

Ser Pro Met Val Val Thr Gly Asn His Arg Ala Trp Arg Tyr Val Ile
180 185 190

Lys Asn Arg Trp His Glu Ala Ala Asp Ala Glu Ile Arg Glu Leu Ala
195 200 205

Gly Glu Leu Leu Arg Gln Leu Arg Glu Ile Ala Pro Asn Thr Tyr Gln
210 215 220

Asp Ile Pro Thr Glu Pro Tyr Ser Tyr Gly Gly
225 230 235

<210> 15

<211> 250

<212> PRT

<213> Mycobacterium tuberculosis

<400> 15

Met Ala Glu Gln Val Lys Leu Ser Val Glu Leu Ile Ala Cys Ser Ser
1 5 10 15

Phe Thr Pro Pro Ala Asp Val Glu Trp Ser Thr Asp Val Glu Gly Ala
20 25 30

Glu Ala Leu Val Glu Phe Ala Gly Arg Ala Cys Tyr Glu Thr Phe Asp
35 40 45

Lys Pro Asn Pro Arg Thr Ala Ser Asn Ala Ala Tyr Leu Arg His Ile
50 55 60

Met Glu Val Gly His Thr Ala Leu Leu Glu His Ala Asn Ala Thr Met
65 70 75 80

Tyr Ile Arg Gly Ile Ser Arg Ser Ala Thr His Glu Leu Val Arg His
85 90 95

Arg His Phe Ser Phe Ser Gln Leu Ser Gln Arg Phe Val His Ser Gly
100 105 110

Glu Ser Glu Val Val Val Pro Thr Leu Ile Asp Glu Asp Pro Gln Leu
115 120 125

Arg Glu Leu Phe Met His Ala Met Asp Glu Ser Arg Phe Ala Phe Asn
130 135 140

Glu Leu Leu Asn Ala Leu Glu Glu Lys Leu Gly Asp Glu Pro Asn Ala
145 150 155 160

Leu Leu Arg Lys Lys Gln Ala Arg Gln Ala Ala Arg Ala Val Leu Pro
165 170 175

Asn Ala Thr Glu Ser Arg Ile Val Val Ser Gly Asn Phe Arg Thr Trp
 180 185 190

Arg His Phe Ile Gly Met Arg Ala Ser Glu His Ala Asp Val Glu Ile
 195 200 205

Arg Glu Val Ala Val Gly Cys Leu Arg Lys Leu Gln Val Ala Ala Pro
 210 215 220

Thr Val Phe Gly Asp Phe Glu Ile Glu Thr Leu Ala Asp Gly Ser Gln
 225 230 235 240

Met Ala Thr Ser Pro Tyr Val Met Asp Phe
 245 250

<210> 16
 <211> 250
 <212> PRT
 <213> Corynebacterium glutamicum

<400> 16
 Met Ala Glu Gln Val Lys Leu Ser Val Glu Leu Ile Ala Cys Ser Ser
 1 5 10 15

Phe Thr Pro Pro Ala Asp Val Glu Trp Ser Thr Asp Val Glu Gly Ala
 20 25 30

Glu Ala Leu Val Glu Phe Ala Gly Arg Ala Cys Tyr Glu Thr Phe Asp
 35 40 45

Lys Pro Asn Pro Arg Thr Ala Ser Asn Ala Ala Tyr Leu Arg His Ile
 50 55 60

Met Glu Val Gly His Thr Ala Leu Leu Glu His Ala Asn Ala Thr Met
 65 70 75 80

Tyr Ile Arg Gly Ile Ser Arg Ser Ala Thr His Glu Leu Val Arg His
 85 90 95

Arg His Phe Ser Phe Ser Gln Leu Ser Gln Arg Phe Val His Ser Gly
 100 105 110

Glu Ser Glu Val Val Val Pro Thr Leu Ile Asp Glu Asp Pro Gln Leu
 115 120 125

Arg Glu Leu Phe Met His Ala Met Asp Glu Ser Arg Phe Ala Phe Asn
 130 135 140

Glu Leu Leu Asn Ala Leu Glu Glu Lys Leu Gly Asp Glu Pro Asn Ala
 145 150 155 160

Leu Leu Arg Lys Lys Gln Ala Arg Gln Ala Ala Arg Ala Val Leu Pro
 165 170 175

Asn Ala Thr Glu Ser Arg Ile Val Val Ser Gly Asn Phe Arg Thr Trp
 180 185 190

Arg His Phe Ile Gly Met Arg Ala Ser Glu His Ala Asp Val Glu Ile

Thr Val Phe Gly Asp Phe Glu Ile Glu Thr Leu Ala Asp Gly Ser Gln
225 230 235 240

Met Ala Thr Ser Pro Tyr Val Met Asp Phe
245 250

<211> 243

<212> PRT

<213> Mycobacteriophage 15

Met Lys Ala Lys Leu Ile Ala Ala Thr Glu Ile Asp Pro Gly Ala Leu
1 5 10 15

Arg Asp Ile Gly Phe Glu Val Asp Asp Phe Glu Glu Ser Lys Asp Glu
20 25 30

Asp Pro Tyr Phe Gly Asp Phe Asp Ala Asp Glu Leu Ala Glu Phe Ala
35 40 45

Gly Arg Asn Cys Tyr Arg Ser Phe His Arg Pro Asn Pro Ala Thr Ala
50 55 60

Glu Asn Glu Asp Tyr Leu Asn His Ile Ile Asp Leu Gly His Glu Ser
65 70 75 80

Val Phe Glu His Ala Ser Ala Thr Phe Tyr Ile Glu Ala Ser Arg Ser
85 90 95

Val Leu Thr Glu Leu Glu Arg His Arg His Leu Ser Phe Ser Val Val
100 105 110

Ser Gln Arg Tyr Val Asp Pro Thr Asp Leu Gly Ile His Leu Pro Pro
115 120 125

Ala Leu Phe Lys Leu His Pro Asp Asp Arg Asp Asp Leu Val His Ile
130 135 140

```
Met Glu Ser Val Ser Ser Glu Ile Asp Ala Val Tyr Glu His Ile Val
145                      150                      155                      160
```

Asn Arg Leu Ala Asp Arg Gly Leu Pro Arg Lys Gln Ala Arg Glu Ala
165 170 175

Ala Arg Ala Val Leu Pro Asn Met Thr Asn Ser Pro Met Val Val Thr
180 185 190

Gly Asn His Arg Ala Trp Arg Tyr Val Ile Lys Ala Arg Trp His Glu
195 200 205

Ala Ala Asp Ala Glu Ile Arg Glu Leu Ala Gly Glu Leu Leu Arg Gln
210 215 220

Leu Arg Gln Ile Ala Pro Asn Thr Tyr Gln Asp Ile Pro Asp Val Pro
 225 230 235 240

Tyr Ser Tyr

<210> 18

<211> 237

<212> PRT

<213> Bacteriophage phi-C31

<400> 18

Met Lys Val Asn Val Leu Ala Thr Thr Ala Leu Asn Pro Ser Pro Leu
 1 5 10 15

Leu Asp Ala Tyr Glu Tyr Arg Val Ser Gly Ala Ala Tyr Asn Arg Asp
 20 25 30

Arg Pro Thr Asp Ala Asp Ala Leu Gly Glu Ala Ala Gly Arg Ile Cys
 35 40 45

Tyr Lys Ser Phe Glu Arg Lys Asn Pro Ala Thr Ala Ser Asn Pro Gly
 50 55 60

Tyr Leu Gly Asn Ile Leu Ala Gln Gly His Phe Ser Val Leu Glu His
 65 70 75 80

Ala Ser Val Thr Phe Leu Val Arg Asp Val Ser Arg Ala Leu Leu Thr
 85 90 95

Glu Leu Ser Arg His Arg His Leu Ser Phe Ser Val Val Ser Gln Arg
 100 105 110

Tyr Val Asp His Ala Asp Thr Glu Pro Val Val Pro Pro Ala Ile Arg
 115 120 125

Gly Thr Glu Leu Glu Lys Pro Phe Arg Glu Asp Tyr Ala Glu Ala Leu
 130 135 140

Gln Ala Tyr Asp Ala Gly Val Lys Leu Leu Arg Ala Arg Gly Tyr Gly
 145 150 155 160

Arg Lys Gln Ala Arg Glu Ala Ala Arg Ala Leu Leu Pro Asn Ala Ala
 165 170 175

Pro Val Asp Met Val Val Thr Gly Asn Leu Arg Ala Trp Arg Asp Val
 180 185 190

Leu Gly Lys Arg Trp His Val Ala Ala Asp Ala Glu Ile Arg Glu Phe
 195 200 205

Ala Gly Arg Val Leu Asp His Leu His Ala Val Ala Pro Asn Ser Val
 210 215 220

Gln Asp Met Pro Thr Ser Pro Phe Gly Ser Asp Gly Lys
 225 230 235

<210> 19
 <211> 327
 <212> PRT
 <213> Aeropyrum pernix

<400> 19
 Met Ser Leu Ala Ala Ser Leu Glu Lys Ala Gly Leu Gly Ile Ser Val
 1 5 10 15
 Arg Leu Leu Glu Tyr Thr Gly Asp Gly Glu Arg Ile Val Ala Val Ala
 20 25 30
 Ser Lys Val Ser Leu Ser Arg Ser Pro Ala Glu Arg Leu Leu Ala Ile
 35 40 45
 Gly Glu Asp Glu Val Glu Thr Trp Ile Leu Glu Thr Phe Arg Arg Gln
 50 55 60
 His Phe Ser Pro Trp Glu His Ser Val Tyr Thr Phe Met Val Glu Gly
 65 70 75 80
 Leu Ser Arg Val Ala Ser His Gln Leu Val Arg His Arg Val Ala Ser
 85 90 95
 Tyr Thr Gln Leu Ser His Arg Tyr Ser Glu Gly Tyr Leu Arg Glu Ala
 100 105 110
 Ala Leu Lys Ala Cys Glu Ser Ile Gly Leu Asp Cys Pro Ser Lys Pro
 115 120 125
 Ala Glu Thr Glu Gly Gly Arg Lys Ala Ala Tyr Arg Leu Tyr Ser Gln
 130 135 140
 Ala Leu Glu Arg Ala Ala Arg Asp Phe Gly Ala Ser Glu Arg Phe Ala
 145 150 155 160
 Ile Ala Ala Lys Ala Phe Val Ile Pro Pro Thr Ile Leu Ala Arg Gly
 165 170 175
 Asp Gly Gly Asp Gly Val Val Glu Ala Tyr Leu Arg Ser Ala Ala Ile
 180 185 190
 Tyr Tyr Ser Leu Leu Ser Arg Gly Ala Arg Arg Glu Asp Ala Arg Tyr
 195 200 205
 Ile Leu Pro Asp Ala Leu Arg Thr Arg Ile Val Val Thr Met Asn Ala
 210 215 220
 Arg Glu Leu Ile Gln Val Phe Phe Pro Leu Arg Met Cys Thr Arg Ala
 225 230 235 240
 Gln Trp Glu Ile Arg His Ile Ala Trp Leu Leu Trp Arg Glu Leu Ser
 245 250 255
 Arg Val His Pro Arg Leu Phe Arg Trp Ala Gly Pro Ser Cys Val Leu
 260 265 270
 Arg Glu Asn Thr Leu Arg Thr Thr Pro Ala Ser Leu Tyr Ser Tyr Leu
 275 280 285
 Glu Gly Val Glu Arg Phe Thr Gln Pro Arg Cys Pro Glu Leu Val Glu

290

295

300

Asn Lys Ala Ile Pro Gly Cys Leu Arg Gln Ala Ala Ser Val Ala Pro
 305 310 315 320

Pro Gly Asp Gly Glu Tyr Glu
 325

<210> 20

<211> 269

<212> PRT

<213> Treponema denticola

<400> 20

His Ser Pro Met Ala His Cys Ile Ala Pro Glu Ala Glu Lys Ile Leu
 1 5 10 15

Asp Lys Glu Phe Lys Val Leu Asp Lys Gly Phe Ile Arg Leu Val Asp
 20 25 30

Tyr Met Gly Thr Asp Ala Arg Ile Val Gln Ser Ala Arg Val Ser Tyr
 35 40 45

Gly Glu Gly Thr Lys Thr Val Arg Glu Asp Ala Ala Leu Ile Asp Tyr
 50 55 60

Leu Leu Arg Asn Lys His Thr Ser Pro Phe Glu Gln Val Val Phe Thr
 65 70 75 80

Phe His Val Lys Leu Pro Ile Phe Val Ala Arg Gln Trp Ile Arg His
 85 90 95

Arg Thr Ala Arg Leu Asn Glu Ile Ser Gly Arg Tyr Ser Ile Leu Lys
 100 105 110

Ala Glu Phe Tyr Val Pro Ala Gly Lys Asp Ile Ala Leu Gln Ser Ser
 115 120 125

Asp Asn Lys Gln Gly Arg Met Asn Glu Ala Val Pro Gln Asp Leu Gln
 130 135 140

Asn Glu Val Ile Thr Ser Leu Gln Lys Gln Gln Glu Glu Ile Tyr Ala
 145 150 155 160

Gly Tyr Ser Lys Leu Leu Asp Lys Asn Ile Ala Arg Glu Leu Ala Arg
 165 170 175

Ile Asn Leu Pro Leu Ser Thr Tyr Thr Glu Trp Tyr Trp Gln Ile Asp
 180 185 190

Leu His Asn Leu Phe His Phe Leu Arg Leu Arg Met Asp Ala His Ala
 195 200 205

Gln Lys Glu Ile Arg Asp Tyr Ala Glu Val Met Phe Glu Ile Cys Lys
 210 215 220

Thr Val Thr Pro Leu Ala Cys Ala Ser Phe Glu Arg His Glu Lys Asn
 225 230 235 240

Ala Gly Lys Asp Ser Gly Leu Lys Gly Lys Glu Leu Glu
260 265

```
<210> 21
<211> 231
<212> PRT
<213> Helicobacter pylori
```

```
<400> 21
Met Trp Ile Thr Gln Glu Thr Trp Leu Lys Ala Leu Pro Trp Asn Lys
      1             5              10             15
```

Lys Arg Tyr Arg Ser Gln Ile Met Glu Val Ile Cys Lys His Tyr Thr
20 25 30

Pro Leu Asp Ile Ala Ser Gln Ala Ile Arg Thr Cys Trp Gln Ser Phe
35 40 45

Glu Tyr Ser Asp Asp Gly Gly Cys Lys Asp Lys Glu Leu Ile His Arg
50 55 60

Val Gly Asn Ile Phe Arg His Ser Ser Thr Leu Glu His Leu Tyr Tyr
65 70 75 80

Asn Phe Glu Ile Lys Gly Leu Ser Arg Gly Ala Leu Gln Glu Leu Ser
85 90 95

Arg His Arg Ile Ala Ser Leu Ser Val Lys Ser Ser Arg Tyr Thr Leu
100 105 110

Arg Glu Leu Lys Glu Val Glu Ser Phe Leu Pro Leu Asn Glu Thr Asn
115 120 125

Leu Glu Arg Ala Lys Glu Phe Leu Val Phe Val Asp Asn Glu Lys Val
130 135 140

Asn Ala Met Ser Val Leu Ala Leu Glu Asn Leu Arg Ile Leu Leu Ser
145 150 155 160

Glu His Asn Ile Lys Asn Asp Leu Ala Lys Tyr Ala Met Pro Glu Ser
165 170 175

Tyr Lys Thr His Leu Ala Tyr Ser Ile Asn Ala Arg Ser Leu Gln Asn
180 185 190

Phe Leu Thr Leu Arg Ser Ser Asn Lys Ala Leu Lys Glu Met Gln Asp
195 200 205

Leu Ala Lys Ala Leu Phe Asp Ala Leu Pro Gly Glu His Gln Tyr Leu
210 215 220

Phe Glu Asp Cys Leu Lys His
225 230

<210> 22

<211> 241

<212> PRT

<213> *Pyrococcus furiosus*

<400> 22

Met Val Arg Val Thr Leu Val Asn Tyr Thr Lys Arg Pro Leu Glu Thr
 1 5 10 15

Ile Thr Trp Ala Ala Leu Ile Ser Tyr Trp Gly Glu Trp Ser Thr Glu
 20 25 30

Ser Phe Glu Arg Ile Ser Glu Asn Asp Val Glu Lys His Leu Pro Arg
 35 40 45

Ile Leu Gly Tyr Gly His Glu Ser Ile Leu Glu His Ala Thr Phe Thr
 50 55 60

Phe Ser Ile Glu Gly Cys Ser Arg Val Cys Thr His Gln Leu Val Arg
 65 70 75 80

His Arg Ile Ala Ser Tyr Thr Gln Gln Ser Gln Arg Tyr Ile Val Leu
 85 90 95

Asp Glu Glu Asn Val Glu Glu Thr Phe Val Ile Pro Glu Ser Ile Lys
 100 105 110

Lys Asp Arg Glu Leu Tyr Glu Lys Trp Lys Lys Val Met Ala Glu Thr
 115 120 125

Ile Ser Leu Tyr Lys Glu Ser Ile Asn Arg Gly Val His Gln Glu Asp
 130 135 140

Ala Arg Phe Ile Leu Pro Gln Ala Val Lys Thr Lys Ile Ile Val Thr
 145 150 155 160

Met Asn Leu Arg Glu Leu Lys His Phe Phe Gly Leu Arg Leu Cys Glu
 165 170 175

Arg Ala Gln Trp Glu Ile Arg Glu Val Ala Trp Lys Met Leu Glu Glu
 180 185 190

Met Ala Lys Arg Asp Asp Ile Arg Pro Ile Ile Lys Trp Ala Lys Leu
 195 200 205

Gly Pro Arg Cys Ile Gln Phe Gly Tyr Cys Pro Glu Arg Asp Leu Met
 210 215 220

Pro Pro Gly Cys Leu Lys Lys Thr Arg Lys Lys Trp Glu Lys Val Ala
 225 230 235 240

Glu

<210> 23

<211> 282

<212> PRT

<213> *Rhodobacter capsulatus*

<400> 23

Thr Thr Leu Arg Ala Thr Ser Ala Gly Met Glu Ala Arg Leu Tyr Gln
 1 5 10 15

Ala His Pro Val Leu Asp His Gly Leu Ile Arg Val Ile Asp Tyr Met
 20 25 30

Gly Asp Asp Ala Ala Ile Cys Gln Ala Ala Arg Val Ser Tyr Gly Arg
 35 40 45

Gly Thr Lys Ala Val Ser Asp Asp Arg Gly Leu Ile Arg Tyr Leu Met
 50 55 60

Arg His Trp His Ser Thr Pro Phe Glu Met Cys Glu Val Lys Phe His
 65 70 75 80

Val Lys Leu Pro Ile Phe Val Ala Arg Gln Trp Ile Arg His Arg Thr
 85 90 95

Ala Asn Val Asn Glu Tyr Ser Ala Arg Tyr Ser Val Met Asp Arg Glu
 100 105 110

Phe Tyr Ile Pro Ala Pro Glu His Leu Ala Ala Gln Ser Thr Val Asn
 115 120 125

Asn Gln Gly Arg Gly Gln Val Leu Glu Gly Ala Glu Ala Ala Arg Val
 130 135 140

Leu Asp Leu Leu Arg Glu Asp Ala Met Arg Ala Tyr Asp His Tyr Glu
 145 150 155 160

Asp Met Leu Thr Pro Asp Ala Asp Ala Gly Lys Leu Gly Leu Ala Arg
 165 170 175

Glu Leu Ala Arg Met Asn Leu Pro Ala Asn Val Tyr Thr Gln Trp Tyr
 180 185 190

Trp Lys Ile Asp Leu His Asn Leu Phe His Phe Leu Arg Leu Arg Ala
 195 200 205

Asp Val His Ala Gln Tyr Glu Ile Arg Val Tyr Ala Gln Ile Met Gly
 210 215 220

Asp Ile Val Lys Asp Trp Val Pro Gln Ala Tyr Glu Ala Phe Glu Asp
 225 230 235 240

Tyr Arg Leu Gly Ala Val Ser Val Ser Ala Lys Ala Lys Glu Val Leu
 245 250 255

Lys Arg Arg Leu Ala Gly Glu Val Val Thr Ala Glu Thr Ser Gly Met
 260 265 270

Ser Lys Gly Glu Trp Arg Glu Phe Val Glu
 275 280

<210> 24

<211> 247

<212> PRT

<213> Halobacterium salinarum

<400> 24

Met Arg Val Arg Leu Leu Glu Ala Thr Glu Asn Pro Glu Glu Leu Ile
 1 5 10 15

Cys Gln Ser Ala Arg Asn Asp Tyr Met Ser Asp Trp Val Gly Asp Thr
 20 25 30

Pro Leu Asp Thr Ala Met Ala Ser Val Asp Gly Asp Thr Thr Asp Glu
 35 40 45

Lys Leu Ser Asn Leu Ile Ala Gln Leu Leu Thr Arg Gly His Tyr Gly
 50 55 60

Pro Phe Glu His Pro Ser Ala Thr Phe Ala Ile Glu Gly Val Ser Arg
 65 70 75 80

Ser Cys Met Ala Gln Leu Thr Arg His Arg His Ala Ser Phe Asp Val
 85 90 95

Gln Ser Met Arg Tyr Val Ala Phe Asp Asp Val Asp Pro Ala Ala Val
 100 105 110

Ala Glu Gly Glu Leu Val Val Thr Pro Pro Ser Ala Thr Asp Pro Asp
 115 120 125

Trp Val Gly Arg Asn Gln Asp Ala Gly Asp Ile Asp Glu Glu Thr Met
 130 135 140

Ala Glu Arg Gln Ala Val Phe Gln Ala Ser Val Arg Arg Ala Val Glu
 145 150 155 160

Asp Tyr Gln Glu Leu Leu Gly Leu Gly Met Pro Pro Val Asp Ala Arg
 165 170 175

Phe Val Leu Pro Ile Gly Thr Glu Val Asn Val Val Ile Thr Leu Asn
 180 185 190

Pro Arg Ser Leu Met His Val Ala Asp Met Arg Ala Ala Ala Asp Ala
 195 200 205

Gln Trp Glu Ile Arg Glu Leu Thr Glu Gln Leu Leu Asp Ala Ala Ala
 210 215 220

Gln Trp Cys Pro His Thr Phe Glu Tyr Tyr Asp Ala Glu Met Lys His
 225 230 235 240

Arg Lys Asn Arg Leu Ala Pro
 245

<210> 25

<211> 254

<212> PRT

<213> Mycobacterium leprae

<400> 25

Met Ala Gln Ile Ala Pro Leu Arg Val Gln Leu Ile Ala Lys Thr Glu
 1 5 10 15

Phe Leu Ala Pro Pro Asp Val Ser Trp Thr Thr Asp Ala Asp Gly Gly
 20 25 30
 Ser Ala Leu Val Glu Phe Ala Gly Arg Ala Cys Tyr Gln Ser Trp Ser
 35 40 45
 Lys Pro Asn Pro Arg Thr Ala Thr Asn Ala Ala Tyr Ile Lys His Ile
 50 55 60
 Ile Asp Val Gly His Val Ala Val Leu Glu His Ala Ser Val Ser Phe
 65 70 75 80
 Tyr Ile Ser Gly Ile Ser Arg Ser Cys Thr His Glu Leu Ile Arg His
 85 90 95
 Arg His Phe Ser Tyr Ser Gln Leu Ser Gln Arg Tyr Val Pro Glu Lys
 100 105 110
 Asp Ala Gln Val Val Val Pro Pro Asp Met Glu Asp Asp Asp Glu Leu
 115 120 125
 Gln Gln Ile Leu Ile Ala Ala Val Glu Ala Ser Arg Ala Thr Tyr Thr
 130 135 140
 Glu Leu Leu Val Lys Leu Asn Ala Lys Leu Met Ala Gly Glu Leu Gly
 145 150 155 160
 Gly Asn Arg Ala Val Leu Arg Arg Lys Gln Ala Arg Gln Ala Ala His
 165 170 175
 Ala Val Leu Pro Asn Ala Asn Glu Thr Arg Ile Val Val Thr Gly Asn
 180 185 190
 Tyr Arg Ala Trp Arg His Phe Ile Ala Met Arg Ala Ser Glu His Ala
 195 200 205
 Asp Val Glu Ile Arg Arg Leu Ala Ile Val Cys Leu Arg Arg Leu Val
 210 215 220
 Asp Val Ala Pro Ala Val Phe Ala Asp Phe Glu Ile Thr Ala Leu Ala
 225 230 235 240
 Asp Gly Thr Glu Val Ala Thr Ser Pro Leu Ala Thr Glu Ala
 245 250

<210> 26

<211> 208

<212> PRT

<213> Helicobacter pylori

<400> 26

Met Glu Val Ile Cys Lys His Tyr Thr Pro Leu Asp Ile Ala Ser Gln
 1 5 10 15

Ala Ile Arg Thr Cys Trp Gln Ser Phe Glu Tyr Ser Asp Asp Gly Gly
 20 25 30

Cys Lys Asp Arg Asp Leu Ile His Arg Val Gly Asn Ile Phe Arg His
 35 40 45

Ser Ser Thr Leu Glu His Leu Tyr Tyr Asn Phe Glu Ile Lys Gly Leu
 50 55 60
 Ser Arg Gly Ala Leu Gln Glu Leu Ser Arg His Arg Ile Ala Ser Leu
 65 70 75 80
 Ser Val Lys Ser Ser Arg Tyr Thr Leu Arg Glu Leu Lys Glu Val Glu
 85 90 95
 Ser Phe Leu Pro Leu Asn Glu Thr Asn Leu Glu Arg Ala Lys Glu Phe
 100 105 110
 Leu Val Phe Val Asp Asp Glu Lys Val Asn Glu Met Ser Val Leu Ala
 115 120 125
 Leu Glu Asn Leu Arg Val Leu Leu Ser Glu His Asn Ile Lys Asn Asp
 130 135 140
 Leu Ala Lys Tyr Ala Met Pro Glu Ser Tyr Lys Thr His Leu Ala Tyr
 145 150 155 160
 Ser Ile Asn Ala Arg Ser Leu Gln Asn Leu Leu Thr Leu Arg Ser Ser
 165 170 175
 Asn Lys Ala Leu Lys Glu Met Gln Asp Leu Ala Lys Ala Leu Phe Asp
 180 185 190
 Ala Leu Pro Tyr Glu His Gln Tyr Leu Phe Glu Asp Cys Leu Lys His
 195 200 205

<210> 27
 <211> 207
 <212> PRT
 <213> Campylobacter jejuni

<400> 27
 Met Gln Ile Thr Leu Leu Phe His Thr Pro Leu Ser Val Cys Ser His
 1 5 10 15
 Ala Thr Arg Thr Cys Trp Gln Ser Phe Glu Lys Gly Asp Cys Gly Gly
 20 25 30
 Glu Lys Asp Lys Glu Leu Ile Asp Arg Val Gly Asn Lys Phe Lys His
 35 40 45
 Ala Ser Thr Leu Glu His Leu Asn Tyr Thr Phe Tyr Ile Gln Gly Ile
 50 55 60
 Ser Arg Ala Cys Leu Gln Glu Val Ala Arg His Arg His Thr Ser Pro
 65 70 75 80
 Ser Val Lys Ser Thr Arg Tyr Thr Leu Lys Glu Leu Arg Asn Glu Ala
 85 90 95
 Glu Phe Lys Ile Gly Asp Phe Glu Asn Ala Ser Arg Tyr Leu Val Leu

100 105 110
 Cys Gly Asn Glu Glu Val Asp Asn Ala Ser Ile Lys Ala Leu Glu Asn
 115 120 125
 Leu Arg Thr Ile Leu Gln Lys Ser Ile Ser Leu Asp Ile Ala Lys Tyr
 130 135 140
 Cys Leu Pro Glu Ser Tyr Lys Thr Glu Leu Thr Leu Thr Ile Asn Ala
 145 150 155 160
 Arg Ser Leu Gln Asn Phe Ile Ser Leu Arg Ser Ser Lys Ser Ala Leu
 165 170 175
 Trp Glu Ile Arg Asn Leu Ala Asn Ala Leu Phe Glu Ala Leu Pro Gln
 180 185 190
 Glu His Lys Phe Ile Phe Glu His Cys Leu His Lys Asp Ile Glu
 195 200 205

 <210> 28
 <211> 260
 <212> PRT
 <213> Sulfolobus solfataricus

 <400> 28
 Met Ile Ser Val Lys Leu Val Ser Tyr Thr Asn Asp Gly Glu Lys Val
 1 5 10 15
 Ile Ala Ile Ala Ala Lys Met Ser Arg Ser Arg Lys Gly Trp Asp Tyr
 20 25 30
 His Glu Lys Asp Met Thr Asp Asp Glu Ile Glu Thr Trp Ile Arg Asp
 35 40 45
 Ala Ile Leu His Gly Tyr Trp Ser Val Leu Glu His Ser Val Tyr Thr
 50 55 60
 Phe Ser Ile Glu Glu Ile Ser Arg Val Ala Ser His Gln Leu Val Arg
 65 70 75 80
 His Arg Ile Ala Ser Tyr Thr Gln Met Ser His Arg Phe Ala Lys Pro
 85 90 95
 Ile Asp Glu Tyr Tyr Lys Pro Ile Ile Pro Pro Ser Ala Lys Lys Arg
 100 105 110
 Ser Lys Glu Leu Val Glu Lys Ala Tyr Lys Glu Ala Tyr Asp Asn Tyr
 115 120 125
 Tyr Thr Leu Leu Glu Ser Gly Val Pro Glu Glu Asp Ala Arg Tyr Val
 130 135 140
 Leu Pro Asn Gly Val Asn Thr Asn Ile Val Val Thr Met Asn Ala Arg
 145 150 155 160
 Glu Leu Tyr Asn Phe Phe Ser Leu Arg Leu Cys Ser Arg Ala Gln Trp
 165 170 175

Glu Ile Arg Ala Ile Ala Trp Lys Met Leu Glu Glu Val Lys Lys Val
 180 185 190

His Pro Arg Leu Phe Lys Tyr Thr Gly Pro Asn Cys Ile Ile His Glu
 195 200 205

Asn Phe Ile Arg Asn Glu Asn Glu Ser Ile Thr Leu Glu Asp Ile Phe
 210 215 220

Lys Asp Tyr Lys Leu Glu Phe Ile Ser Gln Arg Cys Ile Glu Gly Val
 225 230 235 240

Leu Arg Asp Gly Ile Lys Lys Cys Ile Ile Asn Ser Arg Ser Val Leu
 245 250 255

Asp Asn Ile Lys
 260

<210> 29

<211> 250

<212> PRT

<213> Mycobacterium bovis

<400> 29

Val Ala Glu Thr Ala Pro Leu Arg Val Gln Leu Ile Ala Lys Thr Asp
 1 5 10 15

Phe Leu Ala Pro Pro Asp Val Pro Trp Thr Thr Asp Ala Asp Gly Gly
 20 25 30

Pro Ala Leu Val Glu Phe Ala Gly Arg Ala Cys Tyr Gln Ser Trp Ser
 35 40 45

Lys Pro Asn Pro Lys Thr Ala Thr Asn Ala Gly Tyr Leu Arg His Ile
 50 55 60

Ile Asp Val Gly His Phe Ser Val Leu Glu His Ala Ser Val Ser Phe
 65 70 75 80

Tyr Ile Thr Gly Ile Ser Arg Ser Cys Thr His Glu Leu Ile Arg His
 85 90 95

Arg His Phe Ser Tyr Ser Gln Leu Ser Gln Arg Tyr Val Pro Glu Lys
 100 105 110

Asp Ser Arg Val Val Val Pro Pro Gly Met Glu Asp Asp Ala Asp Leu
 115 120 125

Arg His Ile Leu Thr Glu Ala Ala Asp Ala Ala Arg Ala Thr Tyr Ser
 130 135 140

Glu Leu Leu Ala Lys Leu Glu Ala Lys Phe Ala Asp Gln Pro Asn Ala
 145 150 155 160

Ile Leu Arg Arg Lys Gln Ala Arg Gln Ala Ala Arg Ala Val Leu Pro
 165 170 175

Asn Ala Thr Glu Thr Arg Ile Val Val Thr Gly Asn Tyr Arg Ala Trp
 180 185 190

Arg His Phe Ile Ala Met Arg Ala Ser Glu His Ala Asp Val Glu Ile
195 200 205

Arg Arg Leu Ala Ile Glu Cys Leu Arg Gln Leu Ala Ala Val Ala Pro
210 215 220

Ala Val Phe Ala Asp Phe Glu Val Thr Thr Leu Ala Asp Gly Thr Glu
225 230 235 240

Val Ala Thr Ser Pro Leu Ala Thr Glu Ala
245 250

<210> 30
<211> 251
<212> PRT
<213> Corynebacterium diphtheriae

<400> 30
Met Gln Gln Ser Leu Leu Asp Val Gln Leu Val Ala Cys Ser Thr Phe
1 5 10 15

Thr Thr Pro Ser Gly Val Asp Trp Lys Val Asp Ser Ala Ala Thr Asp
20 25 30

Ser Glu Ala Leu Val Glu Phe Ala Gly Arg Ala Cys Tyr Glu Thr Phe
35 40 45

Asp Lys Pro Asn Pro Arg Thr Ala Ala Asn Asp Ala Tyr Ile Arg His
50 55 60

Ile Met Glu Val Gly His Met Ala Leu Leu Glu His Pro Thr Ala Thr
65 70 75 80

Val Tyr Ile Arg Gly Leu Ser Arg Ser Ala Thr His Glu Leu Val Arg
85 90 95

His Arg His Phe Ser Phe Ser Gln Leu Ser Gln Arg Phe Val His Ser
100 105 110

Asp Glu Thr His Val Val Ile Pro Pro Leu Ile Asp Asn Asp Pro Gln
115 120 125

Leu Arg Glu Leu Phe Leu Ser Thr Val Asp Glu Val Arg Phe Ala Tyr
130 135 140

Ser Glu Leu Met Thr Ala Leu Asp Asn Lys Leu Ala Asp Glu Pro Asn
145 150 155 160

Ala Ile Leu Arg Arg Lys Gln Ala Arg Gln Ala Ala Arg Ser Ile Leu
165 170 175

Pro Asn Ala Thr Glu Ser Arg Ile Val Val Thr Gly Asn Phe Arg Ala
180 185 190

Trp Arg His Phe Ile Gly Met Arg Ala Thr Glu His Ala Asp Val Glu
195 200 205

Ile Arg Ser Leu Ala Val Arg Cys Leu Glu Ile Leu Lys Glu Lys Ala

210 215 220
 Pro Thr Val Phe Ser Asp Phe Glu Thr Ser Val Leu Ser Asp Gly Ser
 225 230 235 240
 Ile Met Ala Thr Ser Pro Tyr Val Thr Asp Tyr
 245 250

 <210> 31
 <211> 161
 <212> PRT
 <213> Bacillus anthracis

 <400> 31
 Asp Arg Leu Ile Arg His Ile Val Gly Ser Gly His Ala Ser Thr Leu
 1 5 10 15
 Glu His Leu Thr Tyr Thr Phe Ala Val Glu Gly Val Ser Arg Ala Leu
 20 25 30
 Leu Ala Gln Leu Thr Arg His Arg Val Gly Phe Ser Tyr Ser Val Gln
 35 40 45
 Ser Gln Arg Tyr Val Arg Met Gly Ser Asn Asp Lys Ile Gly Gly Phe
 50 55 60
 Asp Tyr Val Val Pro Glu Thr Val Lys Ala Lys Gly Glu Gln Val Val
 65 70 75 80
 Asn Ala Tyr Asn Glu Met Met Tyr Lys Leu Gln Ser Gly Tyr Asp Leu
 85 90 95
 Leu Arg Thr Leu Gly Ile Pro Ala Glu Asp Ala Arg Ser Val Leu Pro
 100 105 110
 Asn Ala Ala Ala Thr Asn Leu Val Leu Thr Val Asn Leu Arg Gly Leu
 115 120 125
 Leu Asp Phe Tyr Asn Lys Arg Arg Lys Gly Lys Gly Ala Gln Ala Glu
 130 135 140
 Ile Ala Glu Leu Ala Glu Gln Leu Arg Gln Glu Val Val Lys Val Glu
 145 150 155 160
 Lys

<210> 32
 <211> 212
 <212> PRT
 <213> Clostridium difficile

<400> 32
 Lys Leu Ile Ser His Thr Pro Glu Pro Glu Lys Val Ile Ala Met Ala
 1 5 10 15
 Ala Lys Leu Cys Tyr Ser Pro Val Gly Thr Asp Glu Ile Glu Lys Asp

20 25 30
 Leu Thr Asp Glu Ser Ile Glu Lys Phe Leu Asn Met Leu Leu Ser Ile
 35 40 45
 Gly His Gly Ser Ile Leu Glu His Ala Ser Phe Thr Phe Ser Ile Glu
 50 55 60
 Gly Ile Ser Arg Ala Cys Ser His Gln Ile Val Arg His Arg Ile Ala
 65 70 75 80
 Ser Phe Ser Gln Gln Ser Gln Arg Tyr Val Lys Leu Glu Gln Phe Glu
 85 90 95
 Tyr Ile Ile Pro Pro Glu Ile Glu Lys Glu Leu Phe Ile Asp Ser Met
 100 105 110
 Lys Lys Asp Gln Glu Asn Tyr Asp Lys Leu Val Glu Ile Leu Phe Glu
 115 120 125
 Asn His Tyr Asn Asp Leu Ile Lys Asn Gly Lys Asn Glu Lys Thr Ala
 130 135 140
 Lys Arg Gln Ala Glu Lys Lys Ala Ile Glu Asp Ala Arg Tyr Val Phe
 145 150 155 160
 Pro Asn Ala Cys Glu Thr Lys Met Val Phe Thr Ile Asn Ala Arg Ser
 165 170 175
 Leu Phe Asn Phe Phe Glu His Arg Cys Cys Glu Arg Ala Gln Trp Glu
 180 185 190
 Ile Arg Asn Leu Ala Val Glu Met Leu Arg Glu Val Lys Lys Val Ala
 195 200 205
 Pro Ile Leu Phe
 210

<210> 33
 <211> 515
 <212> PRT
 <213> Thermoplasma volcanium

<400> 33
 Met Glu Phe Ser Asn Ala Glu Arg Asp Val Phe Leu Ile Lys Ile Glu
 1 5 10 15
 Lys Met Ile Asp Arg Gly Ala Leu Met Ser Arg Tyr Ser Arg Ala Ser
 20 25 30
 Asp Pro Asp Ile Arg Ser Val Tyr Glu Lys Glu Phe Lys Thr Gly Ala
 35 40 45
 Lys Ser Gly Glu Glu Phe Tyr Arg Arg Ile Phe Leu Glu Tyr Gly Asp
 50 55 60
 Glu Ser Ile Ala Glu Leu Thr Thr Ala Gln Met Gly Ile Gln Asn Val
 65 70 75 80

Ser Asn Val Ala Ser Lys Ile Ile Glu Glu Ile Arg Ile Gly Leu Ser
 85 90 95
 Tyr Leu Glu Lys Ser Thr Arg Tyr Val Arg Tyr Asp Lys Lys Val Asp
 100 105 110
 Gly Arg Tyr Leu Tyr Ile Ser Pro Glu Arg Ile Gly Ile Ser Gly Asp
 115 120 125
 Asp Ala Lys Asp Tyr Val Gln Leu Cys Asp Asn Leu Phe Glu Phe Tyr
 130 135 140
 Ser Lys Ala Leu Pro Gln Val Glu Asp Tyr Leu Arg His Lys Phe Pro
 145 150 155 160
 Gln Asp Lys Leu Val Phe Gln Asn Ala Gly Gly Lys Thr Leu Thr Glu
 165 170 175
 Met Asp Gln Asn Glu Lys Lys Ile Ala Glu Arg Ser Tyr Ile Asn Ala
 180 185 190
 Val Arg Ser Arg Ala Leu Asp Asp Val Arg Tyr Ile Leu Pro Ala Ser
 195 200 205
 Thr Leu Thr Asn Ile Gly Ile Ser Gly Asn Gly Arg Ala Leu Ile His
 210 215 220
 Leu Ile Gln Lys Leu Met Glu Tyr Glu Ile Pro Glu Thr Thr Lys Leu
 225 230 235 240
 Ala Lys Asp Ile Tyr Asp Glu Leu Lys Pro Glu Leu Pro Gln Leu Ile
 245 250 255
 Asp Asp Ala Leu Ser Gly His Gly Leu Glu Ile Ile Asn Phe Lys Lys
 260 265 270
 Asn Leu Met Gly Leu Phe Pro Tyr Asp Leu Thr Gly Asn Phe Glu Arg
 275 280 285
 Ile Arg Leu Leu Ser Tyr Gly Glu Glu Asp Lys Glu Leu Arg Lys Val
 290 295 300
 Ala Ser Leu Ile Glu Tyr Pro Phe His Gly Asp Ala Ala Ser Leu Tyr
 305 310 315 320
 Ser Arg Ser Ser Glu Ser Tyr Val Lys Tyr Met Lys Glu Leu Ile Glu
 325 330 335
 Ser Ile Arg Ser Leu Arg Ala Asn Arg Arg Met Lys Pro Gly Arg Ala
 340 345 350
 Phe Glu Ser Val Asn Tyr Val Phe Glu Leu Asn Leu Asn Tyr Gly Ser
 355 360 365
 Phe Arg Asp Leu Gln Arg His Arg Phe Leu Gly Ile Ile Arg Lys Pro
 370 375 380
 Leu Thr Ala Ala Tyr Gly Tyr Asp Thr Pro Pro Val Ile Ser Ala Ile
 385 390 395 400
 Asp Glu Leu Lys Thr Gln Tyr Asp Glu Leu Met Ala Asn Ser Ser Ser

405 410 415
 Phe Tyr Gln Arg Leu Arg Glu Lys Tyr Gly Pro Trp Ile Ser Gln Tyr
 420 425 430
 Val Val Pro Phe Ala Phe Lys Tyr Pro Ile Thr Phe Ser Thr Asn Leu
 435 440 445
 Ser Glu Val Thr Tyr Phe Val Glu Leu Arg Ser Thr Ala Gln Ala His
 450 455 460
 Phe Asp Leu Arg Asp Ile Ala Val Ser Met Tyr Arg Glu Val Ser Lys
 465 470 475 480
 Val His Pro Thr Leu Ser Arg Ile Ile Lys Phe Val Asp Thr Ala Asp
 485 490 495
 Tyr Pro Leu Gly Arg Leu Ser Ala Glu Phe Arg Lys Glu Ser Lys Lys
 500 505 510
 Ala Gly Ile
 515

 <210> 34
 <211> 499
 <212> PRT
 <213> Thermoplasma acidophilum

 <400> 34
 Met Ile Asp Arg Gly Ala Leu Met Ser Arg Tyr Ser Arg Ala Ser Asp
 1 5 10 15
 Pro Asp Ile Arg Ser Val Phe His Arg Glu Phe Glu Gly Asn Gln Lys
 20 25 30
 Arg Ser Glu Asp Phe Tyr Arg Arg Ile Phe Leu Glu Tyr Gly Asp Glu
 35 40 45
 Ser Ile Ala Glu Leu Val Thr Ala Gln Val Gly Ile Gln Asn Val Ser
 50 55 60
 Asn Val Ile Ser Lys Val Ile Glu Glu Ile Arg Ile Gly Leu Ser Tyr
 65 70 75 80
 Leu Glu Lys Ser Thr Arg Tyr Val Ala Tyr Asp Arg Lys Val Asp Gly
 85 90 95
 His Tyr Leu Phe Met Gln Ala Glu Lys Ile Gly Leu Ser Gly Glu Ala
 100 105 110
 Ala Arg Glu Tyr Thr Asp Leu Cys Asn Arg Leu Phe Asp Leu Tyr Ser
 115 120 125
 Ser Thr Leu Pro Arg Ile Glu Glu Glu Ile Ser Arg Gln Trp Pro Ile
 130 135 140
 Glu Ser Phe Asp Phe Asn Ile Asp Gly Asn Pro Arg Asn Tyr Lys Glu
 145 150 155 160

Leu Asp Glu Asn Gly Arg Lys Leu Ala Gln Lys Ser Tyr Arg Ser Ser
 165 170 175
 Val Arg Ser Arg Ala Leu Asp Asp Ala Arg Phe Ile Leu Pro Ala Ser
 180 185 190
 Thr Leu Thr Asn Met Gly Val Ser Gly Asn Gly Arg Ser Phe Ile His
 195 200 205
 Leu Ile Gln Lys Leu Met Glu Tyr Gly Val Pro Glu Ser Glu Arg Leu
 210 215 220
 Ala His Asp Leu Tyr Glu Glu Leu Lys Gly Glu Phe Pro Gln Ile Ile
 225 230 235 240
 Asp Asp Ala Leu Ser Gln His Gly Gln Asp Ile Ile Asn Tyr Lys Arg
 245 250 255
 Ser Leu Ala Ser Leu Phe Pro Tyr Thr Asp Gly Gly Arg Phe Glu Lys
 260 265 270
 Val Arg Leu Ile Lys Tyr Ser Asn Glu Arg Glu Glu Met Gln Lys Val
 275 280 285
 Leu Ala Leu Leu Met Tyr Pro Phe Ala Glu Asp Ala Ser Gly Ile Ile
 290 295 300
 Ser Arg Ile Lys Ala Met Glu Leu Ser Glu Ala Ser Ala Ile Leu Glu
 305 310 315 320
 Arg Ile Arg Asp Leu Arg Lys Asn Arg Arg Met Lys Val Gly Arg Pro
 325 330 335
 Phe Glu Ala Val Asn Tyr Val Phe Glu Val Thr Thr Asn Tyr Gly Ala
 340 345 350
 Phe Arg Asp Leu Gln Arg His Arg Phe Leu Ser Ile Val Arg Lys Pro
 355 360 365
 Leu Thr Val Ser Tyr Gly Phe Asp Val Pro Pro Ile Ile Ala Lys Met
 370 375 380
 Pro Asp Leu Ser Glu Glu Tyr Ala Glu Ala Met Lys Asp Ala Glu Arg
 385 390 395 400
 Val Tyr Arg Ile Ile Lys Glu Arg Tyr Gly Ala Trp Ile Ala Gln Tyr
 405 410 415
 Ala Val Pro Phe Ala Tyr Arg Tyr Pro Val Val Phe Thr Thr Asn Leu
 420 425 430
 Ala Glu Ala Thr Tyr Phe Ile Glu Leu Arg Ser Thr Pro Gln Ala His
 435 440 445
 Phe Asp Leu Arg Asp Ile Ala Ile Arg Met Tyr Asn Glu Ile Lys Ser
 450 455 460
 Val His Pro Ser Leu Ala Gly Ile Ile Lys Phe Val Asp Thr Gly Asp
 465 470 475 480
 Tyr Pro Leu Gly Arg Leu Ser Ala Glu Val Arg Lys Asn Val Lys Ala

485

490

495

Gly Gly Ile

<210> 35

<211> 531

<212> PRT

<213> Chlamydophila pneumoniae

<400> 35

Met Leu Gly Lys Glu Glu Glu Phe Thr Cys Lys Gln Lys Gln Cys Leu
 1 5 10 15

Ser His Phe Val Thr Asn Leu Thr Ser Asp Val Phe Ala Leu Lys Asn
 20 25 30

Leu Pro Glu Val Val Lys Gly Ala Leu Phe Ser Lys Tyr Ser Arg Ser
 35 40 45

Val Leu Gly Leu Arg Ala Leu Leu Leu Lys Glu Phe Leu Ser Asn Glu
 50 55 60

Glu Asp Gly Asp Val Cys Asp Glu Ala Tyr Asp Phe Glu Thr Asp Val
 65 70 75 80

Gln Lys Ala Ala Asp Phe Tyr Gln Arg Val Leu Asp Asn Phe Gly Asp
 85 90 95

Asp Ser Val Gly Glu Leu Gly Gly Ala His Leu Ala Met Glu Asn Val
 100 105 110

Ser Ile Leu Ala Ala Lys Val Leu Glu Asp Ala Arg Ile Gly Gly Ser
 115 120 125

Pro Leu Glu Lys Ser Thr Arg Tyr Val Tyr Phe Asp Gln Lys Val Arg
 130 135 140

Gly Glu Tyr Leu Tyr Tyr Arg Asp Pro Ile Leu Met Thr Ser Ala Phe
 145 150 155 160

Lys Asp Met Phe Leu Gly Thr Cys Asp Phe Leu Phe Asp Thr Tyr Ser
 165 170 175

Ala Leu Ile Pro Gln Val Arg Ala Tyr Phe Glu Lys Leu Tyr Pro Lys
 180 185 190

Asp Ser Lys Thr Pro Ala Ser Ala Tyr Ala Thr Ser Leu Arg Ala Lys
 195 200 205

Val Leu Asp Cys Ile Arg Gly Leu Leu Pro Ala Ala Thr Leu Thr Asn
 210 215 220

Leu Gly Phe Phe Gly Asn Gly Arg Phe Trp Gln Asn Leu Ile His Lys
 225 230 235 240

Leu Gln Gly His Asn Leu Ala Glu Leu Arg Arg Leu Gly Asp Glu Ser
 245 250 255

Leu Thr Glu Leu Met Lys Val Ile Pro Ser Phe Val Ser Arg Ala Glu
 260 265 270
 Pro His His His His His Gln Ala Met Met Gln Tyr Arg Arg Ala Leu
 275 280 285
 Lys Glu Gln Leu Lys Gly Leu Ala Glu Gln Ala Thr Phe Ser Glu Glu
 290 295 300
 Met Ser Ser Ser Pro Ser Val Gln Leu Val Tyr Gly Asp Pro Asp Gly
 305 310 315 320
 Ile Tyr Lys Val Ala Ala Gly Phe Leu Phe Pro Tyr Ser Asn Arg Ser
 325 330 335
 Leu Thr Asp Leu Ile Asp Tyr Cys Lys Lys Met Pro His Glu Asp Leu
 340 345 350
 Val Gln Ile Leu Glu Ser Ser Val Ser Ala Arg Glu Asn Arg Arg His
 355 360 365
 Lys Ser Pro Arg Gly Leu Glu Cys Val Glu Phe Gly Phe Asp Ile Leu
 370 375 380
 Ala Asp Phe Gly Ala Tyr Arg Asp Leu Gln Arg His Arg Thr Leu Thr
 385 390 395 400
 Gln Glu Arg Gln Leu Leu Ser Thr His His Gly Tyr Asn Phe Pro Val
 405 410 415
 Glu Leu Leu Asp Thr Pro Met Glu Lys Ser Tyr Arg Glu Ala Met Glu
 420 425 430
 Arg Ala Asn Glu Thr Tyr Asn Glu Ile Val Gln Glu Phe Pro Glu Glu
 435 440 445
 Ala Gln Tyr Met Val Pro Met Ala Tyr Asn Ile Arg Trp Phe Phe His
 450 455 460
 Val Asn Ala Arg Ala Leu Gln Trp Ile Cys Glu Leu Arg Ser Gln Pro
 465 470 475 480
 Gln Gly His Gln Asn Tyr Arg Thr Ile Ala Thr Gly Leu Val Arg Glu
 485 490 495
 Val Val Lys Phe Asn Pro Met Tyr Glu Leu Phe Phe Lys Phe Val Asp
 500 505 510
 Tyr Ser Asp Ile Asp Leu Gly Arg Leu Asn Gln Glu Met Arg Lys Glu
 515 520 525
 Pro Thr Thr
 530

<210> 36

<211> 529

<212> PRT

<213> Chlamydia trachomatis

<400> 36

```

Met Leu Ser Lys Glu Gly Gly Phe Ser Glu Glu Gln Arg Ala Arg Leu
 1              5              10              15

Ser His Phe Val Thr Asn Leu Asp Ser Pro Ile Phe Ala Leu Lys Asn
      20              25              30

Leu Pro Glu Val Val Lys Gly Ala Leu Phe Ser Lys Tyr Ser Arg Ser
      35              40              45

Thr Leu Gly Leu Arg Ala Leu Leu Leu Lys Glu Phe Leu Asp Gly Glu
 50              55              60

Gly Gly Asn Phe Leu Asp Asp Asp Gln Gln Asp Cys Glu Leu Gly Ile
 65              70              75              80

Gln Lys Ala Ala Asp Phe Tyr Arg Arg Val Leu Asp Asn Phe Gly Asp
      85              90              95

Asp Ser Val Gly Glu Leu Gly Gly Ala His Leu Ala Leu Glu Gln Val
      100              105              110

Ser Met Leu Ala Ala Lys Ile Leu Glu Asp Ala Arg Ile Gly Gly Ser
      115              120              125

Pro Leu Glu Lys Ser Ser Arg Tyr Val Tyr Phe Asp Gln Lys Val Asn
      130              135              140

Gly Glu Tyr Leu Tyr Tyr Arg Asp Pro Ile Leu Met Thr Ser Ala Phe
      145              150              155              160

Lys Asp Val Phe Leu Asp Thr Cys Asp Phe Leu Phe Asn Thr Tyr Ser
      165              170              175

Asp Leu Ile Pro Gln Val Arg Ser His Phe Glu Lys Leu Tyr Pro Lys
      180              185              190

Asp Pro Glu Val Ser Gln Ser Ala Tyr Thr Val Ser Leu Arg Ala Lys
      195              200              205

Val Leu Asp Cys Leu Arg Gly Leu Leu Pro Ala Ala Thr Leu Thr Asn
      210              215              220

Leu Gly Phe Phe Gly Asn Gly Arg Phe Trp Gln Asn Leu Leu His Arg
      225              230              235              240

Leu Gln Asp Asn Ser Leu Val Glu Val Arg Asn Ile Gly Glu Gln Ser
      245              250              255

Leu Thr Glu Leu Met Lys Ile Ile Pro Ser Phe Val Ser Arg Ala Glu
      260              265              270

Ser His His Tyr His His Gln Ala Met Val Asp Tyr Arg Arg Ala Leu
      275              280              285

Lys Glu Gln Leu Lys Ser Phe Ala His Arg Tyr Gly Glu Glu Arg Glu
      290              295              300

Ile Ser Lys Glu Ala Gly Val Lys Leu Val Tyr Gly Asp Pro Asp Gly
      305              310              315              320

```

```
<210> 37
<211> 529
<212> PRT
<213> Chlamydia muridarum
```

```

<400> 37
Met Leu Ser Lys Glu Gly Asp Phe Ser Lys Glu Gln Arg Glu Arg Leu
  1             5             10             15

Ser His Phe Val Thr Asn Leu Asp Ser Pro Ile Phe Ala Leu Lys Asn
      20             25             30

Leu Pro Glu Val Val Lys Gly Ala Leu Phe Ser Lys Tyr Ser Arg Ser
      35             40             45

Ile Leu Gly Leu Arg Ala Leu Leu Leu Lys Glu Phe Leu Asp Gly Glu
  50             55             60

```

Gly Gly Asp Phe Leu Ser Glu Asp Leu Gln Asp Cys Glu Leu Gly Ile
 65 70 75 80
 Gln Lys Ala Ala Asp Phe Tyr Arg Arg Val Leu Asp Asn Phe Gly Asp
 85 90 95
 Asp Ser Val Gly Glu Leu Gly Gly Ala His Leu Ala Ile Glu Gln Val
 100 105 110
 Ser Met Leu Ala Ala Lys Val Leu Glu Asp Ala Arg Ile Gly Gly Ser
 115 120 125
 Pro Leu Glu Lys Ser Ser Arg Tyr Val Tyr Phe Asp Gln Lys Val Asn
 130 135 140
 Gly Glu Tyr Leu Tyr Tyr Arg Asp Pro Ile Leu Met Thr Ser Ala Phe
 145 150 155 160
 Lys Asp Thr Phe Leu Asp Thr Cys Asp Phe Leu Phe Asn Thr Tyr Ser
 165 170 175
 Glu Leu Ile Pro Gln Val Arg Ala Tyr Phe Glu Lys Ile Tyr Pro Lys
 180 185 190
 Asp Pro Glu Val Ser Gln Ser Ala Tyr Thr Val Ser Leu Arg Ala Lys
 195 200 205
 Val Leu Asp Cys Leu Arg Gly Leu Leu Pro Ala Ala Thr Leu Thr Asn
 210 215 220
 Leu Gly Phe Phe Gly Asn Gly Arg Phe Trp Gln Asn Leu Leu His Arg
 225 230 235 240
 Leu Gln Asp Asn Asn Leu Val Glu Val Arg Asn Ile Gly Glu Gln Ala
 245 250 255
 Leu Thr Glu Leu Met Lys Ile Ile Pro Ser Phe Val Ser Arg Ala Glu
 260 265 270
 Pro His His His His His Gln Ala Met Val Asp Tyr His Leu Gly Leu
 275 280 285
 Arg Glu Gln Leu Lys Ser Phe Ala Gln Arg Tyr Gly Glu Glu Arg Glu
 290 295 300
 Pro Ser Leu Glu Lys Gly Val Lys Leu Val Tyr Gly Asp Pro Asp Gly
 305 310 315 320
 Leu Tyr Lys Ile Ala Ala Ala Ser Met Phe Pro Tyr Ser Glu His Thr
 325 330 335
 Tyr Ala Asp Leu Leu Asp Ile Cys Arg Lys Ile Pro Asp Glu Asp Leu
 340 345 350
 Met Leu Ile Leu Glu Ser Ser Ala Ser Ser Arg Glu Asn Arg Arg His
 355 360 365
 Lys Ser Pro Arg Gly Leu Glu Cys Ala Glu Phe Ala Phe Asp Ile Thr
 370 375 380

39/48

Ala Asp Phe Gly Ala Tyr Arg Asp Leu Gln Arg His Arg Ile Leu Thr
 385 390 395 400

Gln Glu Arg Gln Leu Leu Thr Thr Lys Leu Gly Tyr Ser Ile Pro Gln
 405 410 415

Gln Leu Leu Asp Thr Pro Met Glu Ala Pro Phe Arg Glu Ala Met Glu
 420 425 430

Lys Ala Asp Gln Ala Tyr Arg Leu Ile Ala Ala Glu Phe Pro Glu Glu
 435 440 445

Ala Gln Tyr Val Val Pro Leu Ala Tyr Asn Ile Arg Trp Leu Phe His
 450 455 460

Ile Asn Thr Arg Gly Leu Gln Trp Leu Cys Glu Leu Arg Ser Gln Pro
 465 470 475 480

Gln Gly His Glu Ser Tyr Arg Gln Ile Ala Ile Asp Met Ala Lys Glu
 485 490 495

Val Ile Gln Phe His Pro Ala Tyr Lys Ser Phe Leu Lys Phe Val Asp
 500 505 510

Tyr Ser Glu Thr Asp Leu Gly Arg Leu Lys Gln Glu Ser Arg Arg Lys
 515 520 525

Ala

<210> 38

<211> 24

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 38

gccatgggga tcaccaaga aact

24

<210> 39

<211> 24

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 39

atgcttcaaa caatcttcaa acaa

24

<210> 40

<211> 24

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 40

gccatgggttc gagttacgct cgtc

24

<210> 41

<211> 24

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 41

taagggtttt tcgagaagtt tcat

24

<210> 42

<211> 24

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 42

gccatggaaa tcaactttact tttt

24

<210> 43

<211> 24

<212> ADN

<213> Séquence artificielle

<220>

<223> Description de la séquence artificielle:amorce

<400> 43

ttctatatct ttatgtaggc aatg

24

<210> 44

<211> 693

<212> ADN

<213> Helicobacter pylori

<400> 44

atgtggatca cccaagaaac ttggctaaaa gcgttaccgt ggaataaaaa gcgatatagg 60
agtcaaataa tggaagtgat ttgtaagcac tatacccctt tagacattgc gagccaagcg 120
atccgcactt gctggcagag ctttgaatac agcgatgatg gaggctgtaa ggataaagaa 180
ttgatccaca ggggtggggaa tatttttagg cattcttcca ctttagagca tctttattac 240
aattttgaaa tcaagggttt gagcaggggg gcgttgcaag aattgagccg gcatagaata 300
gcgagcttga gcgtgaaatc aagccgttac actttgaggg aattgaaaga agtggagagc 360
tttttaccctt ttaatgaaac gaatttagaa agagcgaaa agtttttagt ttttgtggat 420
aatgaaaaag tgaatgcaat gagcgtttta gctttggaaa atctaaggat cttattgagc 480
gagcataaca ttaaaaaacga tttagccaaa tacgccatgc ctgaaagcta taaaacgcat 540
ttggcctata gtattaacgc taggagcttg caaaatttct tgactttaag aagcagtaat 600
aaagccttaa aagaaatgca agatttagcc aaagccttat ttgacgcttt acctggcgag 660
catcagtatt tgtttgaaga ttgttgaag cat 693

<210> 45
 <211> 624
 <212> ADN
 <213> *Helicobacter pylori*

<400> 45
 atggaagtga tttgtaagca ttacacccct ttagacattg cgagccaagc gatccgcact 60
 tgctggcaga gttttgaata cagcgatgat ggaggttgta aggataggga tttgatccac 120
 aggggtgggga atattttcag gcattcttcc acttttagagc atctttatta caattttgaa 180
 atcaaggggt tgagtagggg ggcgttgcaa gaattgagcc gacatcgaat agcgagcttg 240
 agcgtgaaat caagccgtta cactttaagg gaattgaaag aagtggagag ctttttacc 300
 cttaatgaaa cgaatttaga aagagctaaa gagtttttag tttttgtaga tgatgaaaaa 360
 gtgaatgaaa tgagcgtttt agctttggaa aatctcaggg ttttattgag cgagcataac 420
 attaaaaacg atttagccaa atacgccatg cctgaaagct ataaaacgca tttagcctat 480
 agcattaacg ctaggagctt gcaaaattta ttgactttta gaagcagcaa taaagcctta 540
 aaagaaatgc aagatttagc caaagcctta tttgacgcct taccttatga gcatcaatat 600
 ttgtttgagg attgtttgaa gcac 624

<210> 46
 <211> 621
 <212> ADN
 <213> *Campylobacter jejuni*

<400> 46
 atgcaaatca ctttactttt tcacactcca ttatctgttt gttctcatgc aacaagaact 60
 tgttggcaaa gctttgaaaa aggcgattgc ggtggcgaaa aagataaaga acttatcgat 120
 cgcgtgggta ataaattcaa acacgcttca accttagaac atctaaacta tactttttat 180
 atacaaggaa tttcaagagc ttgcttataa gaagttgcaa gacacggtca cactagtcct 240
 agtgtaaaaa gcacgcgtta taccttaaaa gaacttcgca atgaagctga atttaaaata 300
 ggagattttg aaaatgcaag ccgttatctt gtgctttgtg gtaatgaaga ggttgataac 360
 gcaagcatta aagccttaga aaattttacgc actattttac aaaaaagcat tagtctagat 420
 atagccaaat actgcttacc agaaagctat aaaaccgaac ttactactaac gattaatgca 480
 agaagtttac aaaattttat atctttgcgt agttcaaaat cagctctttg ggagattaga 540
 aatttagcaa atgctttatt tgaagcctta ccacaagaac acaaatttat atttgaacat 600
 tgcctacata aagatataga a 621

<210> 47
 <211> 882
 <212> ADN
 <213> *Rickettsia prowazekii*

<400> 47
 atgcacaata cgacaaaaag agtaacagta ccagcacttg aagcaatggt atatgagact 60
 atcaaagttt tagatcatgg ttttattagg gtaatcgatt atatgggtga tgatagttcg 120
 atagtacagg cagctcgtgt ttcttatggg aaaggtacta agcagttaaa ccaagataaa 180
 ggattaataa attacttgct gcgtcattat catacgacac cttttgagat gtgtgacatc 240
 aaatttcaca ttaaattacc tatattttatt gcacggcagt ggattaggca taggactgct 300
 agtggttaacg aatattcagc aaggtattct attttaggaa atgaatttta tttacctgac 360
 cctgccataa ttgcttctca atctgttgta aataaacaat gtagagcagg ggatagcgta 420
 ccgaaaaaag tatctgaaaa agttctcgca attttagaag aagatgctag acggtgctac 480
 aggcattata aggagctcat gaatgctgat gaagatggaa atattctaga tgagaatggt 540
 tcaggcatag caagagaact tgctcgtatt aatttaactt tgaattatta tacggaatgg 600
 tattggaaga ttaatttaca taatttgctt ctttttttaa gattacgaac tgatcctaag 660
 gcacaatatg aaattagagt ttatgccgaa aagatacttg acatagtaaa agcttgggta 720
 ctttttactt acgaagcgtt tgaagagtat cgtttgcaag gcgcgaatat ttcacgtaaa 780
 ggttttagagg taattaaaaa aatgataaaa ggtgaaaaag tgatccatga aactagtggg 840
 atgaataaaa gagaatggga ggaattggta aagattttta gg 882

<210> 48
 <211> 660
 <212> ADN
 <213> *Thermotoga maritima*

<400> 48
 atgaagatcg atatcttggga caaaggattc gttgaacttg tggatgtgat gggaaacgac 60
 ctctccgctg tacgggctgc ccgcgtttct ttcgatatgg gtttgaaaga cgaagaaagg 120
 gatagacatc tcatcgaata tctcatgaaa cacggtcacg agacaccttt cgaacatatt 180
 gtcttcactt tccacgtgaa agcccctata ttcgtggcga ggcagtgggt cagacacagg 240
 atcgcttctt acaacgaact gagcggcaga tactcgaagc tctcctacga attctacatt 300
 ccctctcccg aacgcctgga agggtaaaaa acgaccatcc ctcccgaacg ggtgacggag 360
 aagatctcag aaatagtcga taaagcgtat cgaacgtatc tggagtgtgat agaaagcggg 420
 gttcctcgag aagtggcaag gatagtgtc cctttgaatc tgtacacgag gtttttctgg 480
 actgtgaacg caagaagtct catgaacttt ttgaacctga gggcagattc tcacgctcag 540
 tgggaaattc aacagtacgc tctggcaatc gcgagaattt tcaaagagaa gtgtccctgg 600
 acttttgagg catttcttaa gtatgcttat aaaggagata tcctgaagga ggtacaggta 660

<210> 49
 <211> 729
 <212> ADN
 <213> *Pyrococcus horikoshii*

<400> 49
 atggttaagg taaaactgat aaattacact cccaaaccct tggagacagt tacttgggct 60
 gcacttataa gctattggga tgggtggagt accgaagctt ttgagaagat atccccta 120
 gatgttgaaa ttcacctacc tagaatccta agctacggac acgaatcaat cctggagcat 180
 gcaactttca cattctccat agagggatgc tcaaggggtgt gcacgcatca gctcgtagg 240
 cataggatag caagttagac ccaacagagt caaagggtaca ttaaaataaa ccctgaagat 300
 gttgaggaaa ccttcgctcat accggagtc ataaaaaag attcggagtt attaaaagag 360
 tggaaaagag tacttaagag atcttttaga ttgtacgaga aaagcattga aagggggata 420
 catcaagagg atgcaagatt catccttccc caatcagtta aaacgaaaat tgtagtaacg 480
 atgaacctaa gggaaactaaa gcacttcttt ggattgagat tatgcgagag ggcccaatgg 540
 gagatcaggg aagttgcttg gaagatgctt gaggaattg ccaaaaggaa agagctcaag 600
 ccgataattg agtgggctaa gctagggcca agatgcattc aacttggtta ctgtccagag 660
 agggaaactta tgccaccagg ttgcttaaaa agaacgaggg aaaggtggaa aaacttattg 720
 gaaaagtat 729

<210> 50
 <211> 750
 <212> ADN
 <213> *Mycobacterium tuberculosis*

<400> 50
 gtggccgaga ccgcgccgct gcgcgtgcaa ctgatcgcca agaccgactt cttggcccca 60
 cccgacgtgc cctggaccac cgacgccgac ggccgacccg cgctggctga gttcgccggc 120
 cgggcctgct atcagagctg gtccaagccc aatccaaga ccgccacca cgccggctac 180
 ctccggcaca tcatcgacgt cggacatttc tcgggtgctag agcatgccag cgtgtcgttc 240
 tacatcaccg ggatctcgcg atcgtgcacc cacgagctga tccgccaccg gcatttctcc 300
 tactcgcagc tctcccagcg ctacgtaccc gagaaggact cgcgggtcgt cgtgccgccc 360
 ggcattggagg acgacgccga cctgcgccac atcctgaccg aggccgccga cgccgccgcg 420
 gccacctaca gcgagctgct ggccaagctg gaagccaagt tcgccgacca acccaacgcg 480
 atcctgcgcc gcaagcaggc ccgccaagcc gcccgccgcg tgctgcccga cgccaccgaa 540
 acccgcatcg tggtagaccg caactaccgg gcctggcggc acttcacgc aatgcggggc 600
 agcgagcacg ccgacgtgga aatccggcga ctggccatcg aatgcctgcg ccagctcgcg 660
 gccgtggccc ccgcggtgtt cgccgacttc gaggtgacca ccctggccga cggcaccgag 720
 gtggcgacca gcccggtggc gaccgaagcc 750

<210> 51
 <211> 750
 <212> ADN
 <213> *Mycobacterium tuberculosis*

<400> 51
 gtggccgaga ccgcgcgcgt gcgcgtgcaa ctgatcgcca agaccgactt cttggcccca 60
 cccgacgtgc cctggaccac cgacgcccgc ggcggaccgc cgctggtcga gttcgccggc 120
 cgggcctgct atcagagctg gtccaagccc aatcccaaga ccgccacca cgcgggctac 180
 ctccggcaca tcatcgacgt cggacatttc tcggtgctag agcatgccag cgtgtcggtc 240
 tacatcacccg ggatctcgcg atcgtgcacc caccgagctga tccgccaccg gcattttctc 300
 tactcgcagc tctcccagcg ctacgtaccc gagaaggact cgcgggctcg cgtgccgccc 360
 ggcattggagg acgacgccga cctgcgccac atcctgaccg aggccgccga cgcgcgccgc 420
 gccacctaca gcgagctgct ggccaagctg gaagccaagt tcgccgacca acccaacgcg 480
 atcctgcgcc gcaagcaggc ccgccaagcc gccgcgcgcg tgctgcccga cgcaccgaa 540
 acccgcatcg tggtagccgg caactaccgg gcctggcggc acttcacgcg aatgcggggc 600
 agcgagcacg ccgacgtgga aatccggcga ctggccatcg aatgcctgcg ccagctcgcc 660
 gccgtggccc ccgcggtgtt cgcgcacttc gaggtgacca ccctggccga cggcaccgag 720
 gtggcgacca gcccgttggc gaccgaagcc 750

<210> 52
 <211> 762
 <212> ADN
 <213> *Mycobacterium leprae*

<400> 52
 gtggcccaga tcgcgcgcgt acgcgtgcaa ctaattgcc aagactgagtt tctggcgcct 60
 cctgacgtgt cgtggactac cgacgcccgc ggtggttccg cgctggtcga attcgccggt 120
 cgcgcctggt atcagagctg gtcgaagccc aatccccgga ccgcgaccga cgcgcgctac 180
 attaaacaca tcattgacgt cgggcatgtt gcggtgctcg agcatgccag cgtttcggtc 240
 tatatcagcg gcatctcgcg atcatgcact caccgagctga tccgacatcg gcattttctc 300
 tactctcagc tgtcgccagc ctacgtgccg gaaaaggatg cccaggctcg tgtgccaccg 360
 gacatggagg atgacgacga acttcaacag attctgattg cggccgtaga ggccagccgg 420
 gccacctaca ctgaactgct ggtcaagctg aacgccaagt taatggccgg tgagctcggc 480
 ggggaataggg cgggtgttgcg gcgcaagcaa gctcgccaag ctgcccacgc ggtgctgccc 540
 aacgccaacg agaccgcgat cgttgtgacc gggaactacc gggcatggcg gcacttcac 600
 gccatgcggg ccagcgagca cgcgcagctc gaaatccggc ggctggccat tgtctgcctg 660
 cgcgcgctcg tcgacgttgc gccgcagta ttcgctgatt tcgagatcac cgcacttgct 720
 gatggtactg aggtcgcaac tagtccttta gccaccgaag ct 762

<210> 53
 <211> 780
 <212> ADN
 <213> *Sulfolobus solfataricus*

<400> 53
 atgatctctg ttaaattagt ctcatatacg aatgatggag agaaagtcac tgctattgct 60
 gctaagatga gtaggagtag aaaagggttg gattatcatg agaaagacat gactgatgat 120
 gaaattgaga catggatcag agatgcaatc cttcacgggt attggagtgt tcttgagcat 180
 agcgtttata ctttttctat agaagaaatt tctcgtgtag cttcacatca gcttgtgagg 240
 cataggattg caagttatac gcaaatgagt cataggtttg ctaagcctat agacgaatac 300
 taaaaaccaa ttattccgcc atctgccaa gaaaagaagta aagagttagt ggagaaagcg 360
 tataaggaag cctatgacaa ctattacaca cttttagaaa gcggtgtacc agaggaagat 420
 gcaagatatg tactaccaa tggagtaaat acaaatattg tcgtcacaat gaatgctaga 480
 gagttatata acttcttctc gcttcgactg tgttcaaggg cacagtggga gattagagca 540
 atagcttgga aaatgttaga ggaggttaag aaagtcatc cccgactatt caaatatact 600
 ggaccaaatt gtataatata tgaaaacttc ataagaaatg aaaaatgaatc tataacttta 660

gaagatatat ttaaagatta caaattagag ttcatatctc aaagatgcat cgagggagta 720
 ttgagagacg gaataaaaaa atgcatcata aattcgcgtt ctgtattgga taatattaaa 780

<210> 54

<211> 951

<212> ADN

<213> Aquifex aeolicus

<400> 54

ttgatgaaaa tctacttaat gggctcggac cagaggatag ttaggtgtgc gaggggtgtcc 60
 tttgcaaagg actcatatgt cgatgaaaag agggacaaaa gactcattcg ctacctcttc 120
 aagcacaggc acgcgtctcc ctttgagcac aacatcattg cctttgagtg gaaaaaggag 180
 aagtggatag aactcctttc aaagctggaa aaccctaccg ttcaggttta ctactcaaat 240
 ggctttgttt ttcttaactt aaggaacgct atcaacgtat gggaactttt gcccgacgca 300
 gtcaaagaaa ggataaagga ggcttttccct accacttacg gagtaattca gcggagaggg 360
 gaaattgagg atgaggaact ttattccctt ccctacacta aagataaagc ttacgtaaag 420
 gaaaagattg agacgagttc tggctggatc gggcttgttg ataagctgga actggaaact 480
 gatatggatt tttacacttt tgtgggtgaa tgccctcttt ttgtagcccg tcagtggatg 540
 aggcacaggt tcggctctta caacgaggtt tctaaaaggt acgtgggaaa ggagtttctg 600
 gaattttatc tccctaaata cataagaaag caagcggaga aaaacaaaaca ggcttctgtg 660
 gatgaaccta tttctgagag tgaagttttt ataaaaaaa tagaaaattt gataagcaaa 720
 agcgtaaaagc tctacgagga gataatagaa aaggcgggtg caaaagagct cgcaagggga 780
 gtgcttcccc agtttatgaa aacgaggttt tactggaccg ttccgaggat ttctcttgac 840
 aacttcataa ccttcagaac tcacgaaggt gcacaaaaag agataaggga atttgcagaa 900
 gctataaagg aatggtagg atacagggga actgataaga agaacgtcat t 951

<210> 55

<211> 981

<212> ADN

<213> Aeropyrum pernix

<400> 55

gtgtccttgg cagcgtccct ggaaaaagct ggccctcggca tctcggtcag gctactcgag 60
 tatactgggtg atggagagcg tattgtcgct gttgctagta aggtaagcct ctcccgcagc 120
 ccggcggaga ggctgctggc cataggggaa gatgaagttg agacttggat actcgagacg 180
 tttagaaggc agcatttcag cccctgggag cacagcgtct acacattcat gggtgagggg 240
 ctctcgaggg tggctagcca ccagctagtt agacataggg ttgctagcta cacacaacta 300
 agccataggt atagcagagg ctacctcagg gaggcgcgcc tcaaggcctg cgaatccatc 360
 ggccctagact gtccatcgaa accggctgag acggagggcg ggcggaagc cgcatacagg 420
 ctgtactctc aggcctctgga gagagctgct cgagatttcg gggctagcga gaggtttgct 480
 atagcggcga aggccttctg aataccgcct accattcttg cgaggggtga tggcgggtgat 540
 ggtgtcgtgg aggcgtatct acggctctgca gcgatatact atagcctact ctcccggggg 600
 gctaggcgtg aggatgcgcg gtatatcctg cccgatgcgc tgagaaccag aatcgttgtt 660
 actatgaatg cacgggagct tatacaggtc ttcttcccc tgaggatgtg cactagagcc 720
 cagtgggaga tacgccacat agcctggctg ctatggcgag agctatccag gggtcatcca 780
 aggcgtgttca ggtgggcccgg gccagctgt gtgttaaggg agaacaccct taggacgacg 840
 cccgccagct tatacagcta cctggagggc gtagagcggg ttacacagcc ccgctgccct 900
 gagcttggtg agaacaaggc tataccggga tgcctcaggc aggcggcatc ggtggcgccc 960
 cccggagacg gggagtagga a 981

<210> 56

<211> 798

<212> ADN

<213> Halobacterium sp.

<400> 56

atggtaccgg cgcgtgggtt cggagtgttt ttgccaccgg caggcacacc gtctagcatg 60
 cgcgtccgtc tcctcgaagc gactgagaac ccggaggaaac tcactctgtca gagcgccccg 120

```

aacgactaca tgtccgactg ggtcgggtgac acgccactcg acaccgggat ggcggtccgtg 180
gacggcgaca ccacagacga aaagctgtcg aacctcatcg cacagctgct cactcgccgc 240
cactacggcc ccttcgagca tccgtcggcg acgttcgcca tcgaggggtg gagccggctg 300
tgtatggcgc agctcactcg ccaccgccac gccagcttcg acgtgcagtc gatgcgctac 360
gtcgcgttcg acgacgtcga tccggcgggcg gttgcggagg gcgagctggg ggtgacgccg 420
ccgtcggcga ccgaccccgga ctgggtcggc cgcaaccagg acgcgggcga catcgacgag 480
gagaccatgg ccgaacgcga ggcggtcttc caggcgtcgg tgcggcgcg cgtcgaagac 540
taccaggaac tgctggggct cgggatgccc ccggaggacg cccggttcgt gctccccatc 600
gggaccgaag tgaacgtggg gatcacgctg aaccgcgggt cgtgatgca cgtcgcggac 660
atgcgggccg ccgccgacgc acagtgggag atccgcgagc tcaccgaaca gctactcgat 720
gcggccgctc agtgggtgtc ccacaccttc gaattattcg acgcggagat gaaacaccgg 780
aagaaccggc tcgcgccc                                     798

```

<210> 57

<211> 795

<212> ADN

<213> *Borrelia burgdorferi*

<400> 57

```

ttgaataaag aatataaaat tttggataat ggttttttaa aacttattga tttcatggga 60
gatgatagga gaatagttaa ggctgcaagg atttcttctc gagaagagag cgttaaaaga 120
aaagatgcgg aacttattga ctatttaata agaaatgggc acacaagccc attggagcag 180
gtggttttta catttcatgt taaagctcca atatttggtg caaggcaatg gatgaggcat 240
agaacggcaa ggattaatga agtttctgga tgctacagct tggcaagaga ggaattttat 300
gtccctttag aagaagattt aaagtgtcaa acttctagta atagctctga aaaagagttt 360
aagtctttgg aaaaattgtc ggacaaaata aagcatcatc aaaaacattc ttatgagctt 420
tatcaagata tgatcaatgc taatattcca aaagaactct caagaatagt tttgccctta 480
agtttatata ccgaatggta ttggcaaatt gatttaaata atctttttca ttttattaaa 540
ttgcgattag ccctagacag tccaaaagaa attaaagaaa attcgccaaa agaaatgcgc 600
gaatatgcta aagcattgat aagcatagta agagaaattg tgcccatcgc ttttaacagt 660
tttgaaaatc attttttaag aggaaagaga ttttcccacg aagagataat tgcaattatt 720
aatgcttttg atttaataaa gcttagtatg gatgctgaaa aattgaactt attgaaagat 780
aagctaggaa ttgat                                     795

```

<210> 58

<211> 915

<212> ADN

<213> *Treponema pallidum*

<400> 58

```

gtgacgttgc gtacgcttca agccgggtgtg gcggtcagta tcgctctgga tcgtgtgtgc 60
tttttctggt ataacggggc ggtggcacac tgtgtagtag aagctgccga agatattttg 120
gaccggcggt tttctgtatt ggataagggt ttcgtgcgtt tgatagatta cctgggaggg 180
gatgcacgca ttgtgcaggc agcgcgtggt tcttacgggt cggggactag gactgcgcgt 240
gacgatgcgg cgttatcga ttttctttta cgcaataagc atacgtctcc ttttgagcag 300
gtggtcctta ccttccatgt acgtgcaccg atttttgtcg cgcgtcagtg gatgcggcat 360
cgcactgctc gcactcagtg ggtgtctagt cgttattcgc ttcttagtca tgactgttat 420
gtccgcagg aaacttcagt tgcagttcag tccacgcgta acaagcaggg ccgcgcgtcc 480
gaaggtatct ctctgaaca gcagcaggaa gtgcgggcag cgtttgaagc tcagcagaaa 540
gcggcgtgtg ccgcttacga cgcattgatt caaaagaaca tcgcgcggga gctagcgcgt 600
attaacgtgc cgttttcgct ttacaccgag tggattggc agattgattt acacaatctt 660
tttcattttt tgcgtttacg tgcgagcgt catgcgcaag cagagattcg tgcgtatgca 720
gaggtaatca ttgaaattac ccgtgcagtt gcgcgtgcg ctaccgcctc ttttgaaaat 780
catgaaaaag atgggggtgca gttttcaggg cgggagtttg ctgcgcttaa ggccttactg 840
gctggagagg gtctctccct tgaggggaag gaacgtgcgc gctttgaaga aaaattacgc 900
tctggcctgc agcag                                     915

```

<210> 59

<211> 1545
<212> ADN
<213> *Thermoplasma volcanium*

<400> 59
atggagtttt cgaacgccga gagagatgta ttcttaataa aaatagagaa aatgatagac 60
agaggtgccc tgatgtcccg ttacagtagg gcatctgato ctgatatacg atccgtttat 120
gaaaaagagt tcaaaacagg ggcgaagagc ggcgaagagt ttacaggag gattttcctt 180
gaatatggcg atgaatcgat tgccgaactt acaactgcac agatgggcat acagaacgtt 240
tcaaacgtag catccaagat aattgaggag ataaggatcg gcctttctta tcttgaaaaa 300
tcgacaagat atgtacgtta cgacaaaaag gttgatggca gataccttta catttcacca 360
gagaggatag gaatatcggg ggatgatgct aaggattacg tgcaactatg cgacaatctc 420
ttcgaattct attccaaggc attgcctcag gtagaagatt accttaggca taagtttccc 480
caagacaagt tggattacca aaatgctggg ggaaagactc taaccgaaat ggatcaaaac 540
gagaaaaaaa tagcagaaag atcctatatc aatgcagtta ggtcacgtgc gctggatgat 600
gtaaggtata ttttgccagc atcgacgctg acgaacatcg gtattagcgg caacggaagg 660
gccctaattc atcttataca gaagttaatg gaatatgaaa tcccagaaac taaaagctt 720
gcaaaggaca tatacgatga actaaagcca gaattgccac agcttataga tgatgcacta 780
tctggccacg gccttgagat cataaatttt aagaagaatc tcatgggctt gtttccctat 840
gatctcactg gcaattttga aagaattcgc ctactctcat atggcgaaga ggataaagag 900
ctcaggaagg tggcatccct catcgaatat ccattccatg gggacgcggc ttacttttac 960
tcccgcagct ctgaatccta cgtaaaatat atgaaagagc tcatagagtc gataaggtcg 1020
ttgagggcaa acaggaggat gaaaccagga agggcattcg aatcagtaaa ttatgttttt 1080
gagctaaact taaactacgg ctcatctcgt gatctacaaa gacaccgctt tcttggtata 1140
attaggaaac cgcttacggc agcatacggg tacgacactc ctccagtgat ctctgctata 1200
gatgaattga aaacgcagta tgatgagttg atggcgaatt cttcttcctt ttaccaaagg 1260
ctaagggaaa agtatgggccc ctggatatca cagtatgttg ttccgttcgc ttttaaatac 1320
ccgatcacgt tttccacgaa tctctctgaa gtaacgtatt ttgtagaact caggagtact 1380
gcgcaggccc atttcgatct aagagacata gcggtaagca tgtacaggga agtctcgaaa 1440
gtgcacccaa ctctgtccag aatcataaaa tttgttgata ctgcagatta tccgcttggc 1500
aggttatctg ctgagtttag gaaggaatcg aagaaggccg gtatt 1545

<210> 60
<211> 732
<212> ADN
<213> *Pyrococcus abyssi*

<400> 60
atggttcgag ttacgctcgt caattacaca aggagacccc tcgaaacgat aacatgggag 60
gccctcgtaa gttactggga cgagtggagc actgagtcac ttgaaaagat caacgaggat 120
gatgtaaaaag ctacactccc caggatactt ggttatggac acgagagcat tttagagcac 180
gcaacgttca cattctcaat agagggctgc agtagggttt gcacgcacca gttggtgagg 240
cataggatag ccagctacac tcagcaaagc caacgctaca ttgttctcaa cgaggagaac 300
gttgaagaaa ccttcgtgat accagagtcg ataaagaagg acagagaact ctatgaaaaa 360
tggaagaaaag ctatggcgga gacaataaag ctctacaagg aaagttaaaa gagaggtatt 420
caccaagaag atgctagggt catccttccc caggcgggta ggagtataat agtcgttacg 480
atgaacctta gggagctcaa gcacttcttc ggcttaagat tgtgcgagag ggcacagtgg 540
gaaatcaggg aagttgcctg gaagatgctt gaggaaatcg caaagaggga agagctaagg 600
cctataataa agtgggctaa gcttgggcct cgttgtattc agttagggtta ttgccccgaa 660
agggaaattaa tgccccagg atgcttcaag agaacaaggg agagatggat gaaacttctc 720
gaaaaaccct ta 732

<210> 61
<211> 1587
<212> ADN
<213> *Chlamydia trachomatis*

<400> 61
atgttgagca aagaggggtg tttttctgag gagcaaagag cgcgtttatc gcattttgtg 60

```

acgaatttag actcgcctat atttgctttg aaaaaccttc cagaagtggg taaaggcgct 120
ttattttcaa aatattccag atcgactctg gggttgcgag cgcttctttt gaaagaattt 180
ttagatgggg aaggcggtaa tttccttgat gatgaccaac aagattgtga gttggggatc 240
caaaaagctg cggacttcta tcgtcgcgtt ttagacaact ttggtgatga ttctgttga 300
gagttgggag gagcgcatct tgctctggaa caagtatcca tgctcgcagc aaaaatttta 360
gaagatgctc ggattggagg gtccccctta gaaaaatcgt ctagatacgt ttatttcgat 420
caaaaagtta acggggagta tttatattac cgagacccta ttttgatgac ctcggccttt 480
aaagacgtct ttttgatata ttgtgatttc ctattcaaca catactccga tcttatccct 540
caggttcggt occatttcga gaaactatac cctaaagatc cagaagtttc tcaatcagcg 600
tatacagttt ctttacgagc taaagtatta gactgtttac gaggtttgct acctgcagcg 660
aactcaciaa atttagggtt ttttggtaat ggccgggttt ggccagaactt gctacaccgt 720
ttgcaagaca atagtttggt tgaggtacgc aatattggag agcagtcctt aacagaatta 780
atgaaaataa ttccctcttt tgtaagccgc gcagagtctc atcattatca tcaccaagct 840
atggtggatt accgtcgggc tttaaaagaa caattaaaaa gttttgcaca tcgttacggg 900
gaagagagag aaatttcgaa agaggctggg gtaaaattag tatacggaga tccagacggg 960
ttatacaaaa ttgctgcagc ctacatgttc ccctactcgg aacacactta tgcagagctg 1020
ttagatattt gtgcgaatat tcctaataaa gatctcatgc gtatcttaga gtcgggagct 1080
tctttccgag agaatcggcg gcacaaatcc cctcgcggat tggaatgtgc tgagtttgct 1140
ttgatatta cagcggattt tggagcctat cgggatttac aaagacatcg tatcctaact 1200
caagaagaag agcttttgac gacaaaattg ggttacacga tgccctcaca attgatcgac 1260
actcctatgg aagctccctt cagagaagct atggaaaaag ctgatcaagc gtatcgtcta 1320
atagcagaag agttcccaga agaagcacia tatgtggttc ctttagctta caatattcga 1380
tggtttttcc atatcaacgc tagaggtttg cagtggcttt gtgagttacg ctctcaacca 1440
caagggcatg aaagctatcg taaaattgct atagatatgg ctagagaggt tattcagttt 1500
catccagctt acgagctgtt cttgaagttt gtcgactact cagagactga cctaggaaga 1560
ttacaacaag aatcgcgtaa aaagtct 1587

```

<210> 62

<211> 1481

<212> ADN

<213> Dictyostelium discoideum

<400> 62

```

atgggtcttg atattcaaac agaaatcgat aaaattgtaa ttgaaaaggt taaaccagaa 60
gttgaatact atgacgttat gggagggtca catagatggg aagtcaaagt gcacgaccat 120
ggtaaagttg cattagtcca tactatgcca agattagcac cagttgggtca aaccgccgat 180
ttctcaattt gtcaagcagc aagagtgtca tatggtgccg gtactaaaaa agtactgaa 240
gataaaggtt taattcgtta tctttataga catcaacata cttgtaagta tataaataaa 300
taaataaata atgaaaaaaa aaaaaaaaaa aaaaaaaaag tttattaaca atttttaatt 360
tttaattttt tagcaccatt tgaaatggta gaatttaaatt ttcattgtgt aatgccagta 420
tttatcgcac gtcaatggat tagacataga acagcaaattg tcaatgaata tagtgcacgt 480
tattcagtat taccagataa attttatcat ccatcaattg aggaagttag aaagcaatca 540
acatcaaata gacaaggttg tgaagaagca ttggagccaa aaaccgcca agagttttta 600
gactatttag ataaagttga agaaaactat aaaacctata atgaactctt agagaaaggt 660
ttatcacgtg aattgggtcg tattgggtta ccagttagta tctacactga atggtattgg 720
aagatcgatt tacataatct tttccatttc ttaagacttc gtatggatag tcattctcaa 780
aaagagatta gagattatgc aaatacaata tttgctctca ttcgtccaat tgtaccagtt 840
gcttggtgaag gcatttatag attatgcttt tgaaagttta aacttacacg tttagaaatt 900
gaagcattcg tactggtagt ccaactcaata ctacaaataa aagagaaata gaagaatttg 960
aagaaaagaa gaaattatta ttcccaaata ctcaagccta aaaaaaacac aaaacacaaa 1020
cacaattatt taattaatta aataatctca tctttttttg tatataaaat aaataaaaca 1080
aatattatta atgtattgtt cattttgtta tgaaacaatt cattttctaa ataaaatata 1140
tatgttatca caataaaatt atttatttat atttattttt taaaaaatta atcatcaaga 1200
atgatttcag aatcatcttg gtattgttgt tgttgaggat tttgttggtt ttgaatgcta 1260
ttattattac gttgatttaa ttgttctttt tgtaattctt cttaattttc agaatagggt 1320
gaaacatcca attcttttaa agtttcattt gtaatttcag tatcttgggt tattattttt 1380
attaaattat catagatttt attttgtgga ttatcaaatg ctgggtggtaa attggataat 1440
ggactgtttg aatttggtta tattcaatat gtattcttta c 1481

```

<210> 63
 <211> 669
 <212> ADN
 <213> Roseophage

<400> 63
 atgacacaga ttgaagcaac atacatcgac cacatgggtt cagacctttc agtcgtcaac 60
 gcagcacggg ttagcttttg taagaagagt gagtgggttt actgtggtca gtcagacggc 120
 agagacaaag gtttatctgg ccgtgacacc aagctcatca agtacctagc caagcacaag 180
 cacatcagcc ccttcggaca tgccttcgca agcttccacg ttaaggctcc aatctttgta 240
 gcacggcagt tggatgaagc taagtctcta cggttgaatg agatcagccg ccgttacgtt 300
 gatgatgaac ctgagttcta cacacctgat gtatggcgtg gacgtagtgc tgacaagaag 360
 cagggtagtg atgggtgtag taacctcagag tataaccccc aatacctaga caacaaaatc 420
 aagtttgcct atctacaagc acttgacata ggtattttcac ctgaacaagc acggatgctg 480
 ttgccgcagt ccacaatgac tgagtgggtat tggtcaggta gccttgatgc ctttgctgac 540
 atgtgccgcc ttcggttgtaa ggaagacaca cagtatgaga gccgtgttgt ggctgaccag 600
 atcagtgaga agatggcaga cctgtatcct gtctcatggg ctgcacttat ggaaggagaa 660
 aagcaatga 669

<210> 64
 <211> 650
 <212> ADN
 <213> Chorella virus

<400> 64
 atgtccgcaa agctcatttc cgttaccaag ccagtcgtgg aggggtgttaa cactgcagag 60
 gaactaattg cgtatgctgc ccgtgtctcc aaccccgaaa accaaattaa caacaagact 120
 gcatcagggc ttctaaagta ttgtatccgc cacaagcatt ggtcaatctt tgagactgcg 180
 ttcattgactc ttgaactgaa gacgtctcgc ggtatcgagc ctcagggttct tcgccatcgg 240
 agcttccact tccaggaatt ttcccagagg tacgcatctg tgatggaaac tccaaccact 300
 catcaagcac gattccaaga tcataaaaat cgccaaaatt ctctggacac cgtcccggaa 360
 gatgatcaaa cgtgggtgggc aaccgaacaa gaaaaactgt atgcacagag catggagctc 420
 tataacaagg ctctcgaaaa gggaattgca aaagaatgtg caaggtttat tcttctcttg 480
 agtacaccaa ctactattta catgtcgggt acgatcaggg attggatcca ttacatcgaa 540
 ctgcgcactt caaacgggac acaacgagaa cacattgatc ttgcaaatgc ttgcaaagaa 600
 attttcatta aggaattccc cagcattgca aaagcacttg attgggtctg 650